М.М.Постников

АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ

М.: 1973 «НАУКА», Главная редакция физико-математической литературы

СОДЕРЖАНИЕ

Предисловие 7

Глава 1. **Векторное исчисление** **11**

§ 1*. Понятие вектора 11

1. Предварительное определение вектора 11
2. Отношения эквивалентности 13
3. Окончательное определение вектора 15
4. Векторы на прямой, на плоскости и в пространстве 19

§ 2*. Векторы на прямой 21

1. Ориентации прямой 21
2. Длина и величина вектора на прямой 25
3. Отношение векторов на прямой 26
4. Сложение векторов на прямой. Лемма Шаля 29
5. Алгебраические свойства линейных операций 30
6. Теорема об изоморфизме 32

§ 3*. Линейные операции над векторами на плоскости и в пространстве 34

1. Определение линейных операций 34
2. Алгебраические свойства линейных операций 37
3. Линейная зависимость 38
4. Геометрический смысл линейной зависимости 42
5. Базисы и координаты 46
6. Проекции и координаты 50
7. Преобразование координат при замене базиса 54

Дополнение. Теорема о ранге матрицы 58

§ 4*. Ориентации прямой, плоскости и пространства 63

1. Понятие ориентации 63
2. Правые и левые ориентации 67
3. Произведения ориентации 68
4. Стороны прямой на плоскости и плоскости в пространстве 70
5. Деформации базисов и ориентации 73
6. Резюме 79

Дополнение. О понятии угла 80

§ 5*. Метрическая теория векторов 82

1. Длина вектора и угол между векторами 82
2. Скалярное произведение векторов 86
3. Применение скалярного умножения к доказательству геометрических теорем 88
4. Выражение скалярного произведения в координатах 90
5. Ортонормированные базисы 94
6. Ортогональные матрицы 98

§ 6. Поливекторы 105

1. Бивекторы 105
2. Линейные операции над бивекторами 110
3. Линейная теория бивекторов 116
4. Метрическая теория бивекторов 121
5. Тривекторы 126
6. Векторное и смешанное произведения 133
§ 7. Линейные операторы 135
1. Отображения и преобразования 135
2. Кольцо линейных операторов 138
3. Описание линейных операторов 140
4. Обратимые линейные операторы 145
5. Операторы, действующие по равенству координат 149
6. Обратимые линейные операторы и ориентации 151
7. Изометричные операторы 155
8. Свойства изометричных операторов 157
Глава 2. **Метод координат** **164**
§ 1*. Координаты на прямой, в плоскости и в пространстве 164
1. Аффинные координаты 164
2. Замена аффинных координат 169
3. Деление отрезка в данном отношении 172
4. Прямоугольные координаты 173
5. Полярные, сферические и цилиндрические координаты 175
6. Однородные координаты 177
§ 2. Уравнения линий и поверхностей 184
1. Задание линий и поверхностей уравнениями 184
2. Алгебраические линии 187
3. Параметрические уравнения линий и поверхностей 191
§ 3. Координатно-аксиоматическое построение геометрии 195
1. Основные положения аксиоматического метода 195
2. Аксиоматика евклидовой геометрии 198
3. Аксиоматика аффинной геометрии 208
4. Афинная геометрия над полем комплексных чисел 217
5. Вещественно-комплексная геометрия 221
Дополнение. Аксиоматика Гильберта 226
Глава 3. **Линии и поверхности первого порядка** **239**
§ 1*. Прямая на плоскости 239
1. Прямая как линия первого порядка 239
2. Параметрические и канонические уравнения прямой 242
3. Взаимное расположение прямых на плоскости 247
4. Полуплоскости, на которые прямая разбивает плоскость 251
5. Прямая на евклидовой плоскости 253
§ 2*. Плоскость в пространстве 258
1. Плоскость как поверхность первого порядка 258
2. Параметрические уравнения плоскости 260

3. Взаимное расположение плоскостей в пространстве 263
4. Полупространства, на которые плоскость разбивает пространство 266
5. Плоскость в евклидовом пространстве 268
§ 3*. Прямая в пространстве 271
1. Прямая в аффинном пространстве 271
2. Взаимное расположение прямых и плоскостей 275
3. Прямая в евклидовом пространстве 278
4. Расстояние между двумя прямыми в пространстве 282
***Глава 4.* Геометрия прямых, плоскостей и окружностей 290**
§ 1. Геометрия прямых на плоскости 290
1*. Пучки прямых 290
2. Расширенная плоскость 294
3. Полнота и непротиворечивость аксиом геометрии расширенной плоскости 298
4. Координаты на расширенной плоскости 302
5. Проективная плоскость 307
6. Интерпретации проективной геометрии и их применения 316
7. Конфигурационная геометрия 322
Дополнение. Трилинейные координаты 326
§ 2. Геометрия плоскостей в пространстве 331
1*. Пучки плоскостей 332
2*. Связки плоскостей 333
3. Расширенное пространство 337
4. Проективное пространство 341
Дополнение. О геометрии прямых в пространстве 346
§ 3. Геометрия окружностей на плоскости 349
1. Степень точки относительно окружности 349
2. Связки окружностей 353
3. Пучки окружностей 362
4. Пучки как пересечения связок 370
5. Прямые как окружности 377
6. Окружности на вещественно-комплексной плоскости 386
Дополнение. Геометрии параболической и гиперболической связок 396
***Глава 5.* Элементарная теория линий и поверхностей второго порядка 398**
§ 1*. Линии второго порядка 398
1. Параболы 398
2. Эллипсы 401
3. Гиперболы 411
4. Уравнения эллипса, параболы и гиперболы в полярных координатах 423
§ 2. Некоторые дополнительные свойства линий второго порядка 426
1. Эллипс, парабола и гипербола как конические сечения 426
2. Взаимное расположение конических сечений и прямых 430
3. Прямые, касающиеся конических сечений 436

4. Семейства софокусных эллипсов и гипербол 443
5. Диаметры конических сечений 447
6. Теоремы Аполлония 455
§ 3*. Поверхности второго порядка 460
1. Эллипсоиды 460
2. Двуполостные гиперболоиды 463
3. Однополостные гиперболоиды 466
4. Прямолинейные образующие однополостного гиперболоида 469
5. Гиперболические параболоиды 480
6. Эллиптические параболоиды 488
7. Конусы второго порядка 490
8. Цилиндры второго порядка 496
***Глава 6.* Общая теория линий второго порядка** **499**
§ 1*. Классификация линий второго порядка 499
1. Линии второго порядка на евклидовой плоскости 499
2. Инварианты уравнений линий второго порядка 509
3. Определение вида линии второго порядка по инвариантам ее уравнения 513
4. Линии второго порядка на аффинной плоскости. Теорема единственности 519
5. Центры линий второго порядка 527
6. Асимптоты и диаметры линий второго порядка 533
7. Приведение уравнений линий второго порядка к простейшему виду 540
8. Главные направления и диаметры линий второго порядка 544
Дополнение. Классификация поверхностей второго порядка 548
§ 2. Проективная теория линий второго порядка 554
1. Линии второго порядка на аффинно-проективной и проективной плоскостях 554
2. Пересечение прямой и линии второго порядка 558
3. Поляры и полюсы 563
Дополнение. Поляры в пространстве 568
4. Теорема Безу 569
Дополнение. Дифференциально-геометрическое истолкование кратности точки пересечения 577
§ 3. Геометрия линий второго порядка 579
1. Пучки линий второго порядка 579
2. Описание пучков линий второго порядка 583
Дополнение. Еще раз о пучках окружностей 593
3. Линии второго порядка, проходящие через пять точек 594
Дополнение. Поверхности второго порядка, проходящие через девять точек 596
4. Теорема Штурма 598
5. Теорема Паскаля 601

6. Квадратичные пучки прямых 603
7. Фокусы линий второго порядка 612
Глава 7. **Геометрические преобразования** **617**
§ 1. Аффинные, проективные и ортогональные преобразования 617
1. Аффинные преобразования 617
2. Линейный оператор, индуцированный аффинным преобразованием 621
3. Общий вид афинных преобразований 628
4. Проективные преобразования 633
5. Ортогональные преобразования 637
§ 2. Разложение аффинных и ортогональных преобразований в композицию более простых 644
1. Аффинные и ортогональные преобразования прямой 644
2. Разложение аффинных преобразований плоскости и пространства 647
3. Разложение ортогональных преобразований плоскости 652
4. Разложение ортогональных преобразований пространства 658
5. Представление движений пространства с помощью кватернионов 665
§ 3. Конформные преобразования 677
1. Инверсия относительно окружности 677
2. Пополненная плоскость 679
3. Свойства конформных преобразований 686
4. Конформные преобразования и ориентации 694
5. Конформная геометрия 704
§ 4. Группы и геометрии 707
1. Геометрии с данной группой автоморфизмов 707
2. Простейшие геометрии аффинного типа 711
3. Геометрии Галилея и Пуансо 718
4. Геометрия Минковского 727
5. Комплексная евклидова геометрия 734
6. Унитарная геометрия 738
7. Геометрия Пуанкаре — Лобачевского 740

ПРЕДИСЛОВИЕ

Эта книга отличается от традиционных учебников аналитической геометрии по крайней мере в двух отношениях. Во-первых, в ней сделана попытка привести изложение аналитической геометрии на уровень строгости и формализации, давно уже достигнутый в учебниках алгебры и анализа. Во-вторых, помимо общеобязательных, стандартных вещей, в ней изложено довольно много материала либо никогда ранее в учебники не включавшегося, либо давно из учебников исключенного.

В современной математике проективная геометрия, на которую, собственно говоря, и ориентирован традиционный курс аналитической геометрии, потеряла уже свое ведущее значение. Поэтому роль «научной базы» курса аналитической геометрии перешла к линейной алгебре. Это ведет к все более тесному слиянию курса аналитической геометрии с курсом линейной алгебры, слиянию, получившему уже организационные формы (читаются «объединенные курсы аналитической геометрии и линейной алгебры» и издаются соответствующие учебники). Тем самым курс аналитической геометрии как курс геометрии фактически ликвидируется, в результате чего студент университета может получить диплом математика, не имея, по существу, никакого представления о комплексе идей и методов классической геометрии плоскости и пространства. Для преодоления этого необходимо заново насытить собственно курс аналитической геометрии (первый семестр) живым и ярким геометрическим материалом, перенеся из него весь, какой только возможно, линейно-алгебраический материал в последующий «объединенный курс аналитической геометрии и линейной алгебры». Конечно, курс аналитической геометрии должен быть в первую очередь по силам студенту — бывшему школьнику. Вместе с тем он должен иметь общематематическое значение, развивать математическую культуру студента и подготавливать его к усвоению программ следующих семестров. Как бы ни была красива и интересна та или иная геометрическая конструкция, если с точки зрения современной математики она ведет в тупик и не находит в ней дальнейшего широкого развития, тратить на нее драгоценное лекционное время, конечно, нельзя.

Окончательное решение вопроса о том, что именно должно быть включено в программу, чрезвычайно ответственно и, по-видимому, может быть принято только апостериори на основе тщательного анализа нескольких лет экспериментального преподавания по меняющимся программам. Основная цель, которую ставил перед собой автор при написании этой книги, и состояла в том, чтобы обеспечить такого рода преподавание учебным пособием. На основе этой книги можно составить много разнообразных программ различной степени трудности и по-разному ориентированных.

Таким образом, настоящая книга не является учебником аналитической геометрии, следующим той или иной определенной программе. Ни при каких условиях все содержание этой книги не может быть включено в один лекционный курс: ее объем слишком велик, а содержание слишком разнообразно. Она не предназначена также ни для первоначального изучения аналитической геометрии, ни для самообразования. Наибольшую пользу от нее получит способный и любознательный студент, уже прослушавший курс аналитической геометрии и желающий углубить и систематизировать свои знания (а не заинтересованный только в сдаче экзамена). Можно надеяться также, что ее с пользой и интересом прочтут многочисленные любители геометрии.

Тем не менее, те отделы книги, в которых излагается стандартный обязательный материал, вполне доступны и «среднему» студенту. Эти отделы (отмеченные в оглавлении звездочками) изложены по возможности элементарно и со всеми подробностями. Отделы, посвященные факультативным вопросам, написаны более сжато и предъявляют к читателю бо́льшие (хотя и вполне посильные) требования.

Наиболее важные абстрактные понятия вводятся в книге постепенно, часто сначала без названия и только после того, как подробно и с различных точек зрения рассмотрены соответствующие конкретные примеры.

Грамотное изложение аналитической геометрии нуждается в простейшем алгебраическом аппарате: теории определителей и правилах действия над матрицами. Предполагается, что эти вещи читателю знакомы (их изложение было бы напрасным дублированием курса алгебры). К концу книги постепенно используются основные понятия теории групп. Предварительного знакомства с группами от читателя формально не предполагается, но, конечно, оно по меньшей мере очень желательно. (Заметим, что все, что нам нужно из теории групп, обычно с избытком излагается на первом семестре университета в курсе алгебры.)

Как уже было сказано, особое внимание в книге уделено повышению уровня строгости и формализации. Конечно, начинать курс, скажем, с аксиом, а теорию кривых второго порядка

с самого начала строить в комплексной плоскости было бы неуместно. Однако изложить в своем месте аксиоматику и дать четкое определение комплексной плоскости представляется необходимым.

В книге используется мелкий шрифт двух видов: в элементарных (в основном начальных) главах книги один из них употребляется для выделения более трудного материала, другой используется с вспомогательными целями (упражнения и т. п.).

Текст сопровождается «заданиями» (предназначенными исключительно для самоконтроля читателя) и более трудными «упражнениями» (решение которых не обязательно).

Особенностью книги является экономное использование чертежей: они помещены, как правило, только в наиболее трудных местах и в тексте книги на них ссылок нет. Предполагается, что все остальные чертежи читатель должен восстановить самостоятельно (что, в частности, может служить хорошим средством самоконтроля).

Порядок изложения в книге не следует лекционному. Многие вопросы целесообразно излагать на лекциях значительно раньше, чем это сделано в книге (а другие, наоборот, несколько позже).

Содержание книги ясно из оглавления. Обратим все же внимание на § 6 гл. 1, посвященный бивекторам и тривекторам. Обычно эти понятия изучаются в завуалированной форме под видом векторного и смешанного произведений векторов, что, в частности, полностью смазывает их аффинный характер. Необходимость в их явном введении уже давно назрела и сделалась настоятельной после того, как в «объединенном курсе» появились поливекторы. Без предварительного знакомства с бивекторами и тривекторами в трехмерном пространстве общее понятие поливектора вызывает, как правило, большие трудности. История повторяется: теперешнее сопротивление введению в курс аналитической геометрии бивекторов полностью аналогично тому, как пятьдесят лет назад сопротивлялись введению понятия вектора.

В пп. 4, 5 § 3 гл. 2 тщательно проанализировано понятие комплексной плоскости (в аффинном и евклидовом вариантах). Этот материал — один из важнейших в книге. В той или иной форме он обязательно должен быть включен в лекционный курс.

Существенным новшеством является подробное рассмотрение в § 3 гл. 4 геометрии окружностей на плоскости (пучки и связки окружностей). Хотя сама эта геометрия может быть и не очень интересна, она служит хорошим примером для иллюстрации общего принципа изоморфизма (она изоморфна геометрии точек пространства, лежащих вне некоторого параболоида вращения), а также обладает известной поучительностью ввиду дальнейшего развития некоторых из ее идей в рамках алгебраической геометрии.

В теории кривых второго порядка на проективной плоскости (§§ 2, 3 гл. 6) новшеством является изложение теории пересечений и теории пучков кривых второго порядка. Элементы этих теорий входили обязательной составной частью в курсы аналитической геометрии лет сорок назад. В связи с общим возрождением в последнее время интереса к алгебраической геометрии восстановление этого материала в курсе аналитической геометрии представляется оправданным.

Теория поверхностей второго порядка ограничена в книге, по существу, только исследованием их формы. Из теории приведения уравнений поверхностей к каноническому виду приведены без доказательств лишь формулировки теорем. Подробное изложение этой теории было бы излишним дублированием курса линейной алгебры, в рамках которого этот материал излагается сразу для пространств любой размерности.

Рукопись книги была дважды внимательно прочитана И. М. Ягломом, сделавшим большое число замечаний и предложившим еще большее число усовершенствований. Я приношу И. М. Яглому свою сердечную благодарность (хотя и не всегда следовал его советам).

Я искренне благодарен также редактору этой книги Ф. И. Кизнер за большую работу, ею проделанную, а также за выполнение всех чертежей.

М. М. Постников

27 октября 1972 г.

Глава 1

ВЕКТОРНОЕ ИСЧИСЛЕНИЕ

§ 1. ПОНЯТИЕ ВЕКТОРА

В школьном курсе математики понятие вектора вводится наглядным, не строгим образом. В этом вводном параграфе мы обсудим «школьное» определение вектора и попытаемся сделать его по возможности строгим.

1. Предварительное определение вектора

Пусть A и B — две точки (пока различные). Эти точки определяют отрезок вместе с направлением на нем (а именно, направлением от A к B), т. е. *направленный отрезок с началом A и концом B*. Этот направленный отрезок мы будем обозначать символом $\overrightarrow{AB}$.

Направленные отрезки называются также *векторами*. О векторе $\overrightarrow{AB}$ говорят, что он *отложен* от точки A (или *приложен* к этой точке).

Векторы $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ называются равными, если один из них может быть получен из другого параллельным переносом, т. е. движением, при котором вектор $\overrightarrow{AB}$ перемещается параллельно самому себе так, что точка A движется по отрезку $\overline{AA_1}$, а точка B — по отрезку $\overline{BB_1}$.

Эти определения вполне наглядны, но с логической точки зрения мало удовлетворительны, поскольку они используют понятие «движения», относящееся скорее к механике, чем к геометрии. Конечно, это понятие можно определить и чисто геометрически (что мы в дальнейшем и сделаем), но при этом оказывается, что строгое определение «движения» основывается на понятии вектора. Поэтому желательно дать определение равенства векторов, не использующее «механических» понятий.

Имея это в виду, заметим, что если равные векторы $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ расположены на различных прямых, то при описанном вы-

ше движении вектор $\overrightarrow{AB}$ заметает параллелограмм ABA_1B_1, в котором векторы $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ являются противоположными сторонами. Таким образом, *векторы $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$, не лежащие на одной прямой, равны, если четырехугольник ABA_1B_1 является параллелограммом.*

Это определение уже вполне удовлетворительно, но, к сожалению, оно неприменимо к векторам, лежащим на одной прямой. Поэтому равенство таких векторов нуждается в отдельном описании.

Такое описание легко дать: например, *векторы, лежащие на одной прямой, равны, если их длины одинаковы и они направлены в одну сторону.* Однако тогда мы получим, что равенство векторов будет по-разному определяться в случае, когда эти векторы лежат на разных прямых, и в случае, когда они лежат на одной и той же прямой. Это не очень удобно, потому что из-за этого во всех доказательствах придется рассматривать много частных случаев в зависимости от возможного расположения соответствующих векторов. Поэтому хотелось бы иметь одно общее определение, пригодное во всех случаях взаимного расположения данных векторов.

Чтобы дать такое определение, заметим, что четырехугольник тогда и только тогда является параллелограммом, когда его диагонали делятся в их точке пересечения пополам. Следовательно, векторы $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$, расположенные на различных прямых, тогда и только тогда равны, когда середины отрезков $\overline{AB_1}$ и $\overline{A_1B}$ совпадают. С другой стороны, ясно, что это верно и для векторов, расположенных на одной прямой (при $A = B_1$ серединой «отрезка» $\overline{AB_1}$, по определению, считается точка A). Тем самым мы приходим к следующему определению:

Определение 1. Векторы $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ называются *равными*, если середины отрезков $\overline{AB_1}$ и $\overline{A_1B}$ совпадают.

В дальнейшем именно это определение равенства векторов мы будем рассматривать как основное.

Замечание 1. Выше мы исключили из рассмотрения случай, когда $A = B$. В этом случае, мы, собственно говоря, уже не можем говорить об отрезке $\overline{AB}$. Однако, чтобы не допускать исключений (что в математике всегда плохо), мы условимся говорить, что при $A = B$ мы имеем дело с *вырожденным* направленным отрезком $\overrightarrow{AA}$. *Все вырожденные направленные отрезки* (приложенные в различных точках) *равны.* Действительно, при $A = B$ и $A_1 = B_1$ отрезки $\overline{AB_1}$ и $\overline{A_1B}$ совпадают, а потому и совпадают их середины. Заметим еще, что никакой невырожденный направленный отрезок $\overrightarrow{AB}$ не может быть равен вырожденному

отрезку $\overrightarrow{CC}$, поскольку середины отрезков $\overline{AC}$ и $\overline{BC}$, где A и B различны, совпадать не могут.

Вырожденный направленный отрезок называется также *нулевым вектором* и обозначается символом **0**.

2. Отношения эквивалентности

Обратим внимание на то, что равенство векторов мы ввели специальным определением. Этим равенство векторов в принципе отличается от «равенства в смысле тождественного совпадения», в специальном определении не нуждающегося.

Чтобы иметь возможность обсудить возникающую здесь ситуацию, мы напомним в этом пункте основные общематематические (логические) определения, связанные с различными видами «равенств», рассматриваемыми в математике. При этом термин «равенство» мы будем пока понимать лишь в смысле тождественного совпадения.

Общее понятие равенства тесно связано с понятием «свойства» Что такое «свойство», столь же трудно определить, как и что такое «множество». Проще всего считать его одним из основных, не определяемых, понятий математики. Однако с некоторой точки зрения его все же можно свести к понятию множества (или, точнее, подмножества).

Пусть X — произвольное множество и пусть $\mathscr{P}$ — некоторое свойство его элементов. Тогда в множестве X определено подмножество $X_{\mathscr{P}}$, состоящее из всех элементов этого множества, обладающих свойством $\mathscr{P}$. Два свойства $\mathscr{P}$ и $\mathscr{Q}$ называются *равнообъемными*, если $X_{\mathscr{P}} = X_{\mathscr{Q}}$, т. е. если элемент множества X тогда и только тогда обладает свойством $\mathscr{P}$, когда он обладает свойством $\mathscr{Q}$. Для всех практических целей равнообъемные свойства можно не различать. Поэтому в математике принято считать их совпадающими (тождественными). В силу этого соглашения, свойства элементов множества X находятся в естественном биективном [1]) соответствии с его подмножествами (именно, свойству $\mathscr{P}$ соответствует подмножество $X_{\mathscr{P}}$, а подмножеству $Y \subset X$ — свойство «принадлежать Y»).

Ситуация, когда некоторые объекты одной природы находятся в естественном биективном соответствии с некоторыми объектами другой природы, весьма часто возникает в математике (мы не будем здесь обсуждать смысл прилагательного «естественный»). Опыт конструирования математических теорий показывает, что во всех таких ситуациях объекты, естественно соответствующие друг другу, целесообразно рассматривать как одинаковые. На основании этого общего принципа отождествления мы, в частности, можем отождествлять свойства элементов множества X с подмножествами этого множества. Таким образом, с этой точки зрения свойства элементов множества — это не что иное как его подмножества.

Поскольку для любых двух различных элементов множества X всегда существует подмножество, содержащее один из этих элементов и не содер-

[1]) Термин «биективный» мы употребляем как синоним термина «взаимно однозначный».

жащее другой, мы видим, таким образом, что *два элемента множества X тогда и только тогда равны, когда все их свойства одинаковы.*

Однако весьма часто мы интересуемся только некоторыми свойствами элементов множества X, игнорируя все остальные возможные их свойства. В соответствии с этим возникает новое понятие ревенства, учитывающее лишь интересные для нас свойства. Формально это понятие может быть определено следующим образом.

Пусть $\mathfrak{P}$ — произвольный набор (или, как говорят, семейство) некоторых свойств элементов множества X. Элементы a и b множества X называются *равными по отношению к свойствам семейства* $\mathfrak{P}$, если для любого свойства $\mathscr{P}$ из семейства $\mathfrak{P}$ элемент a тогда и только тогда обладает свойством $\mathscr{P}$, когда им обладает элемент b. Если семейство $\mathfrak{P}$ содержит все свойства, то это есть равенство в смысле тождественного совпадения. Если семейство $\mathfrak{P}$ пусто, то любые два элемента множества X равны по отношению к свойствам этого семейства.

Можно сказать, что, фиксируя семейство свойств $\mathfrak{P}$, мы получаем из элементов множества X некоторые новые математические объекты, отличающиеся от этих элементов игнорированием их свойств, не принадлежащих семейству $\mathfrak{P}$. Каждый такой объект a' определяется некоторым элементом $a \in X$, причем элементы $a \in X$ и $b \in X$ тогда и только тогда определяют один и тот же объект a', когда они равны по отношению к свойствам семейства $\mathfrak{P}$.

Каждый элемент $a \in X$ определяет некоторое подмножество $\mathfrak{P}_a$, состоящее из всех элементов b, равных элементу a по отношению к свойствам семейства $\mathfrak{P}$. Ясно, что, введенные выше объекты a' находятся в естественном биективном соответствии с подмножествами $\mathfrak{P}_a$ (объект a', определяемый элементом a, соответствует подмножеству $\mathfrak{P}_a$, и обратно). В соответствии с объясненным выше принципом отождествления объекты a' отождествляются поэтому с подмножествами $\mathfrak{P}_a$.

Все это можно описать несколько иначе, если воспользоваться понятием «отношения». Наглядно, «отношение» между элементами множества X — это некоторое их «взаимное свойство», т. е. свойство пар вида (a, b), где $a \in X$, $b \in X$. Другими словами, *отношение* на множестве X — это некоторое множество R таких пар. Если $(a, b) \in R$, то пишут $a \underset{R}{\sim} b$ (или просто $a \sim b$) и говорят, что элементы a и b находятся в отношении R. Отношение называется *рефлексивным*, если $a \sim a$ для всех $a \in X$, *симметричным*, если для любых $a, b \in X$ из $a \sim b$ следует $b \sim a$, и *транзитивным*, если для любых $a, b, c \in X$ из $a \sim b$ и $b \sim c$ следует $a \sim c$. Рефлексивное, симметричное и транзитивное отношение называется *отношением эквивалентности* или просто *эквивалентностью*. Если множество X представлено в виде объединения непустых непересекающихся подмножеств, то на X возникает отношение эквивалентности, в котором $a \sim b$ тогда и только тогда, когда a и b принадлежат одному и тому же из этих подмножеств. Обратно, задание на X некоторого отношения эквивалентности определяет разбиение этого множества на непустые непересекающиеся подмножества (*классы эквивалентности*), состоящие из попарно эквивалентных элементов. Класс эквивалентности, содержащий данный элемент a, обозначается обычно символом $\{a\}$. Таким образом, $\{a\} = \{b\}$ тогда и только тогда, когда $a \sim b$. Множество всех классов эквив-

валентности называется *фактормножеством* множества X по данному отношению эквивалентности R и обозначается символом X/R.

Ясно, что для любого семейства свойств $\mathfrak{P}$ отношение «равны по отношению к свойствам семейства $\mathfrak{P}$» является эквивалентностью. При этом соответствующие классы эквивалентности совпадают с введенными выше подмножествами $\mathfrak{P}_a$. Обратно, легко видеть, что для любого отношения эквивалентности существует такое семейство свойств $\mathfrak{P}$, что элементы тогда и только тогда эквивалентны, когда они равны по отношению к свойствам семейства $\mathfrak{P}$. Таким семейством является, например, семейство свойств $\mathscr{P}_a$, занумерованных элементами множества X и определенных условием, что элемент $b \in X$ тогда и только тогда обладает свойством $\mathscr{P}_a$, когда $b \sim a$ (заметим, что свойство $\mathscr{P}_a$, рассматриваемое как подмножество множества X, является не чем иным, как классом эквивалентности $\{a\}$ элемента a).

Таким образом, каждое отношение эквивалентности на множестве X сводится к тому, что мы фиксируем внимание лишь на некоторых свойствах элементов множества X и игнорируем все другие их свойства.

Заметим, что различные семейства свойств могут определять одно и то же отношение эквивалентности. Такие «равнообъемные» семейства можно не различать.

3. Окончательное определение вектора

При определении равенства векторов в п. 1 мы пренебрегли некоторыми свойствами направленных отрезков (например, свойством иметь начало в данной точке, поскольку векторы могут быть равны и иметь начало в различных точках). Это означает, что на самом деле мы имеем здесь дело с равенством по отношению к некоторому семейству свойств, т. е. с некоторым отношением эквивалентности. Чтобы не путать это отношение с «настоящим» (тождественным) равенством, мы будем называть теперь его «эквиполлентностью». Таким образом, мы принимаем следующие определения:

Определение 1. Направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ называются *эквиполлентными,* если середины отрезков $\overline{AB_1}$ и $\overline{A_1B}$ совпадают.

Определение 2. Класс эквиполлентных направленных отрезков называется *вектором.* В частности, *нулевой вектор* — это класс, содержащий все вырожденные отрезки.

Таким образом, в теперешнем уточненном понимании термины «вектор» и «направленный отрезок» относятся к различным математическим понятиям. Поэтому, строго говоря, мы уже не можем обозначать векторы теми же символами $\overrightarrow{AB}$, что и направленные отрезки. В соответствии с общими обозначениями п. 2 вектор, содержащий направленный отрезок $\overrightarrow{AB}$, следует обозначать символом $\{\overrightarrow{AB}\}$. Равенство векторов $\{\overrightarrow{AB}\} = \{\overrightarrow{A_1B_1}\}$ (понимаемое в смысле тождественного совпадения) равносильно эквиполлентности направленных отрезков $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$.

Впрочем, оказывается удобным упростить обозначения и употреблять для вектора $\{\overrightarrow{AB}\}$ прежний символ $\overrightarrow{AB}$. Но при этом формула $\overrightarrow{AB} = \overrightarrow{A_1B_1}$ становится двусмысленной: она либо означает равенство (совпадение) направленных отрезков $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ (т. е. тот факт, что $A = A_1$ и $B = B_1$), либо равенство соответствующих векторов (эквиполлентность направленных отрезков $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$). В контексте каждый раз ясно, в каком смысле эта формула понимается, так что эта двусмысленность к недоразумениям не приводит. Конечно, если такое недоразумение возможно, необходимо сделать соответствующие оговорки или использовать более педантичное обозначение $\{\overrightarrow{AB}\} = \{\overrightarrow{A_1B_1}\}$.

В согласии с этим сокращенным обозначением термины «вектор» и «направленный отрезок» вновь можно употреблять как синонимы.

Таким образом, мы снова возвращаемся к обозначениям (и терминологии) из п. 1, но уже «на высшем уровне». Мы знаем, что они на самом деле означают, и можем каждый раз, когда это необходимо, вернуться к более четким и недвусмысленным формулировкам.

Обозначать векторы мы будем также жирными строчными буквами латинского алфавита: $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{x}$, ...

Внимательный читатель должен уже заметить, что изложенное «окончательное» определение вектора нуждается в обосновании. Действительно, чтобы иметь право говорить о классах эквиполлентных направленных отрезков, необходимо предварительно доказать следующее

Предложение 1. *Отношение эквиполлентности является на множестве всех направленных отрезков отношением эквивалентности, т. е. оно рефлексивно, симметрично и транзитивно.*

Доказательство. Рефлексивность и симметричность эквиполлентности очевидны (если $A = A_1$ и $B = B_1$, то отрезки $\overline{AB_1}$ и $\overline{A_1B}$ совпадают и потому совпадают и их середины; аналогично, если середины отрезков $\overline{AB_1}$ и $\overline{A_1B}$ совпадают, то середины отрезков $\overline{A_1B}$ и $\overline{AB_1}$ также совпадают). Для доказательства же транзитивности нам будет нужна следующая элементарно-геометрическая

Лемма. *Пусть A, B, C и D — четыре произвольные точки и пусть*

M — *середина отрезка* $\overline{AB}$,

N — *середина отрезка* $\overline{CD}$,

P — *середина отрезка* $\overline{AD}$,

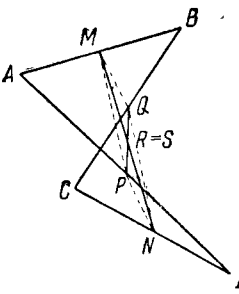
Q — *середина отрезка* $\overline{BC}$,

Тогда середины R и S отрезков $\overrightarrow{MN}$ и $\overrightarrow{PQ}$ совпадают.

Доказательство. Если $B = D$, то $\overline{AB} = \overline{AD}$ и $\overline{BC} = \overline{CD}$ Поэтому $M = P$, $N = Q$ и утверждение леммы очевидно. Случай $A = C$ рассматривается аналогично.

Пусть $B \neq D$ и $A \neq C$. Легко видеть, что тогда $M \neq P$ (ибо если $M = P$, то $B = D$). Аналогично, $N \neq Q$, $M \neq Q$ и $N \neq P$.

Поскольку $M \neq P$ и $B \neq D$, определены прямые MP и BD. Ясно, что эти прямые параллельны [1]). Действительно, если точки A, B, D не лежат на одной прямой, то MP — средняя линия треугольника ABD; если же точки A, B, D лежат на одной прямой, то точки M и P также лежат на этой прямой. Аналогично, прямая NQ также параллельна прямой BD Следовательно, прямые MP и NQ параллельны. Аналогично доказывается, что прямые MQ и NP также параллельны.



Таким образом, если точки M, N, P, Q не лежат на одной прямой, то они являются вершинами параллелограмма со сторонами $\overline{PN}$, $\overline{NQ}$, $\overline{QM}$ и $\overline{MP}$. Отрезки $\overline{MN}$ и $\overline{PQ}$, являясь диагоналями этого параллелограмма, делятся их точкой пересечения пополам. Но это и означает, что середины этих отрезков совпадают.

Пусть теперь точки M, N, P, Q принадлежат одной прямой α. В этом случае (как и в предыдущем) длина $|MP|$ отрезка $\overline{MP}$ равна половине длины $|BD|$ отрезка $\overline{BD}$. Действительно, если точка A не лежит на прямой BD, то это очевидно, поскольку отрезок $\overline{MP}$ является средней линией треугольника ABD. Если же точка A лежит на прямой BD, то для доказательства следует поочередно рассмотреть пять случаев: точка B лежит между точками A и D; точка A лежит между точками B и D; точка D лежит между точками A и B; точка A совпадает с точкой B; точка A совпадает с точкой D. В каждом из этих случаев равенство $|MP| = \frac{1}{2}|BD|$ очевидно.

Аналогично устанавливается, что $|NQ| = \frac{1}{2}|BD|$. Следовательно,

$$|MP| = |NQ|.$$

Точно так же доказывается, что имеет место и равенство

$$|MQ| = |NP|.$$

Заметим теперь, что при $M \neq N$ эти равенства однозначно характеризуют точку Q, т. е. если $|MQ'| = |NP|$, $|MP| = |NQ'|$ для некоторой точки Q' прямой α, то $Q = Q'$. В самом деле, тогда

$$|MQ| = |MQ'|, \quad |NQ| = |NQ'|$$

и, следовательно, при $Q \neq Q'$ точки M и N обе являются серединами отрезка $\overline{QQ'}$, и значит, вопреки предположению, совпадают.

Пусть теперь Q' — такая точка, что точка R является серединой отрезка PQ'. Очевидно, что точка Q' лежит на прямой α. Так как относительно точки R точки M и N, так же как и точки P и Q', симметричны, то отрезок $\overline{MP}$ симметричен отрезку $\overline{NQ'}$, а отрезок $\overline{MQ'}$ симметричен отрезку $\overline{NP}$.

---

[1]) Мы называем прямые *параллельными*, если они либо «параллельны в школьном смысле» (лежат в одной плоскости и не имеют общих точек), либо совпадают. Эта модификация терминологии оказывается очень удобной, поскольку только при таком понимании термина «параллельность» отношение параллельности прямых является отношением эквивалентности. Соответствующие классы эквивалентности называются иногда *направлениями*.

Аналогичное значение термин «параллельные» будет иметь у нас и по отношению к прямым и плоскостям в пространстве.

Поэтому

$$|MQ'| = |NP|, \quad |MP| = |NQ'|$$

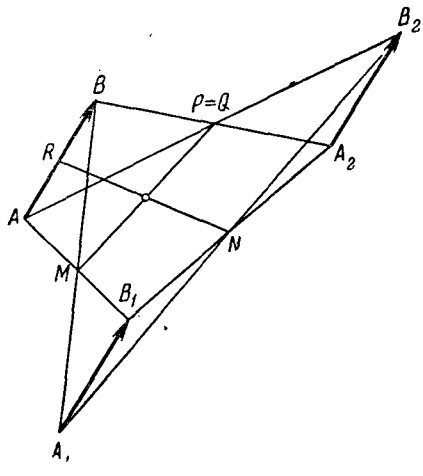
и значит, по доказанному, $Q = Q'$. Таким образом, при $M \neq N$ середина R отрезка $\overline{MN}$ является также и серединой S отрезка $\overline{PQ}$.

Случай $P \neq Q$ рассматривается аналогично.

Предположим, наконец, что $M = N$ и $P = Q$. Так как точки A и B симметричны относительно точки M, точки C и D симметричны относительно точки N и $N = M$, то отрезки $\overline{AD}$ и $\overline{BC}$ симметричны относительно точки M. Поскольку $P = Q$, это возможно только тогда, когда все точки M, N, P, Q совпадают. Поэтому середины отрезков $\overline{MN}$ и $\overline{PQ}$ также совпадают.

Тем самым лемма полностью доказана.

Пусть теперь $\overrightarrow{AB}$, $\overrightarrow{A_1B_1}$ и $\overrightarrow{A_2B_2}$ — такие направленные отрезки, что отрезок $\overrightarrow{AB}$ эквиполлентен отрезку $\overrightarrow{A_1B_1}$, а отрезок $\overrightarrow{A_1B_1}$ эквиполлентен отрезку $\overrightarrow{A_2B_2}$. Нам надо доказать, что отрезок $\overrightarrow{AB}$ эквиполлентен отрезку $\overrightarrow{A_2B_2}$, т. е. что середина P отрезка $\overrightarrow{AB_2}$ совпадает с серединой Q отрезка $\overrightarrow{A_2B}$. Рассмотрим с этой целью середину R отрезка $\overline{AB}$. Пусть M — общая середина отрезков $\overline{AB_1}$ и $\overline{A_1B}$, а N — общая середина отрезков $\overline{A_1B_2}$ и $\overline{A_2B_1}$. Применив к точкам A, B, A_2 и B_1 нашу лемму, мы немедленно получим, что середины отрезков RN и QM совпадают. Аналогично, применив эту лемму к точкам A, B, A_1 и B_2, мы получим, что совпадают середины отрезков RN и PM. Следовательно, середины отрезков QM и PM совпадают. Кроме того, эти отрезки имеют общую концевую точку M. Поэтому эти отрезки совпадают, а значит, и совпадают точки Q и P.



Тем самым предложение 1 доказано, и наше определение вектора полностью обосновано.

На достигнутом нами сейчас уровне строгости мы обязаны доказывать утверждения, наглядно вполне очевидные. Например мы должны доказать, что каждый вектор можно единственным образом отложить от произвольной точки. Другими словами, мы должны доказать следующее

Предложение 2. *Для любого направленного отрезка $\overrightarrow{AB}$ и любой точки A_1 существует одна и только одна такая точка B_1, что отрезок $\overrightarrow{A_1B_1}$ эквиполлентен отрезку $\overrightarrow{AB}$.*

Доказательство. Единственность точки B_1 очевидна: если направленные отрезки $\overrightarrow{A_1B_1}$ и $\overrightarrow{A_1B_1'}$ эквиполлентны, то середина отрезка $\overline{A_1B_1'}$ совпадает с серединой отрезка $\overline{A_1B_1}$ и потому эти отрезки (имея общую концевую точку A_1) совпадают.

Для доказательства существования точки B_1 мы рассмотрим середину C отрезка A_1B. Если $C = A$, то за точку B_1 мы можем, очевидно, принять точку C. При $C \neq A$ мы рассмотрим на прямой AC точку B_1, обладающую тем свойством, что точка C является серединой отрезка $\overline{AB_1}$. Ясно, что направленный отрезок $\overrightarrow{A_1B_1}$ эквиполлентен направленному отрезку $\overrightarrow{AB}$.

Наш анализ понятия вектора можно еще углубить, заметив, что лежащее в его основе понятие направленного отрезка определено нами без должной четкости. Действительно, в этом определении мы употребляем термин «направление», интуитивно ясный, но строго не определенный. Эту трудность можно преодолеть, заметив, что направленный отрезок $\overrightarrow{AB}$ однозначно определяется двумя точками A и B, взятыми в определенном порядке, т. е. *упорядоченной парой точек*. На этом основании мы имеем право принять следующее, уже формально безупречное

Определение 3. *Направленным отрезком* называется упорядоченная пара точек. Первая из этих точек называется *началом* отрезка, а вторая — его *концом*. Направленный отрезок с началом A и концом B обозначается символом $\overrightarrow{AB}$.

Заметим, что в этом определении вырожденные направленные отрезки ничем особым не выделяются.

Окончательный вывод. Резюмируя, мы видим, что строгим образом векторы определяются в несколько этапов:

1) Сначала (определение 3) вводятся направленные отрезки.

2) Затем (определение 1) вводится отношение эквиполлентности направленных отрезков.

3) Доказывается (предложение 1), что отношение эквиполлентности является отношением эквивалентности.

4) Наконец (определение 2), вводятся векторы.

4. Векторы на прямой, на плоскости и в пространстве

Обратим внимание, что во всем сказанном выше точки A, B и т. д. могли быть произвольными точками пространства. Чтобы подчеркнуть это обстоятельство, мы будем рассмотренные выше векторы называть *векторами в пространстве*. Множество всех векторов в пространстве будем обозначать символом $\mathrm{Vect}(3)$.

С другой стороны, мы могли бы ограничиться рассмотрением лишь точек, принадлежащих некоторой фиксированной плоскости Π. Получающиеся в этом случае векторы называются

векторами на плоскости. Их множество мы будем обозначать символом $\mathrm{Vect}(2)$ (или символом $\mathrm{Vect}_\Pi(2)$, когда нужно подчеркнуть его зависимость от плоскости Π).

Аналогично, можно ограничиться рассмотрением лишь точек некоторой фиксированной прямой α и получить *векторы на прямой.* Их множество мы будем обозначать символом $\mathrm{Vect}(1)$ (или символом $\mathrm{Vect}_\alpha(1)$).

Если плоскость Π рассматривать независимо от объемлющего пространства (как это делается в планиметрии), то множества $\mathrm{Vect}(3)$ и $\mathrm{Vect}(2)$ формально между собой никак не связаны. Напротив, если считать, что плоскость Π определенным образом вложена в пространство, то каждый направленный отрезок плоскости одновременно будет и направленным отрезком в пространстве. В плоскости этот отрезок определяет некоторый вектор $\boldsymbol{a}$, а в пространстве — некоторый вектор $\boldsymbol{a}'$ (векторы $\boldsymbol{a}$ и $\boldsymbol{a}'$ различны хотя бы потому, что среди направленных отрезков, принадлежащих вектору $\boldsymbol{a}'$, заведомо имеются отрезки, не принадлежащие вектору $\boldsymbol{a}$). С другой стороны, ясно, что если на плоскости направленные отрезки эквиполлентны, то они эквиполлентны и как направленные отрезки в пространстве. Это означает, что вектор $\boldsymbol{a}'$ зависит только от вектора $\boldsymbol{a}$ и не зависит от того, от какого направленного отрезка, принадлежащего вектору $\boldsymbol{a}$, мы отправляемся. Тем самым возникает естественное биективное соответствие $\boldsymbol{a} \leftrightarrow \boldsymbol{a}'$ между множеством $\mathrm{Vect}_\Pi(2)$ и некоторым подмножеством множества $\mathrm{Vect}(3)$. Поэтому (в силу общего принципа отождествления; см. п. 2) мы можем векторы $\boldsymbol{a}$ и $\boldsymbol{a}'$ отождествлять и тем самым *множество* $\mathrm{Vect}_\Pi(2)$ *считать подмножеством множества* $\mathrm{Vect}(3)$. Аналогично, для любой прямой α множество $\mathrm{Vect}_\alpha(1)$ можно считать подмножеством множества $\mathrm{Vect}(3)$, а также и подмножеством множества $\mathrm{Vect}_\Pi(2)$ (в случае, когда прямая α принадлежит плоскости Π). Эти вложения, очевидно, *согласованы* в том смысле, что если вектор $\boldsymbol{a} \in \mathrm{Vect}_\alpha(1)$ рассматривать как вектор из $\mathrm{Vect}_\Pi(2)$, а потом как вектор из $\mathrm{Vect}(3)$, то получится тот же самый элемент множества $\mathrm{Vect}(3)$, как если вектор $\boldsymbol{a}$ с самого начала рассматривать как вектор из $\mathrm{Vect}(3)$.

Определение 1. Направленный отрезок $\overrightarrow{AB}$ называется *параллельным* плоскости Π, если прямая, на которой он расположен, параллельна этой плоскости. Вырожденный направленный отрезок, по определению, параллелен любой плоскости.

Ясно, что если направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ эквиполлентны и отрезок $\overrightarrow{AB}$ параллелен плоскости Π, то отрезок $\overrightarrow{A_1B_1}$ также параллелен этой плоскости [1]). Другими словами, для лю-

[1]) Обратим внимание, что это верно только при обобщенном понимании термина «параллельны», — см. сноску на стр. 17.

бого вектора a (в пространстве) либо все направленные отрезки, составляющие этот вектор, параллельны плоскости П, либо ни один из них не параллелен этой плоскости. Когда имеет место первый случай, мы будем говорить, что вектор a *параллелен* плоскости П.

Очевидно, что

вектор a *тогда и только тогда параллелен плоскости* П, *когда, отложив его от произвольной точки плоскости* П, *мы получим направленный отрезок, расположенный в этой плоскости.*

Это показывает, что

множество всех векторов в пространстве, параллельных плоскости П, *совпадает с введенным выше подмножеством* $\mathrm{Vect}_{\Pi}(2)$ *множества* $\mathrm{Vect}(3)$.

Аналогично определяется понятие вектора (в плоскости или в пространстве), *параллельного данной прямой* α. Множество всех таких векторов совпадает с множеством $\mathrm{Vect}_{\alpha}(1)$ (рассматриваемым как подмножество соответственно множеств $\mathrm{Vect}(2)$ или $\mathrm{Vect}(3)$).

Подмножества $\mathrm{Vect}_{\Pi}(2)$ и $\mathrm{Vect}_{\Pi'}(2)$ множества $\mathrm{Vect}(3)$, отвечающие двум различным плоскостям П и П′, вообще говоря, различны (они совпадают тогда и только тогда, когда плоскости П и П′ параллельны). Их пересечением (при непараллельных плоскостях П и П′) является множество $\mathrm{Vect}_{\alpha}(1)$, где α — прямая пересечения плоскостей П и П′.

Аналогично, подмножества $\mathrm{Vect}_{\alpha}(1)$ и $\mathrm{Vect}_{\alpha'}(1)$, отвечающие двум прямым α и α', тогда и только тогда совпадают, когда прямые α и α' параллельны. В противном случае эти подмножества пересекаются лишь по нулевому вектору $\mathbf{0}$.

§ 2. ВЕКТОРЫ НА ПРЯМОЙ

Нашей основной целью является исследование векторов на плоскости и в пространстве. Однако оказывается, что для этого необходимо изучить предварительно векторы на прямой. Этому изучению и будет посвящен настоящий параграф. Результаты этого изучения послужат нам также образцом для обобщения на случай плоскости и пространства.

Во всем этом параграфе мы будем считать фиксированной некоторую прямую α и все построения будем относить только к этой прямой. Когда нам придется рассматривать и другие прямые, это будет специально оговорено.

1. Ориентации прямой

На прямой α мы можем перемещаться в одном из двух противоположных направлений. Выбор одного из этих направлений мы будем называть *ориентацией* прямой, а прямую с заданной на ней ориентацией мы будем называть *ориентированной прямой* или *осью.*

Здесь мы опять пользуемся наглядным, но не строгим (пока) понятием движения. Чтобы ввести понятие ориентации вполне строгим образом, нам понадобится общематематическое понятие «упорядоченного множества».

Пусть X — произвольное множество (с элементами A, B, C...). Заданное на X отношение (обозначаемое символом $\prec$) называется *отношением порядка*, когда оно обладает следующими свойствами:

1) если $A \prec B$, то $A \neq B$ (свойство антирефлексивности),

2) если $A \prec B$ и $B \prec C$, то $A \prec C$ (свойство транзитивности),

3) если $A \neq B$, то либо $A \prec B$, либо $B \prec A$ (свойство полноты).

Отношение порядка называется *отношением предшествования*, когда выполнено следующее дополнительное свойство:

4) если $A \prec B$, то существуют такие элементы C_1, C_2 и D, что

$$C_1 \prec A \prec D \prec B \prec C_2.$$

По любому отношению порядка $\prec$ мы можем построить новое отношение порядка $\prec'$ (называемое отношением, *противоположным* данному), считая, по определению, что

$$A \prec' B \text{ тогда и только тогда, когда } B \prec A.$$

Ясно, что если $\prec$ — отношение предшествования, то $\prec'$ — также отношение предшествования.

Мы принимаем в качестве аксиомы следующее утверждение.

У. *На любой прямой существуют два взаимно противоположных отношения предшествования.*

Наглядно, каждое из этих отношений определяется заданием на прямой некоторого направления движения. Именно, $A \prec B$ в одном из этих отношений, если точка B получается из точки A движением в данном направлении.

Теперь мы уже можем дать вполне строгое определение.

Определение 1. Каждое из предусмотренных аксиомой У отношений предшествования называется *ориентацией* прямой. Прямая с заданной на ней ориентацией называется *ориентированной прямой* или *осью*.

Другие понятия, относящиеся к взаимному расположению точек на прямой, могут быть определены через понятие ориентации. Например, точка C называется *лежащей между* точками A и B, если $A \prec C \prec B$ в той ориентации, в которой $A \prec B$. Множество точек, лежащих между точками A и B, пополненное самими этими точками, называется *отрезком* $\overline{AB}$ с концевыми точками A и B. Согласно свойству 4) отношений предшествования отрезок $\overline{AB}$ при $A \neq B$ содержит точки, отличные от точек A и B (*внутренние точки*), и, как можно легко показать, содержит даже бесконечно много таких точек. Поскольку существуют только две ориентации и эти ориентации противоположны, каждая точка, лежащая между точками A и B, лежит также между точками B и A, т. е. $\overline{AB} = \overline{BA}$, и т. д.

Определение 2. Невырожденные направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ прямой α называются *одноименными*, если $A \prec B$ в некоторой ориентации тогда и только тогда, когда $A_1 \prec B_1$ в той же ориентации. Ясно, что это определение корректно (не зависит от выбора ориентации).

Очевидно, что отношение одноименности рефлексивно и симметрично. Кроме того, из свойства 2) отношений порядка легко вытекает, что оно транзитивно. Таким образом,

отношение одноименности направленных отрезков прямой является отношением эквивалентности.

Ясно, что соответствующие классы эквивалентности находятся в естественном биективном соответствии с ориентациями прямой: класс, соответствующий данной ориентации, состоит из всех направленных отрезков $\overrightarrow{AB}$, для которых $A \prec B$ в этой ориентации. Поэтому

ориентации на прямой можно отождествить с классами одноименных направленных отрезков.

Чтобы получить аналогичные результаты для векторов, необходимо предварительно доказать следующее

Предложение 1. *Эквиполлентные невырожденные направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ прямой α одноименны.*

Доказательство этого предложения сводится к проверке его справедливости во всех возможных случаях взаимного расположения точек A, B, A_1 и B_1. На наглядно геометрическом уровне эта проверка совершенно очевидна: достаточно сделать чертеж.

Чтобы привести его формальное доказательство, мы примем в качестве аксиом следующие свойства середины отрезка:

С1. *Для любых точек A и B середина C отрезка $\overline{AB}$ принадлежит этому отрезку и при $A \neq B$ является его внутренней точкой.*

Иными словами, если $A = B$, то $C = A = B$. Если же $A \neq B$, то $A \prec C \prec B$ в ориентации, в которой $A \prec B$.

С2. *Пусть точка C является серединой отрезков $\overline{AB}$ и $\overline{A_1B_1}$. Тогда если $A \prec A_1 \prec C$, то $C \prec B_1 \prec B$.*

В аксиоме С2 имеется в виду любая из двух возможных ориентаций.

С3. *Для любых точек A и C существует одна и только одна точка B, обладающая тем свойством, что точка C является серединой отрезка $\overline{AB}$.*

Предложение 1 из этих аксиом выводится следующим образом.

Пусть C — общая середина отрезков $\overline{AB_1}$ и $\overline{A_1B}$.

Случай 1. $C \prec A$ (в ориентации, в которой $A \prec B$). В этом случае $C \prec A \prec B$ и потому к отрезкам $\overline{BA_1}$ и $\overline{AB_1}$ применима аксиома С2 (по отношению к противоположной ориентации). Следовательно, согласно этой аксиоме $A_1 \prec B_1 \prec C$. В частности, $A_1 \prec B_1$.

Случай 2. $C = A$. Тогда $A = B_1$ (и $C = B_1$), но $A_1 \neq B$ (ибо в противном случае оба отрезка $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ были бы вырождены). Кроме того,

легко видеть, что $A_1 \prec B$ (в ориентации, в которой, $A \prec B$). Действительно, если $B \prec A_1$, то $B \prec C \prec A_1$ и потому, вопреки выбору рассматриваемой ориентации, $B \prec A (= C)$. Но если $A_1 \prec B$, то $A_1 \prec C \prec B$, т. е. $A_1 \prec B_1 \prec B$ и, в частности, $A_1 \prec B_1$.

Случай 3а. $A \prec C$ и $A_1 \prec A$. В этом случае аксиома С2 применима к отрезкам $\overline{A_1B}$, $\overline{AB_1}$ и потому $C \prec B \prec B_1$, т. е. $A_1 \prec A \prec C \prec B_1$. Следовательно, $A_1 \prec B_1$.

Случай 3б. $A \prec C$ и $A_1 = A$. В этом случае $B_1 = B$, и наше утверждение тривиально.

Случай 3в′. $A \prec C$, $A \prec A_1$ и $A_1 \prec C$. В этом случае аксиома С2 применима к отрезкам $\overline{AB_1}$, $\overline{A_1B}$ и потому $C \prec B \prec B_1$, т. е. $A_1 \prec C \prec B_1$, и снова $A_1 \prec B_1$.

Случай 3в″. $A \prec C$, $A \prec A_1$ и $A_1 = C$. Тогда $A_1 = B$, $C = B$ и $A \prec B_1$ (ср. случай 2). Поэтому $A \prec C \prec B_1$, т. е. $A_1 \prec B_1$.

Случай 3в‴. $A \prec C$, $A \prec A_1$ и $C \prec A_1$. Поскольку $C \prec A_1$, то обязательно $B \prec C$ (аксиома С1). Следовательно, $A \prec B \prec C$ и аксиома С2 применима к отрезкам $\overline{AB_1}$ и $\overline{BA_1}$. Поэтому $C \prec A_1 \prec B_1$ и, в частности, $A_1 \prec B_1$.

Теперь мы можем сформулировать наше основное

Определение 3. Отличные от нуля векторы $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ прямой α называются *одноименными*, если одноименны направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$, т. е. если $A_1 \prec B_1$ в ориентации, в которой $A \prec B$.

Согласно предложению 1 это определение корректно.

Ясно, что, так же как и отношение одноименности направленных отрезков,

отношение одноименности отличных от нуля векторов прямой является отношением эквивалентности.

Отношение одноименности разбивает все множество отличных от нуля векторов прямой α на два класса, находящихся в естественном биективном соответствии с ориентациями этой прямой. Поэтому

ориентации прямой можно отождествить с классами одноименных отличных от нуля векторов этой прямой.

Определение 4. Каждый отличный от нуля вектор прямой α называется *базисом* на этой прямой (или, точнее, базисом соответствующего множества $\mathrm{Vect}_\alpha(1)$).

Таким образом, мы можем сказать, что

ориентациями прямой являются классы одноименных базисов на этой прямой.

Базис $\overrightarrow{AB}$ называется *положительно ориентированным* (по отношению к данной ориентации прямой α), если он определяет эту ориентацию, т. е. если $A \prec B$ в данной ориентации. В противном случае базис $\overrightarrow{AB}$ называется *отрицательно ориентированным.*

2. Длина и величина вектора на прямой

Определение 1. *Длиной* $|\overrightarrow{AB}|$ направленного отрезка $\overrightarrow{AB}$ мы будем называть длину $|AB|$ соответствующего ненаправленного отрезка $\overline{AB}$. (Здесь мы не будем вдаваться в подробности строгого, аксиоматического, определения длины, поскольку это увело бы нас далеко в сторону; напомним только, что при определении длины предполагается выбранным некоторый фиксированный отрезок — *эталон длины*, длина которого, по определению, равна единице.)

Нам будут нужны только следующие свойства длины, которые можно считать аксиомами:

Д1. *Если* C *— середина отрезка* $\overline{AB}$, *то* $|AC| = |CB|$.
Д2. *Если* $A < C < B$, *то* $|AB| = |AC| + |CB|$.

Величиной (или *алгебраическим значением*) направленного отрезка $\overrightarrow{AB}$ на ориентированной прямой α мы будем называть его длину, если этот отрезок ориентирован положительно, и его длину, взятую со знаком минус, если этот отрезок ориентирован отрицательно. Величину отрезка $\overrightarrow{AB}$ мы будем обозначать символом (AB). При $A = B$ (и только при $A = B$) величина (AB) равна нулю. Ясно, что

$$(AB) + (BA) = 0.$$

Замечание 1. Обратим внимание на то, что о длине направленного отрезка имеет смысл говорить и в случае отрезков не на данной прямой, а в плоскости или в пространстве, тогда как величина направленного отрезка определена лишь для направленных отрезков прямой и притом ориентированной.

Предложение 1. *Величины эквиполлентных направленных отрезков* $\overrightarrow{AB}$ *и* $\overrightarrow{A_1B_1}$ *совпадают.*

Доказательство. Ясно, что нам достаточно рассмотреть только случай, когда отрезки $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ невырождены. Но в этом случае совпадение знаков величин направленных отрезков $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ непосредственно вытекает из предложения 1 п. 1. Поэтому нам нужно только доказать, что совпадают их длины. Пусть C — общая середина отрезков $\overline{AB_1}$ и $\overline{A_1B}$. Тогда $|AC| = |CB_1|$ и $|A_1C| = |CB|$. Ясно, что без ограничения общности мы можем предполагать, что $A < A_1$ в ориентации, в которой $A < B$ (при $A = A_1$ доказывать нечего, а при $A_1 < A$ достаточно поменять местами отрезки $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$, чтобы свести этот случай к случаю, когда $A < A_1$). Но при $A < A_1$ возможны лишь следующие два случая:

Случай 1. $A < A_1 < C < B < B_1$. В этом случае $|AB| = |AC| + |CB|$ и $|A_1B_1| = |A_1C| + |CB_1|$. Следовательно, $|AB| - |A_1B_1| = (|AC| - |CB_1|) + (|CB| - |A_1C|) = 0 + 0 = 0$, т. е. $|AB| = |A_1B_1|$.

Случай 2. $A < B < C < A_1 < B_1$. В этом случае $|AB| = |AC| + |CB|$ и $|A_1B_1| = |A_1C| + |CB_1|$. Следовательно, $|AB| - |A_1B_1| = (|AC| - |CB_1|) + (|CA_1| - |BC|) = 0$, т. е. снова $|AB| = |A_1B_1|$.

Определение 2. *Величиной* (соответственно, *длиной*) вектора $a = \{\overrightarrow{AB}\}$ называется величина (соответственно длина) направленного отрезка $\overrightarrow{AB}$.

Согласно предложению 1 это определение корректно.

Величина вектора a обозначается символом (a), а длина — символом $|a|$.

Замечание 2. Легко видеть, что векторы a и b (на прямой α) тогда и только тогда равны, когда их величины совпадают:

$$a = b \Leftrightarrow (a) = (b)$$

(т. е. когда эти векторы одноименны и имеют равные длины). Это показывает, что, сопоставив каждому вектору $a \in \mathrm{Vect}(1)$ его величину (a), мы получим биективное соответствие между множеством $\mathrm{Vect}(1)$ всех векторов на прямой и полем $\mathbb{R}$ вещественных чисел. Однако было бы неосторожно воспользоваться этим соответствием и отождествить множества $\mathrm{Vect}(1)$ и $\mathbb{R}$, поскольку это соответствие свойством естественности не обладает (оно зависит и от ориентации прямой и от выбора эталона длины).

3. Отношение векторов на прямой

Пусть $e = \overrightarrow{A_0B_0}$ — произвольный отличный от нуля вектор прямой α (базис). Приняв отрезок $\overrightarrow{A_0B_0}$ за эталон длины и выбрав ориентацию прямой α так, чтобы вектор e был положительно ориентирован, мы согласно сказанному выше для любого вектора $a = \overrightarrow{AB}$ можем говорить о его величине (a) (в частности, $(e) = 1$).

Определение 1. Величина (a) вектора a, получающаяся при указанном выборе эталона длины и ориентации, называется *отношением* векторов a и e и обозначается символом $a : e$ (или $\overrightarrow{AB} : \overrightarrow{A_0B_0}$).

При произвольном эталоне длины число $a : e$ равно, очевидно, отношению величин (a) и (e):

$$a : e = (a) : (e).$$

При этом ясно, что эта формула сохраняет свою силу и тогда, когда мы сменим ориентацию прямой α на противоположную. Следовательно, отношение $a : e = (a) : (e)$ не зависит ни от выбора эталона длины, ни от выбора ориентации прямой α (хотя от этого зависят величины (a) и (e)), т. е. определяется исключительно векторами a и e.

Очевидно, что

а) *если* $\boldsymbol{a} : \boldsymbol{e} = \boldsymbol{b} : \boldsymbol{e}$, *то* $\boldsymbol{a} = \boldsymbol{b}$;

б) *для любого вектора* $\boldsymbol{e} \neq \mathbf{0}$ *и любого числа* $a \in \mathbb{R}$ *существует такой вектор* $\boldsymbol{a}$, *что* $\boldsymbol{a} : \boldsymbol{e} = a$.

Определение 2. Если $\boldsymbol{a} : \boldsymbol{e} = a$, то мы будем говорить, что вектор $\boldsymbol{a}$ получен *умножением* вектора $\boldsymbol{e}$ на число a и будем этот вектор обозначать символом $a\boldsymbol{e}$.

Свойства а) и б) означают, что операция умножения (отличных от нуля) векторов на вещественные числа однозначно определена, т. е. для любого вектора $\boldsymbol{e} \neq \mathbf{0}$ и любого числа $a \in \mathbb{R}$ существует один и только один вектор $\boldsymbol{a}$, обладающий тем свойством, что $\boldsymbol{a} = a\boldsymbol{e}$.

Мы доопределим эту операцию при $\boldsymbol{e} = \mathbf{0}$, полагая, по определению,

$$a\mathbf{0} = \mathbf{0}$$

для любого числа $a \in \mathbb{R}$.

Ясно, что

$$(a\boldsymbol{e}) = a\,(\boldsymbol{e}).$$

В частности,

$$|\,a\boldsymbol{e}\,| = |\,a\,| \cdot |\,\boldsymbol{e}\,|.$$

Заметим, что при $a > 0$ (и при $\boldsymbol{e} \neq \mathbf{0}$) векторы $\boldsymbol{e}$ и $a\boldsymbol{e}$ одноименны, а при $a < 0$ — разноименны.

Отметим, что для любого вектора $\boldsymbol{a}$

$$0\boldsymbol{a} = \mathbf{0}, \quad 1\boldsymbol{a} = \boldsymbol{a}.$$

Вектор $(-1)\boldsymbol{e}$ обозначается также символом $-\boldsymbol{e}$ и называется вектором, *противоположным* вектору $\boldsymbol{e}$. Очевидно, что если $\boldsymbol{e} = \overrightarrow{A_0B_0}$, то $-\boldsymbol{e} = \overrightarrow{B_0A_0}$.

Замечание 1. Определение вектора $-\boldsymbol{e}$ как вектора $\overrightarrow{B_0A_0}$ имеет смысл и для векторов на плоскости или в пространстве [1]), тогда как вектор $(-1)\boldsymbol{e}$ определен у нас пока лишь для векторов на прямой.

Определение 3. *Координатой* вектора $\boldsymbol{a}$ в базисе $\boldsymbol{e}$ называется отношение $\boldsymbol{a} : \boldsymbol{e}$. Другими словами, координата вектора $\boldsymbol{a}$ в базисе $\boldsymbol{e}$ — это такое число a, что $\boldsymbol{a} = a\boldsymbol{e}$, или, по-другому, координата вектора $\boldsymbol{a}$ в базисе $\boldsymbol{e}$ — это величина $(\boldsymbol{a})$ вектора $\boldsymbol{a}$ при условии, что ориентация и эталон длины на прямой выбраны так, что $(\boldsymbol{e}) = 1$.

Сопоставление вектору $\boldsymbol{a}$ его координаты $a = \boldsymbol{a} : \boldsymbol{e}$ определяет биективное соответствие между множеством $\mathrm{Vect}(1)$ векторов на прямой и полем $\mathbb{R}$ вещественных чисел. Это соответствие зависит от вектора $\boldsymbol{e}$ и потому не является естественным.

---

[1]) Конечно, здесь следует проверить корректность, т. е. тот факт, что если направленные отрезки $\overrightarrow{A_0B_0}$ и $\overrightarrow{A_1B_1}$ эквиполлентны, то направленные отрезки $\overrightarrow{B_0A_0}$ и $\overrightarrow{B_1A_1}$ также эквиполлентны; но это очевидно.

Оно по существу совпадает с соответствием, описанным в замечании 2 п. 2, поскольку задание вектора $e = \overrightarrow{A_0B_0}$ равносильно заданию эталона длины $\overline{A_0B_0}$ и ориентации прямой α (а именно, ориентации, в которой вектор $\overrightarrow{A_0B_0}$ положительно ориентирован).

Легко видеть, что

для любых трех векторов $\boldsymbol{a} \neq \mathbf{0}$, $\boldsymbol{b} \neq \mathbf{0}$ *и* $\boldsymbol{c}$ *имеет место равенство*

$$\boldsymbol{c} : \boldsymbol{a} = (\boldsymbol{c} : \boldsymbol{b}) \cdot (\boldsymbol{b} : \boldsymbol{a}). \tag{1}$$

Действительно, по определению, $\boldsymbol{c} : \boldsymbol{a} = (\boldsymbol{c}) : (\boldsymbol{a})$, $\boldsymbol{c} : \boldsymbol{b} = (\boldsymbol{c}) : (\boldsymbol{b})$ и $\boldsymbol{b} : \boldsymbol{a} = (\boldsymbol{b}) : (\boldsymbol{a})$. Поэтому

$$\boldsymbol{c} : \boldsymbol{a} = \frac{(\boldsymbol{c})}{(\boldsymbol{a})} = \frac{(\boldsymbol{c})}{(\boldsymbol{b})} \cdot \frac{(\boldsymbol{b})}{(\boldsymbol{a})} = (\boldsymbol{c} : \boldsymbol{b}) \cdot (\boldsymbol{b} : \boldsymbol{a}).$$

Пусть $\boldsymbol{c} : \boldsymbol{b} = k$ и $\boldsymbol{b} : \boldsymbol{a} = l$, т. е. пусть $\boldsymbol{c} = k\boldsymbol{b}$ и $\boldsymbol{b} = l\boldsymbol{a}$. Тогда по доказанному $\boldsymbol{c} : \boldsymbol{a} = kl$, т. е. $\boldsymbol{c} = (kl)\,\boldsymbol{a}$. В этой форме равенство (1) остается, очевидно, верным как при $\boldsymbol{a} = \mathbf{0}$, так и при $\boldsymbol{b} = \mathbf{0}$. Таким образом,

для любого вектора $\boldsymbol{a}$ *и любых чисел* k *и* l *имеет место равенство*

$$(kl)\,\boldsymbol{a} = k\,(l\boldsymbol{a}) \tag{2}$$

Пусть, в частности, $\boldsymbol{a}$ есть базис $\boldsymbol{e}$. Обозначая вектор $\boldsymbol{b}$ (освободившимся теперь) символом $\boldsymbol{a}$ и замечая, что число l является не чем иным, как координатой a вектора $\boldsymbol{a}$ в базисе $\boldsymbol{e}$, а число kl — координатой вектора $\boldsymbol{c} = k\boldsymbol{a}$, мы получаем, таким образом, что

если в базисе $\boldsymbol{e}$ *вектор* $\boldsymbol{a}$ *имеет координату* a, *то для любого числа* k *вектор* $k\boldsymbol{a}$ *будет иметь координату* ka.

Другими словами,

при умножении вектора на число его координата умножается на это же число.

Формулу (2) (или, скорее, формулу (1)) можно прочитать и по-иному, приняв за $\boldsymbol{b}$ некоторый базис $\boldsymbol{e}$, а за $\boldsymbol{a}$ — некоторый другой базис $\boldsymbol{e}'$. Обозначая вектор $\boldsymbol{c}$ символом $\boldsymbol{x}$, мы можем теперь переписать формулу (1) (переставив множители) в следующем виде:

$$\boldsymbol{x} : \boldsymbol{e}' = (\boldsymbol{e} : \boldsymbol{e}')\,(\boldsymbol{x} : \boldsymbol{e}).$$

Но число $\boldsymbol{x} : \boldsymbol{e}'$ является координатой x' вектора $\boldsymbol{x}$ в базисе $\boldsymbol{e}'$, а число $\boldsymbol{x} : \boldsymbol{e}$ — его координатой x в базисе $\boldsymbol{e}$. Таким образом,

при замене базиса $\boldsymbol{e}$ *базисом* $\boldsymbol{e}'$ *координата* x *произвольного вектора* $\boldsymbol{x}$ *заменяется координатой* x', *связанной с координатой* x *формулой*

$$x' = cx,$$

где $c = e : e'$ *— координата «старого» базиса* e *в «новом» базисе* e'.

Обратим внимание на то, что все последние утверждения являются просто тривиальными переформулировками одного и того же геометрического факта.

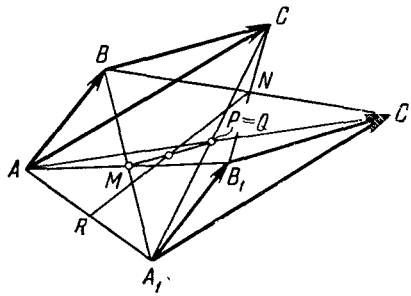
4. Сложение векторов на прямой. Лемма Шаля

Пусть $a = \overrightarrow{AB}$ и $b = \overrightarrow{A'B'}$ — произвольные векторы. Отложив вектор b от точки B, т. е. построив направленный отрезок $\overrightarrow{BC}$, эквиполлентный отрезку $\overrightarrow{A'B'}$, мы получим некоторую точку C.

Определение 1. Вектор $\overrightarrow{AC}$ называется *суммой* векторов $a = \overrightarrow{AB}$ и $b = \overrightarrow{BC}$ и обозначается символом $a + b$.

Это определение нуждается конечно, в проверке корректности, т. е. в доказательстве того факта, что вектор $a + b$ не зависит от имеющегося в этом определении произвола (состоящего в выборе точки A). Другими словами, нам нужно доказать следующее

Предложение 1. *Если направленный отрезок* $\overrightarrow{AB}$ *эквиполлентен направленному отрезку* $\overrightarrow{A_1B_1}$, *а направленный отрезок* $\overrightarrow{BC}$ *эквиполлентен направленному отрезку* $\overrightarrow{B_1C_1}$, *то направленный отрезок* $\overrightarrow{AC}$ *эквиполлентен направленному отрезку* $\overrightarrow{A_1C_1}$.



Доказательство. Пусть P и Q — середины отрезков $\overline{AC_1}$ и $\overline{A_1C}$. Мы должны показать, что $P = Q$. Для этого мы воспользуемся леммой из п. 3 § 1.

Пусть M — общая середина отрезков $\overline{AB_1}$ и $\overline{A_1B}$, а N — общая середина отрезков $\overline{BC_1}$ и $\overline{B_1C}$. Кроме того, пусть R — середина отрезка $\overline{AA_1}$. Применив указанную лемму к точкам A, B_1, C, A_1, мы немедленно получим, что середины отрезков $\overline{MQ}$ и $\overline{NR}$ совпадают. Аналогично, применив эту лемму к точкам A, C_1, B, A_1, мы получим, что совпадают середины отрезков $\overline{MP}$ и $\overline{NR}$. Следовательно, середины отрезков $\overline{MQ}$ и $\overline{MP}$ также совпадают. Имея общую середину и общую концевую точку, эти отрезки совпадают. Поэтому $P = Q$.

Замечание 1. Данное нами определение суммы векторов пригодно не только для векторов на прямой, но и для векторов в плоскости и в пространстве. То же самое справедливо, конечно, и по отношению к доказательству его корректности.

Покажем, что для суммы векторов (на прямой) имеет место следующее свойство дистрибутивности:

для любых трех векторов $\boldsymbol{a}$, $\boldsymbol{b}$ *и* $\boldsymbol{e} \neq 0$ *справедливо равенство*

$$(\boldsymbol{a} + \boldsymbol{b}) : \boldsymbol{e} = \boldsymbol{a} : \boldsymbol{e} + \boldsymbol{b} : \boldsymbol{e}. \qquad (1)$$

Другими словами,

координата суммы векторов (*в произвольном базисе* $\boldsymbol{e}$) *равна сумме координат слагаемых* (*в том же базисе*).

Ясно, что для доказательства этого соотношения достаточно доказать, что

$$(\boldsymbol{a} + \boldsymbol{b}) = (\boldsymbol{a}) + (\boldsymbol{b}),$$

т. е. что величина суммы двух векторов равна сумме величин слагаемых (конечно, в предположении, что прямая α ориентирована и на ней выбран эталон длины; заметим, что формула (1) в этом предположении не нуждалась).

Пусть $\boldsymbol{a} = \overrightarrow{AB}$ и $\boldsymbol{b} = \overrightarrow{BC}$, так что $\boldsymbol{a} + \boldsymbol{b} = \overrightarrow{AC}$. Тогда последнее равенство может быть переписано в следующем виде:

$$(AC) = (AB) + (BC). \qquad (2)$$

Таким образом, нам нужно показать, что

для любых трех точек A, B *и* C *прямой* α *имеет место равенство* (2).

В этой форме свойство дистрибутивности известно как лемма Шаля.

Если хотя бы две из точек A, B и C совпадают, то лемма Шаля очевидна (она сводится либо к тождеству $(AB) = (AB)$, либо к уже известному нам равенству $(AB) + (BA) = 0$). Пусть все точки A, B, C попарно различны. Ясно, что без ограничения общности мы можем предполагать, что $A < C$ (в заданной на прямой α ориентации), так как в противном случае достаточно сменить на α ориентацию, отчего обе стороны равенства (2) умножатся на -1.

Случай 1. $B < A < C$. В этом случае $(AC) = |AC| = |BC| - |BA|$, тогда как $(BC) = |BC|$ и $(AB) = -|BA|$. Следовательно, $(AC) = (AB) + (BC)$.

Случай 2. $A < B < C$. В этом случае, $(AC) = |AC| = |AB| + |BC|$, тогда как $(AB) = |AB|$ и $(BC) = |BC|$. Следовательно, $(AC) = (AB) + (BC)$.

Случай 3. $A < C < B$. В этом случае $(AC) = |AC| = |AB| - |CB|$, тогда как $(AB) = |AB|$ и $(BC) = -|CB|$. Следовательно, $(AC) = (AB) + (BC)$.

Тем самым лемма Шаля полностью доказана.

5. Алгебраические свойства линейных операций

Введенные операции над векторами (умножение на числа и сложение) называются *линейными*. Их основные алгебраические свойства перечислены в следующей теореме:

Теорема 1. *Для любых векторов* $\boldsymbol{a}$, $\boldsymbol{b}$ *и* $\boldsymbol{c}$ *и любых вещественных чисел* k *и* l *имеют место равенства*

1. $(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}=\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c})$,
2. $\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{b}+\boldsymbol{a}$,
3. $\boldsymbol{a}+0=\boldsymbol{a}$,
4. $\boldsymbol{a}+(-\boldsymbol{a})=0$,
5. $(k+l)\,\boldsymbol{a}=k\boldsymbol{a}+l\boldsymbol{a}$,
6. $(kl)\,\boldsymbol{a}=k\,(l\boldsymbol{a})$,
7. $k\,(\boldsymbol{a}+\boldsymbol{b})=k\boldsymbol{a}+k\boldsymbol{b}$,
8. $1\boldsymbol{a}=\boldsymbol{a}$.

Доказательство. 1. Отложим от произвольной точки A вектор $\boldsymbol{a}$, от его конца B — вектор $\boldsymbol{b}$, а от конца C вектора $\boldsymbol{b}$ — вектор $\boldsymbol{c}$. Пусть D — конец этого вектора. Согласно определению направленный отрезок $\overrightarrow{AC}$ определяет вектор $\boldsymbol{a}+\boldsymbol{b}$, направленный отрезок $\overrightarrow{BD}$ определяет вектор $\boldsymbol{b}+\boldsymbol{c}$, а направленный отрезок $\overrightarrow{AD}$ определяет, с одной стороны, вектор $(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}$, а с другой стороны, — вектор $\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c})$. Следовательно, $(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}=\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c})$.

2. Отложим векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ от произвольной точки O. Пусть A и B — концы этих векторов. От точки A мы отложим вектор $\boldsymbol{b}$, а от точки B — вектор $\boldsymbol{a}$. Пусть C и D — концы этих векторов. По определению, направленные отрезки $\overrightarrow{OC}$ и $\overrightarrow{OD}$ определяют соответственно векторы $\boldsymbol{a}+\boldsymbol{b}$ и $\boldsymbol{b}+\boldsymbol{a}$. Таким образом, нам нужно только показать, что $C=D$. Но так как направленные отрезки $\overrightarrow{OA}$ и $\overrightarrow{BD}$ эквиполлентны, то середины отрезков $\overline{OD}$ и $\overline{AB}$ совпадают. Аналогично, так как направленные отрезки $\overrightarrow{OB}$ и $\overrightarrow{AC}$ эквиполлентны, то совпадают и середины отрезков $\overline{OC}$ и $\overline{AB}$. Следовательно, отрезки $\overline{OC}$ и $\overline{OD}$ имеют общую концевую точку и общую середину. Поэтому эти отрезки совпадают и, в частности, $C=D$.

3. Очевидно.

4. Если $\boldsymbol{a}=\overrightarrow{AB}$, то $-\boldsymbol{a}=\overrightarrow{BA}$ и поэтому $\boldsymbol{a}+(-\boldsymbol{a})=\overrightarrow{AA}=0$.

5. При $\boldsymbol{a}=0$ это свойство очевидно. Пусть $\boldsymbol{a}\neq 0$. Тогда, по определению,

$$(k+l)\,\boldsymbol{a}:\boldsymbol{a}=k+l$$

и

$$k\boldsymbol{a}:\boldsymbol{a}=k,\quad l\boldsymbol{a}:\boldsymbol{a}=l,$$

так что

$$(k+l)\,\boldsymbol{a}:\boldsymbol{a}=k\boldsymbol{a}:\boldsymbol{a}+l\boldsymbol{a}:\boldsymbol{a}.$$

С другой стороны, согласно лемме Шаля (или, точнее, равенству (1) п. 4)

$$(k\boldsymbol{a}+l\boldsymbol{a}):\boldsymbol{a}=k\boldsymbol{a}:\boldsymbol{a}+l\boldsymbol{a}:\boldsymbol{a}.$$

Следовательно,

$$(k+l)\boldsymbol{a} : a = (k\boldsymbol{a} + l\boldsymbol{a}) : \boldsymbol{a}$$

и потому

$$(k+l)\boldsymbol{a} = k\boldsymbol{a} + l\boldsymbol{a}.$$

6. Это уже известно (см. п. 3).

7. Пусть $\boldsymbol{e}$ — произвольный базис. Тогда $\boldsymbol{a} = a\boldsymbol{e}$ и $\boldsymbol{b} = b\boldsymbol{e}$, где a и b — координаты векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ в базисе $\boldsymbol{e}$. Следовательно, в силу уже доказанных свойств

$$k(\boldsymbol{a}+\boldsymbol{b}) = k(a\boldsymbol{e} + b\boldsymbol{e}) = k(a+b)\boldsymbol{e} = (ka + kb)\boldsymbol{e} =$$
$$= (ka)\boldsymbol{e} + (kb)\boldsymbol{e} = k(a\boldsymbol{e}) + k(b\boldsymbol{e}) = k\boldsymbol{a} + k\boldsymbol{b}.$$

8. Очевидно.

Доказанная теорема означает, что к векторам применимы обычные вычислительные приемы арифметики: раскрытие скобок, перенесение слагаемого из одной части равенства в другую (с изменением знака), умножение обеих частей равенства на одно и то же (отличное от нуля) число и т. п.

Чтобы облегчить ссылки, мы будем называть перечисленные в теореме 1 восемь свойств линейных операций над векторами их *стандартными свойствами.*

Множества, в которых определены операции сложения и умножения на вещественные числа, обладающие стандартными свойствами 1—8, часто встречаются в математике. Они называются *линейными пространствами* или, короче, *линеалами.*

В этой терминологии доказанная теорема утверждает, что *множество* $\mathrm{Vect}(1)$ *всех векторов на прямой является линеалом.*

Задание. Докажите, что в любом линеале для любых $\boldsymbol{a}$ и k имеют место равенства $0\boldsymbol{a} = 0$, $k0 = 0$.

Упражнение. Докажите, что свойство 2 линейных операций вытекает из остальных семи стандартных свойств.

6. Теорема об изоморфизме

Предполагая, что на прямой α выбран базис $\boldsymbol{e}$, вернемся к биективному соответствию между множеством $\mathrm{Vect}(1)$ всех векторов на прямой α и полем $\mathbb{R}$ вещественных чисел, получающемуся при сопоставлении каждому вектору $\boldsymbol{a} \in \mathrm{Vect}(1)$ его координаты $a = \boldsymbol{a} : \boldsymbol{e}$ в базисе $\boldsymbol{e}$ (см. п. 3). Согласно результатам пп. 3 и 4 при умножении вектора на число его координата умножается на то же число, а при сложении векторов их координаты складываются. Чтобы описать эту ситуацию в общих терминах, удобно ввести понятие «изоморфизма».

Понятие изоморфизма является одним из важнейших общематематических понятий. Грубо говоря, два множества, наделенные некоторым математическим строением, называются *изоморфными*, если между ними можно

установить биективное соответствие, сохраняющее это строение (*изоморфизм*). Чтобы придать этому определению точный смысл, необходимо, конечно, предварительно определить, что такое «математическое строение» и что понимается под отображением, «сохраняющим строение». Мы не будем этого делать, поскольку у нас сейчас нет ни места, ни времени, чтобы этим заниматься. Вместо этого мы предпочтем в каждой конкретной ситуации каждый раз заново определять соответствующее понятие изоморфизма.

Например, два множества X и Y, для которых определены операции сложения элементов и умножения элементов на вещественные числа, мы будем называть *изоморфными*, если существует такое биективное отображение $\varphi\colon X \to Y$ множества X на множество Y (*изоморфизм*), что

$$\varphi(x_1 + x_2) = \varphi(x_1) + \varphi(x_2)$$

и

$$\varphi(kx) = k\varphi(x)$$

для любых элементов x_1, x_2 множества X и любого вещественного числа k.

К множествам, для которых определены операция сложения элементов и операция их умножения на вещественные числа, принадлежат, в частности, множество векторов $\mathrm{Vect}(1)$ и поле вещественных чисел $\mathbb{R}$. Поэтому высказывание об их изоморфности или неизоморфности имеет смысл. С другой стороны, если мы обозначим координату вектора $\boldsymbol{x}$ в базисе $\boldsymbol{e}$ символом $\varphi(\boldsymbol{x})$, то тот факт, что координата суммы векторов является суммой координат слагаемых, запишется в виде формулы

$$\varphi(\boldsymbol{x}_1 + \boldsymbol{x}_2) = \varphi(\boldsymbol{x}_1) + \varphi(\boldsymbol{x}_2),$$

а тот факт, что при умножении вектора на число его координата умножается на то же число, запишется в виде формулы

$$\varphi(k\boldsymbol{x}) = k\varphi(\boldsymbol{x}).$$

Эти формулы означают, что биективное отображение φ является изоморфизмом. Тем самым нами доказана следующая теорема об изоморфизме:

Теорема 1. *Множество* $\mathrm{Vect}(1)$ *векторов на прямой изоморфно полю* $\mathbb{R}$ *вещественных чисел (по отношению к операциям сложения и умножения на вещественные числа). Соответствующий изоморфизм определяется выбором некоторого базиса* $\boldsymbol{e}$ *(и зависит от этого выбора).*

Изоморфизм $\mathrm{Vect}(1) \to \mathbb{R}$, определенный базисом $\boldsymbol{e}$, мы будем называть *координатным изоморфизмом.*

Значение понятия изоморфизма состоит в том, что *все свойства изоморфных множеств одинаковы* (конечно, имеются в виду лишь свойства, относящиеся к рассматриваемому математическому строению, например, к операциям сложения и умножения

на вещественные числа). Таким образом, если некоторые свойства уже доказаны для одного из изоморфных множеств, то мы можем быть уверены, что они выполнены и для другого.

В частности, все свойства 1—8, перечисленные в теореме 1 п. 5 («стандартные свойства»), как известно, выполнены для вещественных чисел. Поэтому они выполнены и для операций над векторами. Таким образом, теорема 1 п. 5 является непосредственным следствием теоремы о изоморфизме. Мы все же предпочли дать ей прямое доказательство, чтобы четче выяснить геометрическую суть дела (впрочем, при доказательстве свойства 7 мы на самом деле воспользовались координатным изоморфизмом).

Подчеркнем, что хотя изоморфные множества и обладают одинаковыми свойствами, отождествлять их целесообразно только тогда, когда между ними существует единственный (естественный) изоморфизм. В частности, как мы уже неоднократно подчеркивали, отождествление векторов на прямой с вещественными числами неправомерно.

§ 3. ЛИНЕЙНЫЕ ОПЕРАЦИИ НАД ВЕКТОРАМИ НА ПЛОСКОСТИ И В ПРОСТРАНСТВЕ

В этом параграфе мы определяем и изучаем линейные операции (сложение и умножение на числа) для векторов на плоскости или в пространстве.

1. Определение линейных операций

Операция сложения векторов на плоскости и в пространстве у нас фактически уже определена, поскольку все сказанное в п. 4 § 2 пригодно без всяких изменений и для векторов на плоскости и в пространстве. Таким образом, *суммой* $\boldsymbol{a}+\boldsymbol{b}$ двух векторов $\boldsymbol{a}=\overrightarrow{AB}$ и $\boldsymbol{b}=\overrightarrow{BC}$ называется вектор $\overrightarrow{AC}$, и это определение корректно. В случае, когда векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ не параллельны одной прямой, направленный отрезок $\overrightarrow{AC}$ представляет собой диагональ некоторого параллелограмма. Поэтому описанное правило сложения векторов называется часто «правилом параллелограмма».

С операцией умножения на число дело обстоит сложнее.

Пусть $\boldsymbol{a}=\overrightarrow{AB}$ — произвольный вектор (на плоскости или в пространстве), а k — произвольное вещественное число. При $\boldsymbol{a}=\boldsymbol{0}$ мы, по определению, положим

$$k\boldsymbol{0}=\boldsymbol{0}.$$

Пусть теперь $\boldsymbol{a}\neq\boldsymbol{0}$. Тогда определена прямая $\alpha=AB$ (ибо $A\neq B$), и вектор $\boldsymbol{a}$ мы можем рассматривать как вектор этой прямой ($\boldsymbol{a}\in \mathrm{Vect}_{\alpha}(1)$). Поэтому согласно сказанному в пре-

дыдущем пункте определен вектор $k\boldsymbol{a} \in \mathrm{Vect}_\alpha(1)$. Этот вектор, рассматриваемый как вектор на плоскости (или, соответственно, в пространстве), мы и примем за произведение $k\boldsymbol{a}$ вектора $\boldsymbol{a}$ на число k.

Другими словами, мы принимаем следующее

Определение 1. *Произведением* $k\boldsymbol{a}$ вектора $\boldsymbol{a} = \overrightarrow{AB}$, $\boldsymbol{a} \neq 0$, на число k называется вектор $\overrightarrow{CD}$, удовлетворяющий следующим двум условиям:

а) точки C и D принадлежат прямой $\alpha = AB$;

б) отношение $\overrightarrow{CD} : \overrightarrow{AB}$ (оно определено, поскольку направленные отрезки $\overrightarrow{CD}$ и $\overrightarrow{AB}$ принадлежат одной прямой и $\overrightarrow{AB} \neq 0$) равно k:

$$\overrightarrow{CD} : \overrightarrow{AB} = k.$$

Это определение нуждается, конечно (при $\boldsymbol{a} \neq 0$), в проверке корректности, т. е. в доказательстве независимости вектора $k\boldsymbol{a}$ от выбора направленного отрезка $\overrightarrow{AB}$. Поскольку при $k = 0$ это очевидно, нам следует при этом рассмотреть лишь случай, когда $k \neq 0$ (и $\boldsymbol{a} \neq 0$). Другими словами, нужно доказать следующее

Предложение 1. *Пусть* $\overrightarrow{AB}$ *и* $\overrightarrow{A_1B_1}$ *— эквиполлентные невырожденные направленные отрезки и пусть* $\overrightarrow{CD}$ *и* $\overrightarrow{C_1D_1}$ *— такие невырожденные направленные отрезки, что*

1) *отрезок* $\overrightarrow{CD}$ *расположен на той же прямой, что и отрезок* $\overrightarrow{AB}$ *(т. е. прямые AB и CD совпадают; эти прямые определены, поскольку* $A \neq B$ *и* $C \neq D$*), а отрезок* $\overrightarrow{C_1D_1}$ *расположен на той же прямой, что и отрезок* $\overrightarrow{A_1B_1}$;

2) *имеет место числовое равенство*

$$\overrightarrow{C_1D_1} : \overrightarrow{A_1B_1} = \overrightarrow{CD} : \overrightarrow{AB}.$$

Тогда отрезки $\overrightarrow{CD}$ *и* $\overrightarrow{C_1D_1}$ *эквиполлентны.*

Доказательство. Пусть α — прямая AB и α_1 — прямая A_1B_1. Поскольку отрезки $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ эквиполлентны, прямые α и α_1 параллельны. Если они совпадают, то наше предложение сводится к уже известному нам утверждению о корректности умножения на число векторов данной прямой. Поэтому мы можем предполагать, что $\alpha \neq \alpha_1$ (и, в частности, $A \neq A_1$). Кроме того, мы без ограничения общности можем считать, что $A = C$ и $A_1 = C_1$.

Поскольку направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ эквиполлентны (и расположены на различных прямых), четырехугольник

ABA_1B_1 является параллелограммом. Следовательно, прямые AA_1 и BB_1 параллельны. Нам надо доказать, что направленные отрезки $\overrightarrow{CD} = \overrightarrow{AD}$ и $\overrightarrow{C_1D_1} = \overrightarrow{A_1D_1}$ эквиполлентны, т. е. что четырехугольник ADA_1D_1 также является параллелограммом. Ясно, что для этого достаточно доказать, что прямая DD_1 параллельна прямой AA_1.

Рассмотрим с этой целью *проектирование* θ прямой α на прямую α_1 параллельно прямой $\beta = AA_1$. По определению, это проектирование переводит произвольную точку M прямой α в точку M', по которой прямая α_1 пересекается с прямой, параллельной прямой β и проходящей через точку M. В частности, $\theta(A) = A_1$ и $\theta(B) = B_1$.

Заметим, что проектирование θ определено и для непараллельных прямых α и α_1 (нужно лишь, чтобы прямая β не была параллельна ни прямой α, ни прямой α_1).

Мы будем пользоваться следующими известными свойствами параллельного проектирования (справедливыми и для непараллельных прямых α и α_1):

П1. *Если точка M лежит на прямой α между точками P и Q, то точка $M' = \theta(M)$ лежит на прямой α_1 между точками $P' = \theta(P)$ и $Q' = \theta(Q)$.*

П2. *Для любых четырех точек P, Q, M и $N \neq M$ прямой α имеет место равенство*

$$|P'Q'| : |M'N'| = |PQ| : |MN|.$$

Из свойств П1 и П2 немедленно вытекает, что

при параллельном проектировании отношение направленных отрезков инвариантно, т. е. для любых четырех точек P, Q, M и $N \neq M$ прямой α имеет место равенство

$$\overrightarrow{P'Q'} : \overrightarrow{M'N'} = \overrightarrow{PQ} : \overrightarrow{MN},$$

где $P' = \theta(P)$, $Q' = \theta(Q)$, $M' = \theta(M)$, $N' = \theta(N)$.

Действительно, свойство П2 обеспечивает сохранение абсолютной величины отношения, а из свойства П1 следует, что его знак также остается прежним.

Вернемся теперь к доказательству предложения 1. Пусть $D' = \theta(D)$. По только что сказанному

$$\overrightarrow{A_1D'} : \overrightarrow{A_1B_1} = \overrightarrow{AD} : \overrightarrow{AB}.$$

Но по условию

$$\overrightarrow{A_1D_1} : \overrightarrow{A_1B_1} = \overrightarrow{AD} : \overrightarrow{AB}$$

(напомним, что $A = C$ и $A_1 = C_1$). Следовательно, $\overrightarrow{A_1D} : \overrightarrow{A_1B_1} = \overrightarrow{A_1D_1} : \overrightarrow{A_1B_1}$ и потому $\overrightarrow{A_1D'} = \overrightarrow{A_1D_1}$.

Таким образом, $D' = D_1$ и, следовательно, прямая DD_1 параллельна прямой $\beta = AA_1$.

Тем самым корректность определения вектора $k\boldsymbol{a}$ полностью доказана.

2. Алгебраические свойства линейных операций

Теорема 1. *Теорема* 1 п. 5 § 2 *справедлива и для векторов на плоскости или в пространстве, т. е. перечисленные в этой теореме восемь стандартных свойств сложения и умножения на число имеют место и для векторов на плоскости или в пространстве.*

Доказательство. Изложенные в п. 5 § 2 доказательства свойств 1—4 сохраняются без каких-либо изменений. С другой стороны, все фигурирующие в свойствах 5—6 (а также в тривиальном свойстве 8) векторы параллельны одной и той же прямой, и потому, если эти свойства верны для векторов на прямой, то они верны и для векторов на плоскости или в пространстве. Таким образом, в доказательстве нуждается лишь свойство 7 (и только для случая, когда векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ не параллельны одной прямой, а $k \neq 0$).

Пусть $\boldsymbol{a} = \overrightarrow{AB}$, $\boldsymbol{b} = \overrightarrow{BC}$ и, следовательно, $\boldsymbol{a} + \boldsymbol{b} = \overrightarrow{AC}$. Далее, пусть $k\boldsymbol{a} = \overrightarrow{AB_1}$ (ясно, что без ограничения общности мы можем считать векторы $\boldsymbol{a}$ и $k\boldsymbol{a}$ отложенными от одной точки), $k\boldsymbol{b} = \overrightarrow{B_1C_1}$ и $k(\boldsymbol{a} + \boldsymbol{b}) = \overrightarrow{AC'}$. Нам нужно доказать, что $C_1 = C'$.

Рассмотрим с этой целью проектирование θ прямой $\alpha = AB$ на прямую $\alpha_1 = AC$ параллельно прямой $\beta = BC$ (поскольку векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ мы предполагаем не параллельными одной прямой, прямые α, α_1 и β определены и прямая β не параллельна прямым α и α_1). При этом проектировании точка A переходит сама в себя, а точка B переходит в точку C. Далее, тругольники ABC и AB_1C_1 подобны (их углы при вершинах B и B_1 одинаковы и они имеют по паре пропорциональных сторон), и потому их углы при вершине A равны. Поскольку прямая AB совпадает с прямой AB_1, этим доказано, что последние углы либо совпадают, либо являются вертикальными углами. В обоих случаях точка C_1 принадлежит прямой AC. Поскольку, по построению, прямая B_1C_1 параллельна прямой β, это означает, что $\theta(B_1) = C_1$.

Поэтому, согласно сказанному в предыдущем пункте,

$$\overrightarrow{AC_1} : \overrightarrow{AC} = \overrightarrow{AB_1} : \overrightarrow{AB} = k.$$

Но, по определению,

$$\overrightarrow{AC'} : \overrightarrow{AC} = k.$$

Следовательно,

$$\overrightarrow{AC_1} = \overrightarrow{AC'},$$

т. е. $C' = C_1$.

Тем самым наша теорема полностью доказана.

В терминологии, введенной в конце п. 5 § 2, теорема 1 означает, что

множества $\mathrm{Vect}(2)$ *и* $\mathrm{Vect}(3)$ *векторов на плоскости и в пространстве являются линеалами.*

Замечание 1. Обратим внимание на тот важный факт, что описанные в п. 4 § 1 *вложения* $\mathrm{Vect}_\alpha(1) \subset \mathrm{Vect}_\Pi(2) \subset \mathrm{Vect}(3)$ *согласованы с линейными операциями*, т. е. результат применения линейных операций, скажем, к векторам из $\mathrm{Vect}_\Pi(2)$ не зависит от того, рассматриваются ли эти векторы как векторы на плоскости или как векторы в пространстве. Для сложения этот факт вытекает из того, что оно определяется единым образом как для векторов на прямой, так и для векторов на плоскости или в пространстве, а для умножения на число — из того, что, по определению, оно сводится к умножению на число векторов на прямой.

Иначе этот факт можно выразить, сказав, что подмножества $\mathrm{Vect}_\alpha(1)$ и $\mathrm{Vect}_\Pi(2)$ множества $\mathrm{Vect}(3)$ (а также подмножества $\mathrm{Vect}_\alpha(1)$ множества $\mathrm{Vect}(2)$) *замкнуты* относительно линейных операций, действующих в множестве $\mathrm{Vect}(3)$ (соответственно, в множестве $\mathrm{Vect}(2)$), т. е.

если векторы $\boldsymbol{a}$ *и* $\boldsymbol{b}$ *из* $\mathrm{Vect}(3)$ (*из* $\mathrm{Vect}(2)$) *принадлежат, скажем, подмножеству* $\mathrm{Vect}_\alpha(1)$ *то их сумма* $\boldsymbol{a}+\boldsymbol{b}$ *и произведение* $k\boldsymbol{a}$ *вектора* $\boldsymbol{a}$ *на произвольное число* k *также принадлежат подмножеству* $\mathrm{Vect}_\alpha(1)$.

Еще иначе, мы можем сказать, что

естественное биективное отображение, скажем, множества $\mathrm{Vect}_\Pi(2)$ *в множество* $\mathrm{Vect}(3)$ *является изоморфизмом* (*по отношению к линейным операциям*) *множества* $\mathrm{Vect}_\Pi(2)$ *на подмножество множества* $\mathrm{Vect}(3)$, *состоящее из векторов, параллельных плоскости* Π.

3. Линейная зависимость

Определение 1. Семейство $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ векторов (произвольного линеала) называется *линейно зависимым*, если существуют такие числа $k_1, \ldots, k_N$, не все равные нулю, что

$$k_1\boldsymbol{a}_1 + \ldots + k_N\boldsymbol{a}_N = \boldsymbol{0}.$$

Допуская определенную вольность речи, часто в этой ситуации говорят, что линейно зависимы векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$. (Воль-

ность здесь состоит в том, что на самом деле линейно зависимы не векторы, а линейно зависимо их семейство, рассматриваемое как единый математический объект.)

Выражение $k_1\boldsymbol{a}_1 + \ldots + k_N\boldsymbol{a}_N$ называется *линейной комбинацией* векторов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ *с коэффициентами* $k_1, \ldots, k_N$. Если $k_1 = \ldots = k_N = 0$, то линейная комбинация называется *тривиальной*, в противном случае — *нетривиальной*. Таким образом, векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ линейно зависимы, если существует нетривиальная линейная комбинация этих векторов, равная нулю.

Если вектор $\boldsymbol{a}$ равен некоторой линейной комбинации векторов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$, то говорят, что он *линейно выражается* через эти векторы, а также что он *линейно зависит* от этих векторов. В частности, нулевой вектор $\mathbf{0}$ линейно выражается через любое семейство векторов, поскольку он равен тривиальной линейной комбинации этих векторов.

Очевидно, что

если вектор $\boldsymbol{a}$ *линейно выражается через векторы* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$, *а каждый из этих векторов линейно выражается через векторы* $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_M$, *то вектор* $\boldsymbol{a}$ *линейно выражается через векторы* $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_M$.

Несмотря на свою тривиальность, это свойство отношения «линейно выражается» играет очень важную роль. Оно называется свойством транзитивности.

Во избежание излишних оговорок, удобно ввести в рассмотрение также пустое семейство векторов. По определению, от пустого семейства векторов линейно зависит только нулевой вектор. *Пустое семейство линейно независимо.* Действительно, семейство векторов $\{\boldsymbol{a}_i\}$ линейно независимо, если при $\sum\limits_i k_i\boldsymbol{a}_i = \mathbf{0}$ не существует такого индекса i, что $k_i \neq 0$. Ясно, что эта формулировка сохраняет смысл (и верна) и тогда, когда множество индексов i пусто.

Легко видеть, что

любое семейство $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$, *содержащее линейно зависимое подсемейство* $\{\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_n}\}$, *само линейно зависимо.*

Действительно, по условию, существует равная нулю линейная комбинация векторов подсемейства, не все коэффициенты которой равны нулю. Добавив к этой линейной комбинации остальные векторы семейства с нулевыми коэффициентами, мы, очевидно, и получим равную нулю нетривиальную линейную комбинацию векторов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$.

С другой стороны, ясно, что

семейство, состоящее из одного вектора $\boldsymbol{a}$, *тогда и только тогда линейно зависимо, когда этот вектор равен нулю.*

Следовательно, в частности,

любое семейство, содержащее нулевой вектор, линейно зависимо.

Далее, легко видеть, что

векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ *тогда и только тогда линейно зависимы, когда один из них линейно выражается через остальные.*

Действительно, если $k_1\boldsymbol{a}_1 + \ldots + k_N\boldsymbol{a}_N = \boldsymbol{0}$ и, скажем, $k_1 \neq 0$, то $\boldsymbol{a}_1 = \left(-\frac{k_2}{k_1}\right)\boldsymbol{a}_2 + \ldots + \left(-\frac{k_N}{k_1}\right)\boldsymbol{a}_N$. Обратно, если $\boldsymbol{a}_1 = l_2\boldsymbol{a}_2 + \ldots + l_N\boldsymbol{a}_N$, то $(-1)\boldsymbol{a}_1 + l_2\boldsymbol{a}_2 + \ldots + l_N\boldsymbol{a}_N = \boldsymbol{0}$, причем $-1 \neq 0$.

Важным уточнением последнего утверждения является следующее

Предложение 1. *Векторы* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ *тогда и только тогда линейно зависимы, когда один из них линейно выражается через векторы с меньшими номерами.*

Доказательство. Если $\boldsymbol{a}_1 = \boldsymbol{0}$, то этот вектор линейно выражается через пустое множество векторов с меньшими номерами. Поэтому в этом случае предложение справедливо.

Пусть $\boldsymbol{a}_1 \neq \boldsymbol{0}$. Если векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ линейно зависимы (и $\boldsymbol{a}_1 \neq \boldsymbol{0}$), то существует такое число $n > 1$, что векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ линейно зависимы, а векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}$ уже линейно независимы. Пусть $k_1, \ldots, k_n$ — такие числа, не все равные нулю, что

$$k_1\boldsymbol{a}_1 + \ldots + k_n\boldsymbol{a}_n = \boldsymbol{0}.$$

Поскольку векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}$ линейно независимы, число k_n должно быть отлично от нуля. Следовательно,

$$\boldsymbol{a}_n = \left(-\frac{k_1}{k_n}\right)\boldsymbol{a}_1 + \ldots + \left(-\frac{k_{n-1}}{k_n}\right)\boldsymbol{a}_{n-1},$$

так что вектор $\boldsymbol{a}_n$ действительно является линейной комбинацией векторов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}$ с меньшими номерами.

Обратное утверждение очевидно.

Следствие. *Векторы* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ *тогда и только тогда линейно зависимы, когда либо линейно зависимы векторы* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{N-1}$, *либо вектор* $\boldsymbol{a}_N$ *линейно выражается через векторы* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{N-1}$.

Предложение 2. *Векторы* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ *тогда и только тогда линейно независимы, когда любой вектор* $\boldsymbol{a}$, *линейно выражающийся через эти векторы, выражается через них единственным образом.*

Доказательство. Если векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ линейно зависимы, то, скажем, нулевой вектор двумя разными способами линейно выражается через эти векторы: в виде тривиальной и нетривиальной линейных комбинаций.

Обратно, если для некоторого вектора $\boldsymbol{a}$ имеют место равенства

$$\boldsymbol{a} = k_1\boldsymbol{a}_1 + \ldots + k_N\boldsymbol{a}_N,$$
$$\boldsymbol{a} = l_1\boldsymbol{a}_1 + \ldots + l_N\boldsymbol{a}_N,$$

где хотя бы для одного индекса i коэффициент k_i отличен от коэффициента l_i, то

$$(k_1 - l_1)\boldsymbol{a}_1 + \ldots + (k_N - l_N)\boldsymbol{a}_N = \boldsymbol{0}$$

и, следовательно, векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ линейно зависимы.

Докажем в заключение следующую трудную, но важную теорему о зависимости:

Теорема 1. *Пусть каждый вектор семейства*

$$\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N \tag{1}$$

линейно выражается через векторы

$$\boldsymbol{b}_1, \ldots, \boldsymbol{b}_M. \tag{2}$$

Тогда, если $N > M$, то векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ линейно зависимы.

Доказательство. Прежде чем доказывать эту теорему, мы докажем следующую вспомогательную лемму:

Лемма. *Если условия теоремы 1 выполнены и семейство (1) линейно независимо, то для любого $s = 0, \ldots, M$ существует семейство векторов*

$$\boldsymbol{c}_1^{(s)}, \ldots, \boldsymbol{c}_M^{(s)}, \tag{3}$$

обладающее следующими свойствами:

а) *каждый вектор семейства (1) линейно выражается через векторы (3);*

б) *первые s векторов семейства (3) совпадают с первыми s векторами семейства (1):*

$$\boldsymbol{c}_1^{(s)} = \boldsymbol{a}_1, \ldots, \boldsymbol{c}_s^{(s)} = \boldsymbol{a}_s.$$

Мы докажем эту лемму индукцией по числу s. При $s = 0$ она очевидна (за семейство (3) можно принять семейство (2)). Пусть она уже доказана для числа s. Рассмотрим семейство

$$\boldsymbol{a}_{s+1}, \boldsymbol{c}_1^{(s)}, \ldots, \boldsymbol{c}_M^{(s)}. \tag{4}$$

Это семейство линейно зависимо, ибо вектор $\boldsymbol{a}_{s+1}$ линейно выражается через векторы $\boldsymbol{c}_1^{(s)}, \ldots, \boldsymbol{c}_M^{(s)}$. Поэтому согласно предложению 1 один из векторов семейства (4) линейно выражается через предыдущие. Этим вектором не может быть вектор $\boldsymbol{a}_{s+1}$, ибо он по условию отличен от нуля (семейство (1) линейно независимо). Им не может быть и ни один из векторов $\boldsymbol{c}_1^{(s)}, \ldots, \boldsymbol{c}_s^{(s)}$, ибо это снова противоречит линейной независимости векторов (1) (см. условие б)). Следовательно, не-

который вектор $c_i^{(s)}$, $i > s$, линейно выражается через векторы a_{s+1}, $c_1^{(s)}$, ..., $c_{i-1}^{(s)}$. Удалим теперь из семейства (4) вектор $c_i^{(s)}$. Полученное семейство обладает, очевидно, тем свойством, что через него линейно выражается любой вектор семейства (3), а следовательно, в силу свойства транзитивности, и любой вектор семейства (1). Другими словами, полученное семейство (состоящее из M векторов) обладает свойством а). Кроме того, оно содержит первые $s+1$ векторов семейства (1) (ибо $i > s$), т. е. по отношению к числу $s+1$ оно обладает и свойством б) (после, конечно, соответствующей перестановки его членов). Тем самым по индукции лемма полностью доказана.

Перейдем теперь непосредственно к доказательству теоремы 1. Предположим, что эта теорема неверна, т. е. что семейство (1) линейно независимо. Тогда к этому семейству будет применима доказанная лемма. Но при $s = M$ эта лемма дает, что каждый вектор семейства (1) и, в частности, вектор a_N линейно выражается через векторы a_1, ..., a_M. Поскольку $N > M$, это означает, что система (1) линейно зависима. Полученное противоречие доказывает теорему.

Замечание 1. Во всем этом пункте мы пользовались лишь восемью стандартными свойствами линейных операций. Поэтому все полученные здесь результаты справедливы для элементов любого линеала.

4. Геометрический смысл линейной зависимости

Определение 1. Векторы на плоскости или в пространстве называются *коллинеарными*, если они параллельны (см. п. 4 § 1) одной и той же прямой. Все векторы на прямой коллинеарны по определению.

Предложение 1. *Два вектора* $\boldsymbol{a}$ *и* $\boldsymbol{b}$ (*на прямой, на плоскости или в пространстве*) *тогда и только тогда коллианеарны, когда они линейно зависимы.*

Доказательство. В случае векторов на прямой для любых векторов $\boldsymbol{a} \neq \boldsymbol{0}$ и $\boldsymbol{b}$ определено их отношение $\boldsymbol{b} : \boldsymbol{a} = k$. Поэтому $\boldsymbol{b} = k\boldsymbol{a}$ и, следовательно, векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ линейно зависимы. Если $\boldsymbol{a} = \boldsymbol{0}$, то векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ линейно зависимы по общим соображениям. Таким образом, любые два вектора на прямой линейно зависимы. С другой стороны, любые два вектора на прямой, по определению, коллинеарны. Таким образом, для векторов на прямой наше предложение справедливо.

Рассмотрим теперь случай плоскости или пространства. Если векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ параллельны некоторой прямой, то, как мы знаем, их можно отождествить с векторами на этой прямой. Следовательно, по доказанному, эти векторы линейно зависимы (на прямой, а потому и на плоскости или, соответственно, в

пространстве). Обратно, если векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ линейно зависимы, то один из них (скажем, вектор $\boldsymbol{b}$) выражается через другой вектор (вектор $\boldsymbol{a}$), т. е. существует такое число k, что $\boldsymbol{b} = k\boldsymbol{a}$. Пусть α — такая прямая, что $\boldsymbol{a} \in \mathrm{Vect}_\alpha(1)$. Тогда, по определению, $k\boldsymbol{a} \in \mathrm{Vect}_\alpha(1)$, т. е. $\boldsymbol{b} \in \mathrm{Vect}_\alpha(1)$. Таким образом, векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ параллельны одной и той же прямой α.

Согласно общим результатам п. 3 доказанное предложение может быть сформулировано также в следующем виде:

Векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ тогда и только тогда коллинеарны, когда существует такое число k, что либо $\boldsymbol{b} = k\boldsymbol{a}$, либо $\boldsymbol{a} = k\boldsymbol{b}$.

В частности, мы видим, что на прямой отношение линейной зависимости «вырождается», в том смысле, что любое семейство векторов на прямой, состоящее более чем из одного вектора, линейно зависимо. Вместе с тем, на прямой существуют линейно независимые семейства, состоящие из одного вектора (таковыми являются семейства, состоящие из одного отличного от нуля вектора).

Посмотрим теперь, как обстоит дело на плоскости. Во-первых, ясно, что

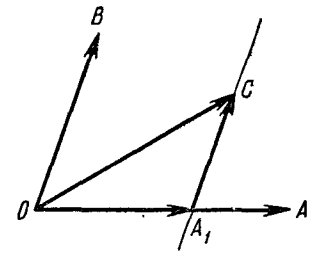
на плоскости существуют линейно независимые семейства, состоящие из двух векторов.

Действительно, чтобы построить такое семейство, достаточно взять три точки O, A и B, не лежащие на одной прямой, и рассмотреть векторы $\boldsymbol{a} = \overrightarrow{OA}$ и $\boldsymbol{b} = \overrightarrow{OB}$. По построению эти векторы не коллинеарны. Следовательно, они линейно независимы.

С другой стороны, легко видеть, что

любое семейство векторов на плоскости, состоящее более чем из двух векторов, линейно зависимо.

Согласно результатам п. 3 это утверждение достаточно доказать для семейства, состоящего из трех векторов. Пусть $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ — произвольные векторы на плоскости. Если векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ коллинеарны, то они линейно зависимы и потому линейно зависимы и векторы $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$. Следовательно, без ограничения общности мы можем предполагать, что векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ не коллинеарны (и, по аналогичным соображениям, что не коллинеарны векторы $\boldsymbol{a}$ и $\boldsymbol{c}$, а также векторы $\boldsymbol{b}$ и $\boldsymbol{c}$).



Пусть O — произвольная точка рассматриваемой плоскости и пусть $\boldsymbol{a} = \overrightarrow{OA}$, $\boldsymbol{b} = \overrightarrow{OB}$ и $\boldsymbol{c} = \overrightarrow{OC}$. Так как векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ не коллинеарны, то точки O, A и B не расположены на одной прямой, т. е. прямые $\alpha = OA$ и $\beta = OB$ различны. Кроме того, точка C не принадлежит ни одной из этих прямых (ибо в против-

ном случае вектор $\boldsymbol{c}$ был бы коллинеарен либо вектору $\boldsymbol{a}$, либо вектору $\boldsymbol{b}$). Пусть β_1 — прямая, проходящая через точку C параллельно прямой β, а A_1 — ее точка пересечения с прямой α (ясно, что прямые α и β_1 не параллельны). По определению,

$$\boldsymbol{c} = \overrightarrow{OC} = \overrightarrow{OA_1} + \overrightarrow{A_1C}.$$

С другой стороны, поскольку точка A_1 принадлежит прямой α, а вектор $\boldsymbol{a}$ отличен от нуля, существует такое число k, что $\overrightarrow{OA_1} = k\boldsymbol{a}$. Аналогично, поскольку вектор $\overrightarrow{A_1C}$ параллелен прямой β, существует такое число l, что $\overrightarrow{A_1C} = l\boldsymbol{b}$. Следовательно,

$$\boldsymbol{c} = k\boldsymbol{a} + l\boldsymbol{b},$$

и векторы $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ линейно зависимы.

Определение 2. Векторы в пространстве называются *компланарными*, если они параллельны одной и той же плоскости. Все векторы на прямой или на плоскости компланарны по определению.

Очевидно, что любые два вектора компланарны. Для трех векторов это уже не так.

Предложение 2. *Три вектора $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ (на прямой, на плоскости или в пространстве) тогда и только тогда компланарны, когда они линейно зависимы.*

Доказательство. Тот факт, что три компланарных вектора линейно зависимы, непосредственно вытекает из того, что любые три вектора плоскости (или прямой) линейно зависимы. Поэтому нам нужно доказать лишь обратное, т. е. что три линейно зависимых вектора компланарны. При этом нам достаточно рассмотреть лишь случай пространства.

Итак, пусть $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ — три линейно зависимых вектора в пространстве. Один из этих векторов (скажем, вектор $\boldsymbol{c}$) линейно выражается через другие (т. е. через векторы $\boldsymbol{a}$ и $\boldsymbol{b}$). Таким образом, существуют такие числа k и l, что

$$\boldsymbol{c} = k\boldsymbol{a} + l\boldsymbol{b}.$$

Для векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ существует плоскость, которой они параллельны, т. е. такая плоскость Π, что $\boldsymbol{a} \in \mathrm{Vect}_\Pi(2)$ и $\boldsymbol{b} \in \mathrm{Vect}_\Pi(2)$. Но тогда $k\boldsymbol{a} \in \mathrm{Vect}_\Pi(2)$, $l\boldsymbol{b} \in \mathrm{Vect}_\Pi(2)$ и потому $k\boldsymbol{a} + l\boldsymbol{b} \in \mathrm{Vect}_\Pi(2)$, т. е. $\boldsymbol{c} \in \mathrm{Vect}_\Pi(2)$. Следовательно, все три вектора $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ параллельны одной и той же плоскости Π, т. е. эти векторы компланарны.

Из предложения 2 немедленно вытекает, что

в пространстве существуют линейно независимые семейства, состоящие из трех векторов.

Действительно, чтобы построить такое семейство, достаточно взять четыре точки O, A, B и C, не расположенные на одной плоскости, и рассмотреть векторы $\boldsymbol{a} = \overrightarrow{OA}$, $\boldsymbol{b} = \overrightarrow{OB}$ и $\boldsymbol{c} = \overrightarrow{OC}$. Эти векторы не компланарны и потому, по доказанному, линейно независимы.

Вместе с тем, как мы сейчас покажем,

любое семейство векторов в пространстве, состоящее более чем из трех векторов, линейно зависимо.

Достаточно доказать это утверждение для семейства из четырех векторов $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ и $\boldsymbol{d}$, никакие три из которых не компланарны.

Пусть $\boldsymbol{a} = \overrightarrow{OA}$, $\boldsymbol{b} = \overrightarrow{OB}$, $\boldsymbol{c} = \overrightarrow{OC}$ и $\boldsymbol{d} = \overrightarrow{OD}$. Проведем через точку D прямую, параллельную вектору $\boldsymbol{c}$. Эта прямая пересекает плоскость OAB в некоторой точке C_1. По определению

$$\boldsymbol{d} = \overrightarrow{OD} = \overrightarrow{OC_1} + \overrightarrow{C_1D}.$$

Вектор $\overrightarrow{OC_1}$ расположен в плоскости OAB и потому согласно доказанному выше линейно выражается через неколлинеарные векторы $\boldsymbol{a} = \overrightarrow{OA}$ и $\boldsymbol{b} = \overrightarrow{OB}$. С другой стороны, вектор $\overrightarrow{C_1D}$ параллелен прямой OC и потому линейно выражается через вектор $\boldsymbol{c} = \overrightarrow{OC}$. Следовательно, вектор $\boldsymbol{d}$ линейно выражается через векторы $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, так что векторы $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{d}$ линейно зависимы.

Замечание 1. Доказанные результаты делают совершенно тривиальной теорему о зависимости из п. 3. Действительно, например, для векторов в пространстве эта теорема при $M > 2$ автоматически верна, ибо любая система векторов, состоящая более чем из трех векторов, линейно зависима. При $M = 2$ она утверждает (в не ограничивающем общности предположении линейной независимости векторов $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_M$), что любая система компланарных векторов, состоящая более чем из двух векторов, линейно зависима, а при $M = 1$, — что линейно зависима любая система коллинеарных векторов, состоящая более чем из одного вектора.

Следует, однако, иметь в виду, что в п. 3 мы доказали теорему о зависимости, пользуясь только стандартными свойствами 1—8 (и потому можем быть уверены в ее справедливости для любого линеала), тогда как теперь мы существенно использовали геометрические соображения.

Подводя итоги, мы можем сформулировать полученные результаты в виде следующей окончательной теоремы:

Теорема 1. *В линеале* $\mathrm{Vect}(n)$, $n = 1, 2, 3$, *существует n линейно независимых векторов, тогда как любое семейство, состоящее из большего числа векторов, линейно зависимо.*

Здесь $n = 1$ в случае прямой, $n = 2$ в случае плоскости и $n = 3$ в случае пространства.

5. Базисы и координаты

Определение 1. Семейство n линейно независимых векторов линеала $\mathrm{Vect}(n)$ называется *базисом* (соответственно — на прямой, на плоскости или в пространстве) [1]).

Таким образом, на прямой (в $\mathrm{Vect}(1)$) базис состоит из одного отличного от нуля вектора, на плоскости (в $\mathrm{Vect}(2)$) — из двух неколлинеарных векторов, а в пространстве (в $\mathrm{Vect}(3)$) — из трех некомпланарных векторов.

Определение 2. Семейство

$$\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$$

векторов некоторого линеала мы назовем *полным*, если любой вектор из этого линеала линейно выражается (вообще говоря, не единственным образом) через векторы этого семейства.

Легко видеть, что

любое полное семейство векторов из $\mathrm{Vect}(n)$ *содержит не менее n векторов.*

Действительно, пусть

$$\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$$

— данное полное семейство векторов и пусть

$$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$$

— произвольный базис в $\mathrm{Vect}(n)$. По условию, каждый вектор базиса $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ линейно выражается через векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$. Поэтому, если $N < n$, то в силу теоремы о зависимости векторы $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ были бы линейно зависимы, что противоречит определению базиса. Следовательно, $N \geqslant n$.

Далее, легко видеть, что

если полное семейство векторов

$$\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N \tag{1}$$

линейно зависимо, то из него можно удалить один вектор так, чтобы оставшееся семейство было также полно.

Действительно, если семейство линейно зависимо, то хотя бы один вектор этого семейства линейно выражается через остальные. Пусть, для определенности, это — вектор $\boldsymbol{a}_N$. Тогда каждый вектор семейства (1) будет линейно выражаться через векторы семейства

$$\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{N-1}. \tag{2}$$

Поэтому в силу транзитивности свойства «линейно выражаться» (см. п. 3 § 3) любой вектор, линейно выражающийся через се-

---

[1]) Обратим внимание на то, что при $n = 1$ это определение формально отличается от принятого в п. 1 § 2, где базисом на прямой назывался произвольный отличный от нуля вектор на этой прямой, тогда как теперь базисом называется семейство векторов на прямой, состоящее из одного отличного от нуля вектора. С точностью до этого формального различия наше общее определение базиса при $n = 1$ совпадает с определением из п. 1 § 2.

мейство (1), будет линейно выражаться и через семейство (2). Следовательно, если семейство (1) полно, то семейство (2) также полно.

Рассмотрим теперь следующие три свойства семейств векторов из $\mathrm{Vect}(n)$:

1) семейство линейно независимо;

2) семейство полно;

3) семейство состоит из n векторов.

Оказывается, что

любые два из трех свойств 1), 2), 3) *влекут за собой третье.*

Действительно, если выполнены свойства 1) и 2), то по теореме из п. 4 число N векторов этого семейства не больше n (в силу линейной независимости), а по только что доказанному утверждению оно и не меньше N (в силу полноты). Следовательно, $N = n$.

Пусть теперь выполнены свойства 1) и 3). Тогда для любого вектора $\boldsymbol{a} \in \mathrm{Vect}(n)$ семейство $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n, \boldsymbol{a}$ линейно зависимо и потому хотя бы один вектор этого семейства линейно выражается через векторы с меньшими номерами (предложение 1 п. 3). Поскольку этим вектором не может быть ни один из векторов $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ (ибо векторы $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ по условию линейно независимы), то, следовательно, вектор $\boldsymbol{a}$ линейно выражается через векторы $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. Таким образом, семейство $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ полно.

Наконец, если для семейства векторов выполнены свойства 2) и 3), то оно не может быть линейно зависимым, потому что в противном случае, удалив из этого семейства некоторый вектор, мы получим полное семейство, состоящее из $n-1$ векторов, что невозможно.

По определению, семейство, обладающее свойствами 1) и 3), является базисом. Следовательно, доказанное утверждение можно переформулировать в виде следующего предложения:

Предложение 1. *Семейство векторов из* $\mathrm{Vect}(n)$ *является базисом, если оно либо*

а) *линейно независимо и состоит из n векторов,*

либо

б) *полно и состоит из n векторов,*

либо

в) *линейно независимо и полно.*

Определение 3. Пусть $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ — произвольный базис в $\mathrm{Vect}(n)$. *Координатами* вектора $\boldsymbol{a} \in \mathrm{Vect}(n)$ в этом базисе называются такие числа [1]) $a^1, \ldots, a^n$, что

$$\boldsymbol{a} = a^1\boldsymbol{e}_1 + \ldots + a^n\boldsymbol{e}_n.$$

[1]) Здесь верхние индексы являются номерами, но не показателями степени.

Существование координат обеспечивается полнотой базиса, а их единственность — линейной независимостью базиса (предложение 2 п. 3).

Таким образом, при $n=1$

$$\boldsymbol{a}=a^1\boldsymbol{e}_1,$$

при $n=2$

$$\boldsymbol{a}=a^1\boldsymbol{e}_1+a^2\boldsymbol{e}_2,$$

и при $n=3$

$$\boldsymbol{a}=a^1\boldsymbol{e}_1+a^2\boldsymbol{e}_2+a^3\boldsymbol{e}_3.$$

Вектор $\boldsymbol{a}$ с координатами $a^1, \ldots, a^n$ мы будем иногда обозначать символом $\boldsymbol{a}(a^1, \ldots, a^n)$ или просто $(a^1, \ldots, a^n)$.

Подчеркнем, что это обозначение правомерно только при фиксированном (и явно указанном) базисе.

Если $\boldsymbol{a}=(a^1, \ldots, a^n)$ и $\boldsymbol{b}=(b^1, \ldots, b^n)$, т. е. если

$$\boldsymbol{a}=a^1\boldsymbol{e}_1+\ \ldots\ +a^n\boldsymbol{e}_n,$$
$$\boldsymbol{b}=b^1\boldsymbol{e}_1+\ \ldots\ +b^n\boldsymbol{e}_n,$$

то

$$\boldsymbol{a}+\boldsymbol{b}=a^1\boldsymbol{e}_1+\ \ldots\ +a^n\boldsymbol{e}_n+b^1\boldsymbol{e}_1+\ \ldots\ +b^n\boldsymbol{e}_n=$$
$$=(a^1+b^1)\,\boldsymbol{e}_1+\ \ldots\ +(a^n+b^n)\,\boldsymbol{e}_n,$$

и

$$k\boldsymbol{a}=k\,(a^1\boldsymbol{e}_1+\ \ldots\ +a^n\boldsymbol{e}_n)=$$
$$=(ka^1)\,\boldsymbol{e}_1+\ \ldots\ +(ka^n)\,\boldsymbol{e}_n.$$

Это показывает, что числа $a^1+b^1, \ldots, a^n+b^n$ являются координатами вектора $\boldsymbol{a}+\boldsymbol{b}$, а числа $ka^1, \ldots, ka^n$ — координатами вектора $k\boldsymbol{a}$. Таким образом,

при сложении векторов их координаты складываются, а при умножении вектора на число — умножаются на это число.

Чтобы описать эту ситуацию в терминах изоморфизма (ср. п. 6 § 2), мы введем в рассмотрение множество $\mathbb{R}^n$, элементами которого являются n-членные последовательности $(a^1, \ldots, a^n)$ вещественных чисел.

Таким образом, множество $\mathbb{R}^2$ состоит из пар (a^1, a^2), а множество $\mathbb{R}^3$ — из троек (a^1, a^2, a^3). Что же касается множества $\mathbb{R}^1$, то оно состоит из одночленных последовательностей вида (a^1). Но каждую такую последовательность мы можем отождествить с соответствующим числом a^1 и, следовательно, множество $\mathbb{R}^1$ мы можем считать совпадающим с множеством $\mathbb{R}$ вещественных чисел.

Мы введем в множестве $\mathbb{R}^n$ операции «покомпонентного сложения» и «покомпонентного умножения на число»:

$$(a^1, \ldots, a^n)+(b^1, \ldots, b^n)=(a^1+b^1, \ldots, a^n+b^n),$$
$$k\,(a^1, \ldots, a^n)=(ka^1, \ldots, ka^n).$$

Легко видеть, что эти операции обладают всеми восемью стандартными свойствами линейных операций над векторами, т. е. относительно этих операций *множество* $\mathbb{R}^n$ *является линеалом* [1]).

Замечание 1. Обратим внимание на то, что множество $\mathbb{R}^n$ имеет смысл рассматривать при любом натуральном n, хотя нам здесь нужны только значения $n = 1$, 2 и 3.

Линеал $\mathbb{R}^n$ мы будем называть *стандартным* (или *арифметическим*) *линеалом*. Он обладает естественным базисом, состоящим из последовательностей (1, 0, 0, ..., 0), (0, 1, 0, ..., 0), ..., (0, 0, ..., 0, 1). Этот базис мы будем называть *стандартным базисом* линеала $\mathbb{R}^n$.

Например, при $n = 3$ стандартный базис имеет вид

$$(1, 0, 0), \quad (0, 1, 0), \quad (0, 0, 1),$$

а при $n = 2$ — вид

$$(1, 0), \quad (0, 1).$$

При $n = 1$ стандартный базис состоит из одноэлементной последовательности (1), т. е. из числа 1.

Тот факт, что стандартный базис действительно является базисом, немедленно вытекает из очевидного соотношения (мы ограничиваемся случаем $n = 3$)

$$(a^1, a^2, a^3) = a^1(1, 0, 0) + a^2(0, 1, 0) + a^3(0, 0, 1).$$

Сопоставление произвольному вектору $\boldsymbol{a} \in \mathrm{Vect}(n)$ последовательности $(a^1, \ldots, a^n)$ его координат (в некотором базисе $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$) определяет, очевидно, биективное отображение множества $\mathrm{Vect}(n)$ на множество $\mathbb{R}^n$, переводящее, согласно сказанному выше, сумму в сумму и произведение на число в произведение на то же число. Другими словами, это соответствие является *изоморфизмом* линеала $\mathrm{Vect}(n)$ на линеал $\mathbb{R}^n$. Мы будем называть его *координатным изоморфизмом*, определенным базисом

$$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n.$$

Этот изоморфизм переводит базис $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ линеала $\mathrm{Vect}(n)$ в стандартный базис линеала $\mathbb{R}^n$.

Таким образом, справедлива следующая теорема об изоморфизме, обобщающая аналогичную теорему из п. 6 § 2:

---

[1]) Поэтому, в частности, для элементов множества $\mathbb{R}^n$ имеют смысл введенные в п. 3 понятия (линейной зависимости, независимости и т. п.), причем доказанные в п. 3 свойства этих понятий остаются справедливыми и для элементов множества $\mathbb{R}^n$ (при любом n).

Это замечание будет нам полезно ниже при оперировании с матрицами, поскольку строки и столбцы матриц можно рассматривать как элементы из $\mathbb{R}^n$ (при соответствующем n).

Теорема 1. *Для любого* $n = 1, 2, 3$ *линеал* $\mathrm{Vect}(n)$ *изоморфен линеалу* $\mathbb{R}^n$. *Соответствующий изоморфизм* $\mathrm{Vect}(n) \to \mathbb{R}^n$ *определяется выбором в* $\mathrm{Vect}(n)$ *некоторого базиса* $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ (*и зависит от этого базиса*).

Поскольку координатный изоморфизм не является естественным (он зависит от выбора базиса), отождествлять с его помощью линеалы $\mathrm{Vect}(n)$ и $\mathbb{R}^n$, вообще говоря, неправомерно. Однако, если базис фиксирован, такое отождествление иногда бывает удобно.

Упражнение. Докажите, что любой изоморфизм линеала $\mathrm{Vect}(n)$ на линеал $\mathbb{R}^n$ является координатным изоморфизмом, определенным некоторым базисом.

6. Проекции и координаты

Мы ввели координаты векторов довольно формальным алгебраическим определением [1]). Дадим теперь их описание, имеющее более геометрический характер.

Рассмотрим сначала векторы на плоскости.

Пусть α и β — две непараллельные прямые на плоскости.

Для произвольного вектора $\boldsymbol{a} = \overrightarrow{AB}$ рассмотрим две прямые, параллельные прямой β, одна из которых проходит через точку A, а другая — через точку B. Пусть A' и B' — точки пересечения этих прямых с прямой α.

Определение 1. Вектор $\overrightarrow{A'B'}$ называется *проекцией* вектора $\boldsymbol{a}$ на прямую α параллельно прямой β и обозначается символом $\mathrm{pr}_\alpha^\beta \boldsymbol{a}$.

Это определение нуждается, конечно, в проверке корректности, т. е. в доказательстве того, что вектор $\mathrm{pr}_\alpha^\beta \boldsymbol{a}$ не зависит от выбора направленного отрезка $\overrightarrow{AB}$, представляющего вектор $\boldsymbol{a}$.

Мы докажем корректность определения 1 алгебраическим способом, что позволит нам, кроме того, связать проекции с координатами.

Упражнение. Докажите корректность определения 1 прямым геометрическим построением, аналогично тому, как в п. 1 была доказана корректность определения вектора $k\boldsymbol{a}$.

Как мы знаем, линеалы $\mathrm{Vect}_\alpha(1)$ и $\mathrm{Vect}_\beta(1)$, рассматриваемые как подмножества линеала $\mathrm{Vect}(2)$, пересекаются только по нулевому вектору (см. п. 4 § 1). Отсюда следует, что

если некоторый вектор $\boldsymbol{a}$ *разложен в сумму*

$$\boldsymbol{a} = \boldsymbol{a}_1 + \boldsymbol{a}_2, \quad \boldsymbol{a}_1 \in \mathrm{Vect}_\alpha(1), \quad \boldsymbol{a}_2 \in \mathrm{Vect}_\beta(1), \tag{1}$$

то это разложение единственно.

---

[1]) Зато все результаты п. 5 справедливы для любого линеала, обладающего (по отношению к данному числу n) свойством, описанным в теореме 1 п. 4.

Действительно, если

$$a_1 + a_2 = a_1' + a_2',$$

где $a_1,\ a_1' \in \mathrm{Vect}_\alpha(1)$, а $a_2,\ a_2' \in \mathrm{Vect}_\alpha(1)$, то

$$a_1' - a_1 = a_2 - a_2'.$$

Но $a_1' - a_1 \in \mathrm{Vect}_\alpha(1)$ и $a_2 - a_2' \in \mathrm{Vect}_\beta(1)$. Таким образом, вектор $a_1' - a_1 = a_2 - a_2'$ принадлежит обоим линеалам $\mathrm{Vect}_\alpha(1)$ и $\mathrm{Vect}_\beta(1)$ и потому равен нулю. Следовательно,

$$a_1' = a_1, \quad a_2' = a_2.$$

Вернемся теперь к вектору $\mathrm{pr}_\alpha^\beta a = \overrightarrow{A'B'}$. По определению суммы векторов

$$a = \overrightarrow{AB} = \overrightarrow{A'B'} + (\overrightarrow{B'B} - \overrightarrow{A'A}). \tag{2}$$

Так как вектор $\overrightarrow{A'B'} = \mathrm{pr}_\alpha^\beta a$ принадлежит линеалу $\mathrm{Vect}_\alpha(1)$, а вектор $\overrightarrow{B'B} - \overrightarrow{A'A}$ — линеалу $\mathrm{Vect}_\beta(1)$ (по построению, линеалу $\mathrm{Vect}_\beta(1)$ принадлежат как вектор $\overrightarrow{B'B}$, так и вектор $\overrightarrow{A'A}$), то разложение (2) имеет вид (1) и, следовательно, по доказанному, однозначно определено. В частности, вектор $\overrightarrow{A'B'}$ однозначно определен вектором a. Тем самым корректность определения 1 полностью доказана. Одновременно мы доказали, что *проекция* $\mathrm{pr}_\alpha^\beta a$ *совпадает с вектором* a_1 *в разложении* (1).

Что же касается второго слагаемого a_2, то по аналогичным соображениям этот вектор совпадает с проекцией $\mathrm{pr}_\beta^\alpha a$ вектора a на прямую β параллельно прямой α. Таким образом, справедливо следующее

Предложение 1. *Для любого вектора* $a \in \mathrm{Vect}(2)$ *имеет место разложение*

$$a = \mathrm{pr}_\alpha^\beta a + \mathrm{pr}_\beta^\alpha a. \tag{3}$$

Проекция $\mathrm{pr}_\alpha^\beta a$ *обладает тем свойством, что вектор*

$$a - \mathrm{pr}_\alpha^\beta a$$

параллелен прямой β. *Этим свойством проекция* $\mathrm{pr}_\alpha^\beta a$ *характеризуется однозначно.*

Замечание 1. Мы видим, в частности, что проекция $\mathrm{pr}_\alpha^\beta a$ зависит только от направления прямых α и β, т. е. не изменится, если мы эти прямые заменим параллельными. Поэтому часто говорят о проекции не на прямую α, а *на направление* прямой α (см. сноску на стр. 17).

Направление прямой α однозначно задается произвольным базисом e_1 линеала $\mathrm{Vect}_\alpha(1)$, т. е. произвольным отличным от

нуля вектором, параллельным этой прямой. Поэтому проекцию $\mathrm{pr}_{\alpha}^{\beta}\boldsymbol{a}$ часто называют проекцией *на направление вектора* $\boldsymbol{e}_1$.

Аналогично, направление прямой β однозначно задается произвольным базисом $\boldsymbol{e}_2$ линеала $\mathrm{Vect}_{\beta}(1)$, и потому проекцию $\mathrm{pr}_{\alpha}^{\beta}\boldsymbol{a}$ можно называть проекцией параллельно вектору $\boldsymbol{e}_2$ (или проекцией *по* направлению вектора $\boldsymbol{e}_2$).

Введя в рассмотрение вместо прямых α и β векторы $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$, естественно обозначать проекцию $\mathrm{pr}_{\alpha}^{\beta}\boldsymbol{a}$ символом $\mathrm{pr}_{\boldsymbol{e}_1}^{\boldsymbol{e}_2}\boldsymbol{a}$.

Векторы $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$ составляют, очевидно, базис на плоскости (их два и они неколлинеарны, т. е. линейно независимы).

Резюмируя все сказанное, мы получаем, что

задание на плоскости произвольного базиса $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ *позволяет сопоставить каждому вектору* $\boldsymbol{a}$ *два вектора*

$$\mathrm{pr}_{\boldsymbol{e}_1}^{\boldsymbol{e}_2}\boldsymbol{a} \quad \textit{и} \quad \mathrm{pr}_{\boldsymbol{e}_2}^{\boldsymbol{e}_1}\boldsymbol{a}$$

(*проекции этого вектора на направление одного из векторов базиса параллельно другому вектору*), *сумма которых равна данному вектору*:

$$\boldsymbol{a} = \mathrm{pr}_{\boldsymbol{e}_1}^{\boldsymbol{e}_2}\boldsymbol{a} + \mathrm{pr}_{\boldsymbol{e}_2}^{\boldsymbol{e}_1}\boldsymbol{a}. \tag{4}$$

Сопоставим теперь разложение (4) с разложением

$$\boldsymbol{a} = a^1\boldsymbol{e}_1 + a^2\boldsymbol{e}_2 \tag{5}$$

вектора $\boldsymbol{a}$ по векторам базиса. Поскольку $a^1\boldsymbol{e}_1 \in \mathrm{Vect}_{\alpha}(1)$ и $a^2\boldsymbol{e}_2 \in \mathrm{Vect}_{\beta}(1)$, разложения (4) и (5) должны совпадать:

$$\mathrm{pr}_{\boldsymbol{e}_1}^{\boldsymbol{e}_2}\boldsymbol{a} = a^1\boldsymbol{e}_1, \quad \mathrm{pr}_{\boldsymbol{e}_2}^{\boldsymbol{e}_1}\boldsymbol{a} = a^2\boldsymbol{e}_2.$$

Таким образом,

координата a^1 *вектора* $\boldsymbol{a}$ *в базисе* $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ *равна отношению его проекции на направление вектора* $\boldsymbol{e}_1$ *параллельно вектору* $\boldsymbol{e}_2$ *к вектору* $\boldsymbol{e}_1$:

$$a^1 = (\mathrm{pr}_{\boldsymbol{e}_1}^{\boldsymbol{e}_2}\boldsymbol{a}) : \boldsymbol{e}_1.$$

Приняв вектор $\boldsymbol{e}_1$ за эталон длины (и направления) на прямой α, мы можем это утверждение сформулировать также следующим образом:

координата a^1 *равна величине проекции вектора* $\boldsymbol{a}$ *на направление вектора* $\boldsymbol{e}_1$ *параллельно вектору* $\boldsymbol{e}_2$.

Конечно, аналогичные утверждения справедливы и для второй координаты a_2. В частности,

$$a^2 = (\mathrm{pr}_{\boldsymbol{e}_2}^{\boldsymbol{e}_1}a) : \boldsymbol{e}_2.$$

В пространстве возможны два вида проекций: проекция на плоскость параллельно прямой и проекция на прямую параллельно плоскости.

Пусть Π и α — плоскость и не параллельная ей прямая и пусть $\boldsymbol{a} = \overrightarrow{AB}$ — произвольный вектор в пространстве. Проведем через точки A и B плоскости, параллельные плоскости Π. Пусть A_1 и B_1 — точки пересечения этих плоскостей с прямой α.

Определение 2. Вектор $\overrightarrow{A_1B_1}$ называется *проекцией* вектора $\boldsymbol{a}$ на прямую α параллельно плоскости Π и обозначается символом $\mathrm{pr}_{\alpha}^{\Pi}\boldsymbol{a}$.

Аналогично, если мы проведем через точки A и B прямые, параллельные прямой α, и рассмотрим точки пересечения A_2 и B_2 этих прямых с плоскостью Π, то мы получим вектор $\overrightarrow{A_2B_2}$.

Определение 3. Вектор $\overrightarrow{A_2B_2}$ называется *проекцией* вектора $\boldsymbol{a}$ на плоскость Π параллельно прямой α и обозначается символом $\mathrm{pr}_{\Pi}^{\alpha}\boldsymbol{a}$.

Предложение 2. *Для любого вектора* $\boldsymbol{a} \in \mathrm{Vect}(3)$ *имеет место разложение*

$$\boldsymbol{a} = \mathrm{pr}_{\alpha}^{\Pi}\boldsymbol{a} + \mathrm{pr}_{\Pi}^{\alpha}\boldsymbol{a}. \tag{6}$$

Проекция $\mathrm{pr}_{\alpha}^{\Pi}\boldsymbol{a}$ *обладает тем свойством, что вектор*

$$\boldsymbol{a} - \mathrm{pr}_{\alpha}^{\Pi}\boldsymbol{a}$$

параллелен плоскости Π, *а проекция* $\mathrm{pr}_{\Pi}^{\alpha}\boldsymbol{a}$ — *тем свойством, что вектор*

$$\boldsymbol{a} - \mathrm{pr}_{\Pi}^{\alpha}\boldsymbol{a}$$

параллелен прямой α. *Этими свойствами проекции* $\mathrm{pr}_{\alpha}^{\Pi}\boldsymbol{a}$ *и* $\mathrm{pr}_{\Pi}^{\alpha}\boldsymbol{a}$ *характеризуются однозначно.*

Доказательство этого предложения полностью аналогично доказательству предложения 1 и поэтому мы его опустим.

Обратим внимание на то, что предложение 2 обеспечивает, в частности, корректность определений 2 и 3.

Упражнение. Докажите корректность определений 2 и 3 прямым геометрическим построением.

Проекции $\mathrm{pr}_{\alpha}^{\Pi}\boldsymbol{a}$ и $\mathrm{pr}_{\Pi}^{\alpha}\boldsymbol{a}$ зависят только от направлений прямой α и плоскости Π, т. е. не меняются при замене этой прямой или этой плоскости параллельными прямой или плоскостью. Поэтому вместо задания прямой α достаточно задать произвольный ее базис $\boldsymbol{e}_1$, а вместо задания плоскости Π — произвольный ее базис $\boldsymbol{e}_2$, $\boldsymbol{e}_3$. В этом случае проекция $\mathrm{pr}_{\alpha}^{\Pi}\boldsymbol{a}$ обозначается символом $\mathrm{pr}_{\boldsymbol{e}_1}^{\boldsymbol{e}_2,\,\boldsymbol{e}_3}\boldsymbol{a}$, а проекция $\mathrm{pr}_{\Pi}^{\alpha}\boldsymbol{a}$ — символом $\mathrm{pr}_{\boldsymbol{e}_2,\,\boldsymbol{e}_3}^{\boldsymbol{e}_1}\boldsymbol{a}$.

Для того чтобы проекции $\mathrm{pr}_{\boldsymbol{e}_1}^{\boldsymbol{e}_2,\,\boldsymbol{e}_3}\boldsymbol{a}$ и $\mathrm{pr}_{\boldsymbol{e}_2,\,\boldsymbol{e}_3}^{\boldsymbol{e}_1}\boldsymbol{a}$ были определены, необходимо и достаточно, чтобы векторы $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ образовывали базис.

С координатами a^1, a^2, a^3 вектора $\boldsymbol{a}$ в базисе $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ проекции $\mathrm{pr}_{\boldsymbol{e}_1}^{\boldsymbol{e}_2,\,\boldsymbol{e}_3}\boldsymbol{a}$ и $\mathrm{pr}_{\boldsymbol{e}_2,\,\boldsymbol{e}_3}^{\boldsymbol{e}_1}\boldsymbol{a}$ связаны формулами

$$\mathrm{pr}_{\boldsymbol{e}_1}^{\boldsymbol{e}_2,\,\boldsymbol{e}_3} = a^1\boldsymbol{e}_1,$$

$$\mathrm{pr}_{\boldsymbol{e}_2,\,\boldsymbol{e}_3}^{\boldsymbol{e}_1}\boldsymbol{a} = a^2\boldsymbol{e}_2 + a^3\boldsymbol{e}_3.$$

Таким образом, в частности,

координата a^1 вектора $\boldsymbol{a}$ в базисе $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ равна отношению его проекции на направление вектора $\boldsymbol{e}_1$ параллельно векторам $\boldsymbol{e}_2$ и $\boldsymbol{e}_3$ к вектору $\boldsymbol{e}_1$:

$$a^1 = (\mathrm{pr}_{\boldsymbol{e}_1}^{\boldsymbol{e}_2,\,\boldsymbol{e}_3}\boldsymbol{a}) : \boldsymbol{e}_1.$$

Конечно, аналогичные формулы

$$a^2 = (\mathrm{pr}_{\boldsymbol{e}_2}^{\boldsymbol{e}_1,\,\boldsymbol{e}_3}\boldsymbol{a}) : \boldsymbol{e}_2,$$

$$a^3 = (\mathrm{pr}_{\boldsymbol{e}_3}^{\boldsymbol{e}_1,\,\boldsymbol{e}_2}\boldsymbol{a}) : \boldsymbol{e}_3$$

имеют место и для остальных двух координат.

Из установленной связи между проекциями и координатами немедленно вытекают следующие свойства проекции:

1. *Проекция суммы векторов равна сумме проекций слагаемых*:

$$\mathrm{pr}\,(\boldsymbol{a} + \boldsymbol{b}) = \mathrm{pr}\,\boldsymbol{a} + \mathrm{pr}\,\boldsymbol{b}.$$

2. *Проекция произведения вектора на число равна произведению его проекции на это число*:

$$\mathrm{pr}\,(k\boldsymbol{a}) = k\,\mathrm{pr}\,\boldsymbol{a}.$$

В этих свойствах имеются в виду все три типа проекций (pr_α^β, pr_α^Π, pr_Π^α).

Упражнение. Докажите свойства 1 и 2 прямым геометрическим построением и затем выведите отсюда свойства координат.

7. Преобразование координат при замене базиса

В заключение этого параграфа рассмотрим вопрос о том, как изменяются координаты векторов при изменении базиса.

Пусть $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ и $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$ — два базиса [1]) в $\mathrm{Vect}\,(n)$ (где, как всегда, $n = 1$, 2 или 3). Поскольку $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ является базисом, каждый вектор из $\mathrm{Vect}\,(n)$ и, в частности, каждый вектор $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$ линейно выражается (и притом единственным образом) через векторы $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.

[1]) Обратите внимание на положение штрихов.

Это означает, что существуют такие (однозначно определенные) числа $c^i_{i'}$, $i, i' = 1, \ldots, n$, что

$$\begin{aligned} \boldsymbol{e}_{1'} &= c^1_{1'}\boldsymbol{e}_1 + \ \ldots\ + c^n_{1'}\boldsymbol{e}_n, \\ &\ldots\ldots\ldots\ldots\ldots \\ \boldsymbol{e}_{n'} &= c^1_{n'}\boldsymbol{e}_1 + \ \ldots\ + c^n_{n'}\boldsymbol{e}_n. \end{aligned} \tag{1}$$

Таким образом, например, при $n = 2$

$$\begin{aligned} \boldsymbol{e}_{1'} &= c^1_{1'}\boldsymbol{e}_1 + c^2_{1'}\boldsymbol{e}_2, \\ \boldsymbol{e}_{2'} &= c^1_{2'}\boldsymbol{e}_1 + c^2_{2'}\boldsymbol{e}_2. \end{aligned}$$

При фиксированном $i' = 1, \ldots, n$ числа $c^i_{i'}$ представляют собой координаты вектора $\boldsymbol{e}_{i'}$ в базисе $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.

Числа $c^i_{i'}$ мы будем записывать в виде матрицы

$$C = \begin{pmatrix} c^1_{1'} & \ldots & c^1_{n'} \\ \ldots & \ldots & \ldots \\ c^n_{1'} & \ldots & c^n_{n'} \end{pmatrix},$$

располагая координаты векторов $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$ в базисе $\boldsymbol{e}_1 \ldots, \boldsymbol{e}_n$ по столбцам.

Определение 1. Матрица C называется *матрицей перехода* от «старого» базиса $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ к «новому» базису $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$.

При $n = 1$ матрица перехода сводится к одному числу $c^1_{1'}$ (ср. конец п. 3 § 2), при $n = 2$ имеет вид

$$C = \begin{pmatrix} c^1_{1'} & c^1_{2'} \\ c^2_{1'} & c^2_{2'} \end{pmatrix},$$

а при $n = 3$ — вид

$$C = \begin{pmatrix} c^1_{1'} & c^1_{2'} & c^1_{3'} \\ c^2_{1'} & c^2_{2'} & c^2_{3'} \\ c^3_{1'} & c^3_{2'} & c^3_{3'} \end{pmatrix}.$$

Используя матрицу перехода C, можно формулы (1) «перехода от старого базиса к новому» записать в компактном виде, если ввести в рассмотрение однострочечные матрицы

$$\boldsymbol{e} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n), \quad \boldsymbol{e}' = (\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}),$$

состоящие из векторов данных базисов. По правилу умножения матриц [1]) произведение $\boldsymbol{e}C$ матриц $\boldsymbol{e}$ и C совпадает, очевидно, с матрицей $\boldsymbol{e}'$. Таким образом,

$$\boldsymbol{e}' = \boldsymbol{e}C.$$

[1]) Произведение матрицы с векторными элементами на обыкновенную матрицу (с числовыми элементами) определяется той же (известной из алгебры) формулой, что и произведение числовых матриц.

Отсюда сразу же вытекает, что
для любых трех базисов

(i) $e_1, \ldots, e_n$,

(ii) $e_{1'}, \ldots, e_{n'}$

и

(iii) $e_{1''}, \ldots, e_{n''}$

матрица перехода C'' от базиса (1) *к базису* (3) *равна произведению CC' матрицы перехода C от базиса* (1) *к базису* (2) *на матрицу перехода C' от базиса* (2) *к базису* (3).

Действительно, так как

$$e' = eC, \quad e'' = e'C',$$

то

$$e'' = (eC)\,C' = e\,(CC'),$$

и потому $CC' = C''$, поскольку равенство $e'' = eC''$ однозначно определяет матрицу C''.

В частности, мы видим, что матрица C' перехода от базиса $e_{1'}, \ldots, e_{n'}$ к базису $e_1, \ldots, e_n$ связана с матрицей C перехода от базиса $e_1, \ldots, e_n$ к базису $e_{1'}, \ldots, e_{n'}$ формулой

$$CC' = E$$

(поскольку единичная матрица E является, очевидно, матрицей перехода от базиса $e_1, \ldots, e_n$ к самому себе).

Аналогично показывается, что

$$C'C = E.$$

Это означает, что C' является обратной матрицей C^{-1}.

Таким образом,

матрицей перехода от базиса $e_{1'}, \ldots, e_{n'}$ к базису $e_1, \ldots, e_n$ является матрица C^{-1}, обратная к матрице перехода C от базиса $e_1, \ldots, e_n$ к базису $e_{1'}, \ldots, e_{n'}$.

Поскольку обратная матрица может существовать только у невырожденных матриц (имеющих отличный от нуля определитель), тем самым, в частности, доказано, что

любая матрица перехода от одного базиса к другому является невырожденной матрицей.

Является ли условие невырожденности достаточным для того, чтобы данная квадратная матрица C порядка n была бы матрицей перехода между некоторыми базисами? Ответ на этот вопрос является утвердительным; более того, оказывается, что первый базис $e_1, \ldots, e_n$ можно при этом взять произвольно. Именно, легко видеть, что

для любого базиса $e_1, \ldots, e_n$ и любой невырожденной матрицы $C = (c^i_{i'})$ порядка n существует такой базис $e_{1'}, \ldots, e_{n'}$, что матрица C является матрицей перехода от базиса $e_1, \ldots, e_n$ к базису $e_{1'}, \ldots, e_{n'}$.

Действительно, по данным векторам $e_1, \ldots, e_n$ и данной матрице C определим векторы $e_{1'}, \ldots, e_{n'}$ формулами (1). Достаточно доказать, что полученные таким образом векторы линейно независимы. Но это ясно, поскольку в противном случае столбцы матрицы C были ли линейно зависимы, что ввиду ее невырожденности невозможно.

Замечание 1. В последнем рассуждении мы использовали тот известный из теории определителей факт, что *определитель с линейно зависимыми столбцами (или строками) равен нулю.*

Обратим внимание на то, что теперь мы можем доказать и обратное утверждение, т. е. что

определитель с линейно независимыми столбцами (или строками) отличен от нуля.

Действительно, пусть $C = (c^i_{i'})$ — матрица данного определителя. Рассмотрим в $\mathrm{Vect}(n)$ произвольный базис $e_1, \ldots, e_n$ и построим векторы $e_{1'}, \ldots, e_{n'}$ по формулам (1). По условию, эти векторы линейно независимы. Следовательно, они образуют базис (их n). Таким образом, матрица C является матрицей перехода от одного базиса к другому и потому невырождена.

Конечно, это рассуждение годится лишь для «геометрических» значений n, т. е. при $n = 1, 2, 3$.

Более подробно этот вопрос мы рассмотрим в дополнении к этому параграфу.

Пользуясь матричным умножением, можно в компактном виде записать также формулу разложения

$$x = x^1 e_1 + \ldots + x^n e_n$$

произвольного вектора x по векторам базиса. Действительно, эта формула равносильна, очевидно, формуле

$$x = ex,$$

где x — одностолбцовая матрица

$$\begin{pmatrix} x^1 \\ \cdot \\ \cdot \\ \cdot \\ x^n \end{pmatrix},$$

составленная из координат вектора x.

Пусть теперь x' — одностолбцовая матрица, составленная из координат $x^{1'}, \ldots, x^{n'}$ вектора x в новом базисе $e_{1'}, \ldots, e_{n'}$. Тогда

$$x = e'x'.$$

Но

$$e' = eC.$$

Поэтому

$$\boldsymbol{x} = \boldsymbol{e}C\boldsymbol{x}',$$

т. е.

$$\boldsymbol{e}x = \boldsymbol{e}C\boldsymbol{x}'.$$

Следовательно,

$$x = Cx',$$

или в развернутом виде

$$\begin{aligned} x^1 &= c^1_{1'}x^{1'} + \ldots + c^1_{n'}x^{n'}, \\ &\ldots\ldots\ldots\ldots \\ x^n &= c^n_{1'}x^{1'} + \ldots + c^n_{n'}x^{n'}. \end{aligned} \qquad (2)$$

Например, при $n = 2$

$$\begin{aligned} x^1 &= c^1_{1'}x^{1'} + c^1_{2'}x^{2'}, \\ x^2 &= c^2_{1'}x^{1'} + c^2_{2'}x^{2'}. \end{aligned}$$

Обратим внимание, что формулы (2) выражают «старые» координаты через «новые», тогда как для векторов базиса ситуация была как раз обратной. Кроме того, в этих формулах координаты векторов $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$ в базисе $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ (столбцы матрицы C) оказались расположенными по строкам.

С другой стороны, как в формуле (1), так и в формуле (2) суммирование происходит по индексам, один раз встречающимся вверху, а один раз внизу.

Замечание об обозначениях. Использованные обозначения (в частности, соглашение о положении штрихов и индексов), хорошо приспособлены для рассмотрения случая произвольного n (и для доказательства общих утверждений). Поэтому естественно, что при фиксированном n они не очень удобны. Они неудобны также и при конкретных вычислениях.

Например, при $n = 1$ (случай прямой) проще всего не писать вообще никаких индексов (как это мы делали в § 2). При $n = 2$ (случай плоскости) удобно координаты обозначать двумя различными буквами (скажем, x и y) и соответствующим образом изменить обозначения элементов матрицы перехода. При $n = 3$ (случай пространства) для обозначения координат можно пользоваться буквами x, y, z и т. п.

Это замечание очень существенно, поскольку в дальнейшем мы без оговорок будем часто пользоваться не общими, а специализированными обозначениями, и читатель должен это иметь в виду.

Дополнение. **Теорема о ранге матрицы**

Пусть

$$\begin{pmatrix} a_{11} \ldots a_{1n} \\ \ldots\ldots\ldots \\ a_{m1} \ldots a_{mn} \end{pmatrix} \qquad (1)$$

— произвольная (вообще говоря, прямоугольная) матрица. Выбрав произвольно k строк и k столбцов и взяв элементы, расположенные на пересечении этих строк и столбцов, мы получим некоторый определитель k-го порядка, называемый *минором* матрицы (1), соответствующим данным строкам и столбцам. Рассмотрим значения k, обладающие тем свойством, что все миноры k-го порядка матрицы (1) равны нулю. В частности (в соответствии с общими правилами употребления в математике понятия пустого множества), к этим значениям принадлежат все значения, для которых множество миноров k-го порядка пусто, т. е. значения, большие числа строк или столбцов матрицы (1).

Легко видеть, что

если все миноры k-го порядка матрицы (1) равны нулю, то все миноры любого большего порядка также равны нулю.

Действительно, достаточно доказать равенство нулю любого минора $k+1$-го порядка. Но каждый такой минор, будучи разложен по элементам некоторой его строки (или столбца), является линейной комбинацией миноров k-го порядка. Поскольку все последние миноры равны нулю, данный минор также равен нулю.

Отсюда непосредственно вытекает, что для любой матрицы (1) существует такое число $r \geqslant 0$, что

а) все миноры порядка, большего чем r, равны нулю;

б) среди миноров любого фиксированного порядка, не превосходящего r, существует по крайней мере один отличный от нуля минор.

Определение 1. Число r, обладающее свойствами а) и б), называется *рангом* матрицы (1). Коротко его можно охарактеризовать как наивысший порядок отличных от нуля миноров матрицы (1).

Очевидно, что

если любые p строк (или столбцов) матрицы линейно зависимы, то ее ранг меньше p:

$$r < p.$$

Действительно, если некоторая система строк линейно зависима, то система «укороченных» строк, получающихся из данных выбрасыванием некоторых компонент (имеющих для всех строк одни и те же номера), также линейно зависима. Поэтому в каждом миноре порядка $\geqslant p$ любые p строк (столбцов) линейно зависимы и, следовательно, этот минор равен нулю.

Одним из важнейших результатов линейной алгебры является тот факт, что справедливо и обратное утверждение, иными словами, справедлива следующая теорема, известная как теорема о ранге матрицы:

Теорема 1. *Число p тогда и только тогда обладает тем свойством, что любые p строк матрицы линейно зависимы, когда $p > r$.*

Таким образом, в частности,

любая матрица ранга r содержит r линейно независимых строк (столбцов).

Доказательство теоремы о ранге матрицы отнюдь не элементарно и мы его здесь приводить не будем. Мы ограничимся лишь обсуждением некоторых частных случаев этой теоремы.

Пусть нам дана квадратная матрица

$$\begin{pmatrix} a_{11} \dots a_{1n} \\ \cdot \quad \cdot \quad \cdot \\ a_{n1} \dots a_{nn} \end{pmatrix}$$

порядка n. Наивысшее возможное значение ее ранга равно n. При этом $r = n$ тогда и только тогда, когда определитель

$$\begin{vmatrix} a_{11} \dots a_{1n} \\ \cdot \quad \cdot \quad \cdot \\ a_{n1} \dots a_{nn} \end{vmatrix}$$

матрицы отличен от нуля, т. е. когда матрица невырождена. Таким образом, частным случаем теоремы о ранге матрицы является следующая теорема о невырожденности:

Теорема 2. *Квадратная матрица тогда и только тогда невырождена, когда ее строки (столбцы) линейно независимы.*

В одну сторону («если матрица невырождена, то ее строки линейно независимы») эта теорема является тривиальным следствием известных элементарных свойств определителей. Доказательства требует лишь обратное утверждение («если строки матрицы линейно независимы, то матрица невырождена»).

Для определителей первого порядка теорема о невырожденности сводится к тавтологии. Для определителей второго порядка она утверждает, что определитель

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = a_1 b_2 - a_2 b_1$$

тогда и только тогда равен нулю, когда

$$\frac{a_1}{b_1} = \frac{a_2}{b_2},$$

что, очевидно, верно. Для определителей третьего порядка она нам известна из геометрических соображений (см. стр. 57).

Резюмируя, мы, следовательно, можем сказать, что

теорема о невырожденности заведомо справедлива для матриц порядка не большего трех.

Весьма замечательно, что хотя теорема о невырожденности является довольно частным случаем теоремы о ранге матрицы, тем не менее последняя теорема может быть без особого труда выведена из первой. Действительно, в силу теоремы о зависимости (см. п. 3) теорема о ранге матрицы может быть переформулирована в следующем виде:

в любой матрице ранга r существует r линейно независимых строк (столбцов), через которые линейно выражается любая другая строка (любой другой столбец).

Но, по условию, в матрице существует отличный от нуля минор Δ порядка r. Пусть $\alpha_1, \dots, \alpha_r$ — строки матрицы, участвующие в этом миноре. Ясно, что эти строки линейно независимы. Таким образом, нам достаточно

лишь показать, что любая другая строка β матрицы линейно выражается через эти строки. (Для столбцов рассуждение вполне аналогично.)

С этой целью рассмотрим в матрице минор Δ', получающийся добавлением к минору Δ строки β и некоторого (безразлично какого) столбца. Являясь минором порядка $r+1$, этот минор равен нулю. Следовательно, его строки линейно зависимы (здесь мы используем теорему о невырожденности). Но строки минора Δ линейно независимы. Поэтому строка β (или точнее, ее часть, участвующая в миноре Δ') линейно выражается через остальные строки этого минора. Это означает, что, вычтя из строки β некоторую линейную комбинацию строк $\alpha_1, \ldots, \alpha_r$, мы получим строку (обозначим ее через β'), обладающую тем свойством, что из составляющих ее чисел по крайней мере $r+1$ число равно нулю (именно, равны нулю числа, расположенные в столбцах, участвующих в миноре Δ'). Для завершения доказательства остается лишь показать, что на самом деле строка β' состоит только из нулей.

Пусть это не так, т. е. пусть строка β' содержит некоторое отличное от нуля число b. Пусть это число расположено на i-м месте. Рассмотрим минор Δ'', получающийся из минора Δ добавлением строки β и i-го столбца. Являясь минором порядка $r+1$, минор Δ'' равен нулю. Вычтя из его строки β ту же линейную комбинацию строк, что и выше, мы преобразуем минор Δ'' в определитель вида

$$\begin{vmatrix} & & * \\ & \Delta & \vdots \\ & & * \\ 0 & \ldots\ 0 & b \end{vmatrix}$$

Следовательно, $\Delta b = \Delta'' = 0$. Но, по условию, $\Delta \neq 0$. Таким образом, вопреки предположению, $b = 0$. Тем самым полностью доказано, что

из теоремы о невырожденности вытекает теорема о ранге матрицы.

Более того, поскольку теорему о невырожденности мы использовали только один раз и только применительно к минорам порядка $r+1$, на самом деле доказано, что

из теоремы о невырожденности для матриц порядка $r+1$ вытекает теорема о ранге для матриц ранга r, а также для матриц ранга $r+1$, число строк или столбцов которых равно $r+1$.

В соответствии со сказанным выше мы имеем право, таким образом, считать, что

теорема о ранге матрицы справедлива для матриц ранга, меньшего трех, а также для матриц, число строк или столбцов которых не превосходит трех.

Только для таких матриц эта теорема нам в дальнейшем и понадобится.

Ранг 0 имеет только матрица, состоящая из одних нулей. Ранг 1 имеет матрица, все строки (столбцы) которой пропорциональны.

Заметим, что поскольку в формулировку теоремы о ранге матрицы строки и столбцы входят совершенно равноправно,

максимальное число линейно независимых строк равно максимальному числу линейно независимых столбцов (и равно рангу матрицы).

В частности,

если строки матрицы пропорциональны, то ее столбцы также пропорциональны, и наоборот.

В дальнейшем нам неоднократно придется рассматривать пару матриц, вторая из которых получается добавлением к первой матрице некоторого столбца (или строки). Пусть r — ранг первой, а R — ранг второй («расширенной») матрицы. Тогда

$$r \leqslant R \leqslant r + 1.$$

Действительно, неравенство $r \leqslant R$ очевидно. С другой стороны, каждый минор расширенной матрицы либо является минором исходной матрицы, либо, будучи разложен по элементам дополнительного столбца, может быть представлен в виде линейной комбинации миноров исходной матрицы на единицу меньшего порядка. Следовательно, все миноры расширенной матрицы порядка $r + 2$ равны нулю, и потому $R \leqslant r + 1$.

Полезно также заметить, что

равенство $R = r$ имеет место тогда и только тогда, когда дополнительный столбец линейно выражается через столбцы исходной матрицы.

Действительно, если дополнительный столбец линейно выражается через столбцы исходной матрицы, то все миноры порядка $r + 1$ расширенной матрицы, не являющиеся минорами исходной матрицы, имеют линейно зависимые столбцы и потому равны нулю.

Обратно, пусть $R = r$. Рассмотрим r линейно независимых столбцов исходной матрицы, через которые линейно выражаются все остальные ее столбцы. Поскольку $R = r$, то через эти столбцы должен выражаться и дополнительный столбец. Таким образом, при $R = r$ дополнительный столбец линейно выражается через столбцы исходной матрицы.

Обратим внимание, что в последнем рассуждении мы воспользовались теоремой о ранге матрицы.

Замечание 1. Пусть

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}, \quad \begin{pmatrix} a_{11} & \dots & a_{1n} & b_1 \\ \dots & \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} & a_m \end{pmatrix}$$

— данные матрицы. Тот факт, что дополнительный столбец второй матрицы линейно выражается через столбцы первой матрицы, означает, что система линейных уравнений

$$\begin{aligned} a_{11}x_1 + \dots + a_{1n}x_n &= b_1 \\ \dots\dots\dots\dots\dots\dots & \\ a_{m1}x_1 + \dots + a_{mn}x_n &= b_m \end{aligned}$$

имеет хотя бы одно решение (совместна). Следовательно, доказанное выше утверждение может быть переформулировано в виде следующей теоремы, известной как теорема Кронекера—Капелли:

Теорема 3. *Система линейных уравнений тогда и только тогда совместна, когда ранг матрицы ее коэффициентов не меняется при добавлении столбца свободных членов.*

§ 4. ОРИЕНТАЦИИ ПРЯМОЙ, ПЛОСКОСТИ И ПРОСТРАНСТВА

1. Понятие ориентации

В п. 1 § 2 мы определили ориентацию прямой как выбор на ней определенного направления движения, т. е. одного из двух возможных отношений предшествования точек этой прямой. Затем мы показали, что ориентацию прямой можно с равным правом определить так же и как класс одноименных невырожденных направленных отрезков или как класс одноименных отличных от нуля векторов (т. е. базисов).

Ясно, что первое определение ориентации (выбор отношения предшествования) на плоскость и пространство непосредственно не переносится. Иначе обстоит дело с двумя другими определениями.

Определение 1. Два базиса $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ и $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ на плоскости ($n = 2$) или в пространстве ($n = 3$) называются *одноименными*, если определитель $|C|$ матрицы перехода C от первого базиса ко второму (см. п. 7 § 3) положителен:

$$|C| > 0.$$

Если $|C| > 0$, то базисы называются *разноименными*.

Заметим, что при $n = 1$ это определение согласуется с определением одноименных базисов на прямой. Поэтому в дальнейшем мы случай $n = 1$ исключать не будем, хотя, конечно, ничего нового мы для этого случая не получим.

В первую очередь мы покажем, что

отношение одноименности базисов является отношением эквивалентности.

Матрицей перехода от произвольного базиса к нему самому является единичная матрица, определитель которой равен единице и потому положителен. Следовательно, отношение одноименности рефлексивно.

Если C — матрица перехода от базиса $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ к базису $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$, то, как мы знаем, обратная матрица C^{-1} является матрицей перехода от базиса $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ к базису $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. Поэтому если $|C| > 0$, то и $|C^{-1}| = |C|^{-1} > 0$. Следовательно, отношение одноименности симметрично.

Если C — матрица перехода от базиса $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ к базису $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$, а C' — матрица перехода от базиса $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ к базису $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n$, то, как мы знаем (п. 7 § 3), матрицей перехода от базиса $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ к базису $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n$ является матрица CC'. Но по правилу умножения определителей, $|CC'| = |C| \cdot |C'|$. Поэтому, если $|C| > 0$ и $|C'| > 0$, то и $|CC'| > 0$. Следовательно, отношение одноименности транзитивно.

Поскольку отношение одноименности является отношением эквивалентности, все базисы разбиваются на классы одноименных базисов.

Определение 2. Классы одноименных базисов называются *ориентациями* (соответственно — прямой, плоскости и пространства).

Так как существуют базисы, связанные с данным базисом произвольной невырожденной матрицей перехода (см. п. 7 § 3), и, в частности, произвольной матрицей с отрицательным определителем, мы видим, что

на прямой, на плоскости и в пространстве существует точно две различные ориентации: базисы, принадлежащие одной и той же ориентации, одноименны, а базисы, принадлежащие различным ориентациям, разноименны.

Определение 3. Прямая, плоскость или пространство, для которой(ого) выбрана некоторая ориентация, называется *ориентированной(ым)*. Для каждой ориентации o другая возможная ориентация обозначается символом $-o$ и называется *противоположной* ориентацией. О базисе, принадлежащем некоторой ориентации, говорят, что он *определяет* эту ориентацию. Базис ориентированной(ого) прямой, плоскости или пространства, определяющий данную ориентацию, называется *положительно ориентированным* (по отношению к данной ориентации), а базис, определяющий противоположную ориентацию, называется *отрицательно ориентированным.*

На практике часто возникает вопрос об одноименности базисов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ и $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$, векторы которых заданы их координатами в некотором третьем базисе $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. Пусть Δ_a — определитель, столбцами которого являются столбцы координат векторов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. (Таким образом, Δ_a — это не что иное, как определитель матрицы перехода от базиса $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ к базису $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$.) Аналогично строится определитель Δ_b. Матрица перехода C от базиса $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ к базису $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ является, как мы знаем, произведением матрицы перехода от базиса $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ к базису $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ (эта матрица обратна к матрице перехода от базиса $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ к базису $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ и потому ее определитель равен Δ_a^{-1}) на матрицу перехода от базиса $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ к базису $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ (имеющую, по определению, определитель Δ_b). Поэтому согласно правилу умножения определителей

$$|C| = \Delta_a^{-1} \Delta_b.$$

Отсюда непосредственно вытекает, что

базисы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ и $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ тогда и только тогда одноименны, когда построенные для них определители Δ_a и Δ_b имеют один и тот же знак.

Определение ориентации как класса одноименных направленных отрезков также можно перенести на случай плоскости и пространства.

Определение 4. Упорядоченная тройка (A, B, C) точек плоскости (или пространства) называется *неколлинеарной тройкой*, если точки A, B, C не принадлежат одной прямой. Аналогично, упорядоченная четверка (A, B, C, D) точек пространства называется *некомпланарной четверкой*, если точки A, B, C, D не принадлежат одной плоскости. Такие тройки и четверки называются также тройками и четверками точек *общего положения*.

Замечание 1. Введя, по аналогии, понятие пары (A, B) точек общего положения как упорядоченной пары, состоящей из различных точек A и B, мы немедленно получим, что это понятие в точности совпадает с понятием невырожденного направленного отрезка. Таким образом, понятия неколлинеарной тройки и некомпланарной четверки можно рассматривать как непосредственные обобщения на случай плоскости и пространства понятия невырожденного направленного отрезка.

Для любой неколлинеарной тройки (A, B, C) точек плоскости векторы $\overrightarrow{AB}$ и $\overrightarrow{AC}$ неколлинеарны (т. е. линейно независимы) и потому образуют базис плоскости. Аналогично, для любой некомпланарной четверки (A, B, C, D) точек пространства векторы $\overrightarrow{AB}$, $\overrightarrow{AC}$, $\overrightarrow{AD}$ образуют базис пространства.

Определение 5. Две неколлинеарные тройки (A, B, C) и (A', B', C') точек плоскости (две некомпланарные четверки (A, B, C, D) и (A', B', C', D') точек пространства) называются *одноименными*, если соответствующие базисы $\overrightarrow{AB}$, $\overrightarrow{AC}$ и $\overrightarrow{A'B'}$, $\overrightarrow{A'C'}$ (базисы $\overrightarrow{AB}$, $\overrightarrow{AC}$, $\overrightarrow{AD}$ и $\overrightarrow{A'B'}$, $\overrightarrow{A'C'}$, $\overrightarrow{A'D'}$) одноименны.

Ясно, что отношение одноименности троек (четверок) является отношением эквивалентности и что соответствующие классы эквивалентности находятся в естественном биективном соответствии с классами одноименных базисов, т. е. с ориентациями плоскости (пространства).

Это означает, что

ориентации плоскости (пространства) можно отождествить с классами одноименных неколлинеарных троек (некомпланарных четверок) точек этой плоскости (пространства).

Упражнение. Определите одноименность троек и четверок точек общего положения элементарно-геометрически, не обращаясь к векторам.

Наглядное определение ориентации как направления движения возможно, конечно, не только на прямой, но и на любой линии, например, на окружности. Однако ориентацию окружности нельзя формально охарактеризовать (в отличие от ориентации прямой) как некоторое отношение, в котором участвуют две точки. Это связано с тем, что окружность является замкнутой линией, а прямая — нет.

Аналогом отношения предшествования точек прямой является на окружности отношение между тремя (попарно раз-

личными) точками A, B, C, наглядно состоящее в том, что, передвигаясь по окружности в данном направлении и отправляясь от точки A, мы не можем достичь точки C, не переходя через точку B.

Можно без труда аксиоматически охарактеризовать это отношение между точками окружности, подобно тому как в п. 1 § 2 мы охарактеризовали отношение предшествования точек прямой, а затем перенести на случай окружности описание ориентации прямой как класса одноименных направленных отрезков (при этом роль направленных отрезков будут играть упорядоченные тройки (A, B, C) попарно различных точек окружности, а их отношение одноименности будет наглядно означать, что они определяют на окружности одно и то же направление движения). Такое «независимое» рассмотрение ориентаций окружности, возможно, само по себе и интересное, в достаточной степени громоздко, а с принципиальной стороны не дает ничего нового по сравнению со случаем прямой. Поэтому мы предпочтем здесь описать ориентации окружности не независимо, а пользуясь тем, что она расположена в плоскости.

С этой целью мы заметим, что на неформальном, наглядном, уровне ясно (например, из чертежа), что две тройки (A, B, C) и (A', B', C') точек окружности Σ тогда и только тогда определяют на окружности одно и то же направление движения, когда базисы $\overrightarrow{AB}$, $\overrightarrow{AC}$ и $\overrightarrow{A'B'}$, $\overrightarrow{A'C'}$ плоскости, в которой расположена окружность Σ, одноименны. Таким образом, мы приходим к следующему (уже формальному) определению ориентации окружности:

Определение 6. Упорядоченные тройки (A, B, C) и (A', B', C') точек окружности Σ (состоящие каждая из попарно различных точек) называются *одноименными*, если базисы $\overrightarrow{AB}$, $\overrightarrow{AC}$ и $\overrightarrow{A'B'}$, $\overrightarrow{A'C'}$ плоскости, в которой расположена окружность Σ, одноименны. Класс одноименных троек точек окружности называется ее *ориентацией*, а окружность, для которой выбрана некоторая ориентация, называется *ориентированной*.

Сравнив это определение с определением 5 одноименности неколлинеарных точек плоскости, мы немедленно заметим, что они, по существу, совпадают. Это означает, что

между ориентациями плоскости и ориентациями расположенной в этой плоскости окружности существует естественное биективное соответствие.

Иначе говоря,

на ориентированной плоскости каждая окружность автоматически ориентирована, и, обратно, задание на плоскости ориентированной окружности однозначно определяет некоторую ориентацию плоскости.

Различие между ориентациями плоскости и ориентациями расположенных в этой плоскости окружностей состоит только в

том, что ориентацией плоскости является класс одноименных всевозможных неколлинеарных троек точек плоскости, а ориентацией окружности — класс одноименных неколлинеарных троек точек, принадлежащих этой окружности.

2. Правые и левые ориентации

Обратим внимание, что из двух возможных ориентаций прямой, плоскости или пространства чисто внутренним математическим способом нельзя выделить какую-нибудь определенную. Для этого приходится обращаться к соображениям, математике посторонним (скажем, к анатомии человека).

Например, на прямой, изображенной на чертеже горизонтальной линией, можно выделить направление «слева направо». Это направление для человека наиболее привычно (в частности, большинство культурных народов пишут в этом направлении) и потому ему присвоено наименование «положительного направления» («положительной ориентации»). Однако следует отчетливо понимать, что фиксация этого направления никакого инвариантного (не зависящего от чертежа) смысла не имеет: поверните чертеж «вверх ногами», и «положительное» направление перейдет в «отрицательное». Не нужно также путать понятие «положительного направления» в этом смысле с понятием «положительного направления по отношению к данной ориентации», которое вполне инвариантно, но зависит от выбора ориентации.

На вертикальных прямых принято «положительным» направлением считать направление «снизу вверх».

На плоскости задание ориентации равносильно, как мы знаем, заданию ориентаций расположенных в этой плоскости окружностей, т. е. заданию некоторого «направления вращения» этих окружностей. Наглядно это направление можно описать как направление, в котором следует вращать первый вектор базиса, определяющего данную ориентацию плоскости, чтобы кратчайшим путем совместить его со вторым вектором этого базиса (в этой формулировке предполагается, что векторы базиса имеют одинаковые длины и отложены от *одной* точки). При этом «положительной» ориентацией принято считать ориентацию, соответствующую вращению «против часовой стрелки». Это соглашение, конечно, также не инвариантно: посмотрите на плоскость с другой стороны, и «положительная» ориентация окажется «отрицательной».

Если посмотреть на правую руку со стороны ладони, то большой и указательный пальцы будут образовывать базис, ориентированный «против часовой стрелки». На этом основании ориентация плоскости «против часовой стрелки» обычно называется *правой* ориентацией, а противоположная ее ориентация — *левой*.

Аналогично, в пространстве *правой* ориентацией называется ориентация, определенная «базисом», состоящим из большого,

указательного и среднего пальцев правой руки, а противоположная ориентация называется *левой*. Здесь, к сожалению, нет общепринятого соглашения, какую ориентацию считать «положительной». Большинство авторов (и мы в том числе) считают «положительной» правую ориентацию. Однако во многих (особенно старых) учебниках за положительную ориентацию принимается левая ориентация.

Заметим, что термин «правая» и «левая» по отношению к ориентациям пространства имеют уже инвариантный смысл, поскольку «посмотреть на пространство с другой стороны» мы не можем. Тем не менее определить их внутренним математическим способом нельзя.

3. Произведения ориентаций

Пусть α_1 и α_2 — две ориентированные прямые на плоскости. Предполагая, что эти прямые не параллельны, выберем на каждой из них положительно ориентированный базис. Пусть это будет вектор $\boldsymbol{a}_1$ на прямой α_1 и вектор $\boldsymbol{a}_2$ на прямой α_2. Так как векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ по условию не коллинеарны, они составляют некоторый базис на плоскости. Оказывается, что

ориентация плоскости, определенная базисом $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, *не зависит от выбора векторов* $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ *и определяется исключительно ориентациями прямых* α_1 *и* α_2 (*взятых в данном порядке*).

Действительно, пусть $\boldsymbol{a}_1'$ и $\boldsymbol{a}_2'$ — другие положительно ориентированные базисы прямых α_1 и α_2. Тогда $\boldsymbol{a}_1' = k_1\boldsymbol{a}_1$ и $\boldsymbol{a}_2' = k_2\boldsymbol{a}_2$, где $k_1 > 0$ и $k_2 > 0$. Поэтому матрица перехода от базиса $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ к базису $\boldsymbol{a}_1'$, $\boldsymbol{a}_2'$ имеет вид

$$\begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix}$$

и ее определитель k_1k_2 положителен. Следовательно, базис $\boldsymbol{a}_1'$, $\boldsymbol{a}_2'$ определяет ту же ориентацию плоскости, что и базис $\boldsymbol{a}_1$, $\boldsymbol{a}_2$.

Таким образом,

для задания ориентации плоскости достаточно задать ориентации двух не параллельных прямых этой плоскости.

Определение 1. Построенная ориентация плоскости называется *произведением* данных ориентаций прямых α_1 и α_2 и обозначается символом o_1o_2, где o_1 — ориентация прямой α_1, а o_2 — ориентация прямой α_1.

Очевидно, что

$$o_2o_1 = -o_1o_2$$

и

$$(-o_1)\,o_2 = o_1\,(-o_2) = -o_1o_2,$$

т. е.

при перестановке прямых α_1 *и* α_2 *или при смене ориентации одной из них результирующая ориентация плоскости переходит в противоположную.*

Отсюда, в частности, вытекает, что

для любой ориентации o плоскости и любой ориентации o_2 прямой α_2 существует такая ориентация o_1 прямой α_1, что $o = o_1 o_2$.

Действительно, выберем произвольно некоторую ориентацию o_1' прямой α_1 и построим ориентацию $o_1' o_2$. Если окажется, что $o = o_1' o_2$, то все в порядке: $o_1 = o_1'$. Если же $o_1' o_2 \neq o$, то обязательно $o_1' o_2 = -o$ и потому $o_1 = -o_1'$.

Пусть теперь α и Π — ориентированные прямая и плоскость в пространстве (с ориентациями o_α и o_Π соответственно). Предполагая прямую α и плоскость Π не параллельными, рассмотрим ориентацию пространства, определенную базисом $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$, где $\boldsymbol{a}_1$ — некоторый положительно ориентированный базис прямой α, а $\boldsymbol{a}_2$, $\boldsymbol{a}_3$ — некоторый положительно ориентированный базис плоскости Π.

Если $\boldsymbol{a}_1' = k_1 \boldsymbol{a}_1$ — другой положительно ориентированный базис прямой α (так что $k_1 > 0$), а $\boldsymbol{a}_2' = k_2 \boldsymbol{a}_2 + k_2' \boldsymbol{a}_3$, $\boldsymbol{a}_3' = k_3' \boldsymbol{a}_2 + k_3 \boldsymbol{a}_3$ — другой положительно ориентированный базис плоскости Π $\left(\text{так что } \begin{vmatrix} k_2 & k_2' \\ k_3' & k_3 \end{vmatrix} > 0\right)$, то базис $\boldsymbol{a}_1'$, $\boldsymbol{a}_2'$, $\boldsymbol{a}_3'$ пространства связан с базисом $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$ матрицей перехода

$$\begin{pmatrix} k_1 & 0 & 0 \\ 0 & k_2 & k_2' \\ 0 & k_3' & k_3 \end{pmatrix},$$

определитель которой равен $k_1 \begin{vmatrix} k_2 & k_2' \\ k_3' & k_3 \end{vmatrix}$ и потому положителен. Следовательно,

ориентация пространства, определенная базисом $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$, не зависит от выбора базисов $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$, $\boldsymbol{a}_3$ и определяется исключительно данными ориентациями o_α и o_Π прямой α и плоскости Π.

Таким образом,

для задания ориентации пространства достаточно задать ориентации произвольной прямой и произвольной не параллельной ей плоскости.

Определение 2. Построенная ориентация пространства называется *произведением* ориентаций o_α и o_Π и обозначается символом $o_\alpha o_\Pi$.

Аналогично определяется ориентация $o_\Pi o_\alpha$ (задаваемая базисом $\boldsymbol{a}_2$, $\boldsymbol{a}_3$, $\boldsymbol{a}_1$). Впрочем, ясно, что она совпадает с ориентацией $o_\alpha o_\Pi$:

$$o_\Pi o_\alpha = o_\alpha o_\Pi.$$

Далее, очевидно, что

$$(-o_\alpha)\,o_\Pi = o_\alpha(-o_\Pi) = -o_\alpha o_\Pi,$$

откуда, как и выше, вытекает, что

для любой ориентации o пространства и любой ориентации o_α прямой α (любой ориентации o_Π плоскости Π) существует такая ориентация o_Π плоскости Π (ориентация o_α прямой α), что $o = o_\alpha o_\Pi$.

Аналогичным образом определяется произведение $o_1o_2o_3$ ориентаций трех прямых α_1, α_2, α_3, обладающих тем свойством, что параллельные им прямые, проходящие через одну точку, не лежат в одной плоскости (такие прямые называются *прямыми общего положения*). Впрочем, это произведение сводится к уже определенным произведениям:

$$o_1o_2o_3 = o_1(o_2o_3).$$

Таким образом,

для задания ориентации пространства достаточно задать ориентации трех прямых общего положения (взятых в данном порядке).

Заметим, что при четной перестановке прямых α_1, α_2, α_3 ориентация $o_1o_2o_3$ не меняется, а при нечетной — переходит в противоположную.

4. Стороны прямой на плоскости и плоскости в пространстве

Как известно, любая прямая α на плоскости разбивает эту плоскость на две *полуплоскости*: точки A и B плоскости, не принадлежащие прямой α, тогда и только тогда принадлежат одной полуплоскости, когда отрезок $\overline{AB}$ не пересекает прямую α.

Аналогично, любая плоскость Π в пространстве разбивает пространство на два *полупространства*, причем две точки A, B, не принадлежащие плоскости, тогда и только тогда принадлежат одному полупространству, когда отрезок $\overline{AB}$ не пересекает плоскость.

Определение 1. Говорят, что у прямой α в плоскости (или у плоскости Π в пространстве) выбрана *сторона*, если выбрана одна из полуплоскостей, на которые прямая α разбивает плоскость (соответственно, если выбрано одно из полупространств, на которые плоскость Π разбивает пространство). Выбранная полуплоскость (полупространство) или соответствующая сторона называется при этом *положительной*, а другая полуплоскость (полупространство) — *отрицательной*.

Пусть e — произвольный вектор, не параллельный прямой α (плоскости Π). Выбрав на прямой α (плоскости Π) произволь-

ную точку O, отложим от нее вектор $\boldsymbol{e}$, т. е. построим направленный отрезок $\overrightarrow{OE} = \boldsymbol{e}$. Если точка E окажется при этом в положительной полуплоскости (положительном полупространстве), то мы будем говорить, что вектор $\boldsymbol{e}$ *направлен в положительную сторону* прямой α (плоскости Π). Это определение нуждается, конечно, в проверке корректности, т. е. в доказательстве независимости полуплоскости (полупространства), содержащей точку E, от выбора точки O. Другими словами, мы должны доказать, что

если для точек O и O' прямой α (плоскости Π) и точек E и E', не принадлежащих этой прямой (этой плоскости), имеет место равенство $\overrightarrow{OE} = \overrightarrow{O'E'}$, то точки E и E' принадлежат одной полуплоскости (одному полупространству).

Пусть это не так, т. е. пусть отрезок $\overline{EE'}$ пересекается с прямой α (плоскостью Π). Ясно, что это возможно тогда и только тогда, когда прямые OO' и EE' пересекаются. С другой стороны, поскольку векторы $\overrightarrow{OE}$ и $\overrightarrow{O'E'}$ равны, эти прямые должны быть параллельны (причем совпадать они не могут). Полученное противоречие доказывает, что точки E и E' принадлежат одной и той же полуплоскости (полупространству).

Предложение 1. *Пусть $\boldsymbol{e}$ и $\boldsymbol{e}'$ — два коллинеарных вектора, не параллельных прямой α (плоскости Π), и пусть вектор $\boldsymbol{e}$ направлен в положительную сторону прямой α (плоскости Π). Вектор $\boldsymbol{e}'$ тогда и только тогда обладает тем же свойством, когда отношение $\boldsymbol{e}' : \boldsymbol{e}$ положительно.*

Доказательство. Пусть $\boldsymbol{e} = \overrightarrow{OE}$ и $\boldsymbol{e}' = \overrightarrow{OE'}$, где O — некоторая точка прямой α (плоскости Π). Точки E и E' тогда и только тогда лежат в разных полуплоскостях (полупространствах), когда отрезок $\overline{EE'}$ пересекается с прямой α (плоскостью Π). Но прямая EE' имеет с прямой α (плоскостью Π) одну-единственную точку пересечения O. Следовательно, точки E и E' тогда и только тогда лежат в разных полуплоскостях (полупространствах), когда отрезок $\overline{EE'}$ содержит точку O, т. е. когда $E \prec O \prec E'$ (в ориентации прямой EE', в которой $E \prec E'$). Для завершения доказательства осталось заметить, что $E \prec O \prec E'$ тогда и только тогда, когда отношение $\boldsymbol{e}' : \boldsymbol{e}$ отрицательно.

Определение 2. Мы будем говорить, что ориентированная прямая β *направлена в положительную сторону* прямой α (плоскости Π), если ее положительно ориентированный вектор $\boldsymbol{e}$ направлен в положительную сторону прямой α (плоскости Π).

Согласно предложению 1 это определение корректно.

Если в одной ориентации прямая β направлена в положительную сторону прямой α (плоскости Π), то в другой ориентации она будет направлена, очевидно, в отрицательную сторону. Это означает, что

выбор на прямой α (*плоскости* Π) *положительной стороны определяет на каждой прямой* β, *не параллельной прямой* α (*плоскости* Π), *некоторую ориентацию, а именно, ориентацию, в которой эта прямая направлена в положительную сторону прямой* α (*плоскости* Π).

Другими словами,

ориентацию прямой на плоскости или в пространстве можно определить как выбор стороны на некоторой прямой (*плоскости*), *не параллельной данной прямой.*

Поскольку ориентации непараллельных прямых α и β, лежащих в одной плоскости, определяют (см. п. 3) некоторую ориентацию плоскости и поскольку, согласно только что сказанному, ориентация прямой β определяется заданием стороны прямой α, мы получаем, что

для задания ориентации плоскости достаточно задать в этой плоскости ориентированную прямую и некоторую ее сторону.

Обратно,

для задания ориентации прямой достаточно задать некоторую ориентацию плоскости, содержащей эту прямую, и одну из сторон этой прямой в плоскости.

Аналогично,

для задания ориентации пространства достаточно задать ориентированную плоскость и некоторую ее сторону, а для задания ориентации плоскости достаточно задать ориентацию пространства и некоторую сторону этой плоскости.

Рассмотрим в заключение интересный и тонкий вопрос (часто вызывающий недоразумения), о сравнении ориентаций на двух различных прямых и плоскостях.

Понимая ориентации как классы одноименных базисов, мы видим, что ориентации на прямой, в плоскости или в пространстве определяются исключительно в терминах соответствующего линеала $\mathrm{Vect}(n)$ (так что, собственно говоря, следовало бы говорить об ориентациях именно этого линеала). Отсюда следует, что поскольку для параллельных прямых линеалы $\mathrm{Vect}(1)$ (рассматриваемые как подмножества линеалов $\mathrm{Vect}(2)$ или $\mathrm{Vect}(3)$) совпадают, то

задание ориентации на одной прямой автоматически определяет ориентацию на любой другой параллельной прямой.

Другими словами, это означает, что

ориентации параллельных прямых можно сравнивать, т. е. можно говорить о том, совпадают они или нет.

Конечно, аналогичные утверждения справедливы и для ориентаций параллельных плоскостей.

Совсем иначе дело обстоит со сравнением ориентаций двух непараллельных прямых (или плоскостей). На первый взгляд кажется, что можно предложить следующую процедуру сравне-

ния ориентаций, скажем, двух непараллельных прямых β_1 и β_2 на плоскости:

Пересечем прямые β_1 и β_2 некоторой третьей прямой α. Как мы знаем, ориентации прямых β_1 и β_2 определяются заданием некоторых сторон прямой α. В соответствии с этим условимся называть данные ориентации прямых β_1 и β_2 «одинаковыми», если они определяются одной и той же стороной прямой α.

Однако оказывается, что это определение некорректно (зависит от выбора прямой α).

Задание. Покажите на примере, что при одном выборе прямой α данные ориентации прямых β_1 и β_2 могут быть «одинаковыми», а при другом — «различными».

Таким образом, сравнение ориентаций на непараллельных прямых (а также на непараллельных плоскостях) невозможно.

В этом отношении ориентации прямых в принципе отличаются от ориентаций окружностей, сравнение которых возможно (ориентации двух окружностей, расположенных в одной плоскости, считаются, по определению, «одинаковыми», если они определяют одну и ту же ориентацию плоскости).

Причиной этого различия является то, что ориентация окружности определяет ориентацию плоскости, в которой расположена окружность, а ориентация прямой — нет (для того чтобы по ориентации прямой определить ориентацию содержащей ее плоскости, необходимо, как мы знаем, дополнительно задать сторону прямой в этой плоскости).

Суть дела здесь в том, что две полуплоскости, на которые прямая разбивает плоскость, совершенно равноправны (и потому без дополнительной информации неотличимы друг от друга), тогда как две области, на которые окружность разбивает плоскость («внешняя область» и «внутренняя область»), отличаются друг от друга по их геометрическим свойствам: например, внешняя область содержит целые прямые, тогда как внутренняя область (ограниченный окружностью круг) ни одной прямой не содержит.

К этому вопросу мы еще вернемся в п. 4 § 3 гл. 7.

5. Деформации базисов и ориентации

В этом пункте мы изложим еще один подход к понятию ориентации, отличающийся особой наглядностью.

Рассмотрим произвольную *вектор-функцию* $\boldsymbol{a}(t)$ числовой переменной t (пробегающей, скажем, замкнутый интервал $[0, 1]$ числовой оси $\mathbb{R}$), т. е. отображение, сопоставляющее каждому $t \in [0, 1]$ некоторый вектор $\boldsymbol{a}(t)$ (на прямой, в плоскости или в пространстве). При заданном базисе $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ (где, как всегда, $n = 1$, 2 или 3) вектор-функция $\boldsymbol{a}(t)$ однозначно определяется координатами

$$a^1(t), \ldots, a^n(t) \qquad (1)$$

вектора $\boldsymbol{a}(t)$ в базисе $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, являющимися обыкновенными числовыми функциями переменной t. Таким образом, задание вектор-функции $\boldsymbol{a}(t)$ равносильно (при фиксированном базисе) заданию n обыкновенных (числовых) функций (1).

Определение 1. Вектор-функция $\boldsymbol{a}(t)$ называется *непрерывной*, если все функции (1) непрерывны.

Из формул преобразования координат при замене базиса (п. 7 § 3) немедленно вытекает, что это определение корректно (не зависит от выбора базиса), т. е. что если координаты (1) вектор-функции $\boldsymbol{a}(t)$ непрерывны в одном базисе, то они непрерывны и в любом другом базисе.

Определение 2. Два базиса $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ и $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ (прямой, плоскости или пространства) называются *деформируемыми* друг в друга, если существуют такие векторные функции $\boldsymbol{a}_1(t), \ldots, \boldsymbol{a}_n(t)$ числового параметра $t \in [0,1]$, что:

1. Для любого $i = 1, \ldots, n$ вектор $\boldsymbol{a}_i(0)$ совпадает с вектором $\boldsymbol{a}_i$, а вектор $\boldsymbol{a}_i(1)$ — с вектором $\boldsymbol{b}_i$:

$$\boldsymbol{a}_1(0) = \boldsymbol{a}_1, \ldots, \boldsymbol{a}_n(0) = \boldsymbol{a}_n,$$
$$\boldsymbol{a}_1(1) = \boldsymbol{b}_1, \ldots, \boldsymbol{a}_n(1) = \boldsymbol{b}_n.$$

2. Для любого $i = 1, \ldots, n$ вектор $\boldsymbol{a}_i(t)$ непрерывно зависит от t.

3. Для любого $t \in [0, 1]$ векторы $\boldsymbol{a}_1(t), \ldots, \boldsymbol{a}_n(t)$ линейно независимы, т. е. составляют базис.

В этом случае о векторах $\boldsymbol{a}_1(t), \ldots, \boldsymbol{a}_n(t)$ говорят, что они составляют *деформацию*, связывающую базис $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ с базисом $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$.

Предложение 1. *Отношение деформируемости базисов является на множестве всех базисов отношением эквивалентности.*

Доказательство. Поскольку любой базис $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ может быть продеформирован сам в себя (соответствующую деформацию можно, например, определить формулой $\boldsymbol{a}_i(t) = \boldsymbol{a}_i$, $i = 1, \ldots, n$), отношение деформируемости базисов рефлексивно; оно и симметрично, так как если базис $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ связан с базисом $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ деформацией $\boldsymbol{a}_1(t), \ldots, \boldsymbol{a}_n(t)$, то базис $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ связан с базисом $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, например, деформацией $\boldsymbol{b}_1(t) = \boldsymbol{a}_1(1-t), \ldots, \boldsymbol{b}_n(t) = \boldsymbol{a}_n(1-t)$. Таким образом, нам нужно доказать только транзитивность, т. е. что если существует деформация $\boldsymbol{a}_1(t), \ldots, \boldsymbol{a}_n(t)$, связывающая базис $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ с базисом $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$, и деформация $\boldsymbol{b}_1(t), \ldots, \boldsymbol{b}_n(t)$, связывающая базис $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ с базисом $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n$, то существует и деформация $\boldsymbol{c}_1(t), \ldots, \boldsymbol{c}_n(t)$, связывающая базис $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ с базисом $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n$. Но такую деформацию ничего не стоит построить. Например, ее можно определить формулами

$$\boldsymbol{c}_i(t) = \begin{cases} \boldsymbol{a}_i(2t), & \text{если } 0 \leqslant t \leqslant 1/2, \\ \boldsymbol{b}_i(2t-1), & \text{если } 1/2 \leqslant t \leqslant 1. \end{cases}$$

Тем самым предложение 1 полностью доказано.

Из этого предложения следует, что множество всех базисов распадается на классы деформируемых друг в друга базисов. Оказывается, что эти классы совпадают с классами одноименных базисов, т. е. другими словами, имеет место следующее

Предложение 2. *Два базиса* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ *и* $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ *тогда и только тогда одноименны, когда они деформируемы друг в друга.*

Доказательство. Пусть базисы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ и $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ деформируемы друг в друга и пусть $\boldsymbol{a}_1(t), \ldots, \boldsymbol{a}_n(t)$ — деформация, связывающая базис $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ с базисом $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$. Рассмотрим определитель $\Delta(t)$ матрицы перехода от базиса $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ к базису $\boldsymbol{a}_1(t), \ldots, \boldsymbol{a}_n(t)$. Он непрерывно зависит от t, нигде не обращается в нуль и равен единице при $t=0$. Поэтому для всех значений t этот определитель положителен (ибо непрерывная функция не может сменить знак, не обратившись в нуль). В частности, он положителен при $t=1$. Но определитель $\Delta(1)$ является как раз определителем матрицы перехода от базиса $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ к базису $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$.

Поэтому базисы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ и $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ одноименны.

Докажем теперь обратное утверждение (одноименные базисы деформируемы). Пусть сначала $n=1$ (случай прямой).

Поскольку базисы $\boldsymbol{a}_1$ и $\boldsymbol{b}_1$ одноименны, имеет место равенство

$$\boldsymbol{a}_1 = k\boldsymbol{b}_1,$$

где $k>0$ (см. п. 1 § 2). Определим вектор-функцию $\boldsymbol{a}_1(t)$, $0 \leqslant t \leqslant 1$, формулой

$$\boldsymbol{a}_1(t) = (t + (1-t)k)\,\boldsymbol{b}_1.$$

Ясно, что $\boldsymbol{a}_1(0) = \boldsymbol{a}_1$, $\boldsymbol{a}_1(1) = \boldsymbol{b}_1$ и $\boldsymbol{a}_1(t)$ непрерывно зависит от t. Кроме того, поскольку $t + (1-t)k \neq 0$ при $0 \leqslant t \leqslant 1$, вектор $\boldsymbol{a}_1(t)$ является базисом. Следовательно, вектор-функция $\boldsymbol{a}_1(t)$ является деформацией, связывающей базис $\boldsymbol{a}_1$ с базисом $\boldsymbol{b}_1$.

При $n>1$ нам понадобится следующая

Лемма. *Пусть* $\boldsymbol{a}_1, \boldsymbol{a}_2$ *— произвольный базис на плоскости. Тогда любой другой базис* $\boldsymbol{b}_1, \boldsymbol{b}_2$ *деформируем либо в базис* $\boldsymbol{a}_1, \boldsymbol{a}_2$, *либо в базис* $\boldsymbol{a}_1, -\boldsymbol{a}_2$.

Аналогично, пусть $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ *— произвольный базис в пространстве. Тогда любой другой базис* $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$ *деформируем либо в базис* $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$, *либо в базис* $\boldsymbol{a}_1, \boldsymbol{a}_2, -\boldsymbol{a}_3$.

Предполагая эту лемму справедливой, мы можем без труда завершить доказательство предложения 2. Действительно, пусть базисы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ и $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ одноименны (где $n=2, 3$). Согласно лемме базис $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ деформируем либо в базис $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, либо в базис $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}, -\boldsymbol{a}_n$. Пусть базис $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ деформируем в базис $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}, -\boldsymbol{a}_n$. Тогда по уже доказанному, базис $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}, -\boldsymbol{a}_n$ одноименен с базисом

$\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$, а следовательно, и с базисом $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. Но это невозможно, поскольку матрица перехода от базиса $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}, -\boldsymbol{a}_n$ к базису $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ имеет вид

$$\begin{pmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 1 & 0 \\ 0 & & 0 & -1 \end{pmatrix}$$

и потому ее определитель отрицателен. Следовательно, базис $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ деформируем в базис $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$.

Предложение 2 означает, что

ориентации (прямой, плоскости или пространства) мы можем отождествлять с классами деформируемых друг в друга базисов.

Доказательство леммы. В первую очередь мы рассмотрим случай плоскости.

Отложив векторы $\boldsymbol{a}_1, \boldsymbol{a}_2$ и $\boldsymbol{b}_1, \boldsymbol{b}_2$ от некоторой точки O, мы можем написать, что

$$\boldsymbol{a}_1 = \overrightarrow{OA}_1, \quad \boldsymbol{a}_2 = \overrightarrow{OA}_2, \quad \boldsymbol{b}_1 = \overrightarrow{OB}_1, \quad \boldsymbol{b}_2 = \overrightarrow{OB}_2.$$

Любой поворот вокруг точки O треугольника OB_1B_2 представляет собой, очевидно, некоторую деформацию базиса $\boldsymbol{b}_1, \boldsymbol{b}_2$ (вопрос здесь может возникнуть только в отношении непрерывности; интуитивно она очевидна, а для ее строгого доказательства достаточно написать формулы, выражающие, как меняются при этом повороте векторы $\boldsymbol{b}_1$ и $\boldsymbol{b}_2$; такие формулы далее в своем месте мы получим, а пока мы предоставляем вывести их читателю). В то же время ясно, что таким поворотом мы можем совместить прямую OB_1 с прямой OA_1 и, более того, перевести вектор $\boldsymbol{b}_1$ в вектор прямой OA_1, одноименный с вектором $\boldsymbol{a}_1$.

Сделав это (и обозначив повернутые векторы прежними символами $\boldsymbol{b}_1$ и $\boldsymbol{b}_2$), мы будем теперь, оставляя вектор $\boldsymbol{b}_1$ неподвижным, вращать вокруг точки O лишь один вектор $\boldsymbol{b}_2 = \overrightarrow{OB_2}$, следя за тем, чтобы траектория точки B_2 не пересекала прямую $OB_1 = OA_1$. Ясно, что при этом мы получаем деформацию базиса $\boldsymbol{b}_1, \boldsymbol{b}_2$. Столь же ясно, что этой деформацией мы можем перевести вектор $\boldsymbol{b}_2$ в вектор, коллинеарный вектору $\boldsymbol{a}_2$ и потому одноименный (как вектор прямой OA_2) либо вектору $\boldsymbol{a}_2$, либо вектору $-\boldsymbol{a}_2$.

К получившимся векторам $\boldsymbol{b}_1$ и $\boldsymbol{b}_2$ остается теперь применить (к каждому отдельно) описанную выше (при рассмотрении

случая $n = 1$) деформацию, переводящую их в векторы $\boldsymbol{a}_1$ и $\pm\boldsymbol{a}_2$.

Поскольку комбинация нескольких последовательно выполненных деформаций также является деформацией, наша лемма тем самым для случая плоскости полностью доказана.

Случай пространства рассматривается аналогично. Пусть

$$\boldsymbol{a}_1 = \overrightarrow{OA}_1, \quad \boldsymbol{a}_2 = \overrightarrow{OA}_2, \quad \boldsymbol{a}_3 = \overrightarrow{OA}_3,$$
$$\boldsymbol{b}_1 = \overrightarrow{OB}_1, \quad \boldsymbol{b}_2 = \overrightarrow{OB}_2, \quad \boldsymbol{b}_3 = \overrightarrow{OB}_3.$$

Вращая тетраэдр $OB_1B_2B_3$ вокруг точки O, мы можем совместить плоскость OB_1B_2 с плоскостью OA_1A_2 и, более того, прямую OB_1 с прямой OA_1. Мы можем даже добиться того, чтобы вектор $\boldsymbol{b}_1$ перешел в вектор, одноименный с вектором $\boldsymbol{a}_1$, и чтобы векторы $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ составляли базис плоскости OA_1A_2, одноименный с базисом $\boldsymbol{a}_1$, $\boldsymbol{a}_2$. Проделав затем на плоскости OA_1A_2 описанную выше (при $n = 2$) деформацию (и оставляя вектор $\boldsymbol{b}_3$ неподвижным), мы продеформируем базис $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{b}_3$ в базис вида $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{b}_3$. Остается теперь, не трогая векторов $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$, повернуть вектор $\boldsymbol{b}_3$ так, чтобы он стал коллинеарным вектору $\boldsymbol{a}_3$, и затем применить уже известную нам деформацию на прямой. В результате всех этих деформаций мы и переведем базис $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{b}_3$ в базис вида $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\pm\boldsymbol{a}_3$. Тем самым наша лемма полностью доказана.

Замечание 1. Изложенное доказательство леммы может вызвать возражения по крайней мере по двум пунктам. Во-первых, оно использует «механические» понятия, которые мы в свое время (см. § 1) условились по возможности исключать. Во-вторых, хотя вся излагаемая здесь теория по существу не связана с понятием измерения (носит, как говорят, аффинный характер; см. ниже), в этом доказательстве используются преобразования (вращения), определение которых существенно опирается на понятие длины. Оба эти возражения можно снять, перестроив доказательство на алгебраический лад, для чего достаточно описать все требуемые преобразования формулами, вообще не упоминая об их геометрическом смысле. Впрочем, если уж становиться на этот путь, то можно получить и более простые формулы (но уже не имеющие столь прозрачного элементарно-геометрического истолкования). Мы сделаем это для плоскости, оставляя читателю аналогичную «алгебраизацию» доказательства нашей леммы для случая пространства.

Пусть $\begin{pmatrix} p_1 & q_1 \\ p_2 & q_2 \end{pmatrix}$ — матрица перехода от базиса $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ к базису $\boldsymbol{b}_1$, $\boldsymbol{b}_2$. Таким образом,

$$\boldsymbol{b}_1 = p_1\boldsymbol{a}_1 + p_2\boldsymbol{a}_2,$$
$$\boldsymbol{b}_2 = q_1\boldsymbol{a}_1 + q_2\boldsymbol{a}_2.$$

Рассмотрим определитель

$$\Delta = \begin{vmatrix} p_1 & q_1 \\ p_2 & q_2 \end{vmatrix} = p_1 q_2 - p_2 q_1$$

этой матрицы. Как мы знаем, он отличен от нуля.

Случай 1. Числа $p_1 q_2$ и $p_2 q_1$ имеют разные знаки (и оба отличны от нуля). В этом случае для любого значения параметра t число

$$p_1 q_2 - (1-t)^2 p_2 q_1,$$

очевидно, отлично от нуля. Но это число является определителем матрицы

$$\begin{pmatrix} p_1 & (1-t)\, q_1 \\ (1-t)\, p_2 & q_2 \end{pmatrix}.$$

Поэтому, положив

$$\begin{aligned} \boldsymbol{b}_1(t) &= p_1 \boldsymbol{a}_1 + (1-t)\, p_2 \boldsymbol{a}_2, \\ \boldsymbol{b}_2(t) &= (1-t)\, q_1 \boldsymbol{a}_1 + q_2 \boldsymbol{a}_2, \end{aligned} \qquad 0 \leqslant t \leqslant 1,$$

мы получим деформацию базиса $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ в базис $\boldsymbol{b}_1' = \boldsymbol{b}_1(1)$, $\boldsymbol{b}_2' = \boldsymbol{b}_2(1)$, с которым базис $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ связан диагональной матрицей перехода

$$\begin{pmatrix} p_1 & 0 \\ 0 & q_2 \end{pmatrix}.$$

Случай 2. Числа $p_1 q_2$ и $p_2 q_1$ имеют один и тот же знак (и одно из них, возможно, равно нулю). Если при этом $|p_1 q_2| > |p_2 q_1|$, то будет, как легко видеть, применима та же самая деформация, что и выше, и мы снова сможем продеформировать базис $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ в базис $\boldsymbol{b}_1'$, $\boldsymbol{b}_2'$, связанный с базисом $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ диагональной матрицей перехода. Однако, если $|p_1 q_2| < |p_2 q_1|$ (заметим, что равенство $|p_1 q_2| = |p_2 q_1|$ невозможно), то такую деформацию мы применить не сможем и вместо нее должны будем рассмотреть, скажем, деформацию

$$\begin{aligned} \boldsymbol{b}_1(t) &= (1-t)\, p_1 \boldsymbol{a}_1 + p_2 \boldsymbol{a}_2, \\ \boldsymbol{b}_2(t) &= q_1 \boldsymbol{a}_1 + (1-t)\, q_2 \boldsymbol{a}_2, \end{aligned}$$

переводящую базис $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ в базис $\boldsymbol{b}_1'$, $\boldsymbol{b}_2'$, с которым базис $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ связан матрицей перехода

$$\begin{pmatrix} 0 & q_1 \\ p_2 & 0 \end{pmatrix}.$$

Тем не менее, применив еще дополнительную деформацию

$$\begin{aligned} \boldsymbol{b}_1'(t) &= \left(\cos\frac{\pi}{2}\, t\right) \boldsymbol{b}_1' + \left(-\sin\frac{\pi}{2}\, t\right) \boldsymbol{b}_2', \\ \boldsymbol{b}_2'(t) &= \left(\sin\frac{\pi}{2}\, t\right) \boldsymbol{b}_1' + \left(\cos\frac{\pi}{2}\, t\right) \boldsymbol{b}_2', \end{aligned} \qquad 0 \leqslant t \leqslant 1,$$

мы и в этом случае продеформируем базис $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ в базис, с которым базис $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ связан диагональной матрицей перехода.

Таким образом, мы без ограничения общности можем с самого начала считать, что базис $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ связан с базисом $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ диагональной матрицей перехода:

$$\boldsymbol{b}_1 = p\boldsymbol{a}_1,$$
$$\boldsymbol{b}_2 = q\boldsymbol{a}_2$$

(мы здесь убрали уже ненужные индексы). Но в этом предположении мы можем к каждому из векторов $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ применить деформацию, использованную нами в случае $n = 1$. Например, при $p > 0$ мы можем для вектора $\boldsymbol{b}_1$ рассмотреть деформацию $(t + (1-t)p)\boldsymbol{a}_1$, а при $p < 0$ — деформацию $(-t + (1-t)p)\boldsymbol{a}_1 = (t + (1-t)(-p))(-\boldsymbol{a}_1)$. В результате мы продеформируем базис $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ в базис (векторы которого мы снова будем обозначать через $\boldsymbol{b}_1$, $\boldsymbol{b}_2$), связанный с базисом $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ формулами

$$\boldsymbol{b}_1 = \pm \boldsymbol{a}_1,$$
$$\boldsymbol{b}_2 = \pm \boldsymbol{a}_2.$$

Если $\boldsymbol{b}_1 = \boldsymbol{a}_1$, $\boldsymbol{b}_2 = \boldsymbol{a}_2$ или $\boldsymbol{b}_1 = \boldsymbol{a}_1$, $\boldsymbol{b}_2 = -\boldsymbol{a}_2$, то все в порядке: лемма доказана. Пусть $\boldsymbol{b}_1 = -\boldsymbol{a}_1$, $\boldsymbol{b}_2 = \boldsymbol{a}_2$ или $\boldsymbol{b}_1 = -\boldsymbol{a}_1$, $\boldsymbol{b}_2 = -\boldsymbol{a}_2$. Тогда мы применим еще дополнительную деформацию

$$\boldsymbol{b}_1(t) = (\cos \pi t)\,\boldsymbol{b}_1 + (-\sin \pi t)\,\boldsymbol{b}_2,$$
$$\boldsymbol{b}_2(t) = (\sin \pi t)\,\boldsymbol{b}_1 + (\cos \pi t)\,\boldsymbol{b}_2.$$

Ясно, что это действительно деформация и что она переводит базис $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ в базис $-\boldsymbol{b}_1$, $\boldsymbol{b}_2$, т. е. при наших предположениях либо в базис $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, либо в базис $\boldsymbol{a}_1$, $-\boldsymbol{a}_2$.

Тем самым лемма полностью доказана и этим методом (для случая плоскости).

Задание. Выясните геометрический смысл деформаций, участвующих в «алгебраическом» доказательстве леммы.

Упражнение. Аналогичным образом докажите лемму для случая пространства.

6. Резюме

Подведем итоги.

Наглядный смысл ориентации прямой состоит в выборе некоторого (из двух возможных) направления движения по этой прямой, а ориентации плоскости — в выборе направления вращения в этой плоскости (т. е. задания направления движения по любой окружности в плоскости).

Аналогично, в пространстве ориентации можно представлять себе как выбор одного из двух возможных типов винтового движения («правовинтового» и «левовинтового»).

Формально ориентацию прямой, плоскости или пространства можно определить тремя равносильными способами:

1) либо как класс *одноименных базисов*;

2) либо как класс *одноименных упорядоченных систем точек общего положения*, состоящих в случае прямой из двух точек, в случае плоскости — из трех точек и в случае пространства — из четырех точек;

3) либо как класс *деформируемых друг в друга базисов.*

При этом

1) ориентация прямой в плоскости (в пространстве) однозначно определяется заданием в этой плоскости (в пространстве) некоторой стороны произвольной прямой (плоскости), не параллельной данной прямой;

2) ориентация плоскости (пространства) однозначно определяется заданием в плоскости (в пространстве) произвольной ориентированной прямой (плоскости) вместе с некоторой ее стороной;

3) обратно, ориентация прямой в плоскости (плоскости в пространстве) однозначно определяется заданием стороны этой прямой (плоскости) и некоторой ориентации плоскости (пространства);

4) наконец, сторона прямой в плоскости (плоскости в пространстве) однозначно определяется заданием ориентаций прямой и плоскости (плоскости и пространства).

Свойства 2), 3) и 4) можно объединить в следующей формулировке:

Из трех объектов:

а) *ориентация прямой в плоскости* (*плоскости в пространстве*),

б) *ориентация плоскости* (*пространства*),

в) *сторона прямой в плоскости* (*плоскости в пространстве*)

любые два однозначно определяют третий.

Дополнение. О понятии угла

Известно, что понятие угла довольно трудно строго определить в достаточно общем виде. Возникающие здесь трудности объясняются не только тем, что слово «угол» используется для обозначения целого ряда различных, хотя и связанных друг с другом понятий (фигура, образованная парой лучей с общим началом; часть плоскости, содержащаяся «между» двумя такими лучами; число, являющееся мерой поворота, на который нужно повернуть один луч до совпадения с другим; аргумент тригонометрических функций), но и реальной сложностью понятия угла.

В этом дополнении мы обсудим понятие угла, оставаясь на наглядных, не формальных позициях.

Термин *угол* мы будем использовать в следующих двух основных значениях:

а) фигура, образованная парой лучей с общим началом;

б) число, измеряющее (в радианах) поворот, на который нужно повернуть один луч до совпадения с другим.

Как правило, это двойное понимание термина «угол» к недоразумениям не приводит. В случаях, когда такое недоразумение возможно, мы будем называть угол в смысле б) *мерой угла*.

По определению, мера угла θ заключена в пределах от 0 до π. Случаи, когда она равна 0 или π, не исключаются:

$$0 \leqslant \theta \leqslant \pi.$$

При $\theta = 0$ угол называется *вырожденным*, а при $\theta = \pi$ — *развернутым*.

Пусть две прямые α и β лежат в одной плоскости и пересекаются в (единственной) точке O. Эта точка делит каждую прямую на два луча. Выбрав на каждой из прямых по такому лучу, мы получим некоторый угол. В зависимости от выбора лучей возникает, таким образом, четыре угла. Они называются (элементарно-геометрическими) *углами между прямыми* α *и* β. Эти углы попарно равны (вертикальные углы), так что фактически угол между прямыми имеет только два значения θ_1 и θ_2. Сумма этих углов равна π (смежные углы). В частности, равенство $\theta_1 = \theta_2$ возможно только при $\theta_1 = \theta_2 = \frac{\pi}{2}$, т. е. для перпендикулярных прямых.

Если прямые α и β параллельны (в частности, совпадают), углы θ_1 и θ_2, по определению, считаются равными 0 и π (подчеркнем, что здесь мы нуждаемся в отдельном определении).

Угол между прямыми, не лежащими в одной плоскости, равен, по определению, углу между параллельными им прямыми, лежащими в одной плоскости.

Двузначность угла между прямыми, естественно, очень неудобна. Положение может быть исправлено, если предположить прямые α и β ориентированными. В этом случае на каждой из этих прямых выделяется единственный луч с началом в точке O (так называемый *положительный луч*, точки A которого характеризуются требованием, чтобы вектор $\overrightarrow{OA}$ был положительно ориентирован), и тем самым между прямыми возникает единственный угол θ. Этот угол называется *углом между ориентированными прямыми* α *и* β.

Для параллельных прямых, по определению, полагается $\theta = 0$, если прямые ориентированы одинаково (см. конец п. 4), и $\theta = \pi$, если их ориентации различны.

По определению,

$$0 \leqslant \theta \leqslant \pi$$

и потому *угол* θ *однозначно определен, если известен его косинус* $\cos\theta$.

Если плоскость Π, в которой расположены (ориентированные) прямые α и β, также ориентирована, то определен так называемый *ориентированный угол* φ *от прямой* α *к прямой* β. По определению,

$$\varphi = \pm\theta,$$

причем $\varphi = \theta$, если произведение ориентаций прямых α и β (см. п. 3) совпадает с ориентацией плоскости Π, и $\varphi = -\theta$ в противном случае. Если прямые α и β параллельны, то, по определению, $\varphi = \theta$.

Таким образом,

$$-\pi < \varphi \leqslant \pi$$

и *угол* φ *однозначно определен, если известны его тригонометрические функции* $\sin\varphi$ *и* $\cos\varphi$.

Наглядно φ является углом, на который нужно в заданном на плоскости Π направлении вращения повернуть прямую α так, чтобы она совпала с прямой β (и при этом совпали бы ориентации этих прямых).

В отличие от угла θ угол φ зависит от порядка, в котором рассматриваются прямые α и β: при перестановке этих прямых он меняет знак (за исключением, правда, случая $\varphi = \pi$, когда угол φ остается — в силу принятых выше определений — тем же самым).

Во избежание исключений и утомительных оговорок иногда удобно считать, что угол φ принимает значения из всей числовой оси:

$$-\infty < \varphi < +\infty,$$

предполагая при этом, что значения, отличающиеся на 2π, выражают одну и ту же геометрическую ситуацию. Например, в силу этого соглашения утверждение об изменении знака угла φ при перестановке прямых α и β будет верно уже без всяких исключений.

§ 5. МЕТРИЧЕСКАЯ ТЕОРИЯ ВЕКТОРОВ

В предыдущих параграфах измерение длин играло чисто вспомогательную роль (а измерение углов вообще не встречалось). Более того, при желании понятие длины можно было бы совсем исключить (правда, ценой некоторого усложнения теории). Часть геометрии, не опирающаяся на понятие длины (и величины угла), называется *аффинной геометрией* [1]). Таким образом, можно сказать, что до сих пор мы занимались *аффинной теорией векторов*. Теперь же мы перейдем к их *метрической теории*, существенно опирающейся на понятие измерения.

Во всем этом параграфе мы будем предполагать выбранным и фиксированным некоторый эталон длины (один и тот же для всех прямых).

1. Длина вектора и угол между векторами

Важнейшими метрическими понятиями являются понятия длины вектора и угла между двумя векторами.

Длину $|a|$ вектора $a = \overrightarrow{AB}$ мы определили в п. 2 § 2 для векторов на прямой посредством формулы

$$|a| = |AB|.$$

То же самое определение мы сохраним и для векторов на плоскости или в пространстве. Оно, конечно, нуждается в проверке корректности, т. е. в доказательстве того, что

[1]) Более четкое определение аффинной геометрии мы дадим ниже, когда у нас будет подготовлен необходимый для этого материал.

эквиполлентные направленные отрезки $\overrightarrow{AB}$ *и* $\overrightarrow{A_1B_1}$ *имеют равные длины.*

Но это очевидно. Действительно, если отрезки $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ расположены на одной прямой, то равенство их длин уже доказано в п. 2 § 2. Если же эти отрезки расположены на различных прямых (и потому, в частности, невырождены), то они являются противоположными сторонами некоторого параллелограмма и потому их длины одинаковы.

Заметим, что

длина $|\boldsymbol{a}|$ *вектора* $\boldsymbol{a}$ *тогда и только тогда равна нулю, когда* $\boldsymbol{a} = \boldsymbol{0}$.

Перейдем теперь к углу между векторами.

Пусть $\boldsymbol{a}$ и $\boldsymbol{b}$ — два отличных от нуля вектора на плоскости. Отложив их от некоторой точки O, мы можем написать, что $\boldsymbol{a} = \overrightarrow{OA}$ и $\boldsymbol{b} = \overrightarrow{OB}$. Рассмотрим прямые $\alpha = OA$ и $\beta = OB$, ориентированные таким образом, чтобы векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ (рассматриваемые как векторы на прямых) были положительно ориентированными (см. п. 1 § 2).

Определение 1. Угол θ между ориентированными прямыми α и β (см. дополнение к § 4) называется *углом между векторами* $\boldsymbol{a}$ *и* $\boldsymbol{b}$ и обозначается символом $\widehat{(\boldsymbol{a}, \boldsymbol{b})}$.

Это определение нуждается, конечно, в проверке корректности, т. е. в доказательстве того, что угол $\widehat{(\boldsymbol{a}, \boldsymbol{b})}$ не зависит от выбора точки O. Мы докажем корректность, выразив этот угол через величины, от точки O заведомо не зависящие.

Рассмотрим с этой целью проекцию $\mathrm{pr}_{\alpha}\boldsymbol{b} = \overrightarrow{OB'}$ вектора $\boldsymbol{b}$ на прямую α (на направление вектора $\boldsymbol{a}$) параллельно прямой, перпендикулярной прямой α. (Такого рода проекции называются *ортогональными* или *прямоугольными проекциями*; они должны быть знакомы читателю из школьного курса геометрии. В этом параграфе никаких проекций, кроме ортогональных, мы рассматривать не будем.)

Предположим сначала, что прямые α и β не параллельны и не перпендикулярны. Тогда среди двух элементарно-геометрических углов θ_1 и θ_2 между этими прямыми можно выделить наименьший (острый) угол. Обозначим этот угол символом θ'.

Угол θ' является не чем иным, как углом при вершине O в прямоугольном треугольнике OBB', и потому

$$\cos\theta' = \frac{|OB'|}{|OB|} = \frac{|\mathrm{pr}_{\alpha}\boldsymbol{b}|}{|\boldsymbol{b}|}.$$

С углом θ угол θ' связан, очевидно, соотношениями

$$\theta = \begin{cases} \theta', & \text{если угол } \theta \text{ острый,} \\ \pi - \theta', & \text{если угол } \theta \text{ тупой.} \end{cases}$$

Следовательно,

$$\cos\theta = \begin{cases} \cos\theta', & \text{если угол } \theta \text{ острый,} \\ -\cos\theta', & \text{если угол } \theta \text{ тупой.} \end{cases}$$

С другой стороны, ясно, что если угол θ острый, то базисы $\boldsymbol{a}$ и $\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b}$ прямой α одноименны, а если угол θ тупой, то эти базисы разноименны. (Наглядно этот факт очевиден — достаточно сделать чертеж; формальное доказательство мы опустим.) Другими словами, для величины $(\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b})$ вектора $\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b}$ (см. п. 2 § 2) на ориентированной указанным выше образом прямой α имеют место равенства

$$(\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b}) = \begin{cases} |\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b}|, & \text{если угол } \theta \text{ острый,} \\ -|\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b}|, & \text{если угол } \theta \text{ тупой.} \end{cases}$$

Сопоставляя все сказанное, мы немедленно получаем, что как при тупом, так и при остром угле θ справедливо равенство

$$\cos\theta = \frac{(\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b})}{|\boldsymbol{b}|}. \tag{1}$$

Эта формула остается справедливой и тогда, когда прямые α и β перпендикулярны $\left(\theta = \frac{\pi}{2}\right)$, поскольку в этом случае обе ее стороны равны нулю, а также когда прямые α и β параллельны ($\theta = 0$ или π), поскольку при $\theta = 0$ обе ее стороны равны, очевидно, $+1$, а при $\theta = \pi$ равны -1.

Таким образом, мы доказали, что

для угла θ между любыми двумя (отличными от нуля) векторами $\boldsymbol{a}$ и $\boldsymbol{b}$ справедлива формула (1).

Поскольку, по определению, $0 \leqslant \theta \leqslant \pi$, косинус $\cos\theta$ однозначно определяет угол θ, и следовательно,

формула (1) *однозначно определяет угол между векторами.*

В частности, это доказывает корректность определения угла θ, поскольку величина

$$\frac{(\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b})}{|\boldsymbol{b}|}$$

не зависит от выбора точки O.

Замечание 1. Можно «обратить» предыдущие рассуждения и принять формулу (1) за о п р е д е л е н и е угла θ между векторами. После этого можно определить угол между двумя ориентированными прямыми как угол между их любыми положительно ориентированными базисами (это определение нуждается, конечно, в проверке корректности). Это доставляет нам уже вполне строгое определение угла между прямыми.

Угол можно определить и между векторами на прямой. Именно, угол $\theta = \widehat{(\boldsymbol{a}, \boldsymbol{b})}$ считается, по определению, равным нулю, если векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ одноименны (см. п. 1 § 2), и равным π в противном случае. Формула (1) сохраняется и для векторов на прямой, если под вектором $(\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b})$ понимать вектор $\boldsymbol{b}$.

Таким образом, угол между двумя коллинеарными векторами на плоскости, т. е. векторами, параллельными некоторой прямой α, не зависит от того, рассматриваем ли мы эти векторы как векторы на плоскости или как векторы на прямой α. В обоих случаях этот угол равен нулю, если векторы «одинаково направлены» (их отношение положительно), и равен π, если они «противоположно направлены» (их отношение отрицательно).

Наконец, угол между векторами в пространстве сводится к углу между векторами на плоскости, поскольку любые два вектора параллельны некоторой плоскости. (Независимость этого угла от выбора плоскости очевидна, поскольку для неколлинеарных векторов эта плоскость определена — с точностью до параллельности — единственным образом, а угол между коллинеарными векторами сводится к углу между векторами на прямой.) Для угла между векторами в пространстве формула (1) также сохраняется (только, конечно, под $\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b}$ следует теперь понимать проекцию параллельно плоскости, перпендикулярной прямой $\alpha = OA$).

По самому определению, угол между двумя векторами в пространстве не зависит от того, рассматриваем ли мы эти векторы как векторы в пространстве или как векторы в соответствующей плоскости.

Угол между векторами, среди которых хотя бы один равен нулю, не определяется.

На ориентированной плоскости наряду с углом θ между векторами $\boldsymbol{a} = \overrightarrow{OA}$ и $\boldsymbol{b} = \overrightarrow{OB}$ можно рассматривать также угол φ *от вектора* $\boldsymbol{a}$ *к вектору* $\boldsymbol{b}$. По определению, он равен (ориентированному) углу от прямой $\alpha = OA$ к прямой $\beta = OB$. С углом θ при $0 \leqslant \theta < \pi$ этот угол связан соотношениями

$$\varphi = \pm\theta;$$

при $\theta = \pi$ имеем $\varphi = \pi$. Угол φ однозначно определен, если известны его тригонометрические функции $\cos\varphi$ и $\sin\varphi$.

Пусть α — ориентированная прямая и $\boldsymbol{b}$ — произвольный отличный от нуля вектор. Ясно, что для любого положительно ориентированного вектора $\boldsymbol{a}$ прямой α угол от $\boldsymbol{a}$ к $\boldsymbol{b}$ не зависит от выбора вектора $\boldsymbol{a}$, т. е. определяется исключительно прямой α и вектором $\boldsymbol{b}$. Этот угол называется *наклоном* вектора $\boldsymbol{b}$ к прямой α.

2. Скалярное произведение векторов

По самому определению угол между векторами $\boldsymbol{a}$ и $\boldsymbol{b}$ не зависит от их порядка, т. е. равен углу между векторами $\boldsymbol{b}$ и $\boldsymbol{a}$. Сопоставляя это замечание с формулой (1) предыдущего пункта, мы немедленно получаем, что для любых двух отличных от нуля векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ (на прямой, в плоскости или в пространстве) имеет место соотношение

$$\frac{(\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b})}{|\boldsymbol{b}|}=\frac{(\mathrm{pr}_{\boldsymbol{b}}\boldsymbol{a})}{|\boldsymbol{a}|}.$$

Умножив это равенство на произведение $|\boldsymbol{a}|\cdot|\boldsymbol{b}|$ длин данных векторов, мы получим соотношение

$$|\boldsymbol{a}|\cdot(\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b})=|\boldsymbol{b}|\cdot(\mathrm{pr}_{\boldsymbol{b}}\boldsymbol{a}). \tag{1}$$

Предположим теперь, что один из рассматриваемых векторов (скажем, вектор $\boldsymbol{a}$) равен нулю, тогда как другой вектор (вектор $\boldsymbol{b}$) по-прежнему отличен от нуля. В этом случае имеет смысл только выражение $|\boldsymbol{b}|\cdot(\mathrm{pr}_{\boldsymbol{b}}\boldsymbol{a})$ (причем оно равно нулю), тогда как выражение $|\boldsymbol{a}|\cdot(\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b})$ смысла не имеет (ибо вектор $\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b}$ при $\boldsymbol{a}=0$ не определен). В связи с этим удобно принять следующее общее

Соглашение. Произведение двух величин, одна из которых равна нулю, а другая не определена, считается, по определению, равным нулю.

Это соглашение оправдывается тем, что как бы мы не доопределили второй множитель, произведение все равно будет равно нулю (за счет равенства нулю первого множителя).

В соответствии с этим соглашением мы можем теперь утверждать, что

равенство (1) *справедливо для любых двух векторов* $\boldsymbol{a}$ *и* $\boldsymbol{b}$.

Определение 1. Число $|\boldsymbol{a}|\cdot(\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b})$ (или $|\boldsymbol{b}|\cdot(\mathrm{pr}_{\boldsymbol{b}}\boldsymbol{a})$) называется *скалярным произведением* векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ и обозначается символом $\boldsymbol{ab}$ (а также символом $(\boldsymbol{a},\boldsymbol{b})$):

$$\boldsymbol{ab}=|\boldsymbol{a}|\cdot(\mathrm{pr}_{\boldsymbol{a}}\boldsymbol{b})=|\boldsymbol{b}|\cdot(\mathrm{pr}_{\boldsymbol{b}}\boldsymbol{a}).$$

Замечание 1. Эпитет «скалярное» обязан своим происхождением тому, что числа (а значение скалярного произведения, по определению, является числом) часто называются *скалярами.*

Пользуясь формулой (1) предыдущего пункта, мы можем формулу для скалярного произведения записать в следующем более симметричном виде:

$$\boldsymbol{ab}=|\boldsymbol{a}|\cdot|\boldsymbol{b}|\cdot\cos(\widehat{\boldsymbol{a},\boldsymbol{b}}). \tag{2}$$

В соответствии с принятым выше общим соглашением эта формула справедлива для любых векторов $\boldsymbol{a}$ и $\boldsymbol{b}$.

Таким образом,

скалярное произведение векторов равно произведению их длин на косинус угла между ними.

Если $\boldsymbol{a} = \boldsymbol{b}$, то $(\widehat{\boldsymbol{a}, \boldsymbol{b}}) = 0$, и потому

$$\boldsymbol{a}\boldsymbol{a} = \boldsymbol{a}^2 = |\boldsymbol{a}|^2. \tag{3}$$

Таким образом,

скалярный квадрат вектора равен квадрату его длины.

Эту формулу удобно писать «наоборот»:

$$|\boldsymbol{a}| = \sqrt{\boldsymbol{a}^2}. \tag{3'}$$

Таким образом,

длина вектора равна корню квадратному из его скалярного квадрата.

Аналогично, написав «наоборот» формулу (2), мы получим для угла $\theta = (\widehat{\boldsymbol{a}, \boldsymbol{b}})$ формулу

$$\cos\theta = \frac{\boldsymbol{a}\boldsymbol{b}}{|\boldsymbol{a}| \cdot |\boldsymbol{b}|}. \tag{2'}$$

Таким образом,

косинус угла между (отличными от нуля) векторами равен их скалярному произведению, деленному на произведение их длин.

Замечание 2. Мы видим, что через скалярное произведение выражаются обе основные метрические характеристики векторов (длина и угол).

Векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ называются *ортогональными*, если их скалярное произведение равно нулю:

$$\boldsymbol{a}\boldsymbol{b} = 0.$$

Таким образом,

векторы ортогональны, если хотя бы один из них равен нулю или если эти векторы перпендикулярны (угол между ними равен $\pi/2$).

Докажем теперь следующую важную теорему:

Теорема 1. *Для любых векторов $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ и любого числа k имеют место соотношения*

1. $\boldsymbol{a}\boldsymbol{b} = \boldsymbol{b}\boldsymbol{a}$,
2. $(\boldsymbol{a} + \boldsymbol{b})\,\boldsymbol{c} = \boldsymbol{a}\boldsymbol{c} + \boldsymbol{b}\boldsymbol{c}$,
3. $(k\boldsymbol{a})\,\boldsymbol{b} = k\,(\boldsymbol{a}\boldsymbol{b})$,
4. $\boldsymbol{a}^2 \geqslant 0$, *причем* $\boldsymbol{a}^2 = 0$ *тогда и только тогда, когда* $\boldsymbol{a} = \boldsymbol{0}$.

Доказательство. Свойство 1 очевидно. Свойство 2 непосредственно вытекает из соответствующего свойства величин проекций:

$$(\boldsymbol{a} + \boldsymbol{b})\,\boldsymbol{c} = |\boldsymbol{c}|\,(\mathrm{pr}_{\boldsymbol{c}}(\boldsymbol{a} + \boldsymbol{b})) = |\boldsymbol{c}|\,((\mathrm{pr}_{\boldsymbol{c}}\boldsymbol{a}) + (\mathrm{pr}_{\boldsymbol{c}}\boldsymbol{b})) =$$
$$= |\boldsymbol{c}|\,(\mathrm{pr}_{\boldsymbol{c}}\boldsymbol{a}) + |\boldsymbol{c}|\,(\mathrm{pr}_{\boldsymbol{c}}\boldsymbol{b}) = \boldsymbol{a}\boldsymbol{c} + \boldsymbol{b}\boldsymbol{c}.$$

Свойство 3 доказывается аналогично:

$$(k\boldsymbol{a})\,\boldsymbol{b} = |\,\boldsymbol{b}\,|(\mathrm{pr}_{\boldsymbol{b}}\,k\boldsymbol{a}) = k|\,\boldsymbol{b}\,|\,\mathrm{pr}_{\boldsymbol{b}}\,\boldsymbol{a} = k\,(\boldsymbol{ab}).$$

Наконец, свойство 4 непосредственно вытекает из формулы (3).

Свойство 1 — это свойство коммутативности, свойство 2 — свойство дистрибутивности, свойство 3 — свойство однородности, а свойство 4 — свойство положительной определенности. Вместе свойства 2 и 3 называются свойством линейности. Свойства коммутативности и линейности показывают, что вычисления со скалярным произведением подчиняются обычным правилам арифметики (раскрытие скобок и т. п.).

В дальнейшем мы постараемся (когда это возможно) пользоваться только доказанными четырьмя свойствами скалярного умножения и не будем обращаться к его определению (2).

3. Применение скалярного умножения к доказательству геометрических теорем

Пользуясь скалярным умножением, можно чисто алгебраически доказывать разнообразные теоремы геометрии, связанные с понятиями длины и угла.

Пример 1. Пусть ABC — произвольный прямоугольный треугольник (с прямым углом в вершине C). Рассмотрим векторы $\boldsymbol{a} = \overrightarrow{CB}$, $\boldsymbol{b} = \overrightarrow{AC}$ и $\boldsymbol{c} = \overrightarrow{AB}$. Так как $\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b}$ и $\boldsymbol{ab} = 0$, то $|\,\boldsymbol{c}\,|^2 = (\boldsymbol{a} + \boldsymbol{b})^2 = \boldsymbol{a}^2 + 2\boldsymbol{ab} + \boldsymbol{b}^2 = \boldsymbol{a}^2 + \boldsymbol{b}^2 = |\,\boldsymbol{a}\,|^2 + |\,\boldsymbol{b}\,|^2$, т. е. $|\,AB\,|^2 = |\,CB\,|^2 + |\,AC\,|^2$. Тем самым мы доказали теорему Пифагора (квадрат гипотенузы равен сумме квадратов катетов).

Заметим, что «в векторной форме» теорему Пифагора можно, таким образом, сформулировать в следующем виде:

для любых двух ортогональных векторов $\boldsymbol{a}$ *и* $\boldsymbol{b}$ *имеет место равенство*

$$|\,\boldsymbol{a} + \boldsymbol{b}\,|^2 = |\,\boldsymbol{a}\,|^2 + |\,\boldsymbol{b}\,|^2.$$

В этой форме она, очевидно, справедлива и в случае, когда хотя бы один из векторов $\boldsymbol{a}$ или $\boldsymbol{b}$ равен нулю.

Пример 2. Пусть $ABCD$ — произвольный параллелограмм. Рассмотрим векторы $\boldsymbol{a} = \overrightarrow{AB}$, $\boldsymbol{b} = \overrightarrow{BD}$ и $\boldsymbol{c} = \overrightarrow{AC}$ и $\boldsymbol{d} = \overrightarrow{AD}$. Ясно, что $\boldsymbol{d} = \boldsymbol{a} + \boldsymbol{b}$ и $\boldsymbol{c} = \boldsymbol{d} + \boldsymbol{a}$. Поэтому

$$|\,\boldsymbol{b}\,|^2 = \boldsymbol{b}^2 = (\boldsymbol{d} - \boldsymbol{a})^2 = \boldsymbol{d}^2 - 2\,\boldsymbol{da} + \boldsymbol{a}^2,$$
$$|\,\boldsymbol{c}\,|^2 = \boldsymbol{c}^2 = (\boldsymbol{d} + \boldsymbol{a})^2 = \boldsymbol{d}^2 + 2\,\boldsymbol{da} + \boldsymbol{a}^2,$$

и следовательно,

$$|\,\boldsymbol{b}\,|^2 + |\,\boldsymbol{c}\,|^2 = 2\,(|\,\boldsymbol{d}\,|^2 + |\,\boldsymbol{a}\,|^2),$$

т. е.

$$|\,BD\,|^2 + |\,AC\,|^2 = |\,AD\,|^2 + |\,BC\,|^2 + |\,AB\,|^2 + |\,CD\,|^2.$$

Тем самым мы доказали известную теорему о диагоналях параллелограмма (сумма квадратов диагоналей параллелограмма равна сумме квадратов его сторон).

Заметим, что в «векторной форме» эта теорема сводится к тривиальному алгебраическому тождеству $(\boldsymbol{d}-\boldsymbol{a})^2+(\boldsymbol{d}+\boldsymbol{a})^2=2(\boldsymbol{d}^2+\boldsymbol{a}^2)$.

Пример 3. Рассмотрим два произвольных вектора $\boldsymbol{a}$ и $\boldsymbol{b}$. Возводя равенство (2) п. 2 в квадрат и учитывая, что $\cos^2\theta\leqslant 1$, мы немедленно получим, что

$$(\boldsymbol{ab})^2\leqslant|\boldsymbol{a}|^2\cdot|\boldsymbol{b}|^2,$$

т. е. что

$$|\boldsymbol{ab}|\leqslant|\boldsymbol{a}|\cdot|\boldsymbol{b}|,$$

или иначе — что

$$(\boldsymbol{ab})^2\leqslant\boldsymbol{a}^2\boldsymbol{b}^2. \qquad (1)$$

Это важное неравенство называется неравенством Коши — Буняковского.

Легко видеть, что

неравенство Коши — Буняковского тогда и только тогда вырождается в равенство, когда векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ коллинеарны.

Замечание 1. Мы доказали неравенство Коши — Буняковского, пользуясь определением скалярного произведения. Однако его можно легко доказать и лишь на основе перечисленных в теореме 1 п. 2 свойств. Действительно, рассмотрим вектор $\boldsymbol{a}+t\boldsymbol{b}$, где t — произвольное вещественное число. Согласно свойству 4 для любого t имеет место неравенство $(\boldsymbol{a}+t\boldsymbol{b})^2\geqslant 0$, т. е. неравенство

$$\boldsymbol{a}^2+2(\boldsymbol{ab})t+\boldsymbol{b}^2t^2\geqslant 0.$$

Но из алгебры известно, что если квадратный трехчлен принимает лишь неотрицательные значения, то его дискриминант (подкоренное выражение в формуле для корней квадратного уравнения) не положителен. В нашем случае этот дискриминант равен $(\boldsymbol{ab})^2-\boldsymbol{a}^2\boldsymbol{b}^2$, так что $(\boldsymbol{ab})^2-\boldsymbol{a}^2\boldsymbol{b}^2\leqslant 0$, т. е. $(\boldsymbol{ab})^2\leqslant\boldsymbol{a}^2\boldsymbol{b}^2$.

Из неравенства Коши — Буняковского вытекает, что

$$|\boldsymbol{a}+\boldsymbol{b}|^2=\boldsymbol{a}^2+2\boldsymbol{ab}+\boldsymbol{b}^2\leqslant|\boldsymbol{a}|^2+2|\boldsymbol{a}|\cdot|\boldsymbol{b}|+|\boldsymbol{b}|^2=(|\boldsymbol{a}|+|\boldsymbol{b}|)^2$$

и аналогично

$$|\boldsymbol{a}+\boldsymbol{b}|^2=\boldsymbol{a}^2+2\boldsymbol{ab}+\boldsymbol{b}^2\geqslant|\boldsymbol{a}|^2-2|\boldsymbol{a}|\cdot|\boldsymbol{b}|+|\boldsymbol{b}|^2=(|\boldsymbol{a}|-|\boldsymbol{b}|)^2.$$

Извлекая квадратный корень, мы получаем, что

для любых векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ имеют место неравенства

$$||\boldsymbol{a}|-|\boldsymbol{b}||\leqslant|\boldsymbol{a}+\boldsymbol{b}|\leqslant|\boldsymbol{a}|+|\boldsymbol{b}|. \qquad (2)$$

При этом легко видеть, что

для неколлинеарных векторов оба неравенства (2) — строгие.

Неравенства (2) выражают (для неколлинеарных векторов) известную теорему о сторонах треугольника (длина стороны треугольника меньше суммы длин других его сторон и

больше их разности). Поэтому неравенство (2) называется обычно неравенством треугольника (даже и в случае коллинеарных векторов $\boldsymbol{a}$ и $\boldsymbol{b}$).

Для векторов на прямой неравенство (2) равносильно известному неравенству для абсолютной величины суммы двух вещественных чисел.

4. Выражение скалярного произведения в координатах

Определение 1. Пусть $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ — произвольный базис (на прямой, в плоскости или в пространстве). Попарные скалярные произведения

$$g_{ij} = \boldsymbol{e}_i \boldsymbol{e}_j$$

векторов этого базиса называются его *метрическими коэффициентами.*

Таким образом, при $n = 1$ (случай прямой) имеется только один метрический коэффициент g_{11}, равный квадрату длины единственного базисного вектора $\boldsymbol{e}_1$, при $n = 2$ (случай плоскости) имеется четыре метрических коэффициента $g_{11} = \boldsymbol{e}_1^2$, $g_{12} = \boldsymbol{e}_1 \boldsymbol{e}_2$, $g_{21} = \boldsymbol{e}_2 \boldsymbol{e}_1$, $g_{22} = \boldsymbol{e}_2^2$, которые мы будем записывать в виде матрицы

$$\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix},$$

и, наконец, при $n = 3$ (случай пространства) имеется девять метрических коэффициентов, составляющих матрицу

$$\begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix}.$$

Эти метрические коэффициенты не независимы. Именно, во-первых, ясно, что они симметрично зависят от индексов, т. е.

$$g_{ij} = g_{ji}, \qquad i, j = 1, \ldots, n.$$

Таким образом, при $n = 2$ имеется на самом деле только три метрических коэффициента: g_{11}, g_{22} и $g_{12} = g_{21}$, а при $n = 3$ — только шесть: g_{11}, g_{22}, g_{33} и $g_{12} = g_{21}$, $g_{23} = g_{32}$, $g_{13} = g_{31}$.

Во-вторых, метрические коэффициенты должны удовлетворять некоторым неравенствам, вытекающим из положительной определенности скалярного произведения. Например, поскольку $g_{11} = \boldsymbol{e}_1^2$, то

$$g_{11} > 0. \tag{1}$$

Далее, согласно неравенству Коши — Буняковского $(\boldsymbol{e}_1 \boldsymbol{e}_2)^2 < \boldsymbol{e}_1^2 \boldsymbol{e}_2^2$, т. е.

$$g_{12}^2 < g_{11} g_{22}.$$

Это условие можно переписать также в следующем «детерминантном» виде:

$$\begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} > 0. \tag{2}$$

Можно показать (мы это сделаем позже; см. п. 5 § 6), что при $n = 3$ выполняется аналогичное неравенство

$$\begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} > 0. \tag{3}$$

Оказывается, что этих условий уже достаточно, т. е.

для любой квадратной симметрической матрицы (g_{ij}) *порядка* $n \leqslant 3$, *удовлетворяющей условию* (1) *при* $n = 1$, *условиям* (1) *и* (2) *при* $n = 2$ *и условиям* (1), (2) *и* (3) *при* $n = 3$, *существует базис* (*соответственно на прямой, в плоскости или в пространстве*), *матрицей метрических коэффициентов которого эта матрица является.*

Упражнение. Докажите это утверждение.

Над матрицами с векторными элементами мы можем производить те же действия, что и над векторами. В частности, умея перемножать векторы, мы можем перемножать и матрицы с векторными элементами.

Рассмотрим, например, матрицу-строку

$$e = (e_1, \ldots, e_n),$$

составленную из векторов данного базиса $e_1, \ldots, e_n$, и транспонированную матрицу-столбец

$$e^{\top} = \begin{pmatrix} e_1 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{pmatrix}.$$

Умножая «скалярно» матрицу $e^{\top}$ на матрицу e, мы, очевидно, как раз и получим матрицу метрических коэффициентов $G = (g_{ij})$. Таким образом, мы можем написать, что

$$G = e^{\top} e.$$

Пользуясь этой записью матрицы G, можно сразу же получить ответ на вопрос, как меняется эта матрица при изменении базиса.

Пусть $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ и $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$ — два базиса (на прямой, в плоскости или в пространстве) и пусть G — матрица метрических коэффициентов базиса $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, а G' — матрица метрических коэффициентов базиса $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$. Кроме того, пусть C — матрица перехода от базиса $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ к базису $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$. Тогда

$$G = \boldsymbol{e}^{\top}\boldsymbol{e}, \quad G' = \boldsymbol{e}'^{\top}\boldsymbol{e}'$$

и

$$\boldsymbol{e}' = \boldsymbol{e}C,$$

где

$$\boldsymbol{e} = (\boldsymbol{e}_1, \ldots. \boldsymbol{e}_n), \quad \boldsymbol{e}' = (\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}).$$

Поэтому

$$G' = (\boldsymbol{e}C)^{\top}(\boldsymbol{e}C) = (C^{\top}\boldsymbol{e}^{\top})(\boldsymbol{e}C) = C^{\top}(\boldsymbol{e}^{\top}\boldsymbol{e})C = C^{\top}GC.$$

Таким образом,

$$G' = C^{\top}GC.$$

Это и есть формула преобразования матрицы G при изменении базиса.

Пусть $\boldsymbol{e}_1$ — базис на прямой, а g_{11} — его метрический коэффициент. Тогда для любых двух векторов $\boldsymbol{x} = x^1\boldsymbol{e}_1$ и $\boldsymbol{y} = y^1\boldsymbol{e}_1$

$$\boldsymbol{xy} = (x^1\boldsymbol{e}_1)(y^1\boldsymbol{e}_1) = \boldsymbol{e}_1^2 x^1 y^1 = g_{11}x^1y^1.$$

Аналогично, если $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ — базис на плоскости и g_{11}, g_{22}, $g_{12} = g_{21}$ — его метрические коэффициенты, то для любых векторов $\boldsymbol{x} = x^1\boldsymbol{e}_1 + x^2\boldsymbol{e}_2$ и $\boldsymbol{y} = y^1\boldsymbol{e}_1 + y^2\boldsymbol{e}_2$

$$\begin{aligned}\boldsymbol{xy} &= (x^1\boldsymbol{e}_1 + x^2\boldsymbol{e}_2)(y^1\boldsymbol{e}_1 + y^2\boldsymbol{e}_2) = \\ &= (x^1\boldsymbol{e}_1)(y^1\boldsymbol{e}_1) + (x^1\boldsymbol{e}_1)(y^2\boldsymbol{e}_2) + (x^2\boldsymbol{e}_2)(y^1\boldsymbol{e}_1) + (x^2\boldsymbol{e}_2)(y^2\boldsymbol{e}_2) = \\ &= g_{11}x^1y^1 + g_{12}(x^1y^2 + x^2y^1) + g_{22}x^2y^2.\end{aligned}$$

Наконец, если $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ — базис в пространстве, то в аналогичных обозначениях

$$\begin{aligned}\boldsymbol{xy} &= (x^1\boldsymbol{e}_1 + x^2\boldsymbol{e}_2 + x^3\boldsymbol{e}_3)(y^1\boldsymbol{e}_1 + y^2\boldsymbol{e}_2 + y^3\boldsymbol{e}_3) = \\ &= g_{11}x^1y^1 + g_{22}x^2y^2 + g_{33}x^3y^3 + \\ &+ g_{12}(x^1y^2 + x^2y^1) + g_{13}(x^1y^3 + x^3y^1) + g_{23}(x^2y^3 + x^3y^2).\end{aligned}$$

Таким образом,

скалярное произведение выражается в координатах следующими формулами:

$$\boldsymbol{xy} = g_{11}x^1y^1 \qquad (\textit{на прямой}),$$

$$\boldsymbol{xy} = g_{11}x^1y^1 + g_{12}(x^1y^2 + x^2y^1) + g_{22}x^2y^2 \qquad (\textit{на плоскости}),$$

$$\begin{aligned}\boldsymbol{xy} &= g_{11}x^1y^1 + g_{22}x^2y^2 + g_{33}x^3y^3 + g_{12}(x^1y^2 + x^2y^1) + \\ &+ g_{13}(x^1y^3 + x^3y^1) + g_{23}(x^2y^3 + x^3y^2) \qquad (\textit{в пространстве}).\end{aligned}$$

Эти формулы можно переписать также в следующем более симметричном виде:

$$\begin{aligned}
\boldsymbol{xy} &= g_{11}x^1y^1,\\
\boldsymbol{xy} &= g_{11}x^1y^1 + g_{12}x^1y^2 +\\
&\quad + g_{21}x^2y^1 + g_{22}x^2y^2,\\
\boldsymbol{xy} &= g_{11}x^1y^1 + g_{12}x^1y^2 + g_{13}x^1y^3 +\\
&\quad + g_{21}x^2y^1 + g_{22}x^2y^2 + g_{23}x^2y^3 +\\
&\quad + g_{31}x^3y^1 + g_{32}x^3y^2 + g_{33}x^3y^3.
\end{aligned}$$

Пользуясь знаком $\sum$ суммы, последние формулы можно записать совсем коротко:

$$\boldsymbol{xy} = \sum_{i,\,j=1}^{n} g_{ij}x^iy^j.$$

Можно также (как и в п. 7 § 3) ввести в рассмотрение одностолбцовые матрицы

$$x = \begin{pmatrix} x^1 \\ \cdot \\ \cdot \\ \cdot \\ x^n \end{pmatrix}, \quad y = \begin{pmatrix} y^1 \\ \cdot \\ \cdot \\ \cdot \\ y^n \end{pmatrix},$$

состоящие из координат данных векторов. Тогда предыдущую формулу можно с помощью матричного умножения записать в следующем компактном виде:

$$\boldsymbol{xy} = x^{\top}Gy,$$

где G — матрица метрических коэффициентов, а $x^{\top}$ — однострочечная матрица, полученная транспонированием матрицы x.

В этом виде формула для $\boldsymbol{xy}$ доказывается (для любого n) одной строчкой вычислений:

$$\boldsymbol{xy} = (\boldsymbol{e}x)\,(\boldsymbol{e}y) = (\boldsymbol{e}x)^{\top}\,(\boldsymbol{e}y) = (x^{\top}\boldsymbol{e}^{\top})\,(\boldsymbol{e}y) = x^{\top}\,(\boldsymbol{e}^{\top}\boldsymbol{e})\,y = x^{\top}Gy$$

(здесь мы воспользовались тем, что для матрицы $\boldsymbol{e}x$ первого порядка справедливо равенство $(\boldsymbol{e}x)^{\top} = \boldsymbol{e}x$).

Умея вычислять через координаты скалярное произведение, мы, используя формулы (3′) и (2′) п. 2, можем вычислять через координаты длины векторов и углы между ними. Например, в матричной записи формула для длины вектора имеет вид

$$|\boldsymbol{x}| = \sqrt{x^{\top}Gx}$$

(имеется в виду арифметический корень), а формула для угла между векторами — вид

$$\cos\vartheta = \frac{x^{\top}Gy}{\sqrt{x^{\top}Gx}\,\sqrt{y^{\top}Gy}}.$$

В развернутом виде мы напишем эти формулы только для случая плоскости:

$$|\boldsymbol{x}| = \sqrt{g_{11}(x^1)^2 + 2g_{12}x^1x^2 + g_{22}(x^2)^2},$$

$$\cos\theta = \frac{g_{11}x^1y^1 + g_{12}(x^1y^2 + x^2y^1) + g_{22}x^2y^2}{\sqrt{g_{11}(x^1)^2 + 2g_{12}x^1x^2 + g_{22}(x^2)^2}\sqrt{g_{11}(y^1)^2 + 2g_{12}y^1y^2 + g_{22}(y^2)^2}}.$$

5. Ортонормированные базисы

Формулы для скалярного произведения приобретают особенно простой вид, когда для рассматриваемого базиса матрица метрических коэффициентов является единичной матрицей, т. е. когда

$$g_{ij} = \begin{cases} 1, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases}$$

Определение 1. Базис, матрица метрических коэффициентов которого является единичной, называется *ортонормированным*.

Ясно, что

базис $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ *тогда и только тогда ортонормирован, когда каждый его вектор имеет длину, равную единице, и любые два (различных) его вектора ортогональны.*

Векторы, длина которых равна единице, называются *единичными векторами* или *ортами*. Таким образом, каждый ортонормированный базис состоит из n попарно ортогональных ортов.

Для обозначения векторов ортонормированного базиса принято использовать символы $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_n$. В частности, ортонормированный базис на прямой (т. е. некоторый орт) обозначается символом $\boldsymbol{i}_1$ или просто $\boldsymbol{i}$. (Заметим, что на прямой существует только два ортонормированных базиса $\boldsymbol{i}$ и $-\boldsymbol{i}$.) Аналогично, векторы ортонормированного базиса на плоскости (т. е. два ортогональных орта) обозначаются символами $\boldsymbol{i}_1$ и $\boldsymbol{i}_2$ (иногда используются также обозначения $\boldsymbol{i}$ и $\boldsymbol{j}$), а векторы ортонормированного базиса в пространстве (три ортогональных орта) — символами $\boldsymbol{i}_1$, $\boldsymbol{i}_2$ и $\boldsymbol{i}_3$ (или же символами $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$).

Координаты вектора в ортонормированном базисе принято обозначать, помещая индекс внизу (т. е. принято писать не $x^1, \ldots, x^n$, как в общем случае, а $x_1, \ldots, x_n$). Более того, часто пользуются и безындексным обозначением координат (например, символом x на прямой, символами x, y на плоскости и символами x, y, z в пространстве). Так, например, в пространстве разложение вектора $\boldsymbol{x}$ по векторам ортонормированного базиса обычно пишется в следующем виде:

$$\boldsymbol{x} = x_1\boldsymbol{i} + x_2\boldsymbol{j} + x_3\boldsymbol{k}$$

или

$$\boldsymbol{x} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$$

(впрочем, последняя форма записи редко бывает удобна).

Формула для скалярного произведения в ортонормированном базисе имеет, очевидно, вид

$$\boldsymbol{xy} = x_1y_1 + \ldots + x_ny_n \qquad (n = 1, 2, 3).$$

Для длины вектора и угла между двумя векторами получаем формулы

$$|\boldsymbol{x}| = \sqrt{x_1^2 + \ldots + x_n^2},$$

$$\cos\theta = \frac{x_1y_1 + \ldots + x_ny_n}{\sqrt{x_1^2 + \ldots + x_n^2}\,\sqrt{y_1^2 + \ldots + y_n^2}}.$$

Например, на плоскости

$$\boldsymbol{xy} = x_1y_1 + x_2y_2,$$

$$|\boldsymbol{x}| = \sqrt{x_1^2 + x_2^2},$$

$$\cos\theta = \frac{x_1y_1 + x_2y_2}{\sqrt{x_1^2 + x_2^2}\,\sqrt{y_1^2 + y_2^2}}.$$

Ортонормированные базисы обладают тем приятным свойством, что координаты $x_1, \ldots, x_n$ произвольного вектора в явном виде выражаются через этот вектор. Именно, легко видеть, что справедливо следующее

Предложение 1. *Для любого ортонормированного базиса* $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_n$ *(на прямой, в плоскости или в пространстве) координаты* $x_1, \ldots, x_n$ *произвольного вектора* $\boldsymbol{x}$ *равны скалярным произведениям этого вектора на векторы базиса:*

$$x_1 = \boldsymbol{x}\boldsymbol{i}_1, \ \ldots, \ x_n = \boldsymbol{x}\boldsymbol{i}_n.$$

Доказательство. Скалярно умножая равенство $\boldsymbol{x} = x_1\boldsymbol{i}_1 + \ldots + x_n\boldsymbol{i}_n$, скажем, на $\boldsymbol{i}_1$ и учитывая, что $\boldsymbol{i}_1^2 = 1$ и $\boldsymbol{i}_1\boldsymbol{i}_k = 0$ при $k > 1$, мы немедленно получим, что $x_1 = \boldsymbol{x}\boldsymbol{i}_1$.

Определение 2. Пусть $\boldsymbol{x}$ — произвольный отличный от нуля вектор (на прямой, в плоскости или в пространстве) и пусть $\alpha_1, \ldots, \alpha_n$ — углы, образованные этим вектором с ортами $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_n$ некоторого ортонормированного базиса. Косинусы

$$\cos\alpha_1, \ \ldots, \ \cos\alpha_n$$

этих углов называются *направляющими косинусами* вектора $\boldsymbol{x}$.

Впрочем, рассмотрение этих косинусов по существу интересно только в пространстве (т. е. при $n = 3$). Действительно, на прямой единственный направляющий косинус равен ± 1, а на плоскости имеют место равенства

$$\cos\alpha_1 = \cos\alpha,$$

$$\cos\alpha_2 = \sin\alpha,$$

где α — угол от вектора $i_1 = i$ к вектору x (здесь имеется в виду, что на плоскости выбрана ориентация, определяемая данным базисом i_1, i_2).

В пространстве направляющие косинусы обычно обозначаются символами

$$\cos\alpha, \quad \cos\beta, \quad \cos\gamma.$$

Поскольку $xi_k = |x| \cos\alpha_k$, $k = 1, \ldots, n$, мы видим, что

координаты $x_1, \ldots, x_n$ вектора x в ортонормированном базисе выражаются через его длину и через его направляющие косинусы по формулам

$$x_1 = |x| \cos\alpha_1, \ \ldots, \ x_n = |x| \cos\alpha_n.$$

Таким образом, на плоскости,

$$x_1 = |x| \cos\alpha, \quad x_2 = |x| \sin\alpha,$$

а в пространстве

$$x_1 = |x| \cos\alpha, \quad x_2 = |x| \cos\beta, \quad x_3 = |x| \cos\gamma.$$

Два вектора тогда и только тогда имеют одинаковые направляющие косинусы, когда они отличаются положительным множителем. Для орта (единичного вектора) направляющие косинусы совпадают с его координатами.

Поскольку для любого вектора x имеет место формула $|x|^2 = x_1^2 + \ldots + x_n^2$, то

$$\cos^2\alpha_1 + \ldots + \cos^2\alpha_n = 1.$$

При $n = 1$ это соотношение тривиально, при $n = 2$ оно сводится к известному тригонометрическому тождеству

$$\cos^2\alpha + \sin^2\alpha = 1,$$

а при $n = 3$ имеет вид

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1.$$

Подставив в формулу для косинуса угла между двумя векторами выражения их координат через направляющие косинусы, мы немедленно получим, что

угол θ между векторами x и y с направляющими косинусами $\cos\alpha_1, \ldots, \cos\alpha_n$ и $\cos\beta_1, \ldots, \cos\beta_n$ удовлетворяет соотношению

$$\cos\theta = \cos\alpha_1 \cos\beta_1 + \ldots + \cos\alpha_n \cos\beta_n.$$

В частности, в пространстве (обозначив направляющие косинусы вектора x через $\cos\alpha_1$, $\cos\beta_1$, $\cos\gamma_1$, а направляющие косинусы вектора y через $\cos\alpha_2$, $\cos\beta_2$, $\cos\gamma_2$) мы получаем формулу

$$\cos\theta = \cos\alpha_1 \cos\alpha_2 + \cos\beta_1 \cos\beta_2 + \cos\gamma_1 \cos\gamma_2,$$

а на плоскости (в аналогичных обозначениях) — формулу

$$\cos\theta = \cos\alpha_1 \cos\alpha_2 + \sin\alpha_1 \sin\alpha_2 = \cos(\alpha_2 - \alpha_1).$$

Впрочем, последняя формула нам не дает ничего нового, поскольку угол $\alpha_2 - \alpha_1$ равен, очевидно, углу φ от вектора $\boldsymbol{x}$ к вектору $\boldsymbol{y}$ (в ориентации плоскости, определенной базисом $\boldsymbol{i}_1$, $\boldsymbol{i}_2$) и потому либо он совпадает с углом θ, либо отличается от него знаком.

Тот факт, что $\varphi = \alpha_2 - \alpha_1$, позволяет без труда найти формулы для вычисления угла φ по координатам векторов $\boldsymbol{x}$ и $\boldsymbol{y}$. Действительно,

$$\sin\varphi = \sin(\alpha_2 - \alpha_1) = \cos\alpha_2 \sin\alpha_1 - \sin\alpha_2 \cos\alpha_1 = \frac{x_1 y_2 - x_2 y_1}{|\boldsymbol{x}| \cdot |\boldsymbol{y}|}.$$

Поскольку

$$\cos\varphi = \cos\theta = \frac{x_1 y_1 + x_2 y_2}{|\boldsymbol{x}| \cdot |\boldsymbol{y}|},$$

тем самым доказано, что

для угла φ от вектора $\boldsymbol{x}(x_1, x_2)$ к вектору $\boldsymbol{y}(y_1, y_2)$ (в ориентации, задаваемой базисом $\boldsymbol{i}_1$, $\boldsymbol{i}_2$) имеют место формулы

$$\sin\varphi = \frac{x_1 y_2 - x_2 y_1}{\sqrt{x_1^2 + x_2^2}\,\sqrt{y_1^2 + y_2^2}}, \qquad \cos\varphi = \frac{x_1 y_1 + x_2 y_2}{\sqrt{x_1^2 + x_2^2}\,\sqrt{y_1^2 + y_2^2}}.$$

Напомним, что величины $\sin\varphi$ и $\cos\varphi$ однозначно определяют угол φ.

Пусть λ — произвольная ориентированная прямая (для определенности, в пространстве). Ее направляющими косинусами

$$\cos\alpha, \quad \cos\beta, \quad \cos\gamma$$

называются направляющие косинусы ее произвольного положительно ориентированного вектора $\boldsymbol{a}$. (Очевидно, что это определение корректно.)

По определению, для величины $(\mathrm{pr}_\lambda \boldsymbol{x})$ ортогональной проекции $\mathrm{pr}_\lambda \boldsymbol{x}$ произвольного вектора $\boldsymbol{x}(x_1, x_2, x_3)$ на прямую λ имеют место равенства

$$(\mathrm{pr}_\lambda \boldsymbol{x}) = (\mathrm{pr}_a \boldsymbol{x}) = \frac{\boldsymbol{a}\boldsymbol{x}}{|\boldsymbol{a}|} = \frac{a_1 x_1 + a_2 x_2 + a_3 x_3}{|\boldsymbol{a}|} =$$

$$= x_1 \cos\alpha + x_2 \cos\beta + x_3 \cos\gamma.$$

Таким образом,

величина $(\mathrm{pr}_\lambda \boldsymbol{x})$ ортогональной проекции произвольного вектора $\boldsymbol{x}(x_1, x_2, x_3)$ на ориентированную прямую λ с направляющими косинусами $\cos\alpha$, $\cos\beta$, $\cos\gamma$ выражается формулой

$$(\mathrm{pr}_\lambda \boldsymbol{x}) = x_1 \cos\alpha + x_2 \cos\beta + x_3 \cos\gamma.$$

Аналогичная формула на плоскости имеет вид

$$(\mathrm{pr}_\lambda \boldsymbol{x}) = x_1 \cos\alpha + x_2 \sin\alpha.$$

6. Ортогональные матрицы

Пусть $i_1, \ldots, i_n$ и $i'_1, \ldots, i'_n$ — два ортонормированных базиса (на прямой, в плоскости или в пространстве) и пусть C — матрица перехода от первого базиса ко второму. Как мы знаем (п. 4), матрицы метрических коэффициентов преобразуются при изменении базиса по формуле

$$G' = C^{\top} GC.$$

Но в нашем случае $G' = G = E$. Следовательно,

матрица C перехода от одного ортонормированного базиса к другому удовлетворяет соотношению

$$C^{\top} C = E. \tag{1}$$

Определение 1. Матрицы, удовлетворяющие соотношению (1), называются *ортогональными.*

Таким образом, мы показали, что

матрица перехода от одного ортонормированного базиса к другому является ортогональной матрицей.

Легко видеть, что верно и обратное:

если матрица C перехода от одного базиса к другому ортогональна и если первый базис ортонормирован, то второй базис также ортонормирован.

Действительно, матрица метрических коэффициентов второго базиса имеет вид $C^{\top}GC$, где G — матрица метрических коэффициентов первого базиса. Но по условию $G = E$ и $C^{\top}C = E$. Поэтому эта матрица является единичной матрицей E, и следовательно, второй базис также ортонормирован.

Перейдя в равенстве (1), определяющем ортогональные матрицы, к определителям и вспомнив, что $|C^{\top}| = |C|$ (при транспонировании определитель не меняется), мы немедленно получим, что

определитель произвольной ортогональной матрицы C равен ± 1:

$$|C| = \pm 1.$$

Определение 2. Ортогональные матрицы C, для которых $|C| = 1$, называются *собственными* ортогональными матрицами (а матрицы с $|C| = -1$ соответственно — *несобственными*).

Вспоминая определение одноименных базисов, мы немедленно получаем, что

собственные ортогональные матрицы (и только они) являются матрицами перехода между одноименными ортонормированными базисами.

Умножив равенство (1) справа на C^{-1}, мы немедленно получим, что $C^{\top} = C^{-1}$. Обратно, если $C^{\top} = C^{-1}$, то $C^{\top}C = E$ и, следовательно, матрица C ортогональна. Таким образом,

матрица C тогда и только тогда ортогональна, когда транспонированная матрица $C^{\top}$ является обратной матрицей C^{-1}:

$$C^{\top} = C^{-1}. \tag{2}$$

Умножив равенство (2) слева на C, мы получим, что $CC^{\top} = E$. Обратно, если $CC^{\top} = E$, то $C^{\top} = C^{-1}$ и, следовательно, матрица C ортогональна. Таким образом,

матрица C тогда и только тогда ортогональна, когда она удовлетворяет соотношению

$$CC^{\top} = E. \tag{3}$$

В развернутом виде соотношение (1) означает, что сумма квадратов элементов каждого столбца матрицы C равна единице, а сумма попарных произведений соответственных элементов двух любых столбцов равна нулю. Соотношение (3) означает, что тем же свойством обладают и строки матрицы C.

Ясно, что

для любой ортогональной матрицы C обратная матрица C^{-1} также является ортогональной матрицей.

Действительно, если матрица C была матрицей перехода от одного ортонормированного базиса к другому, то матрица C^{-1} будет матрицей перехода от второго базиса к первому.

Аналогично показывается, что

произведение C_1C_2 двух ортогональных матриц C_1 и C_2 также является ортогональной матрицей.

Последние два утверждения означают, что

совокупность $O(n)$ всех ортогональных матриц порядка n является группой по умножению.

Замечание 1. Это утверждение можно доказать и чисто алгебраически, пользуясь лишь определением ортогональной матрицы как такой матрицы, для которой

$$C^{\top}C = E.$$

Действительно, из курса алгебры известно, что

$$(C_1C_2)^{\top} = C_2^{\top}C_1^{\top}.$$

Поэтому, если

$$C_1^{\top}C_1 = E, \quad C_2^{\top}C_2 = E,$$

то

$$(C_1C_2)^{\top}(C_1C_2) = C_2^{\top}(C_1^{\top}C_1)C_2 = C_2^{\top}C_2 = E.$$

Аналогично, если

$$C^{\top}C = E,$$

то

$$C^{\top}(C^{\top})^{\top} = E,$$

и, следовательно, матрица $C^{\top} = C^{-1}$ ортогональна.

Это доказательство проходит для любого n (в том числе и при $n > 3$).

Поскольку определитель произведения двух матриц равен произведению определителей сомножителей, то

произведение двух ортогональных матриц тогда и только тогда является собственной ортогональной матрицей, когда оба сомножителя являются одновременно либо собственными, либо несобственными ортогональными матрицами.

В частности, мы видим, что

совокупность $O^{+}(n)$ *всех собственных ортогональных матриц является подгруппой группы* $O(n)$.

Эта подгруппа обозначается также символом $SO(n)$.

Замечание 2 (для читателей, знакомых с теорией групп). Очевидно, что группа $SO(n)$ является не только подгруппой, но даже и нормальным делителем группы $O(n)$.

Что же касается совокупности $O^{-}(n)$ всех несобственных ортогональных матриц, то она обладает тем свойством, что для любых ее элементов C_1 и C_2 матрицы $C_1^{-1}C_2$ и $C_1C_2^{-1}$ принадлежат группе $SO(n) = O^{+}(n)$. На языке теории групп это означает, что

множество $O^{-}(n)$ *всех несобственных ортогональных матриц порядка n является двусторонним смежным классом группы* $O(n)$ *по подгруппе* $O^{+}(n)$.

Пусть c_{ij} — элементы ортогональной матрицы C. При $n=1$ условие ортогональности сводится к равенству

$$c_{11}^2 = 1.$$

Следовательно, единственными ортогональными матрицами первого порядка являются числа ± 1. Это вполне согласуется с тем, что на прямой существует только два ортонормированных базиса: i и $-i$.

Числа $+1$ и -1 составляют группу по умножению. Обозначая эту группу символом S^0, мы можем, следовательно, сказать, что

группа $O(1)$ *изоморфна группе* S^0.

В частности, мы видим, что единственной собственной ортогональной матрицей первого порядка является единичная матрица 1.

При $n = 2$ условия ортогональности имеют вид

$$c_{11}^2 + c_{21}^2 = 1,$$

$$c_{11}c_{12} + c_{21}c_{22} = 0,$$

$$c_{12}^2 + c_{22}^2 = 1.$$

Общее решение первого уравнения выражается, очевидно, формулами

$$c_{11} = \cos\alpha, \quad c_{21} = \sin\alpha,$$

где α — некоторый угол, $-\pi < \alpha \leqslant \pi$, а общее решение третьего уравнения — аналогичными формулами

$$c_{12} = \cos\beta, \quad c_{22} = \sin\beta, \quad -\pi < \beta \leqslant \pi.$$

Подставляя эти выражения во второе уравнение, мы получим, что

$$\cos(\beta - \alpha) = 0.$$

Следовательно, $\beta = \alpha \pm \frac{\pi}{2}$ и потому либо $\cos\beta = -\sin\alpha$, $\sin\beta = \cos\alpha$, либо $\cos\beta = \sin\alpha$, $\sin\beta = -\cos\alpha$. В первом случае определитель матрицы C равен $\cos^2\alpha + \sin^2\alpha = 1$, а во втором — равен $-\cos^2\alpha - \sin^2\alpha = -1$.

Таким образом,

произвольная ортогональная матрица второго порядка имеет либо вид

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix},$$

и в этом случае она — собственная, либо вид

$$\begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix},$$

и тогда она несобственная.

Если матрица C является матрицей перехода от базиса i, j к базису i', j', то угол α является не чем иным, как углом φ от вектора i к вектору i' (в ориентации, определенной базисом i, j).

Действительно, по определению

$$i' = i\cos\alpha + j\sin\alpha$$

и потому для угла φ от вектора i к вектору i' имеют место формулы (см. конец п. 5)

$$\sin\varphi = 1 \cdot \sin\alpha - 0 \cdot \cos\alpha = \sin\alpha,$$
$$\cos\varphi = 1 \cdot \cos\alpha + 0 \cdot \sin\alpha = \cos\alpha,$$

и, следовательно, $\varphi = \alpha$.

При вращении плоскости на угол α (в положительном направлении) вектор i переходит, таким образом, в вектор i', а вектор j — в вектор j', если базисы i, j и i', j' одноименны, и в вектор $-j'$, если эти базисы разноименны.

В частности, мы видим, что любые два одноименных базиса плоскости мы можем перевести друг в друга некоторым вращением.

По формуле умножения матриц

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}\begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} =$$

$$= \begin{pmatrix} \cos\alpha\cos\beta - \sin\alpha\sin\beta & -\cos\alpha\sin\beta - \sin\alpha\cos\beta \\ \sin\alpha\cos\beta + \cos\alpha\sin\beta & -\sin\alpha\sin\beta + \cos\alpha\cos\beta \end{pmatrix} =$$

$$= \begin{pmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{pmatrix}.$$

Таким образом,

при умножении собственных ортогональных матриц второго порядка соответствующие углы складываются.

Рассмотрим множество S^1 всех комплексных чисел $z = e^{i\alpha}$, по модулю равных единице. Это множество является группой по умножению, причем

$$e^{i\alpha}e^{i\beta} = e^{i(\alpha+\beta)}.$$

Это показывает, что

соответствие

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \longmapsto e^{i\alpha}$$

представляет собой изоморфизм группы $SO(2)$ *на группу* S^1.

В частности мы видим, что *группа* $SO(2)$ *— абелева.*

При $n = 3$ условия ортогональности имеют вид

$$c_{11}^2 + c_{21}^2 + c_{31}^2 = 1, \quad c_{11}c_{12} + c_{21}c_{22} + c_{31}c_{32} = 0,$$
$$c_{12}^2 + c_{22}^2 + c_{32}^2 = 1, \quad c_{11}c_{13} + c_{21}c_{23} + c_{31}c_{33} = 0,$$
$$c_{13}^2 + c_{23}^2 + c_{33}^2 = 1, \quad c_{12}c_{13} + c_{22}c_{23} + c_{32}c_{33} = 0.$$

Аналитическое исследование этой системы уравнений, хотя и возможно, но крайне канительно. Поэтому мы воспользуемся наглядными геометрическими соображениями. При этом мы будем для простоты считать, что рассматриваемая ортогональная матрица C — собственная, т. е. осуществляет переход от некоторого ортонормированного базиса i, j, k к одноименному ортонормированному базису i', j', k'.

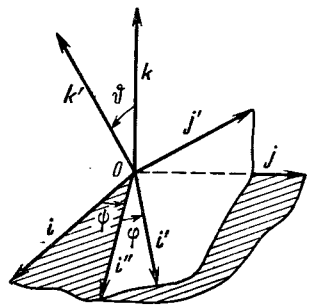
Более формальное (и более изящное) исследование этой системы уравнений мы проведем в п. 4 § 2 гл. 7.

Будем считать, что все векторы i, j, k, i', j', k' отложены от некоторой точки O:

$$i = \overrightarrow{OP}, \quad j = \overrightarrow{OQ}, \quad k = \overrightarrow{OR},$$
$$i' = \overrightarrow{OP'}, \quad j' = \overrightarrow{OQ'}, \quad k' = \overrightarrow{OR'}.$$

Пусть орты k и k' неколлинеарны ($k' \neq \pm k$). Тогда плоскости OPQ и $OP'Q'$ различны и потому пересекаются по некоторой прямой λ. Пусть $i'' = \overrightarrow{OP''}$ — орт этой прямой. Он

ортогонален ортам $\boldsymbol{k}$ и $\boldsymbol{k}'$ и потому им не компланарен, т. е. орты $\boldsymbol{k}$, $\boldsymbol{k}'$, $\boldsymbol{i}''$ составляют некоторый базис пространства (неортонормированный). С равным правом вместо орта $\boldsymbol{i}''$ мы могли бы взять орт $-\boldsymbol{i}''$. Чтобы устранить эту неопределенность, мы потребуем, чтобы базис $\boldsymbol{k}$, $\boldsymbol{k}'$, $\boldsymbol{i}''$ был одноименен с базисом $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$. Тогда орт $\boldsymbol{i}''$ будет однозначно определен. В случае, когда орты $\boldsymbol{k}$ и $\boldsymbol{k}'$ коллинеарны, мы за орт $\boldsymbol{i}''$ примем орт $\boldsymbol{i}'$.



Пусть ψ — угол от орта $\boldsymbol{i}$ до орта $\boldsymbol{i}''$ в плоскости OPQ (ориентация которой определяется базисом $\boldsymbol{i}$, $\boldsymbol{j}$) и пусть $\boldsymbol{j}'' = \overrightarrow{OQ''}$ — такой орт плоскости OPQ, что угол от орта $\boldsymbol{j}$ до орта $\boldsymbol{j}''$ также равен ψ. Иными словами, $\boldsymbol{j}''$ — это такой орт плоскости OPQ, что базис $\boldsymbol{i}''$, $\boldsymbol{j}''$ этой плоскости получается из базиса $\boldsymbol{i}$, $\boldsymbol{j}$ вращением плоскости (вокруг точки O) в положительном направлении на угол ψ. Согласно сказанному выше отсюда следует, что матрица перехода от базиса $\boldsymbol{i}$, $\boldsymbol{j}$ к базису $\boldsymbol{i}''$, $\boldsymbol{j}''$ имеет вид

$$\begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix}.$$

Поэтому в пространстве матрица C_1 перехода от базиса $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ к базису $\boldsymbol{i}''$, $\boldsymbol{j}''$, $\boldsymbol{k}$ выражается формулой

$$C_1 = \begin{pmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad -\pi < \psi \leqslant \pi.$$

Заметим, что базис $\boldsymbol{i}''$, $\boldsymbol{j}''$, $\boldsymbol{k}$ получается из базиса $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ вращением пространства на угол ψ вокруг прямой OR.

Рассмотрим теперь плоскость ORR', перпендикулярную прямой λ. Пусть θ — угол между векторами $\boldsymbol{k}$ и $\boldsymbol{k}'$ этой плоскости ($0 \leqslant \theta \leqslant \pi$). Поворот на угол θ плоскости ORR' (вокруг точки O), совмещающий вектор $\boldsymbol{k}$ с вектором $\boldsymbol{k}'$, переводит вектор $\boldsymbol{j}''$ (также принадлежащий плоскости ORR') в некоторый вектор $\boldsymbol{j}'''$. Отсюда, как и выше, следует, что матрица C_2 перехода от базиса $\boldsymbol{i}''$, $\boldsymbol{j}''$, $\boldsymbol{k}$ к базису $\boldsymbol{i}''$, $\boldsymbol{j}'''$, $\boldsymbol{k}'$ выражается формулой

$$C_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}, \qquad 0 \leqslant \theta \leqslant \pi.$$

При этом базис $\boldsymbol{i}''$, $\boldsymbol{j}'''$, $\boldsymbol{k}'$ получается из базиса $\boldsymbol{i}''$, $\boldsymbol{j}''$, $\boldsymbol{k}$ вращением пространства вокруг прямой λ на угол θ.

Чтобы теперь перейти от базиса i'', j''', k' к нужному нам базису i', j', k', достаточно в плоскости $OP'Q'$ перейти от базиса i'', j''' к (очевидно, одноименному) базису i', j'. Это означает, что переход от базиса i'', j''', k' к базису i', j', k' описывается матрицей

$$C_3 = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad -\pi < \varphi \leqslant \pi,$$

где φ — угол от вектора i'' к вектору i' в ориентированной плоскости $OP'Q'$. При этом базис i', j', k' получается из базиса i'', j''', k' вращением пространства вокруг прямой OR' на угол φ.

Таким образом, мы видим, что базис i', j', k' получается из базиса i, j, k тремя последовательными вращениями. Следовательно, переход от базиса i, j, k к базису i', j', k' описывается произведением $C_1C_2C_3$ матриц перехода C_1, C_2 и C_3.

Тем самым доказано следующее

Предложение 1. *Любая собственная ортогональная матрица C третьего порядка допускает разложение вида*

$$C = \begin{pmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (4)$$

где $-\pi < \psi \leqslant \pi$, $0 \leqslant \theta \leqslant \pi$, $-\pi < \varphi \leqslant \pi$.

Углы ψ, θ, φ называются *углами Эйлера* ортогональной матрицы C (или базиса i, j, k относительно базиса i', j', k'). Они «независимы» в том смысле, что любым их значениям (в указанных пределах) отвечает некоторая (однозначно определенная) ортогональная матрица.

Вообще говоря, нельзя утверждать, что углы ψ, θ, φ однозначно определены матрицей C. Это верно лишь тогда, когда угол θ отличен от нуля и от π. Если же $\theta = 0$ или $\theta = \pi$, то матрица C однозначно определяет соответственно либо сумму $\varphi + \psi$, либо разность $\varphi - \psi$. Поэтому при $\theta = 0$ или $\theta = \pi$ один из углов φ или ψ может иметь (для данной матрицы C) произвольное значение.

Конечно, в формуле (4) можно фактически осуществить умножение матриц и тем самым получить в явном виде выражение элементов матрицы C через углы Эйлера. Мы этого делать не будем, поскольку получающиеся формулы в достаточной мере громоздки и практически не запоминаемы.

Чтобы получить аналогичное описание несобственных ортогональных матриц, достаточно в разложении (1) один из столбцов какой-нибудь матрицы умножить на -1.

§ 6. ПОЛИВЕКТОРЫ

1. Бивекторы

Как мы знаем, направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{A'B'}$ эквиполлентны, если либо

1) $A = B$ и $A' = B'$,

либо

2) $A \neq B$, $A' \neq B'$ и

а) прямая AB параллельна прямой $A'B'$,

б) ориентация прямой AB, определенная направленным отрезком $\overrightarrow{AB}$, совпадает (см. конец п. 4 § 4) с ориентацией прямой $A'B'$, определенной направленным отрезком $\overrightarrow{A'B'}$,

в) длина отрезка $\overline{AB}$ равна длине отрезка $\overline{A'B'}$.

Поскольку направленный отрезок $\overrightarrow{AB}$ является, по определению, не чем иным, как упорядоченной парой точек (A, B), мы можем, по аналогии, ввести следующее

Определение 1. Две упорядоченные тройки точек (A, B, C) и (A', B', C') называются *эквиполлентными*, если либо

1) обе тройки коллинеарны,

либо

2) обе тройки неколлинеарны и

а) плоскость ABC (проходящая через точки A, B, C) параллельна плоскости $A'B'C'$ (проходящей через точки A', B', C'),

б) ориентация плоскости ABC, определенная тройкой (A, B, C) (см. п. 1 § 4), совпадает с ориентацией плоскости $A'B'C'$, определенной тройкой (A', B', C'),

в) площадь треугольника ABC равна площади треугольника $A'B'C'$.

Ясно, что

отношение эквиполлентности упорядоченных троек точек является отношением эквивалентности.

Определение 2. Классы эквиполлентных упорядоченных троек точек называются *бивекторами*. Бивектор, определенный тройкой (A, B, C) (содержащий эту тройку), обозначается символом $\overrightarrow{ABC}$.

По определению, все коллинеарные тройки точек составляют один бивектор. Этот бивектор называется нулевым и обозначается символом 0.

Замечание 1. В изложенных определениях мы неявно считали, что все рассматриваемые точки являются произвольными точками пространства. Получающиеся бивекторы естественно называть поэтому *бивекторами в пространстве*. Множество всех таких бивекторов мы будем обозначать символом $\mathrm{Biv}(3)$.

Вместе с тем мы можем органичиться точками лишь некоторой плоскости Π. Тогда мы получим *бивекторы в плоскости.*

Их множество мы будем обозначать символом $\mathrm{Biv}_{\Pi}(2)$ или просто $\mathrm{Biv}(2)$.

Аналогично можно, конечно, определить множество $\mathrm{Biv}(1)$ *бивекторов на прямой.* Однако этот случай неинтересен, поскольку множество $\mathrm{Biv}(1)$ состоит, очевидно, только из нулевого бивектора $\mathfrak{o}$.

Так же, как и для векторов (см. п. 4 § 1), каждый бивектор на плоскости естественным образом отождествляется с некоторым бивектором в пространстве, так что мы можем считать, что для любой плоскости Π имеет место естественное вложение

$$\mathrm{Biv}_{\Pi}(2) \subset \mathrm{Biv}(3).$$

Конечно, подмножества $\mathrm{Biv}_{\Pi}(2)$, соответствующие различным плоскостям Π, вообще говоря, различны. Они совпадают тогда, когда плоскости параллельны (а в противном случае они имеют только один общий бивектор $\mathfrak{o}$).

Пусть нам даны два вектора $\boldsymbol{a}$ и $\boldsymbol{b}$ (на прямой, в плоскости или в пространстве). Отложив их от одной точки O, т. е. положив

$$\boldsymbol{a} = \overrightarrow{OA}, \quad \boldsymbol{b} = \overrightarrow{OB},$$

мы можем рассмотреть бивектор $\overrightarrow{OAB}$.

Определение 3. Бивектор $\overrightarrow{OAB}$ называется *внешним произведением* векторов $\boldsymbol{a}$, $\boldsymbol{b}$ и обозначается символом $\boldsymbol{a} \wedge \boldsymbol{b}$:

$$\boldsymbol{a} \wedge \boldsymbol{b} = \overrightarrow{OAB}.$$

Конечно, это определение нуждается в проверке корректности, т. е. в доказательстве того, что бивектор $\boldsymbol{a} \wedge \boldsymbol{b}$ определяется исключительно векторами $\boldsymbol{a}$, $\boldsymbol{b}$ (взятыми в данном порядке) и не зависит от выбора точки O. Другими словами, нам нужно доказать, что

если для двух упорядоченных троек точек (O, A, B) *и* (O', A', B') *имеют место равенства*

$$\overrightarrow{OA} = \overrightarrow{O'A'}, \quad \overrightarrow{OB} = \overrightarrow{O'B'},$$

то

$$\overrightarrow{OAB} = \overrightarrow{O'A'B'}.$$

Задание. Докажите корректность определения внешнего произведения.

Ясно, что

векторы $\boldsymbol{a}$ *и* $\boldsymbol{b}$ *тогда и только тогда линейно зависимы* (*коллинеарны*), *когда их внешнее произведение равно нулю*:

$$\boldsymbol{a} \wedge \boldsymbol{b} = \mathfrak{o}.$$

Замечание 2. Очевидно, что любой бивектор $\overrightarrow{ABC}$ является внешним произведением некоторых векторов (например, векторов $\overrightarrow{AB}$ и $\overrightarrow{AC}$). Другими словами, любой бивектор определяется заданием некоторых двух векторов (этим и объясняется термин «бивектор»).

Определение 4. Две упорядоченные пары векторов $(\boldsymbol{a}, \boldsymbol{b})$ и $(\boldsymbol{a}', \boldsymbol{b}')$ называются *эквиполлентными*, если внешние произведения векторов этих пар совпадают:

$$\boldsymbol{a} \wedge \boldsymbol{b} = \boldsymbol{a}' \wedge \boldsymbol{b}'.$$

Ясно, что отношение эквиполлентности пар векторов является отношением эквивалентности.

Столь же ясно, что

пары $(\boldsymbol{a}, \boldsymbol{b})$ *и* $(\boldsymbol{a}', \boldsymbol{b}')$ *тогда и только тогда эквиполлентны, когда либо*

1) *векторы* $\boldsymbol{a}$, $\boldsymbol{b}$ *и векторы* $\boldsymbol{a}'$, $\boldsymbol{b}'$ *линейно зависимы, либо*

2) *векторы* $\boldsymbol{a}$, $\boldsymbol{b}$ *и векторы* $\boldsymbol{a}'$, $\boldsymbol{b}'$ *линейно независимы и*

а) *векторы* $\boldsymbol{a}$, $\boldsymbol{b}$ *и* $\boldsymbol{a}'$, $\boldsymbol{b}'$ *параллельны одной и той же плоскости,*

б) *базисы* $\boldsymbol{a}$, $\boldsymbol{b}$ *и* $\boldsymbol{a}'$, $\boldsymbol{b}'$ *этой плоскости одноименны,*

в) *площадь параллелограмма, построенного на векторах* $\boldsymbol{a}$, $\boldsymbol{b}$, *равна площади параллелограмма, построенного на векторах* $\boldsymbol{a}'$, $\boldsymbol{b}'$.

Замечание 3. Это утверждение можно принять за определение отношения эквиполлентности пар векторов и после этого о п р е д е л и т ь бивекторы как классы эквиполлентных упорядоченных пар векторов.

Определение 5. Мы будем говорить, что отличный от нуля бивектор в пространстве $\mathfrak{a} = \boldsymbol{a} \wedge \boldsymbol{b}$ *параллелен* плоскости П, если векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ параллельны этой плоскости. Согласно условию а) предыдущего утверждения это определение корректно, т. е. не зависит от выбора векторов $\boldsymbol{a}$ и $\boldsymbol{b}$. С точностью до параллельности плоскость П определена бивектором $\mathfrak{a}$ однозначно, а векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ составляют ее базис. Ее ориентация, определенная базисом $\boldsymbol{a}$, $\boldsymbol{b}$, также не зависит от выбора векторов $\boldsymbol{a}$ и $\boldsymbol{b}$. Мы будем говорить, что эта ориентация *определена* бивектором $\mathfrak{a}$.

Ясно, что бивектор $\mathfrak{a}$ тогда и только тогда параллелен плоскости П, когда он принадлежит множеству $\mathrm{Biv}_{\Pi}(2)$ (рассматриваемому как подмножество множества $\mathrm{Biv}(3)$).

Площадь параллелограмма, построенного на векторах $\boldsymbol{a}$ и $\boldsymbol{b}$, называется *площадью* (или *мерой*) бивектора $\mathfrak{a}$ и обозначается символом $|\mathfrak{a}|$. Она также зависит только от $\mathfrak{a}$. Для нулевого бивектора, по определению, полагается $|\mathfrak{a}| = 0$.

Таким образом,

$$|\mathfrak{a}|=0 \quad \textit{тогда и только тогда, когда} \quad \mathfrak{a}=\mathfrak{o}.$$

По определению, нулевой бивектор считается параллельным любой плоскости. Никакой ориентации он на этой плоскости не определяет.

Ясно, что

бивектор $\mathfrak{a}$ *однозначно определен, если известна его площадь и (при* $|\mathfrak{a}|\neq 0$*) известно, какой плоскости он параллелен и какую ориентацию на этой плоскости определяет.*

Наши определения эквиполлентности упорядоченных троек точек или упорядоченных пар векторов используют метрические понятия (площадь). Однако на самом деле можно показать, что отношение эквиполлентности имеет чисто аффинный характер. Покажем это.

Определение 6. Будем говорить, что пара векторов (a', b') получена из пары (a, b) *элементарным преобразованием*, если выполнено одно из следующих двух условий:

1) существует такое число $k\neq 0$, что $a'=ka$ и $b'=\frac{1}{k}b$,

2) существует такое число k, что либо $a'=a+kb$ и $b'=b$, либо $a'=a$ и $b'=b+ka$.

Ясно, что при применении элементарного преобразования к паре (a, b) ни площадь бивектора $\mathfrak{a}=a\wedge b$, ни (при $a\wedge b\neq o$) плоскость, которой он параллелен, ни ориентация этой плоскости, им определенная, не меняются. Другими словами, этот бивектор переходит в равный бивектор. Будучи верным для одного элементарного преобразования, это верно и для любой последовательности элементарных преобразований. Таким образом,

если пара (a', b') *получается из пары* (a, b) *последовательностью элементарных преобразований, то*

$$a'\wedge b'=a\wedge b,$$

т. е. пары (a, b) *и* (a', b') *эквиполлентны.*

Далее, легко видеть, что

каков бы ни был отличный от нуля вектор e*, компланарный с линейно независимыми векторами* a *и* b*, пару* (a, b) *можно последовательностью элементарных преобразований перевести в пару* (e, c)*, первый вектор которой совпадает с данным вектором* e*.*

Действительно, пусть

$$e=ka+lb.$$

Если $k\neq 0$, то

$$(a, b)\sim\left(ka, \frac{1}{k}b\right)\sim\left(ka+lb, \frac{1}{k}b\right)\sim\left(e, \frac{1}{k}b\right).$$

(Знаком $\sim$ мы обозначаем здесь элементарное преобразование.) Аналогично, если $k = 0$, то

$$(a, b) \sim (a - b, b) \sim (a - b, b + (a - b)) \sim$$

$$\sim (a - b, a) \sim (-b, a) \sim \left(lb, -\frac{1}{l} b\right) \sim \left(e, -\frac{1}{l} b\right).$$

Кроме того, нетрудно показать, что

если отличные от нуля бивекторы $a \wedge b$ *и* $a \wedge b'$ *с одинаковым первым вектором* a *равны, то пару* (a, b) *можно перевести в пару* (a, b') *некоторым элементарным преобразованием.*

Действительно, пусть $a = \overrightarrow{OA}$, $b = \overrightarrow{OB}$ и $b' = \overrightarrow{OB'}$. Поскольку базисы a, b и a, b' плоскости OAB одноименны, точки B и B' лежат в этой плоскости по одну сторону прямой OA. Поскольку площадь параллелограмма, построенного на отрезках $\overline{OA}$ и $\overline{OB}$, равна площади параллелограмма, построенного на отрезках $\overline{OA}$ и $\overline{OB'}$, точки B и B' лежат поэтому на прямой, параллельной прямой OA. Следовательно, вектор $\overrightarrow{BB'} = b' - b$ параллелен прямой OA и потому существует такое число k, что $b' - b = ka$. Таким образом, пара (a, b') получается из пары (a, b) элементарным преобразованием.

Полезно доказанное утверждение переформулировать следующим образом:

Предложение 1. *Если для вектора* $a \neq 0$ *и векторов* b *и* b' *имеет место равенство*

$$a \wedge b = a \wedge b',$$

то существует такое число k, *что*

$$b' - b = ka.$$

Заметим, что это предложение справедливо и тогда, когда $a \wedge b = a \wedge b' = 0$ (нужно лишь, чтобы $a \neq 0$).

Теперь уже легко доказать, что

если отличные от нуля бивекторы $a \wedge b$ *и* $a' \wedge b'$ *равны* (*т. е. пары* (a, b) *и* (a', b') *эквиполлентны и линейно независимы*), *то пару* (a', b') *можно получить из пары* (a, b) *последовательностью элементарных преобразований.*

Действительно, согласно доказанному, мы можем пару (a, b) элементарными преобразованиями перевести в пару вида (a', b''). При этом будет иметь место равенство $a' \wedge b' = a' \wedge b''$ и потому еще одним элементарным преобразованием мы сможем пару (a', b'') перевести в нужную нам пару (a', b').

Мы видим, таким образом, что нами доказано следующее предложение, дающее чисто аффинную характеристику равенства бивекторов:

Предложение 2. *Два отличных от нуля бивектора* $a \wedge b$ *и* $a' \wedge b'$ *тогда и только тогда равны, когда пару* (a, b) *можно последовательностью элементарных преобразований перевести в пару* (a', b').

Резюмируя, мы получаем, что

1) бивекторы можно определять либо как классы эквиполлентных троек точек (определения 1 и 2), либо как классы эквиполлентных пар векторов (определение 4 и замечание 3);

2) при этом эквиполлентность пар неколлинеарных векторов (единственный интересный случай) равносильна тому, что одну пару можно последовательностью элементарных преобразований перевести в другую (предложение 2).

2. Линейные операции над бивекторами

Определение 1. Пусть $\mathfrak{a}$ — произвольный бивектор, а k — произвольное вещественное число. Предполагая сначала, что $\mathfrak{a} \neq \mathfrak{o}$ и $k \neq 0$, мы определим *произведение* $k\mathfrak{a}$ бивектора $\mathfrak{a}$ на число k как бивектор, удовлетворяющий следующим условиям:

а) бивектор $k\mathfrak{a}$ параллелен той же плоскости, что и бивектор $\mathfrak{a}$;

б) ориентация этой плоскости, определенная бивектором $k\mathfrak{a}$, совпадает с ориентацией, определенной бивектором $\mathfrak{a}$, если $k > 0$, и противоположна этой ориентации, если $k < 0$;

в) площадь бивектора $k\mathfrak{a}$ выражается формулой

$$|k\mathfrak{a}| = |k| \cdot |\mathfrak{a}|.$$

При $k = 0$ или $\mathfrak{a} = \mathfrak{o}$ мы, по определению, положим

$$0 \cdot \mathfrak{a} = k\mathfrak{o} = \mathfrak{o}.$$

Замечание 1. Строго говоря, это определение относится к бивекторам в пространстве. Для бивекторов на плоскости условие а) излишне.

Из определения 1 непосредственно вытекает, что при умножении одного из векторов, составляющих бивектор, на некоторое число весь бивектор умножается на это число, т. е.

для любых векторов $\boldsymbol{a}$, $\boldsymbol{b}$ *и любого числа* k *имеют место формулы*

$$(k\boldsymbol{a}) \wedge \boldsymbol{b} = \boldsymbol{a} \wedge (k\boldsymbol{b}) = k(\boldsymbol{a} \wedge \boldsymbol{b})$$

(свойство однородности внешнего умножения).

Для доказательства достаточно заметить, что (при $k \neq 0$ и $\boldsymbol{a} \wedge \boldsymbol{b} \neq \mathfrak{o}$) бивектор $k(\boldsymbol{a} \wedge \boldsymbol{b})$ и, скажем, бивектор $(k\boldsymbol{a}) \wedge \boldsymbol{b}$ имеют одну и ту же площадь, параллельны одной и той же плоскости и определяют на этой плоскости одну и ту же ориентацию. При $k = 0$ или $\boldsymbol{a} \wedge \boldsymbol{b} = \mathfrak{o}$ эта формула очевидна.

Замечание 2. Формулу $k(\boldsymbol{a} \wedge \boldsymbol{b}) = (k\boldsymbol{a}) \wedge \boldsymbol{b}$ можно принять за определение произведения $k(\boldsymbol{a} \wedge \boldsymbol{b})$. Такое определение делает непосредственно очевидным аффинный характер операции $k\mathfrak{a}$. Зато оно требует проверки корректности.

По определению, бивектор $-\mathfrak{a} = (-1)\mathfrak{a}$ имеет ту же площадь и параллелен той же плоскости (при $\mathfrak{a} \neq \mathfrak{o}$), что и бивектор $\mathfrak{a}$, но определяет на этой плоскости противоположную ориен-

тацию. Но при $\mathfrak{a} = \boldsymbol{a} \wedge \boldsymbol{b}$ теми же свойствами обладает и бивектор $\boldsymbol{b} \wedge \boldsymbol{a}$. Поэтому

$$\boldsymbol{b} \wedge \boldsymbol{a} = -\boldsymbol{a} \wedge \boldsymbol{b},$$

т. е. *при перестановке множителей внешнее произведение меняет знак.*

Это свойство внешнего умножения называется его антикоммутативностью.

Определение 2. Бивекторы $\mathfrak{a}$ и $\mathfrak{b}$ в пространстве называются *коллинеарными*, если они параллельны одной и той же плоскости. Все бивекторы на плоскости, по определению, коллинеарны. В частности, бивекторы коллинеарны, если хотя бы один из них равен нулю.

Ясно, что аналогично векторам

бивекторы $\mathfrak{a}$ и $\mathfrak{b}$ тогда и только тогда коллинеарны, когда существует такое число k, что либо $\mathfrak{a} = k\mathfrak{b}$, либо $\mathfrak{b} = k\mathfrak{a}$.

Для определения операции сложения бивекторов нам понадобятся следующие две леммы.

Лемма 1. *Для вектора $\boldsymbol{e}$ и бивектора $\mathfrak{a}$ тогда и только тогда существует такой вектор $\boldsymbol{a}$, что*

$$\mathfrak{a} = \boldsymbol{e} \wedge \boldsymbol{a},$$

когда либо бивектор $\mathfrak{a}$ равен нулю, либо вектор $\boldsymbol{e}$ отличен от нуля и параллелен плоскости бивектора $\mathfrak{a}$.

Доказательство. Если $\mathfrak{a} = \boldsymbol{e} \wedge \boldsymbol{a}$, то (при $\mathfrak{a} \neq \mathfrak{o}$) вектор $\boldsymbol{e}$ параллелен, по определению, плоскости бивектора $\mathfrak{a}$. Обратно, пусть вектор $\boldsymbol{e}$ параллелен плоскости бивектора $\mathfrak{a}$ (мы можем считать, что $\mathfrak{a} \neq \mathfrak{o}$, поскольку при $\mathfrak{a} = \mathfrak{o}$ за вектор $\boldsymbol{a}$ можно принять вектор $\boldsymbol{e}$). Искомый вектор $\boldsymbol{a}$ должен удовлетворять следующим условиям:

а) он параллелен плоскости бивектора $\mathfrak{a}$;

б) векторы $\boldsymbol{e}$, $\boldsymbol{a}$ составляют положительно ориентированный базис этой плоскости (по отношению к ориентации, определенной бивектором $\mathfrak{a}$);

в) площадь параллелограмма, построенного на векторах $\boldsymbol{e}$, $\boldsymbol{a}$, равна площади $|\mathfrak{a}|$ бивектора $\mathfrak{a}$.

Ясно, что вектор $\boldsymbol{a}$, удовлетворяющий этим условиям, всегда можно найти (и даже с большим произволом).

Замечание 3. Воспользовавшись произволом в выборе вектора $\boldsymbol{a}$, мы можем его выбрать *ортогональным* вектору $\boldsymbol{e}$. Тогда он будет однозначно определен (вектором $\boldsymbol{e}$ и бивектором $\mathfrak{a}$). Но эта нормализация выбора вектора $\boldsymbol{a}$ уже не будет иметь аффинного характера.

Лемма 2. *Для любых двух бивекторов $\mathfrak{a}$ и $\mathfrak{b}$ существуют такие векторы $\boldsymbol{e}$, $\boldsymbol{a}$ и $\boldsymbol{b}$, что*

$$\mathfrak{a} = \boldsymbol{e} \wedge \boldsymbol{a}, \quad \mathfrak{b} = \boldsymbol{e} \wedge \boldsymbol{b}.$$

Доказательство. Если один из данных бивекторов, скажем, бивектор $\mathfrak{b}$ равен нулю, то за векторы e и a можно взять произвольные векторы, обладающие тем свойством, что $\mathfrak{a} = e \wedge a$, а за вектор b принять, например, вектор e. Если же $\mathfrak{a} \neq 0$ и $\mathfrak{b} \neq 0$, то за вектор e следует принять отличный от нуля вектор, параллельный плоскостям бивекторов $\mathfrak{a}$ и $\mathfrak{b}$. При таком выборе вектора e возможность требуемого представления бивекторов $\mathfrak{a}$ и $\mathfrak{b}$ немедленно обеспечивается леммой 1.

Замечание 4. Ясно, что векторы e, a, b мы всегда можем выбрать так, чтобы вектор e был ортогонален векторам a и b (ср. замечание 3).

Определение 3. *Суммой* бивекторов $\mathfrak{a}$ и $\mathfrak{b}$, представленных в виде $e \wedge a$ и $e \wedge b$, называется бивектор

$$\mathfrak{a} + \mathfrak{b} = e \wedge (a + b).$$

Это определение нуждается, конечно, в проверке корректности, т. е. в доказательстве независимости бивектора от выбора векторов e, a и b.

Случай 1. Бивекторы $\mathfrak{a}$ и $\mathfrak{b}$ неколлинеарны (и, в частности, отличны от нуля). В этом случае, вектор e определен однозначно с точностью до коллинеарности (поскольку он должен быть параллелен плоскостям обоих бивекторов). Другими словами, если

$$\mathfrak{a} = e' \wedge a', \quad \mathfrak{b} = e' \wedge b',$$

то $e' = le$, где l — некоторое отличное от нуля число.

По условию,

$$e' \wedge a' = e \wedge a.$$

В то же время

$$e' \wedge a' = (le) \wedge a' = l(e \wedge a') = e \wedge (la').$$

Таким образом,

$$e \wedge a = e \wedge (la').$$

С другой стороны, в п. 1 было доказано (предложение 1), что при $e \neq 0$ такого рода равенство возможно только тогда, когда

$$la' - a = ke,$$

где k — некоторое вещественное число.

Аналогично показывается, что существует такое число k_1, что

$$lb' - b = k_1 e.$$

Поэтому

$$e' \wedge (a' + b') = e \wedge (l(a' + b')) =$$
$$= e \wedge ((a + b) + (k + k_1)e) = e \wedge (a + b).$$

(Последнее равенство имеет место потому, что переход от пары $(e, (a+b)+(k+k_1)e)$ к паре $(e, a+b)$ является элементарным преобразованием.) Таким образом, в рассматриваемом случае бивектор $\mathfrak{a}+\mathfrak{b}$ определен корректно.

Случай 2. Бивекторы $\mathfrak{a}$ и $\mathfrak{b}$ коллинеарны, т. е. существует такое число l, что, скажем, $\mathfrak{b}=l\mathfrak{a}$. В этом случае

$$e\wedge b=l(e\wedge a)=e\wedge(la)$$

и потому

$$la-b=ke,$$

где k — некоторое вещественное число. Следовательно,

$$e\wedge(a+b)=e\wedge((1+l)a-ke)=e\wedge((1+l)a)=$$
$$=(1+l)(e\wedge a)=(1+l)\mathfrak{a}.$$

Таким образом, при $\mathfrak{b}=l\mathfrak{a}$

$$\mathfrak{a}+\mathfrak{b}=(1+l)\mathfrak{a}. \tag{1}$$

Поскольку правая часть этой формулы не зависит от выбора векторов e, a и b, корректность определения бивектора $\mathfrak{a}+\mathfrak{b}$ доказана и в этом случае.

Определение суммы двух бивекторов можно (несколько меняя обозначения) записать в виде равенства

$$a\wedge(b+c)=a\wedge b+a\wedge c,$$

показывающего, что внешнее умножение обладает по отношению к сложению свойством дистрибутивности.

Замечание 5. Все сказанное (с соответствующими незначительными изменениями) годится, конечно, и для бивекторов на плоскости. Впрочем, в этом случае проще всего определять сумму бивекторов непосредственно формулой (1).

Из этого замечания вытекает, в частности, что для любой плоскости Π вложение множества $\mathrm{Biv}_\Pi(2)$ в множество $\mathrm{Biv}(3)$ *согласовано* с операциями над бивекторами, т. е., например, сумма двух бивекторов из $\mathrm{Biv}_\Pi(2)$ не зависит от того, рассматриваем ли мы их как бивекторы на плоскости (из $\mathrm{Biv}_\Pi(2)$) или в пространстве (из $\mathrm{Biv}(3)$) (ср. для векторов аналогичное замечание 1 в п. 2 § 3).

Теорема 1. *Построенные операции над бивекторами (на плоскости или в пространстве) обладают всеми восемью стандартными свойствами линейных операций над векторами, т. е. множества бивекторов* $\mathrm{Biv}(2)$ *и* $\mathrm{Biv}(3)$ *являются линеалами.*

Доказательство этой теоремы сводится к серии тривиальных проверок, за исключением доказательства ассоциативности сложения (свойство 1) бивекторов в пространстве, которое требует достаточно канительных и сложных геометрических рассуждений. Мы предлагаем читателю самостоятельно найти

это доказательство. Мы же докажем эту теорему косвенным способом, связав бивекторы с векторами. К сожалению, этот косвенный способ будет существенно использовать метрические соображения, хотя сама теорема имеет чисто аффинный характер. Однако, ради простоты изложения, мы за чистотой метода гнаться здесь не будем.

Пусть $e = \overrightarrow{OE}$ — произвольный единичный вектор в пространстве. Предполагая, что пространство ориентировано, мы каждому вектору $a \neq 0$, ортогональному вектору e, сопоставим вектор $e(a)$, обладающий следующими свойствами:

1) длина вектора $e(a)$ равна длине вектора a:

$$|e(a)| = |a|;$$

2) вектор $e(a)$ ортогонален как вектору e, так и вектору a;

3) базис e, a, $e(a)$ пространства положительно ориентирован.

Для нулевого вектора мы положим $e(0) = 0$.

Ясно, что эти условия однозначно определяют вектор $e(a)$.

Наглядно вектор $e(a)$ можно описать как вектор, получающийся вращением вектора a вокруг прямой OE на угол $\pi/2$ в направлении, положительном относительно данной ориентации пространства.

Очевидно, что

для любых векторов a, b, ортогональных вектору e, и любого числа k имеют место равенства

$$e(a+b) = e(a) + e(b),$$
$$e(ka) = ke(a).$$

Пусть $\mathfrak{a}$ — произвольный бивектор (в пространстве). Согласно лемме 1 и замечанию 3 мы можем (вообще говоря, многими способами) представить этот бивектор в виде

$$\mathfrak{a} = e \wedge a,$$

где e — некоторый единичный вектор, а a — вектор, ортогональный вектору e. Сделав это, мы сопоставим бивектору $\mathfrak{a}$ вектор $\mathfrak{a}^\perp$, определенный формулой

$$\mathfrak{a}^\perp = e(a).$$

Определение 4. Вектор $\mathfrak{a}^\perp$ называется вектором, *ассоциированным* с бивектором $\mathfrak{a}$ (по отношению к данной ориентации пространства).

Здесь, конечно, нужно проверить корректность определения, т. е. независимость вектора $\mathfrak{a}^\perp$ от выбора векторов e и a. Но, как легко видеть, $\mathfrak{o}^\perp = 0$, а при $\mathfrak{a} \neq \mathfrak{o}$ вектор $\mathfrak{a}^\perp$ обладает следующими свойствами:

1) длина $|\mathfrak{a}^\perp|$ вектора $\mathfrak{a}^\perp$ равна площади $|\mathfrak{a}|$ бивектора $\mathfrak{a}$:

$$|\mathfrak{a}^\perp| = |\mathfrak{a}|;$$

2) вектор $\mathfrak{a}^\perp$ перпендикулярен каждой плоскости, которой параллелен бивектор $\mathfrak{a}$;

3) произведение ориентации, определенной бивектором $\mathfrak{a}$, на ориентацию, определенную вектором $\mathfrak{a}^\perp$ (это произведение определено, поскольку соответствующие плоскость и прямая не параллельны), совпадает с данной ориентацией пространства.

Поскольку эти свойства, очевидно, однозначно определяют вектор $\mathfrak{a}^\perp$ (вне зависимости от выбора векторов $\boldsymbol{e}$ и $\boldsymbol{a}$), корректность определения вектора $\mathfrak{a}^\perp$ тем самым полностью доказана.

Кроме того, из этих свойств немедленно вытекает, что, *сопоставив произвольному бивектору* $\mathfrak{a}$ *ассоциированный вектор* $\mathfrak{a}^\perp$, *мы получим биективное отображение множества всех бивекторов* Biv(3) *на множество* Vect(3) *всех векторов.*

Более того, оказывается, что

это отображение является изоморфизмом, т. е. для любых бивекторов $\mathfrak{a}$, $\mathfrak{b}$ *и любого числа* k *имеют место соотношения*

$$(\mathfrak{a}+\mathfrak{b})^\perp = \mathfrak{a}^\perp + \mathfrak{b}^\perp,$$
$$(k\mathfrak{a})^\perp = k\mathfrak{a}^\perp.$$

Действительно, согласно лемме 2 и замечанию 4 существует такой единичный вектор $\boldsymbol{e}$ и такие векторы $\boldsymbol{a}$ и $\boldsymbol{b}$, ортогональные вектору $\boldsymbol{e}$, что

$$\mathfrak{a} = \boldsymbol{e} \wedge \boldsymbol{a}, \quad \mathfrak{b} = \boldsymbol{e} \wedge \boldsymbol{b}$$

и потому

$$\mathfrak{a}+\mathfrak{b} = \boldsymbol{e} \wedge (\boldsymbol{a}+\boldsymbol{b}),$$

где вектор $\boldsymbol{a}+\boldsymbol{b}$ также ортогонален вектору $\boldsymbol{e}$. Следовательно,

$$(\mathfrak{a}+\mathfrak{b})^\perp = \boldsymbol{e}(\boldsymbol{a}+\boldsymbol{b}) = \boldsymbol{e}(\boldsymbol{a}) + \boldsymbol{e}(\boldsymbol{b}) = \mathfrak{a}^\perp + \mathfrak{b}^\perp.$$

Аналогично, $k\mathfrak{a} = \boldsymbol{e} \wedge (k\boldsymbol{a})$ и потому

$$(k\mathfrak{a})^\perp = \boldsymbol{e}(k\boldsymbol{a}) = k\boldsymbol{e}(\boldsymbol{a}) = k\mathfrak{a}^\perp.$$

Таким образом, мы доказали следующую теорему об изоморфизме:

Теорема 2. *Линеал* Biv(3) *изоморфен линеалу* Vect(3).

Заметим, что построенный изоморфизм не является естественным: он зависит от выбора ориентации пространства (при переходе к противоположной ориентации ассоциированный вектор $\mathfrak{a}^\perp$ умножается на -1). Кроме того, он имеет существенно метрический характер.

Однако, как бы то ни было, из теоремы об изоморфизме следует, что линейные операции над бивекторами обладают всеми алгебраическими свойствами линейных операций над векторами. В частности, они обладают и интересующими нас восемью свойствами. Тем самым теорему 1 мы можем считать полностью доказанной.

Подчеркнем еще раз, что хотя мы доказали эту теорему, пользуясь метрическими соображениями, по существу она имеет чисто аффинный характер.

Замечание 6. Для бивекторов на плоскости теорема об изоморфизме имеет вид:

Линеал Biv(2) *изоморфен линеалу* Vect(1).

Для доказательства достаточно заметить, что построенный выше изоморфизм $\mathfrak{a} \longmapsto \mathfrak{a}^{\perp}$ переводит бивекторы, параллельные некоторой плоскости, в векторы, параллельные перпендикуляру к этой плоскости.

3. Линейная теория бивекторов

Понятия линейной комбинации и линейной зависимости (независимости) определяются для бивекторов точно так же, как для векторов. Все свойства этих понятий, доказанные в п. 3 § 3, сохраняются и для бивекторов, поскольку в их доказательствах мы пользовались лишь стандартными свойствами линейных операций.

Замечание 1. Все это, конечно, еще проще вытекает из теоремы об изоморфизме. Однако мы предпочли сослаться на теорему 1, чтобы подчеркнуть аффинный характер этих утверждений.

Напомним (определение 3 п. 2), что бивекторы называются коллинеарными, если они параллельны одной и той же плоскости. Выше фактически уже доказано следующее

Предложение 1. *Два бивектора тогда и только тогда коллинеарны, когда они линейно зависимы.*

Определение 1. Бивекторы называются *компланарными*, если они параллельны одной и той же прямой (т. е. этой прямой параллельна каждая плоскость, параллельная хотя бы одному из этих бивекторов). Ясно, что

бивекторы $\mathfrak{a}$, $\mathfrak{b}$ *и* $\mathfrak{c}$ *тогда и только тогда компланарны, когда существует такой отличный от нуля вектор* $\boldsymbol{e}$, *что*

$$\mathfrak{a} = \boldsymbol{e} \wedge \boldsymbol{a}, \quad \mathfrak{b} = \boldsymbol{e} \wedge \boldsymbol{b}, \quad \mathfrak{c} = \boldsymbol{e} \wedge \boldsymbol{c},$$

где $\boldsymbol{a}$, $\boldsymbol{b}$ *и* $\boldsymbol{c}$ *— некоторые векторы.*

Покажем, что справедливо следующее

Предложение 2. *Три бивектора тогда и только тогда компланарны, когда они линейно зависимы.*

Доказательство. Пусть

$$\mathfrak{a} = \boldsymbol{e} \wedge \boldsymbol{a}, \quad \mathfrak{b} = \boldsymbol{e} \wedge \boldsymbol{b}, \quad \mathfrak{c} = \boldsymbol{e} \wedge \boldsymbol{c}.$$

Четыре вектора $\boldsymbol{e}$, $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ пространства обязательно линейно зависимы, т. е. существуют такие числа k_0, k_1, k_2, k_3, не все равные нулю, что

$$k_0\boldsymbol{e} + k_1\boldsymbol{a} + k_2\boldsymbol{b} + k_3\boldsymbol{c} = \boldsymbol{0}. \qquad (1)$$

При этом числа k_1, k_2, k_3 не могут быть все равны нулю, поскольку в противном случае $k_0 \neq 0$ и $k_0 e = 0$, что при $e \neq 0$ невозможно.

Умножив равенство (1) внешним образом на вектор e и воспользовавшись свойствами однородности и дистрибутивности внешнего умножения (а также тем, что $e \wedge (k_0 e) = 0$), мы немедленно получим равенство

$$k_1 (e \wedge a) + k_2 (e \wedge b) + k_3 (e \wedge c) = 0,$$

т. е. равенство

$$k_1 \mathfrak{a} + k_2 \mathfrak{b} + k_3 \mathfrak{c} = 0.$$

Следовательно, бивекторы $\mathfrak{a}$, $\mathfrak{b}$ и $\mathfrak{c}$ линейно зависимы.

Обратно, пусть бивекторы $\mathfrak{a}$, $\mathfrak{b}$ и $\mathfrak{c}$ линейно зависимы. Тогда один из них (скажем, бивектор $\mathfrak{c}$) линейно выражается через другие:

$$\mathfrak{c} = k_1 \mathfrak{a} + k_2 \mathfrak{b}.$$

Пусть e, a и b — такие векторы, что

$$\mathfrak{a} = e \wedge a, \quad \mathfrak{b} = e \wedge b.$$

Тогда

$$\mathfrak{c} = e \wedge (k_1 a + k_2 b).$$

Следовательно, бивекторы $\mathfrak{a}$, $\mathfrak{b}$ и $\mathfrak{c}$ компланарны.

Замечание 2. Легко видеть, что

бивекторы $\mathfrak{a}$, $\mathfrak{b}$ *и* $\mathfrak{c}$ *тогда и только тогда компланарны, когда компланарны ассоциированные с ними векторы* $\mathfrak{a}^\perp$, $\mathfrak{b}^\perp$, $\mathfrak{c}^\perp$.

Геометрически это легко усмотреть, заметив, что бивектор $\mathfrak{a}$ тогда и только тогда может быть приложен к вектору e (т. е. представлен в виде $e \wedge a$), когда вектор $\mathfrak{a}^\perp$ ортогонален вектору e. Более формально это вытекает в силу теоремы об изоморфизме из того, что как для векторов, так и для бивекторов компланарность равносильна линейной зависимости.

Очевидно, что аналогичное замечание справедливо и по отношению к коллинеарности.

Аналогично случаю векторов, семейство $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$ трех некомпланарных (т. е. линейно независимых) бивекторов является базисом линеала $\mathrm{Biv}(3)$. Мы будем его называть *бивекторным базисом* в пространстве. Из теоремы об изоморфизме немедленно вытекает, что

любой бивектор $\mathfrak{a}$ *линейно выражается через базис* $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$, *т. е. существуют такие (однозначно определенные) числа* a^1, a^2, a^3, *что*

$$\mathfrak{a} = a^1 \mathfrak{e}_1 + a^2 \mathfrak{e}_2 + a^3 \mathfrak{e}_3.$$

Чтобы подчеркнуть аффинный характер этого утверждения (и выяснить его геометрический смысл), мы дадим сейчас ему доказательство, не опирающееся на теорему об изоморфизме.

Сначала мы докажем его для бивекторных базисов некоторого специального вида.

Пусть e_1, e_2, e_3 — произвольный векторный базис. Ясно, что бивекторы [1])

$$e_2 \wedge e_3, \quad e_3 \wedge e_1, \quad e_1 \wedge e_2$$

не компланарны и потому составляют некоторый бивекторный базис. Мы будем называть его бивекторным базисом, *соответствующим* данному векторному базису e_1, e_2, e_3. Представив произвольный бивектор $\mathfrak{a}$ в виде $\boldsymbol{a} \wedge \boldsymbol{b}$, разложив векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ по векторам базиса e_1, e_2, e_3 и воспользовавшись свойствами антикоммутативности, однородности и дистрибутивности внешнего умножения, мы немедленно получим, что

$$\mathfrak{a} = (a^2b^3 - a^3b^2)(e_2 \wedge e_3) + (a^3b^1 - a^1b^3)(e_3 \wedge e_1) + \\ + (a^1b^2 - a^2b^1)(e_1 \wedge e_2),$$

т. е. что

$$\boldsymbol{a} \wedge \boldsymbol{b} = \begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix} (e_2 \wedge e_3) - \begin{vmatrix} a^1 & a^3 \\ b^1 & b^3 \end{vmatrix} (e_3 \wedge e_1) + \begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix} (e_1 \wedge e_2), \quad (2)$$

где a^1, a^2, a^3 и b^1, b^2, b^3 — координаты векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ в базисе e_1, e_2, e_3.

Таким образом, для бивекторных базисов, соответствующих векторным базисам, наше утверждение доказано. Чтобы доказать его в полном объеме, достаточно теперь попытаться доказать, что любой бивекторный базис $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$ соответствует некоторому векторному базису e_1, e_2, e_3.

Попробуем это сделать. Пусть Π_1, Π_2 и Π_3 — плоскости, параллельные бивекторам $\mathfrak{e}_1$, $\mathfrak{e}_2$ и $\mathfrak{e}_3$. Эти плоскости попарно не параллельны и потому определены прямая α_1 пересечения плоскостей Π_2 и Π_3, прямая α_2 пересечения плоскостей Π_1 и Π_3 и прямая α_3 пересечения плоскостей Π_1 и Π_2. Пусть $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ и $\boldsymbol{a}_3$ — произвольные базисы на прямых α_1, α_2 и α_3 соответственно. Ясно, что векторы $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$ не компланарны (в противном случае были бы компланарны бивекторы $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$) и потому составляют базис пространства. Кроме того, векторы $\boldsymbol{a}_2$ и $\boldsymbol{a}_3$ составляют базис плоскости Π_1, векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_3$ — базис плоскости Π_2, а векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ — базис плоскости Π_3.

Поскольку вектор $\boldsymbol{a}_2$ параллелен плоскости Π_1, мы можем представить бивектор $\mathfrak{e}_1$ в виде внешнего произведения

$$\mathfrak{e}_1 = \boldsymbol{a}_2 \wedge \boldsymbol{a}'$$

вектора $\boldsymbol{a}_2$ и некоторого вектора $\boldsymbol{a}'$, параллельного плоскости Π_1 и потому выражающегося через векторы $\boldsymbol{a}_2$ и $\boldsymbol{a}_3$:

$$\boldsymbol{a}' = p_1\boldsymbol{a}_2 + q_1\boldsymbol{a}_3.$$

[1]) Обратите внимание на порядок индексов.

Пользуясь свойствами внешнего умножения, мы находим отсюда, что

$$\mathfrak{e}_1 = \boldsymbol{a}_2 \wedge (p_1\boldsymbol{a}_2 + q_1\boldsymbol{a}_3) = q_1(\boldsymbol{a}_2 \wedge \boldsymbol{a}_3).$$

Аналогично показывается, что существуют такие числа q_2 и q_3, что

$$\mathfrak{e}_2 = q_2(\boldsymbol{a}_1 \wedge \boldsymbol{a}_3), \quad \mathfrak{e}_3 = q_3(\boldsymbol{a}_1 \wedge \boldsymbol{a}_2).$$

Постараемся теперь подобрать такие числа k_1, k_2, k_3, чтобы бивекторный базис, соответствующий векторному базису

$$\boldsymbol{e}_1 = k_1\boldsymbol{a}_1, \quad \boldsymbol{e}_2 = k_2\boldsymbol{a}_2, \quad \boldsymbol{e}_3 = k_3\boldsymbol{a}_3,$$

совпал с базисом $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$. Ясно, что для этого необходимо и достаточно, чтобы

$$q_1 = k_2k_3, \quad q_2 = -k_1k_3, \quad q_3 = k_1k_2.$$

Решая эту систему уравнений, мы получим, что

$$k_1 = \frac{\sqrt{-q_1q_2q_3}}{q_1}, \quad k_2 = \frac{\sqrt{-q_1q_2q_3}}{-q_2}, \quad k_3 = \frac{\sqrt{-q_1q_2q_3}}{q_3}.$$

Эти формулы показывают, что наше построение приводит к цели только тогда, когда произведение $q_1q_2q_3$ отрицательно (в противном случае для чисел k_1, k_2, k_3 получаются мнимые значения). Таким образом, наше предположение, что любой бивекторный базис соответствует некоторому векторному базису, вообще говоря, ошибочно (или, по крайней мере, не может быть доказано изложенным методом). Тем не менее наши рассуждения показывают, что

какой бы бивекторный базис $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$ *мы ни взяли, либо он, либо базис* $\mathfrak{e}_1$, $\mathfrak{e}_2$, $-\mathfrak{e}_3$ *соответствует некоторому векторному базису.*

Действительно, если для данного базиса $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$ произведение $q_1q_2q_3$ оказалось положительным, то для базиса $\mathfrak{e}_1$, $\mathfrak{e}_2$, $-\mathfrak{e}_3$ аналогичное произведение будет отрицательным (ибо множитель q_3 сменит знак).

Но этого утверждения нам достаточно, поскольку бивектор, выражающийся через базис $\mathfrak{e}_1$, $\mathfrak{e}_2$, $-\mathfrak{e}_3$, выражается и через базис $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$. Таким образом, любой бивектор действительно можно выразить через базис $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$.

Коэффициенты a^1, a^2, a^3 линейного выражения бивектора $\mathfrak{a}$ через базис $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$, естественно, называются *координатами* бивектора $\mathfrak{a}$. Так же, как для векторов, доказывается, что

координаты суммы бивекторов равны суммам координат слагаемых, а координаты произведения бивектора на число равны произведениям координат бивектора на это число.

Другими словами, выбор базиса $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$ определяет, как и полагается, изоморфизм линеала $\mathrm{Biv}(3)$ всех бивекторов в пространстве на стандартный линеал $\mathbb{R}^3$ троек вещественных чисел.

Замечание 3. На практике часто встает вопрос о вычислении координат внешнего произведения $\boldsymbol{a} \wedge \boldsymbol{b}$ по координатам векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ в некотором базисе $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$. При этом, конечно, имеются в виду координаты бивектора $\boldsymbol{a} \wedge \boldsymbol{b}$ в базисе $\boldsymbol{e}_2 \wedge \boldsymbol{e}_3$, $\boldsymbol{e}_3 \wedge \boldsymbol{e}_1$, $\boldsymbol{e}_1 \wedge \boldsymbol{e}_2$, соответствующем базису $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$. Ответ на этот вопрос у нас фактически уже получен: он дается формулой (2).

Для запоминания этой формулы полезно заметить, что участвующие в ней определители являются минорами второго порядка матрицы

$$\begin{pmatrix} a^1 & a^2 & a^3 \\ b^1 & b^2 & b^3 \end{pmatrix},$$

составленной из координат данных векторов.

Вспомнив формулу разложения определителя третьего порядка по элементам строки (и учтя антикоммутативность внешнего умножения), мы можем формулу (2) переписать в следующем «детерминантном» виде:

$$\boldsymbol{a} \wedge \boldsymbol{b} = \begin{vmatrix} \boldsymbol{e}_2 \wedge \boldsymbol{e}_3 & \boldsymbol{e}_3 \wedge \boldsymbol{e}_1 & \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \\ a^1 & a^2 & a^3 \\ b^1 & b^2 & b^3 \end{vmatrix}.$$

В частности, если нам дан некоторый другой векторный базис $\boldsymbol{e}_{1'}, \boldsymbol{e}_{2'}, \boldsymbol{e}_{3'}$, связанный с базисом $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ соотношениями

$$\begin{aligned} \boldsymbol{e}_{1'} &= c^1_{1'}\boldsymbol{e}_1 + c^2_{1'}\boldsymbol{e}_2 + c^3_{1'}\boldsymbol{e}_3, \\ \boldsymbol{e}_{2'} &= c^1_{2'}\boldsymbol{e}_1 + c^2_{2'}\boldsymbol{e}_2 + c^3_{2'}\boldsymbol{e}_3, \\ \boldsymbol{e}_{3'} &= c^1_{3'}\boldsymbol{e}_1 + c^2_{3'}\boldsymbol{e}_2 + c^3_{3'}\boldsymbol{e}_3, \end{aligned}$$

то для соответствующего бивекторного базиса $\boldsymbol{e}_{2'} \wedge \boldsymbol{e}_{3'}$, $\boldsymbol{e}_{3'} \wedge \boldsymbol{e}_{1'}$, $\boldsymbol{e}_{1'} \wedge \boldsymbol{e}_{2'}$ будут иметь место равенства

$$\boldsymbol{e}_{2'} \wedge \boldsymbol{e}_{3'} = \begin{vmatrix} \boldsymbol{e}_2 \wedge \boldsymbol{e}_3 & \boldsymbol{e}_3 \wedge \boldsymbol{e}_1 & \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \\ c^1_{2'} & c^2_{2'} & c^3_{2'} \\ c^1_{3'} & c^2_{3'} & c^3_{3'} \end{vmatrix},$$

$$\boldsymbol{e}_{3'} \wedge \boldsymbol{e}_{1'} = \begin{vmatrix} c^1_{1'} & c^2_{1'} & c^3_{1'} \\ \boldsymbol{e}_2 \wedge \boldsymbol{e}_3 & \boldsymbol{e}_3 \wedge \boldsymbol{e}_1 & \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \\ c^1_{3'} & c^2_{3'} & c^3_{3'} \end{vmatrix},$$

$$\boldsymbol{e}_{1'} \wedge \boldsymbol{e}_{2'} = \begin{vmatrix} c^1_{1'} & c^2_{1'} & c^3_{1'} \\ c^1_{2'} & c^2_{2'} & c^3_{2'} \\ \boldsymbol{e}_2 \wedge \boldsymbol{e}_3 & \boldsymbol{e}_3 \wedge \boldsymbol{e}_1 & \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \end{vmatrix}.$$

Отсюда вытекает, что

если некоторый бивектор $\mathfrak{x}$ *имел в базисе* $\boldsymbol{e}_2 \wedge \boldsymbol{e}_3$, $\boldsymbol{e}_3 \wedge \boldsymbol{e}_1$, $\boldsymbol{e}_1 \wedge \boldsymbol{e}_2$ *координаты* x^1, x^2, x^3, *то в базисе* $\boldsymbol{e}_{2'} \wedge \boldsymbol{e}_{3'}$, $\boldsymbol{e}_{3'} \wedge \boldsymbol{e}_{1'}$, $\boldsymbol{e}_{1'} \wedge \boldsymbol{e}_{2'}$ *он будет иметь координаты* $x^{1'}$, $x^{2'}$, $x^{3'}$, *связанные с координатами* x^1, x^2, x^3 *соотношениями*

$$x^1 = \begin{vmatrix} x^{1'} & c^2_{1'} & c^3_{1'} \\ x^{2'} & c^2_{2'} & c^3_{2'} \\ x^{3'} & c^2_{3'} & c^3_{3'} \end{vmatrix}, \quad x^2 = \begin{vmatrix} c^1_{1'} & x^{1'} & c^3_{1'} \\ c^1_{2'} & x^{2'} & c^3_{2'} \\ c^1_{3'} & x^{3'} & c^3_{3'} \end{vmatrix}, \quad x^3 = \begin{vmatrix} c^1_{1'} & c^2_{1'} & x^{1'} \\ c^1_{2'} & c^2_{2'} & x^{2'} \\ c^1_{3'} & c^2_{3'} & x^{3'} \end{vmatrix}.$$

Замечание 4. Конечно, аналогичные (только более простые) результаты имеют место и для бивекторов на плоскости.

Например, бивекторный базис на плоскости — это просто произвольный отличный от нуля бивектор. Каждый векторный базис $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ на плоскости определяет бивекторный базис $\boldsymbol{e}_1 \wedge \boldsymbol{e}_2$, причем для любых векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ на плоскости

координатой бивектора $\boldsymbol{a} \wedge \boldsymbol{b}$ *в базисе* $\boldsymbol{e}_1 \wedge \boldsymbol{e}_2$ *является определитель*

$$\begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix},$$

где a^1, a^2 *и* b^1, b^2 — *координаты векторов* $\boldsymbol{a}$ *и* $\boldsymbol{b}$ *в базисе* $\boldsymbol{e}_1$, $\boldsymbol{e}_2$.

При переходе от базиса $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ к другому базису $\boldsymbol{e}_{1'}$, $\boldsymbol{e}_{2'}$ эта координата умножается, очевидно, на определитель $|C|$ соответствующей матрицы перехода C.

4. Метрическая теория бивекторов

Метрическая теория бивекторов изучает вопросы, связанные с площадью бивектора и с углом между бивекторами.

Площадь бивектора мы уже определили. Займемся теперь углом между двумя (отличными от нуля) бивекторами. Ясно, что это понятие (аналогично случаю векторов) сводится к понятию угла между плоскостями. Здесь мы также сталкиваемся с вопросом, какой из двух вертикальных углов между плоскостями мы должны выбрать. Для случая векторов мы преодолели аналогичную трудность, введя в рассмотрение угол между ориентированными прямыми. Сделаем то же самое и для плоскостей.

Пусть Π_1 и Π_2 — две ориентированные плоскости в пространстве. Выбрав единичный вектор $\boldsymbol{e}_1$, параллельный обеим плоскостям, подберем векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ так, чтобы векторы $\boldsymbol{e}$, $\boldsymbol{a}_1$ составляли положительно ориентированный ортонормированный базис плоскости Π_1, а векторы $\boldsymbol{e}$, $\boldsymbol{a}_2$ — положительно ориентированный ортонормированный базис плоскости Π_2. Угол θ между

векторами a_1 и a_2 мы и объявим углом между ориентированными плоскостями Π_1 и Π_2. Корректность этого определения очевидна: единственный произвол состоит в выборе орта e, но при $\theta \neq 0$ его можно заменить только ортом $-e$ и тогда векторы a_1 и a_2 перейдут в векторы $-a_1$ и $-a_2$. Угол же между векторами $-a_1$ и $-a_2$ равен углу между векторами a_1 и a_2. При $\theta = 0$ векторы a_1 и a_2 совпадают при любом выборе вектора e.

Несколько иначе угол между плоскостями Π_1 и Π_2 можно определить, выбрав в пространстве некоторую ориентацию (от которой окончательный результат не будет зависеть) и сопоставив каждой ориентированной плоскости Π перпендикулярный ей вектор n, направленный в положительную сторону, т. е. в такую, что произведение ориентации плоскости Π на ориентацию перпендикулярной прямой, определенной вектором n, совпадает с данной ориентацией пространства (о таком векторе n принято говорить, что он является вектором *положительной нормали* к ориентированной плоскости в ориентированном пространстве). Тогда углом между ориентированными плоскостями Π_1 и Π_2 называется, по определению, угол между соответствующими векторами n_1 и n_2.

Чтобы сравнить это определение с предыдущим, достаточно заметить, что в обозначениях, введенных в п. 2, вектор n_1 является вектором $e(a_1)$, а вектор n_2 — вектором $e(a_2)$, и, следовательно, угол между векторами a_1 и a_2 совпадает с углом между векторами n_1 и n_2.

Определение 1. Пусть $\mathfrak{a}$ и $\mathfrak{b}$ — произвольные отличные от нуля бивекторы. Рассмотрим произвольную плоскость Π_1, которой параллелен бивектор $\mathfrak{a}$, и произвольную плоскость Π_2, которой параллелен бивектор $\mathfrak{b}$. Будем считать эти плоскости снабженными ориентациями, определенными соответственно бивекторами $\mathfrak{a}$ и $\mathfrak{b}$. Тогда угол θ между этими плоскостями мы будем называть *углом* между бивекторами $\mathfrak{a}$ и $\mathfrak{b}$.

Очевидно, что определение 1 корректно (не зависит от выбора плоскостей Π_1 и Π_2).

Теперь мы можем определить *скалярное произведение* $\mathfrak{a}\mathfrak{b}$ бивекторов $\mathfrak{a}$ и $\mathfrak{b}$ той же формулой, что и для векторов:

$$\mathfrak{a}\mathfrak{b} = |\mathfrak{a}||\mathfrak{b}|\cos\theta.$$

Можно без особого труда доказать, что это скалярное произведение обладает всеми алгебраическими свойствами скалярного произведения векторов, перечисленными в теореме п. 2 § 5 (не совсем тривиально лишь доказательство свойства дистрибутивности). После этого весь материал § 5 (по крайней мере с формальной стороны) полностью может быть повторен для бивекторов. В частности, ясно, что понимать под *ортонормированным бивекторным базисом.* Все формулы, связанные с ортонормированными базисами, конечно, также остаются справедливыми. Например, квадрат площади бивектора $\mathfrak{a}$ (совпадаю-

щий с его скалярным квадратом) равен сумме квадратов его координат a_1, a_2, a_3 в произвольном ортонормированном базисе:

$$|\mathfrak{a}|^2 = a_1^2 + a_2^2 + a_3^2.$$

Впрочем, все это можно установить значительно проще, если заметить, что построенный в п. 2 изоморфизм $\mathfrak{a} \mapsto \mathfrak{a}^\perp$ сохраняет скалярное произведение (является изоморфизмом и по отношению к скалярному умножению), т. е. что

для любых бивекторов $\mathfrak{a}$ *и* $\mathfrak{b}$ *имеет место равенство*

$$\mathfrak{a}\mathfrak{b} = \mathfrak{a}^\perp \mathfrak{b}^\perp.$$

Замечание 1. На первый взгляд это дискредитирует всю теорию бивекторов: зачем ее строить, если она изоморфна теории векторов? На самом же деле все обстоит как раз наоборот: установление изоморфизма двух математических теорий не устраняет одну из них, а лишь ликвидирует неприятную необходимость доказывать ее результаты. При этом формально одинаковые выражения могут иметь совершенно различное содержание.

Рассмотрим, например, формулу для скалярного квадрата в ортонормированном базисе. Для векторов — это формула длины, а для бивекторов — это формула площади. Очевидно, что для любого ортонормированного векторного базиса $\boldsymbol{i}_1$, $\boldsymbol{i}_2$, $\boldsymbol{i}_3$ соответствующий бивекторный базис также ортонормирован. Отсюда и из формулы для скалярного квадрата бивектора мы немедленно получаем (вспомнив формулу для координат бивектора $\boldsymbol{a} \wedge \boldsymbol{b}$ в базисе $\boldsymbol{e}_2 \wedge \boldsymbol{e}_3$, $\boldsymbol{e}_3 \wedge \boldsymbol{e}_1$, $\boldsymbol{e}_1 \wedge \boldsymbol{e}_2$), что

квадрат площади s *параллелограмма, построенного на векторах* $\boldsymbol{a}(a_1, a_2, a_3)$ *и* $\boldsymbol{b}(b_1, b_2, b_3)$, *выражается формулой*

$$s^2 = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}^2.$$

Аналогичное утверждение для плоскости гласит, что

площадь s *параллелограмма, построенного на векторах* $\boldsymbol{a}(a_1, a_2)$ *и* $\boldsymbol{b}(b_1, b_2)$, *равна абсолютной величине определителя*

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}.$$

Хотя в этих утверждениях бивекторы явно не упоминаются, по существу любые их доказательства так или иначе должны использовать бивекторы (быть может, в завуалированном виде). При этом каждое такое «искусственное» доказательство волей-неволей будет в достаточной мере длинным и утомительным, тогда как в рамках теории бивекторов это утверждение получается «бесплатно», как изоморфный вариант уже известного факта о векторах.

Однако метрическая теория бивекторов имеет и свои специфические особенности, не имеющие аналогов в теории векторов. Мы имеем в виду, например, формулы, выражающие скалярное произведение бивекторов $a \wedge b$ и $c \wedge d$ через векторы a, b, c, d. Прежде чем выводить эти формулы, мы рассмотрим некоторые необходимые для этого понятия.

Определение 2. Пусть (a, b) и (c, d) — две пары векторов. Их *взаимным определителем Грама* называется определитель

$$\begin{vmatrix} ac & ad \\ bc & bd \end{vmatrix} = (ac)(bd) - (ad)(bc). \qquad (1)$$

Соответствующую матрицу

$$\begin{pmatrix} ac & ad \\ bc & bd \end{pmatrix}$$

полезно представлять себе как «скалярное» произведение

$$\begin{pmatrix} a \\ b \end{pmatrix}(c, d)$$

одностолбцовой матрицы $\begin{pmatrix} a \\ b \end{pmatrix}$ и однострочечной матрицы (c, d).

В случае, когда пары (a, b) и (c, d) совпадают, соответствующий определитель

$$\begin{vmatrix} a^2 & ab \\ ba & b^2 \end{vmatrix} = a^2b^2 - (ab)^2$$

обозначается символом $\Gamma(a, b)$ и называется *определителем Грама* векторов a, b.

В случае, когда векторы a и b линейно независимы, т. е. составляют базис соответствующей плоскости, элементами определителя Грама $\Gamma(a, b)$ являются метрические коэффициенты этого базиса.

Основное свойство определителя Грама выражается следующим предложением:

Предложение 1. *Определитель Грама* (1) *зависит только от бивекторов* $a \wedge b$ *и* $c \wedge d$.

Доказательство. Достаточно заметить, что при элементарных преобразованиях, скажем, пары (a, b) определитель (1) не меняется (при элементарном преобразовании $(a, b) \sim \sim \left(ka, \frac{1}{k} b\right)$ его первая строка умножается на k, а вторая на $\frac{1}{k}$, так что весь определитель умножается на $k \cdot \frac{1}{k} = 1$, а при элементарном преобразовании $(a, b) \sim (a + kb, b)$ к его первой строке прибавляется вторая, умноженная на k, отчего, как известно, определитель не меняется). Кроме того, если

пара $(\boldsymbol{a}, \boldsymbol{b})$ линейно зависима, то определитель (1), очевидно, равен нулю.

Таким образом, определитель Грама (1) является некоторой функцией бивекторов $\boldsymbol{a} \wedge \boldsymbol{b}$ и $\boldsymbol{c} \wedge \boldsymbol{d}$. Оказывается, что он совпадает с их скалярным произведением, так что

для любых векторов $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ *и* $\boldsymbol{d}$ *имеет место формула*

$$(\boldsymbol{a} \wedge \boldsymbol{b})(\boldsymbol{c} \wedge \boldsymbol{d}) = \begin{vmatrix} \boldsymbol{ac} & \boldsymbol{ad} \\ \boldsymbol{bc} & \boldsymbol{bd} \end{vmatrix}. \tag{2}$$

Поскольку обе части формулы (2) зависят только от бивекторов $\boldsymbol{a} \wedge \boldsymbol{b}$ и $\boldsymbol{c} \wedge \boldsymbol{d}$, мы можем в доказательстве этой формулы, без ограничения общности, предполагать, что векторы $\boldsymbol{c}$ и $\boldsymbol{d}$ ортогональны. Кроме того, поскольку при умножении одного из векторов на произвольное число формула (2) также остается справедливой, нам достаточно ее доказать лишь для случая, когда векторы $\boldsymbol{c}$ и $\boldsymbol{d}$ единичны (и ортогональны). Обозначив эти векторы символами $\boldsymbol{i}$ и $\boldsymbol{j}$, дополним их третьим вектором $\boldsymbol{k}$ до ортонормированного базиса пространства. Пусть a_1, a_2, a_3 и b_1, b_2, b_3 — координаты векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ в этом базисе. Тогда, как мы знаем, бивектор $\boldsymbol{a} \wedge \boldsymbol{b}$ в базисе $\boldsymbol{j} \wedge \boldsymbol{k}$, $\boldsymbol{k} \wedge \boldsymbol{i}$, $\boldsymbol{i} \wedge \boldsymbol{j}$ будет иметь координаты

$$\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}, \quad -\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}, \quad \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix},$$

а бивектор $\boldsymbol{c} \wedge \boldsymbol{d} = \boldsymbol{i} \wedge \boldsymbol{j}$ — координаты

$$0, \quad 0, \quad 1.$$

Следовательно, их скалярное произведение будет равно определителю

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \tag{3}$$

(напомним, что бивекторный базис $\boldsymbol{j} \wedge \boldsymbol{k}$, $\boldsymbol{k} \wedge \boldsymbol{i}$, $\boldsymbol{i} \wedge \boldsymbol{j}$ ортонормирован). С другой стороны, в нашем случае

$$\boldsymbol{ac} = a_1, \quad \boldsymbol{ad} = a_2,$$
$$\boldsymbol{bc} = b_1, \quad \boldsymbol{bd} = b_2,$$

так что определитель Грама (1) также равен определителю (3). Тем самым формула (2) полностью доказана.

Замечание 2. В случае коллинеарных бивекторов формула (2) сводится к формуле умножения определителей. Для неколлинеарных бивекторов она представляет собой некоторое алгебраическое тождество (для координат векторов $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{d}$ в ортонормированном базисе), являющееся обобщением на прямоугольные матрицы формулы умножения определителей.

В этой форме оно, конечно, немедленно доказывается прямым вычислением. Мы предпочли дать более геометрическое доказательство, чтобы проиллюстрировать на этом простом примере ряд полезных приемов математического доказательства.

В частности, мы видим, что
определитель Грама $\Gamma(\boldsymbol{a}, \boldsymbol{b})$ *пары векторов равен квадрату площади параллелограмма, построенного на этих векторах*:

$$\Gamma(\boldsymbol{a}, \boldsymbol{b}) = s^2.$$

Поскольку $s^2 \geqslant 0$, отсюда вытекает, что
для любых векторов $\boldsymbol{a}$ *и* $\boldsymbol{b}$ *имеет место неравенство*

$$\Gamma(\boldsymbol{a}, \boldsymbol{b}) \geqslant 0,$$

причем $\Gamma(\boldsymbol{a}, \boldsymbol{b}) = 0$ *тогда и только тогда, когда векторы* $\boldsymbol{a}$ *и* $\boldsymbol{b}$ *линейно зависимы.*

Это неравенство совпадает, очевидно, с неравенством Коши — Буняковского (см. п. 3 § 5). Таким образом, мы заново его доказали. Одновременно мы выяснили его геометрический смысл.

5. Тривекторы

По аналогии с векторами и бивекторами вы вводим следующее

Определение 1. Две упорядоченные четверки точек (A, B, C, D) и (A', B', C', D') называются *эквиполлентными*, если либо

1) обе четверки компланарны,

либо

2) обе четверки некомпланарны и

а) определяют одну и ту же ориентацию пространства,

б) объем тетраэдра $ABCD$ равен объему тетраэдра $A'B'C'D'$.

Ясно, что отношение эквиполлентности четверок является отношением эквивалентности. Соответствующие классы эквивалентности называются *тривекторами.* Тривектор, содержащий четверку (A, B, C, D), обозначается символом $\overrightarrow{ABCD}$.

По определению, все компланарные четверки составляют один тривектор. Этот тривектор называется *нулевым* и обозначается символом $\mathfrak{O}$.

Замечание 1. Тривекторы имеет смысл рассматривать только в пространстве, поскольку на плоскости (и тем более на прямой) имеется только один тривектор — нулевой.

Множество всех тривекторов в пространстве мы будем обозначать символом $\mathrm{Triv}(3)$.

Любой упорядоченной тройке векторов $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ мы можем отнести некоторый тривектор $\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c}$, отложив векторы от одной точки:

$$\boldsymbol{a} = \overrightarrow{OA}, \quad \boldsymbol{b} = \overrightarrow{OB}, \quad \boldsymbol{c} = \overrightarrow{OC}$$

и положив

$$\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c} = \overrightarrow{OABC}.$$

Определение 2. Тривектор $\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c}$ называется *внешним произведением* векторов $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$.

Легко видеть, что внешнее произведение трех векторов определено корректно, т. е. не зависит от выбора точки O.

Ясно, что

векторы $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ тогда и только тогда линейно зависимы (компланарны), когда их внешнее произведение равно нулю:

$$\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c} = \mathfrak{O}.$$

Подобно тому, как бивекторы можно отождествить с классами эквиполлентных пар векторов, тривекторы можно считать не классами упорядоченных четверок точек, а классами упорядоченных троек векторов. При этом две тройки $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ и $(\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{c}')$ тогда и только тогда определяют один тривектор (*эквиполлентны*), когда либо

1) они обе компланарны,

либо

2) они обе некомпланарны и

а) базисы $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ и $\boldsymbol{a}'$, $\boldsymbol{b}'$, $\boldsymbol{c}'$ одноименны,

б) объем параллелепипеда, построенного на векторах $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, равен объему параллелепипеда, построенного на векторах $\boldsymbol{a}'$, $\boldsymbol{b}'$, $\boldsymbol{c}'$.

Об ориентации, определенной базисом $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, говорят, что она *определяется* тривектором $\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c}$, а объем параллелепипеда, построенного на векторах $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, называется *объемом* (или *мерой*) тривектора $\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c}$. Ввиду условий а) и б) эти определения корректны. Объем тривектора $\mathfrak{A}$ обозначается символом $|\mathfrak{A}|$.

Объем нулевого тривектора, по определению, считается равным нулю. Никакой ориентации нулевой тривектор не определяет.

Ясно, что

тривектор $\mathfrak{A}$ однозначно определен, если известны его объем $|\mathfrak{A}|$ и (при $|\mathfrak{A}| \neq 0$) определяемая им ориентация.

Упражнение. Дайте «аффинное» определение равенства тривекторов, вводя для этого соответствующие элементарные преобразования (ср. § 6, п. 1, определение 4 и замечание 3).

Определение 3. *Произведением $k\mathfrak{A}$* тривектора $\mathfrak{A}$ на число k называется тривектор, имеющий объем $|k| \cdot |\mathfrak{A}|$ и (при $|k| \cdot |\mathfrak{A}| \neq 0$) определяющий ту же ориентацию, что и тривектор

$\mathfrak{A}$, когда $k > 0$, и противоположную ориентацию, когда $k < 0$. Тривектор $(-1)\,\mathfrak{A}$ обозначается символом $-\mathfrak{A}$ и называется *противоположным* тривектором (к тривектору $\mathfrak{A}$). Он имеет тот же объем, что и тривектор $\mathfrak{A}$; но (при $|\mathfrak{A}| \neq 0$) определяет противоположную ориентацию.

Ясно, что

для любых двух тривекторов $\mathfrak{A}$ *и* $\mathfrak{B}$ *существует такое число* k, *что либо* $\mathfrak{A} = k\mathfrak{B}$, *либо* $\mathfrak{B} = k\mathfrak{A}$.

Другими словами, можно сказать, что любые два тривектора «коллинеарны».

Без труда проверяется, что

для любых трех векторов $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ *и любого числа* k *имеют место равенства*

$$(k\boldsymbol{a}) \wedge \boldsymbol{b} \wedge \boldsymbol{c} = \boldsymbol{a} \wedge (k\boldsymbol{b}) \wedge \boldsymbol{c} = \boldsymbol{a} \wedge \boldsymbol{b} \wedge (k\boldsymbol{c}) = k(\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c})$$

(свойство однородности внешнего умножения).

Так как при транспозиции векторов базиса его ориентация меняется на противоположную, а объем построенного на них параллелепипеда остается прежним, то

для любых трех векторов $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ *имеют место равенства*

$$\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c} = -\boldsymbol{b} \wedge \boldsymbol{a} \wedge \boldsymbol{c} = -\boldsymbol{a} \wedge \boldsymbol{c} \wedge \boldsymbol{b} = -\boldsymbol{c} \wedge \boldsymbol{b} \wedge \boldsymbol{a} = \\ = \boldsymbol{b} \wedge \boldsymbol{c} \wedge \boldsymbol{a} = \boldsymbol{c} \wedge \boldsymbol{a} \wedge \boldsymbol{b}$$

(свойство антикоммутативности внешнего умножения)

Определение 4. Пусть $\mathfrak{a} = \boldsymbol{a} \wedge \boldsymbol{b}$ и $\boldsymbol{c}$ — произвольные бивектор и вектор. Их *внешним произведением* $\mathfrak{a} \wedge \boldsymbol{c}$ мы будем называть тривектор

$$\mathfrak{a} \wedge \boldsymbol{c} = \boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c}.$$

Кроме того, по определению, мы положим

$$\boldsymbol{c} \wedge \mathfrak{a} = \mathfrak{a} \wedge \boldsymbol{c}.$$

Необходимо, конечно, проверить корректность определения тривектора $\mathfrak{a} \wedge \boldsymbol{c}$, т. е. его независимость от выбора векторов $\boldsymbol{a}$ и $\boldsymbol{b}$. С этой целью мы заметим, что тривектор $\mathfrak{a} \wedge \boldsymbol{c}$ обладает, очевидно, следующими свойствами:

1) его объем выражается формулой

$$|\mathfrak{a} \wedge \boldsymbol{c}| = |\mathfrak{a}| \cdot h,$$

где $|\mathfrak{a}|$ — площадь бивектора $\mathfrak{a}$, а h — длина проекции вектора $\boldsymbol{c}$ параллельно плоскости бивектора $\mathfrak{a}$ на прямую, перпендикулярную этой плоскости;

2) определяемая им ориентация (при $|\mathfrak{a} \wedge \boldsymbol{c}| \neq 0$) является произведением ориентации, определяемой бивектором $\mathfrak{a}$, и ориентации, определяемой вектором $\boldsymbol{c}$.

Поскольку эти свойства однозначно характеризуют тривектор $\mathfrak{a} \wedge \boldsymbol{c}$ и поскольку они не зависят от выбора векторов $\boldsymbol{a}$ и $\boldsymbol{b}$, определение тривектора $\mathfrak{a} \wedge \boldsymbol{c}$ корректно.

Из этого определения непосредственно вытекает следующее предложение, аналогичное предложению 1 п. 1:

Предложение 1. *Равенство* $\mathfrak{a} \wedge \boldsymbol{c} = \mathfrak{a} \wedge \boldsymbol{c}'$ *(при* $\mathfrak{a} \neq 0$*) имеет место тогда и только тогда, когда вектор* $\boldsymbol{c}' - \boldsymbol{c}$ *параллелен плоскости бивектора* $\mathfrak{a}$.

Кроме того, ясно, что аналог леммы 1 из п. 2 также справедлив, т. е.

для любого тривектора $\mathfrak{A}$ *и любого отличного от нуля бивектора* $\mathfrak{a}$ *существует такой вектор* $\boldsymbol{c}$, *что*

$$\mathfrak{A} = \mathfrak{a} \wedge \boldsymbol{c}.$$

Обратим внимание, что для операции внешнего умножения бивектора на вектор свойство однородности также, очевидно, выполнено, т. е. для любого числа k имеют место равенства

$$(k\mathfrak{a}) \wedge \boldsymbol{c} = \mathfrak{a} \wedge (k\boldsymbol{c}) = k(\mathfrak{a} \wedge \boldsymbol{c}).$$

Формулу, определяющую тривектор $\mathfrak{a} \wedge \boldsymbol{c}$, запишем в виде:

$$(\boldsymbol{a} \wedge \boldsymbol{b}) \wedge \boldsymbol{c} = \boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c}.$$

Кроме того, поскольку

$$\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c} = \boldsymbol{b} \wedge \boldsymbol{c} \wedge \boldsymbol{a}$$

и

$$\boldsymbol{a} \wedge (\boldsymbol{b} \wedge \boldsymbol{c}) = (\boldsymbol{b} \wedge \boldsymbol{c}) \wedge \boldsymbol{a} = \boldsymbol{b} \wedge \boldsymbol{c} \wedge \boldsymbol{a},$$

то

$$\boldsymbol{a} \wedge (\boldsymbol{b} \wedge \boldsymbol{c}) = \boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c}.$$

Таким образом,

для любых трех векторов $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ *имеют место равенства*

$$(\boldsymbol{a} \wedge \boldsymbol{b}) \wedge \boldsymbol{c} = \boldsymbol{a} \wedge (\boldsymbol{b} \wedge \boldsymbol{c}) = \boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c}$$

(свойство ассоциативности внешнего умножения).

Заметим, что в последней формуле знак $\wedge$ употребляется в трех различных смыслах.

Определение 5. Пусть $\mathfrak{A}$ и $\mathfrak{B}$ — произвольные тривекторы. Выбрав некоторый отличный от нуля бивектор $\mathfrak{e}$, положим

$$\mathfrak{A} = \mathfrak{e} \wedge \boldsymbol{a}, \quad \mathfrak{B} = \mathfrak{e} \wedge \boldsymbol{b},$$

где $\boldsymbol{a}$ и $\boldsymbol{b}$ — некоторые векторы. Мы определим *сумму* $\mathfrak{A} + \mathfrak{B}$ тривекторов $\mathfrak{A}$ и $\mathfrak{B}$ формулой

$$\mathfrak{A} + \mathfrak{B} = \mathfrak{e} \wedge (\boldsymbol{a} + \boldsymbol{b})$$

(ср. аналогичное определение для бивекторов в п. 2).

Чтобы доказать корректность этого определения, мы вспомним, что существует такое число k, что, скажем $\mathfrak{B} = k\mathfrak{A}$

(причем за исключением случая, когда $\mathfrak{A} = \mathfrak{O}$ и $\mathfrak{B} = \mathfrak{O}$, это число однозначно определено). Тогда

$$\mathfrak{e} \wedge \boldsymbol{b} = k(\mathfrak{e} \wedge \boldsymbol{a}) = \mathfrak{e} \wedge (k\boldsymbol{a})$$

и потому вектор $k\boldsymbol{a} - \boldsymbol{b} = (k+1)\boldsymbol{a} - (\boldsymbol{a} + \boldsymbol{b})$ параллелен плоскости бивектора $\mathfrak{e}$. Следовательно,

$$\mathfrak{e} \wedge (\boldsymbol{a} + \boldsymbol{b}) = \mathfrak{e} \wedge (k+1)\boldsymbol{a}.$$

Заметив, наконец, что

$$\mathfrak{e} \wedge (k+1)\boldsymbol{a} = (k+1)(\mathfrak{e} \wedge \boldsymbol{a}) = (k+1)\mathfrak{A},$$

мы окончательно получим, что при $\mathfrak{B} = k\mathfrak{A}$

$$\mathfrak{A} + \mathfrak{B} = (k+1)\mathfrak{A}. \tag{1}$$

Следовательно, определение тривектора $\mathfrak{A} + \mathfrak{B}$ корректно.

Очевидным образом проверяется (с использованием формулы (1)), что

построенные линейные операции над тривекторами обладают всеми восемью стандартными свойствами линейных операций над векторами, т. е. что множество Triv(3) *всех тривекторов является линеалом.*

Поэтому, в частности, для тривекторов сохраняется вся теория линейной зависимости, построенная в п. 3 § 3. Впрочем, как уже отмечалось, эта теория для тривекторов вырождается в тривиальность, поскольку любые два тривектора линейно зависимы.

Произвольный отличный от нуля тривектор $\mathfrak{E}$ называется (тривекторным) *базисом*. Ясно, что любой другой тривектор $\mathfrak{A}$ единственным образом представляется в виде

$$\mathfrak{A} = a\mathfrak{E}.$$

Число a называется, естественно, *координатой* тривектора $\mathfrak{A}$ в базисе $\mathfrak{E}$. Ясно, что соответствие $\mathfrak{A} \longmapsto a$ определяет изоморфизм линеала Triv(3) на стандартный линеал $\mathbb{R} = \mathbb{R}^1$.

Заметим, что по определению внешнее произведение $\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c}$ обладает свойством дистрибутивности по отношению к последнему множителю:

$$\boldsymbol{a} \wedge \boldsymbol{b} \wedge (\boldsymbol{c}_1 + \boldsymbol{c}_2) = \boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c}_1 + \boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c}_2.$$

Однако, поскольку это произведние антикоммутативно, оно будет обладать этим свойством и по отношению к первым двум множителям:

$$(\boldsymbol{a}_1 + \boldsymbol{a}_2) \wedge \boldsymbol{b} \wedge \boldsymbol{c} = \boldsymbol{a}_1 \wedge \boldsymbol{b} \wedge \boldsymbol{c} + \boldsymbol{a}_2 \wedge \boldsymbol{b} \wedge \boldsymbol{c},$$
$$\boldsymbol{a} \wedge (\boldsymbol{b}_1 + \boldsymbol{b}_2) \wedge \boldsymbol{c} = \boldsymbol{a} \wedge \boldsymbol{b}_1 \wedge \boldsymbol{c} + \boldsymbol{a} \wedge \boldsymbol{b}_2 \wedge \boldsymbol{c}.$$

Отсюда, в частности, следует, что

$$(\mathfrak{a}_1 + \mathfrak{a}_2) \wedge \boldsymbol{c} = \mathfrak{a}_1 \wedge \boldsymbol{c} + \mathfrak{a}_2 \wedge \boldsymbol{c}.$$

Для любого векторного базиса e_1, e_2, e_3 тривектор $e_1 \wedge e_2 \wedge e_3$ отличен от нуля, т. е. является тривекторным базисом. Очевидное вычисление (использующее свойства однородности, дистрибутивности и антикоммутативности внешнего умножения) показывает, что

для любых трех векторов $a = a^1e_1 + a^2e_2 + a^3e_3$, $b = b^1e_1 + b^2e_2 + b^3e_3$ *и* $c = c^1e_1 + c^2e_2 + c^3e_3$ *имеет место равенство*

$$a \wedge b \wedge c = \begin{vmatrix} a^1 & a^2 & a^3 \\ b^1 & b^2 & b^3 \\ c^1 & c^2 & c^3 \end{vmatrix} (e_1 \wedge e_2 \wedge e_3).$$

Из этой связи между тривекторами и определителями можно без труда вывести все основные свойства определителей (только свойство равноправности строк и столбцов получить на этом пути трудно).

Если базис e_1, e_2, e_3 ортонормирован, то объем тривектора $e_1 \wedge e_2 \wedge e_3$ равен единице. Отсюда вытекает, что

объем параллелепипеда, построенного на векторах $a(a_1, a_2, a_3)$, $b(b_1, b_2, b_3)$ *и* $c(c_1, c_2, c_3)$, *равен абсолютной величине определителя*

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}. \qquad (2)$$

Заметим, что в этом утверждении понятие тривектора в явном виде не фигурирует.

Определитель (2) (с точностью до транспонировании) представляет собой определитель матрицы перехода от ортонормированного базиса e_1, e_2, e_3 к базису a, b, c. Таким образом, справедливо следующее

Предложение 2. *Объем параллелепипеда, построенного на векторах некоторого базиса, равен абсолютной величине определителя* $|C|$ *матрицы перехода к данному базису от некоторого ортонормированного базиса.*

Определение 6. Пусть (a, b, c) и (a_1, b_1, c_1) — произвольные упорядоченные тройки векторов. Их *взаимным определителем Грама* называется определитель вида

$$\begin{vmatrix} aa_1 & ab_1 & ac_1 \\ ba_1 & bb_1 & bc_1 \\ ca_1 & cb_1 & cc_1 \end{vmatrix}.$$

Матрицу

$$\begin{pmatrix} aa_1 & ab_1 & ac_1 \\ ba_1 & bb_1 & bc_1 \\ ca_1 & cb_1 & cc_1 \end{pmatrix}$$

этого определителя полезно представлять себе в виде «скалярного» произведения

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} (a_1, b_1, c_1)$$

одностолбцовой матрицы $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ на однострочечную матрицу (a_1, b_1, c_1).

В случае, когда тройки (a, b, c) и (a_1, b_1, c_1) совпадают, соответствующий определитель

$$\begin{vmatrix} a^2 & ab & ac \\ ba & b^2 & bc \\ ca & cb & c^2 \end{vmatrix}$$

называется *определителем Грама* тройки (a, b, c) и обозначается символом $\Gamma(a, b, c)$. Его матрица для линейно независимой тройки (базиса) a, b, c является не чем иным, как матрицей метрических коэффициентов этого базиса.

Определение 7. Скалярным произведением $\mathfrak{AB}$ двух тривекторов называется произведение их длин, взятое со знаком плюс, если эти тривекторы определяют одну и ту же ориентацию пространства, и со знаком минус — в противном случае.

Точно так же, как для бивекторов, доказывается, что

скалярное произведение тривекторов $a \wedge b \wedge c$ *и* $a_1 \wedge b_1 \wedge c_1$ *равно взаимному определителю Грама троек* (a, b, c) *и* (a_1, b_1, c_1):

$$(a \wedge b \wedge c)(a_1 \wedge b_1 \wedge c_1) = \begin{vmatrix} aa_1 & ab_1 & ac_1 \\ ba_1 & bb_1 & bc_1 \\ ca_1 & cb_1 & cc_1 \end{vmatrix}.$$

Впрочем, быть может, это проще доказать, заметив, что скалярное произведение тривекторов, выраженное через координаты в ортонормированном базисе, сводится к умножению определителей.

Применив это предложение к случаю одинаковых тривекторов, мы немедленно получим, что

для любых трех векторов a, b, c *квадрат* v^2 *объема параллелепипеда, построенного на этих векторах, равен их определителю Грама:*

$$v^2 = \Gamma(a, b, c).$$

Отсюда вытекает, что

для любых трех векторов a, b *и* c *имеет место неравенство*

$$\Gamma(a, b, c) \geqslant 0,$$

вырождающееся в равенство тогда и только тогда, когда векторы $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ *линейно зависимы.*

В частности, мы видим, что

для любого базиса пространства определитель матрицы его метрических коэффициентов положителен:

$$\begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} > 0.$$

Об этом факте мы уже имели случай упомянуть в п. 4 § 5.

6. Векторное и смешанное произведения

Воспользовавшись изоморфизмом между бивекторами и векторами, мы при желании можем вообще исключить какое-либо упоминание о бивекторах, заменяя их всюду ассоциированными векторами. Например, вместо того чтобы говорить о бивекторе $\boldsymbol{a} \wedge \boldsymbol{b}$, мы можем говорить о векторе $(\boldsymbol{a} \wedge \boldsymbol{b})^{\perp}$.

Определение 1. Вектор $(\boldsymbol{a} \wedge \boldsymbol{b})^{\perp}$ называется *векторным произведением* векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ и обозначается символом $\boldsymbol{a} \times \boldsymbol{b}$.

Таким образом, векторное произведение линейно зависимых (коллинеарных) векторов равно нулю; если векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ не коллинеарны, векторное произведение $\boldsymbol{a} \times \boldsymbol{b}$ отлично от нуля и определяется следующими условиями:

1) его длина равна площади параллелограмма, построенного на векторах $\boldsymbol{a}$ и $\boldsymbol{b}$;

2) он перпендикулярен плоскости векторов $\boldsymbol{a}$ и $\boldsymbol{b}$;

3) базис $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{a} \times \boldsymbol{b}$ положительно ориентирован (по отношению к данной ориентации пространства).

Свойства векторного умножения немедленно вытекают из соответствующих свойств внешнего умножения. В частности,

векторное умножение антикоммутативно и дистрибутивно относительно сложения.

Для любого положительно ориентированного ортонормированного базиса $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ имеют место, очевидно, формулы

$$\boldsymbol{i} \times \boldsymbol{j} = \boldsymbol{k}, \quad \boldsymbol{j} \times \boldsymbol{k} = \boldsymbol{i}, \quad \boldsymbol{k} \times \boldsymbol{i} = \boldsymbol{j}.$$

Поэтому

в положительно ориентированном ортонормированном базисе вектор $\boldsymbol{a} \times \boldsymbol{b}$ *имеет координаты*

$$\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}, \quad -\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}, \quad \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix},$$

где a_1, a_2, a_3 *— координаты вектора* $\boldsymbol{a}$, *а* b_1, b_2, b_3 *— координаты вектора* $\boldsymbol{b}$.

Воспользовавшись формулой разложения определителя по элементам строки, мы можем этот факт записать в виде следующей компактной формулы:

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}.$$

Полезно также иметь в виду, что

квадрат $|\boldsymbol{a} \times \boldsymbol{b}|^2$ *длины* $|\boldsymbol{a} \times \boldsymbol{b}|$ *векторного произведения* $\boldsymbol{a} \times \boldsymbol{b}$ *равен определителю Грама* $\Gamma(\boldsymbol{a}, \boldsymbol{b})$ *векторов* $\boldsymbol{a}$, $\boldsymbol{b}$.

Векторное произведение имеет то преимущество, что оно оставляет нас в рамках чистой теории векторов. Однако его смысл и значение могут быть полностью выяснены лишь с «бивекторных» позиций. (В частности, прямое доказательство его свойств по существу использует бивекторы, хотя и в неявном виде.) Кроме того, его определение имеет тот неустранимый недостаток, что оно содержит элемент произвола (выбор ориентации пространства). Мы не говорим уже о его сугубо метрическом характере, скрадывающем аффинный характер понятия бивектора [1]).

Аналогично, воспользовавшись изоморфизмом между линеалами $\mathrm{Triv}(3)$ и $\mathbb{R}$, мы можем всюду заменить тривекторы соответствующими числами (координатами). При этом, конечно, возникает элемент произвола, связанный с выбором тривекторного базиса $\mathfrak{E}$, определяющего данный изоморфизм, т. е. с выбором эталона объема и, что более существенно, ориентации пространства.

Определение 2. Координата тривектора $\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c}$ называется *смешанным* (или *тройным*) *произведением* векторов $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ и обозначается символом $\boldsymbol{abc}$.

Смешанное произведение является таким же заменителем внешнего произведения $\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{c}$, каким векторное произведение $\boldsymbol{a} \times \boldsymbol{b}$ является по отношению к внешнему произведению $\boldsymbol{a} \wedge \boldsymbol{b}$. Его свойства немедленно вытекают из соответствующих свойств внешнего произведения. В частности,

смешанное произведение антикоммутативно и дистрибутивно по каждому сомножителю.

В положительно ориентированном ортонормированном базисе смешанное произведение выражается формулой

$$\boldsymbol{abc} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

[1]) Не последнюю роль играет и тот факт, что в n-мерных пространствах при $n > 3$ векторное произведение двух векторов вообще нельзя определить, тогда как теория бивекторов полностью сохраняется.

Впрочем, для справедливости этой формулы ортонормированность базиса является, очевидно, слишком сильным требованием: достаточно, чтобы объем параллелепипеда, построенного на векторах базиса, был бы равен единице.

Упражнение. Докажите, что смешанное произведение следующим образом выражается через векторное и скалярное произведения:

$$abc = a\,(b \times c).$$

(Этим и объясняется термин «смешанное произведение».)

Подчеркнем, что смешанное произведение зависит от выбора ориентации пространства: при смене ориентации оно меняет знак.

§ 7. ЛИНЕЙНЫЕ ОПЕРАТОРЫ

1. Отображения и преобразования

В этом вводном пункте мы напоминаем некоторые общематематические факты, связанные с понятием отображения. Эти факты читателю почти наверняка известны, но мы все же их вкратце изложим, хотя бы для того, чтобы уточнить терминологию.

Определение 1. Пусть X и Y — произвольные непустые множества. Отображением $\varphi\colon X \to Y$ множества X в множество Y называется [1]) произвольное правило, позволяющее каждому элементу $x \in X$ сопоставить некоторый однозначно определенный элемент $y \in Y$. Элемент y называется *образом* элемента x при отображении φ и обозначается символом $\varphi(x)$.

Для отображений $\varphi\colon X \to Y$ и $\psi\colon Y \to Z$ определена их *композиция*

$$\psi \circ \varphi\colon X \to Z$$

(иногда называемая также их *произведением*), задаваемая формулой

$$(\psi \circ \varphi)(x) = \psi(\varphi(x)), \qquad x \in X.$$

Без труда проверяется, что композиция отображений *ассоциативна*, т. е. для любых трех отображений $\varphi\colon X \to Y$, $\psi\colon Y \to Z$ и $\chi\colon Z \to U$ отображения $(\chi \circ \psi) \circ \varphi$ и $\chi \circ (\psi \circ \varphi)$ совпадают:

$$(\chi \circ \psi) \circ \varphi = \chi \circ (\psi \circ \varphi).$$

Поэтому обычно скобки опускают и пишут просто $\chi \circ \psi \circ \varphi$.

Определение 2. Множество всех элементов $y \in Y$, для которых существует такой элемент $x \in X$, что $\varphi(x) = y$, называется *образом* отображения φ и обозначается символом $\varphi(X)$. Если $\varphi(X) = Y$, то отображение φ называется *надъективным*.

Если для любых двух различных элементов $x_1, x_2 \in X$ элементы $\varphi(x_1)$, $\varphi(x_2) \in Y$ также различны, отображение φ называется *инъективным*.

---

[1]) Это, конечно, не определение, а скорее описание. Дать строгое определение понятию отображения по существу так же трудно, как и понятию множества.

Отображение φ называется *биективным* (или *взаимно однозначным*), если оно одновременно инъективно и надъективно. В этом случае для любого элемента $y \in Y$ существует единственный элемент $x \in X$, обладающий тем свойством, что

$$\varphi(x) = y.$$

Поэтому формула

$$\varphi^{-1}(y) = x$$

определяет некоторое отображение

$$\varphi^{-1}\colon Y \to X,$$

называемое отображением, *обратным* к отображению $\varphi\colon X \to Y$.

Обратное отображение также биективно, причем отображением, обратным к нему, служит отображение φ:

$$(\varphi^{-1})^{-1} = \varphi.$$

Примером биективного отображения может служить *тождественное отображение* 1_X (или просто 1) множества X на себя, определяемое формулой

$$1_X(x) = x, \qquad x \in X.$$

Определение 3. Отображение $\varphi\colon X \to Y$ называется *обратимым слева* (соответственно *обратимым справа*), если существует такое отображение $\varphi'\colon Y \to X$ (такое отображение $\varphi''\colon Y \to X$), что $\varphi' \circ \varphi = 1_X$ (соответственно $\varphi \circ \varphi'' = 1_Y$). Отображение называется *обратимым*, если оно обратимо слева и справа.

Легко видеть, что

отображение $\varphi\colon X \to Y$ *тогда и только тогда обратимо слева* (*справа*), *когда оно инъективно* (*надъективно*).

Действительно, пусть отображение φ обратимо слева, т. е. пусть существует такое отображение $\varphi'\colon Y \to X$, что

$$\varphi' \circ \varphi = 1_X,$$

и пусть x_1, x_2 — такие элементы множества X, что

$$\varphi(x_1) = \varphi(x_2).$$

Применяя к обеим частям этого равенства отображение φ' и учитывая, что, по определению, для любого элемента $x \in X$

$$(\varphi' \circ \varphi)(x) = 1_X(x) = x,$$

мы немедленно получаем, что

$$x_1 = x_2.$$

Следовательно, отображение φ инъективно.

Обратно, пусть отображение φ инъективно. Тогда для любого элемента $y \in \varphi(X)$ существует единственный элемент x_y, для которого $\varphi(x_y) = y$. Выбрав в X произвольный элемент x_0, мы определим отображение

$$\varphi'\colon Y \to X$$

формулой

$$\varphi'(y) = \begin{cases} x_y, & \text{если } y \in \varphi(X), \\ x_0, & \text{если } y \notin \varphi(X). \end{cases}$$

Тогда $\varphi' \circ \varphi = 1_X$, так что отображение φ обратимо слева.

Пусть теперь отображение φ обратимо справа, т. е. пусть существует такое отображение φ'': $Y \to X$, что

$$\varphi \circ \varphi'' = 1_Y.$$

Тогда для любого элемента $y \in Y$ элемент $x = \varphi''(y) \in X$ будет обладать тем свойством, что $\varphi(x) = y$, и следовательно, отображение φ надъективно.

Наконец, пусть отображение φ надъективно. Тогда для любого элемента $y \in Y$ существует (вообще говоря, не один) элемент $x \in X$, обладающий тем свойством, что $\varphi(x) = y$. Выбрав (произвольно) один такой элемент, обозначим его символом $\varphi''(y)$. Тем самым мы получим некоторое отображение φ'': $Y \to X$, обладающее тем свойством, что $\varphi \circ \varphi'' = 1_Y$.

В частности, мы видим, что

отображение φ: $X \to Y$ *тогда и только тогда обратимо, когда оно биективно.*

При этом для биективного отображения φ отображения φ' и φ'' определяются единственным образом и совпадают с обратным отображением φ^{-1}. Таким образом, биективное отображение φ: $X \to Y$ можно определить как такое отображение, для которого существует отображение φ^{-1}: $Y \to X$, удовлетворяющее соотношениям

$$\varphi^{-1} \circ \varphi = 1_X, \quad \varphi \circ \varphi^{-1} = 1_Y.$$

Легко видеть, что композиция двух надъективных (инъективных) отображений также надъективна (инъективна). Следовательно,

композиция $\psi \circ \varphi$ *двух биективных отображений является биективным отображением.*

При этом

$$(\psi \circ \varphi)^{-1} = \varphi^{-1} \circ \psi^{-1}.$$

Определение 4. Биективные отображения множества X на себя называются его *преобразованиями.*

Непустое множество преобразований называется *группой преобразований*, если

1) вместе с двумя преобразованиями φ и ψ оно содержит и их композицию $\psi \circ \varphi$;

2) вместе с преобразованием φ оно содержит и обратное преобразование φ^{-1}.

Поскольку группа преобразований непуста, из условия 2) вытекает, что она обязательно содержит тождественное преобразование 1_X.

Замечание 1. Группы преобразований являются частным случаем известного из алгебры общего понятия (абстрактной) группы. В отличие от определения абстрактных групп, в определение групп преобразований включать требование ассоциативности не нужно, поскольку для умножения преобразований оно выполнено автоматически.

Замечание 2. В дальнейшем (например, в гл. 7) мы изредка будем пользоваться некоторыми простейшими понятиями абстрактной теории групп. Напомним их определения, отсылая за подробностями к курсу высшей алгебры. Определение группы мы будем считать известным.

Подмножество H группы G называется

а) *подгруппой*, если $h_1h_2 \in H$ и $h^{-1} \in H$ для любых элементов $h_1, h_2, h \in H$;

б) *нормальным делителем*, если H — подгруппа и если $ghg^{-1} \in H$ для любых элементов $h \in H$, $g \in G$;

в) *смежным классом* (левым) по подгруппе K, если $h_1^{-1}h_2 \in K$ для любых элементов $h_1, h_2 \in H$.

Отображение

$$\varphi\colon G \to G'$$

одной группы в другую называется *гомоморфизмом*, если

$$\varphi(g_1g_2) = \varphi(g_1)\varphi(g_2) \quad \text{для любых} \quad g_1, g_2 \in G.$$

Инъективный гомоморфизм называется *мономорфизмом*, надъективный — *эпиморфизмом*, а биективный — *изоморфизмом*.

Пусть e' — единица группы G'. Множество всех элементов $g \in G$, для которых $\varphi(g) = e'$, называется *ядром* гомоморфизма φ. Оно всегда является нормальным делителем. Гомоморфизм тогда и только тогда инъективен (является мономорфизмом), когда его ядро состоит только из единицы e группы G.

Множество смежных классов группы G по подгруппе H обозначается символом G/H. В случае, когда H является нормальным делителем, умножение смежных классов «по представителям» корректно определено и превращает множество G/H в группу. Эта группа называется *факторгруппой* группы G по нормальному делителю H.

Теорема о гомоморфизмах утверждает, что если существует эпиморфизм $\varphi\colon G \to G'$, то группа G' изоморфна факторгруппе группы G по ядру эпиморфизма φ.

2. Кольцо линейных операторов

Определение 1. Отображения линеала $\mathrm{Vect}(n)$ в себя (где, как всегда, $n = 1, 2, 3$) принято называть *операторами*. Оператор $\boldsymbol{A}$, переводящий вектор $\boldsymbol{x}$ в вектор $\boldsymbol{Ax}$, называется *линейным*, если

$$\boldsymbol{A}(\boldsymbol{x} + \boldsymbol{y}) = \boldsymbol{Ax} + \boldsymbol{Ay}$$

для любых векторов $\boldsymbol{x}$, $\boldsymbol{y}$ (это свойство иногда называют свойством аддитивности), и

$$\boldsymbol{A}(k\boldsymbol{x}) = k(\boldsymbol{Ax})$$

для любого вектора $\boldsymbol{x}$ и любого числа k (это свойство называется свойством однородности).

Примером линейного оператора является *тождественный оператор* E, оставляющий каждый вектор на месте:

$$Ex = x,$$

а также *нулевой оператор* O, переводящий каждый вектор в нулевой вектор:

$$Ox = 0.$$

Более общий пример мы получим, рассмотрев для произвольного вещественного числа k оператор kE, действующий по формуле

$$(kE)x = kx,$$

т. е. умножающий каждый вектор на число k. Этот оператор, очевидно, линеен. При $k = 1$ он превращается в тождественный оператор E, а при $k = 0$ — в нулевой оператор O.

Множество всех линейных операторов, действующих в $\mathrm{Vect}(n)$, мы будем обозначать символом $\mathrm{Lin}(n)$.

Композиция операторов обычно называется их *произведением* и в ее обозначении опускается знак $\circ$. Таким образом, если A и B — операторы, то их произведение AB действует на любом векторе x по формуле

$$(AB)x = A(Bx).$$

Тождественный оператор является единицей этого умножения, т. е.

$$AE = EA = A$$

для любого оператора A.

Ясно, что

произведение AB двух линейных операторов A и B также является линейным оператором.

Действительно,

$$AB(x + y) = A(B(x + y)) = A(Bx + By) = A(Bx) + A(By) = \\ = (AB)x + (AB)y$$

и

$$AB(kx) = A(B(kx)) = A(kBx) = kA(Bx) = k(AB)x.$$

В отличие от отображений произвольных множеств для операторов определена также операция сложения.

Определение 2. *Суммой* $A + B$ двух операторов A и B называется оператор, определенный формулой

$$(A + B)x = Ax + Bx.$$

Легко видеть, что

сумма $A + B$ двух линейных операторов A и B также является линейным оператором.

Действительно,

$$(A+B)(x+y)=A(x+y)+B(x+y)=Ax+Ay+Bx+By=$$
$$=Ax+Bx+Ay+By=(A+B)x+(A+B)y$$

и

$$(A+B)(kx)=A(kx)+B(kx)=kAx+kBx=$$
$$=k(Ax+Bx)=k(A+B)x.$$

Эта операция сложения, очевидно, *ассоциативна* $((A+B)+C=A+(B+C))$, *коммутативна* $(A+B=B+A)$, *обладает нулем* (которым служит нулевой оператор O) и для любого оператора A существует противоположный оператор $-A$ (определяемый формулой $(-A)x=-Ax$). Кроме того, легко видеть, что сложение операторов *дистрибутивно* относительно умножения $((A+B)C=AC+BC$ и $C(A+B)=CA+CB$ для любых операторов A, B, C). Поскольку умножение операторов ассоциативно $((AB)C=A(BC))$, все это на языке алгебры, означает что справедливо следующее

Предложение 1. *Совокупность* $\mathrm{Lin}(n)$ *всех линейных операторов в* $\mathrm{Vect}(n)$ *является кольцом с единицей.*

Подчеркнем, что это кольцо *некоммутативно* (т. е., вообще говоря, $AB\neq BA$).

3. Описание линейных операторов

Пусть

$$e_1, \ldots, e_n$$

— произвольный базис линеала $\mathrm{Vect}(n)$. Линейный оператор A переводит векторы этого базиса в некоторые векторы

$$e'_1=Ae_1, \ldots, e'_n=Ae_n.$$

Легко видеть, что

векторы $e'_1, \ldots, e'_n$ *однозначно определяют линейный оператор* A.

Действительно, любой вектор x однозначно разлагается по векторам базиса:

$$x=x^1e_1+\ldots+x^ne_n.$$

Применив к этому равенству оператор A и воспользовавшись линейностью, мы немедленно получим, что

$$Ax=x^1e'_1+\ldots+x^ne'_n.$$

Таким образом, зная векторы $e'_1, \ldots, e'_n$, мы можем вычислить вектор Ax для любого вектора x. Но это и означает, что векторы $e'_1, \ldots, e'_n$ однозначно определяют оператор A.

Столь же легко показывается, что
для любой системы векторов

$$e'_1, \ldots, e'_n$$

существует (*по доказанному, единственный*) *линейный оператор* A, *для которого*

$$e'_1 = Ae_1, \ \ldots, \ e'_n = Ae_n.$$

Действительно, рассмотрим оператор A, определенный для любого вектора

$$x = x^1 e_1 + \ \ldots \ + x^n e_n$$

формулой

$$Ax = x^1 e'_1 + \ \ldots \ + x^n e'_n.$$

В силу единственности разложения векторов по векторам базиса это определение корректно. Кроме того, ясно, что так построенный оператор обладает тем свойством, что

$$Ae_1 = e'_1, \ \ldots, \ Ae_n = e'_n.$$

Поэтому для завершения доказательства нужно только показать, что построенный оператор A линеен, т. е. для него справедливы соотношения

$$A(x + y) = Ax + Ay,$$
$$A(kx) = k(Ax).$$

Но проверка этих соотношений производится совершенно автоматически: если

$$x = x^1 e_1 + \ \ldots \ + x^n e_n, \quad y = y^1 e_1 + \ \ldots \ + y^n e_n,$$

то

$$x + y = (x^1 + y^1) e_1 + \ \ldots \ + (x^n + y^n) e_n,$$

и потому

$$A(x + y) = (x^1 + y^1) e'_1 + \ \ldots \ + (x^n + y^n) e'_n =$$
$$= (x^1 e'_1 + \ \ldots \ + x^n e'_n) + (y^1 e'_1 + \ \ldots \ + y^n e'_n) = Ax + Ay.$$

Аналогично, так как

$$kx = (kx^1) e_1 + \ \ldots \ + (kx^n) e_n,$$

то

$$A(kx) = (kx^1) e'_1 + \ \ldots \ + (kx^n) e'_n = k(x^1 e'_1 + \ \ldots \ + x^n e'_n) = k(Ax).$$

Резюмируя, мы получаем следующее
Предложение 1. *Выбор базиса*

$$e_1, \ \ldots, \ e_n$$

устанавливает биективное соответствие между линейными операторами и произвольными системами из n векторов

$$e'_1, \ldots, e'_n.$$

Линейный оператор A, *соответствующий такой системе, определяется формулой*

$$Ax = x^1 e'_1 + \ldots + x^n e'_n, \quad x = x^1 e_1 + \ldots + x^n e_n,$$

а система, соответствующая линейному оператору A, *состоит из векторов*

$$e'_1 = Ae_1, \ldots, e'_n = Ae_n.$$

Системе n векторов

$$e'_1, \ldots, e'_n$$

мы сопоставим квадратную матрицу порядка n, столбцы которой состоят из координат этих векторов (в базисе $e_1, \ldots, e_n$). Поскольку это соответствие между системами векторов и матрицами, очевидно, биективно, мы получаем, что

линейные операторы и квадратные матрицы находятся в биективном соответствии.

Определение 1. Матрица A, соответствующая указанным образом линейному оператору A, называется *матрицей оператора* A *в базисе* $e_1, \ldots, e_n$.

По определению, если

$$\begin{aligned} Ae_1 &= a_1^1 e_1 + \ldots + a_1^n e_n, \\ &\ldots\ldots\ldots\ldots \\ Ae_n &= a_n^1 e_1 + \ldots + a_n^n e_n, \end{aligned} \tag{1}$$

то

$$A = \begin{pmatrix} a_1^1 & \ldots & a_n^1 \\ \ldots & \ldots & \ldots \\ a_1^n & \ldots & a_n^n \end{pmatrix}.$$

Подчеркнем, что координаты векторов $Ae_1, \ldots, Ae_n$ расположены по столбцам этой матрицы.

Введя в рассмотрение векторные матрицы-строки

$$e = (e_1, \ldots, e_n),$$
$$Ae = (Ae_1, \ldots, Ae_n),$$

мы можем формулы (1) записать в следующем удобном для запоминания виде:

$$Ae = eA. \tag{1'}$$

Легко видеть, что

для любых двух линейных операторов $\boldsymbol{A}$ *и* $\boldsymbol{B}$ *матрицей оператора* $\boldsymbol{AB}$ *является произведение* AB *матриц* A *и* B *операторов* $\boldsymbol{A}$ *и* $\boldsymbol{B}$, *а матрицей оператора* $\boldsymbol{A}+\boldsymbol{B}$ — *сумма* $A+B$ *матриц* A *и* B.

Действительно, ввиду линейности оператора $\boldsymbol{A}$

$$\boldsymbol{A}(\boldsymbol{e}B)=(\boldsymbol{A}\boldsymbol{e})B.$$

Поэтому

$$(\boldsymbol{AB})\boldsymbol{e}=\boldsymbol{A}(\boldsymbol{B}\boldsymbol{e})=\boldsymbol{A}(\boldsymbol{e}B)=(\boldsymbol{A}\boldsymbol{e})B=\boldsymbol{e}(AB).$$

Аналогично,

$$(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{e}=\boldsymbol{A}\boldsymbol{e}+\boldsymbol{B}\boldsymbol{e}=\boldsymbol{e}A+\boldsymbol{e}B=\boldsymbol{e}(A+B).$$

Доказанное утверждение означает, что справедлива следующая

Теорема 1. *Соответствие*

«*линейный оператор*» $\longmapsto$ «*его матрица*»

устанавливает изоморфизм кольца $\mathrm{Lin}(n)$ *линейных операторов на кольцо всех квадратных матриц порядка* n.

В частности, тождественному оператору $\boldsymbol{E}$ соответствует единичная матрица E, а нулевому оператору $\boldsymbol{O}$ — нулевая матрица O. Вообще, оператору $k\boldsymbol{E}$ умножения на число k соответствует скалярная матрица

$$kE=\begin{pmatrix} k & & 0 \\ & \ddots & \\ 0 & & k \end{pmatrix}.$$

Замечание 1. Мы видим, что матрица оператора $k\boldsymbol{E}$ — одна и та же во всех базисах.

Упражнение. Покажите, что операторы вида $k\boldsymbol{E}$ являются единственными операторами, обладающими тем свойством, что матрица такого оператора — одна и та же во всех базисах.

Формула разложения

$$\boldsymbol{x}=x^1\boldsymbol{e}_1+\ \dots\ +x^n\boldsymbol{e}_n$$

произвольного вектора $\boldsymbol{x}$ по векторам базиса имеет в матричной записи вид

$$\boldsymbol{x}=\boldsymbol{e}x,$$

где x — матрица-столбец

$$x=\begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix}.$$

Аналогично, формула

$$\boldsymbol{Ax} = x^1\boldsymbol{Ae}_1 + \ldots + x^n\boldsymbol{Ae}_n$$

может быть записана в следующем виде:

$$\boldsymbol{Ax} = (\boldsymbol{Ae})\,x.$$

Подставив сюда из формулы (1′) выражение для $\boldsymbol{Ae}$, мы немедленно получим, что

$$\boldsymbol{Ax} = \boldsymbol{e}Ax.$$

Сравнив эту формулу с формулой разложения

$$\boldsymbol{y} = \boldsymbol{e}y$$

вектора

$$\boldsymbol{y} = \boldsymbol{Ax},$$

мы получим, далее, что

$$y = Ax, \tag{2}$$

или, в развернутом виде, что

$$\begin{aligned} y^1 &= a_1^1x^1 + \ldots + a_n^1x^n,\\ &\ldots\ldots\ldots\ldots\\ y^n &= a_1^nx^1 + \ldots + a_n^nx^n. \end{aligned} \tag{2′}$$

Таким образом, мы видим, что координаты преобразованного вектора $y = \boldsymbol{Ax}$ линейно выражаются через координаты исходного вектора x с матрицей коэффициентов A.

Подчеркнем, что коэффициентами выражения i-й координаты являются элементы i-й строки матрицы A.

Резюмируя, мы видим, что предложение 1 может быть переформулировано в следующем виде:

Предложение 2. *Выбор базиса*

$$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$$

устанавливает изоморфизм кольца линейных операторов на кольцо квадратных матриц порядка n. Столбцы матрицы A, соответствующей при этом изоморфизме линейному оператору $\boldsymbol{A}$, *состоят из координат векторов*

$$\boldsymbol{Ae}_1, \ldots, \boldsymbol{Ae}_n.$$

Линейный оператор $\boldsymbol{A}$, *соответствующий матрице A, переводит произвольный вектор* $x = x^1e_1 + \ldots + x^ne_n$ *в вектор* $y = y^1e_1 + \ldots + y^ne_n$ *с координатами* (2′).

При $n = 1$ мы получаем отсюда, что

линейные операторы на прямой (т. е. в линеале Vect(1)) *исчерпываются операторами* $k\boldsymbol{E}$ *умножения на числа.*

Построенный изоморфизм между кольцами линейных операторов и квадратных матриц зависит от выбора базиса. Найдем, как меняется матрица оператора при изменении базиса.

Пусть

$$f_1, \ldots, f_n$$

— другой базис в $\mathrm{Vect}(n)$ и пусть C — матрица перехода от базиса $e_1, \ldots, e_n$ к базису $f_1, \ldots, f_n$. По определению (см. п. 7 § 3) это означает, что векторные матрицы-строки

$$e = (e_1, \ldots, e_n),$$
$$f = (f_1, \ldots, f_n)$$

связаны формулой

$$f = eC.$$

Применив к этой формуле линейный оператор $\boldsymbol{A}$ и заметив, что в силу линейности

$$\boldsymbol{A}(eC) = (\boldsymbol{A}e)\,C,$$

мы получим, что

$$\boldsymbol{A}f = (\boldsymbol{A}e)\,C,$$

т. е. что

$$fA' = (eA)\,C,$$

где A' — матрица оператора $\boldsymbol{A}$ в базисе $f_1, \ldots, f_n$, а A, как и выше, — матрица оператора $\boldsymbol{A}$ в базисе $e_1, \ldots, e_n$.

Подставив сюда выражение для f, мы получим, далее, что

$$e(CA') = e(AC),$$

откуда следует, что

$$CA' = AC,$$

т. е. что

$$A' = C^{-1}AC.$$

Таким образом, мы доказали следующее

Предложение 3. *Если в базисе*

$$e_1, \ldots, e_n$$

линейный оператор $\boldsymbol{A}$ *имеет матрицу* A, *то в базисе*

$$f_1, \ldots, f_n$$

его матрицей будет матрица

$$C^{-1}AC,$$

где C — *матрица перехода от базиса* $e_1, \ldots, e_n$ к базису $f_1, \ldots, f_n$.

4. Обратимые линейные операторы

Определение 1. Линейный оператор $\boldsymbol{A}$ называется *обратимым слева* (*справа*), если существует такой линейный оператор $\boldsymbol{B}$ (линейный оператор $\boldsymbol{C}$), что $\boldsymbol{BA} = \boldsymbol{E}$ (соответственно, $\boldsymbol{AC} = \boldsymbol{E}$). Оператор $\boldsymbol{A}$ называется *обратимым*, если он обратим и слева, и справа.

Обратим внимание на то, что несмотря на полное внешнее сходство определений, это понятие обратимости отличается априори от обратимости в смысле п. 1. Дело здесь в том, что оператор A обратим, например, слева в смысле п. 1 (т. е. инъективен), если существует какой-то оператор B, удовлетворяющий соотношению $BA = E$, тогда как теперь мы требуем, чтобы этот оператор был линейным. Таким образом, если оператор обратим в нашем теперешнем смысле, то он заведомо обратим в смысле п. 1, но обратное пока неясно.

Замечательным фактом теории линейных операторов является то обстоятельство, что на самом деле оба понятия обратимости совпадают. Еще более удивительно то, что для линейных операторов из обратимости слева или справа вытекает обратимость (т. е. обратимость и слева, и справа). Другими словами, справедлива

Теорема 1. *Следующие свойства линейного оператора* A *равносильны:*

1°. *Оператор* A *обратим слева.*
2°. *Оператор* A *инъективен.*
3°. *Оператор* A *обратим справа.*
4°. *Оператор* A *надъективен.*
5°. *Оператор* A *обратим.*
6°. *Оператор* A *биективен.*

Доказательство. Пусть

$$e_1, \ldots, e_n$$

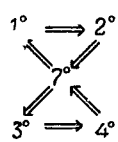
— произвольный базис линеала $\mathrm{Vect}(n)$ и пусть, как и в предыдущем пункте,

$$e'_1 = Ae_1, \ \ldots, \ e'_n = Ae_n.$$

Мы добавим к свойствам 1°—6° еще следующее свойство:

7°. *Векторы* $e'_1, \ldots, e'_n$ *образуют базис.*

Ясно, что теорема будет доказана, если мы докажем следующую систему импликаций:



Импликация $1° \Rightarrow 2°$. Эта импликация имеет общий характер (справедлива для любых отображений, см. п. 1).

Импликация $2° \Rightarrow 7°$. Предположим, что векторы $e'_1, \ldots, e'_n$ не образуют базиса. Тогда они линейно зависимы и потому существуют такие числа $k_1, \ldots, k_n$ (не все равные нулю), что

$$k_1 e'_1 + \ldots + k_n e'_n = 0.$$

Рассмотрим вектор

$$k = k_1 e_1 + \ldots + k_n e_n.$$

В силу линейности оператора A этот вектор обладает тем свойством, что

$$Ak = 0.$$

Но ясно, что $A0 = 0$. Поэтому в силу инъективности оператора A должно иметь место равенство $k = 0$, что невозможно, поскольку векторы $e_1, \ldots, e_n$ линейно независимы. Полученное противоречие доказывает, что $2° \Rightarrow 7°$.

Импликация $7° \Rightarrow 1°$. Определим линейный оператор B требованием, чтобы векторы $e'_1, \ldots, e'_n$ он переводил соответственно в векторы $e_1, \ldots, e_n$. Поскольку векторы $e'_1, \ldots, e'_n$ образуют базис, такой оператор B существует и однозначно определен (см. п. 3). Тогда оператор BA будет переводить базис $e_1, \ldots, e_n$ в себя и потому будет совпадать с тождественным оператором E.

Импликация $3° \Rightarrow 4°$. Подобно импликации $1° \Rightarrow 2°$ эта импликация справедлива из общих соображений (см. п. 1).

Импликация $4° \Rightarrow 7°$. По условию, для любого вектора y существует такой вектор x, что $Ax = y$. Разложим вектор x по векторам базиса:

$$x = x^1 e_1 + \ldots + x^n e_n.$$

Применив к этому равенству оператор A, мы немедленно получим, что

$$y = x^1 e'_1 + \ldots + x^n e'_n.$$

Это показывает, что семейство векторов $e'_1, \ldots, e'_n$ полно. Так как оно содержит n векторов, то, следовательно, оно является базисом.

Импликация $7° \Rightarrow 3°$. Положим $C = B$, где B — линейный оператор, построенный при доказательстве импликации $7° \Rightarrow 1°$. Ясно, что $AC = E$.

Тем самым теорема 1 полностью доказана.

Заметим, что одновременно мы также доказали, что

линейный оператор A тогда и только тогда обратим, когда он обладает свойством 7°.

Поскольку столбцы матрицы оператора A состоят, по определению, из координат векторов $e'_1, \ldots, e'_n$, свойство 7° равносильно следующему свойству:

8°. *Матрица оператора A невырождена.*

Таким образом,

линейный оператор A тогда и только тогда обратим, когда его матрица (в произвольном базисе) невырождена.

Поскольку обратимый оператор $\boldsymbol{A}$ биективен, для него существует обратный оператор $\boldsymbol{A}^{-1}$. Легко видеть, что

обратный оператор $\boldsymbol{A}^{-1}$ линеен.

Действительно, по условию существует такой линейный оператор $\boldsymbol{B}$, что

$$\boldsymbol{BA} = \boldsymbol{E}.$$

Следовательно,

$$\boldsymbol{A}^{-1} = \boldsymbol{EA}^{-1} = (\boldsymbol{BA})\,\boldsymbol{A}^{-1} = \boldsymbol{B}(\boldsymbol{AA}^{-1}) = \boldsymbol{BE} = \boldsymbol{B}.$$

Поэтому оператор $\boldsymbol{A}^{-1}$ линеен.

Впрочем, линейность оператора $\boldsymbol{A}^{-1}$ непосредственно вытекает и из определений. На самом деле, если $\boldsymbol{Ax}_1 = \boldsymbol{y}_1$ и $\boldsymbol{Ax}_2 = \boldsymbol{y}_2$, то $\boldsymbol{A}(\boldsymbol{x}_1 + \boldsymbol{x}_2) = \boldsymbol{y}_1 + \boldsymbol{y}_2$ и потому $\boldsymbol{A}^{-1}(\boldsymbol{y}_1 + \boldsymbol{y}_2) = \boldsymbol{x}_1 + \boldsymbol{x}_2$, т. е.

$$\boldsymbol{A}^{-1}(\boldsymbol{y}_1 + \boldsymbol{y}_2) = \boldsymbol{A}^{-1}\boldsymbol{y}_1 + \boldsymbol{A}^{-1}\boldsymbol{y}_2.$$

Аналогично, если $\boldsymbol{Ax} = \boldsymbol{y}$, то $\boldsymbol{A}(k\boldsymbol{x}) = k\boldsymbol{y}$, и потому $\boldsymbol{A}^{-1}(k\boldsymbol{y}) = k\boldsymbol{x}$, т. е.

$$\boldsymbol{A}^{-1}(k\boldsymbol{y}) = k\boldsymbol{A}^{-1}\boldsymbol{y}.$$

Матрицей оператора $\boldsymbol{A}^{-1}$ является, очевидно, матрица A^{-1}, обратная к матрице A оператора $\boldsymbol{A}$.

Тот факт, что оператор $\boldsymbol{A}^{-1}$ линеен, означает, что *обратимые линейные операторы* (действующие в $\mathrm{Vect}(n)$ при данном n) *образуют группу.* Мы будем обозначать эту группу символом $\mathrm{Op}(n)$.

Из сказанного выше о соответствии между линейными операторами и их матрицами непосредственно вытекает, что

соответствие

«*линейный оператор*» $\longmapsto$ «*его матрица*»

определяет изоморфизм группы $\mathrm{Op}(n)$ *обратимых линейных операторов на группу* $\mathrm{GL}(n)$ *всех невырожденных квадратных матриц порядка* n.

Заметим, что хотя группы операторов и матриц изоморфны, отождествлять их нельзя, поскольку никакого естественного (не зависящего от выбора базиса) изоморфизма между ними не имеется.

Докажем в заключение следующее предложение, часто полезное при доказательстве необратимости конкретных операторов.

Предложение 1. *Линейный оператор $\boldsymbol{A}$ тогда и только тогда необратим, когда существует такой отличный от нуля вектор $\boldsymbol{x}_0$, что*

$$\boldsymbol{Ax}_0 = \boldsymbol{0}. \qquad (1)$$

Доказательство. Если оператор A обратим, то, применив к равенству (1) оператор A^{-1}, мы получим, что

$$x_0 = 0$$

(поскольку оператор A^{-1} линеен, то $A^{-1}0 = 0$). Следовательно, если вектор $x_0 \neq 0$, удовлетворяющий соотношению (1), существует, то оператор A необратим.

Обратно, пусть оператор A необратим. Тогда согласно доказанному выше он обязательно неинъективен, т. е. существуют такие различные векторы x_1 и x_2, что

$$Ax_1 = Ax_2.$$

Положив

$$x_0 = x_2 - x_1,$$

мы и получим вектор $x_0 \neq 0$, удовлетворяющий соотношению (1).

5. Операторы, действующие по равенству координат

Обратим внимание на то, что равенства

$$\begin{aligned} e'_1 &= a_1^1 e_1 + \ldots + a_1^n e_n, \\ &\cdots\cdots\cdots\cdots \\ e'_n &= a_n^1 e_1 + \ldots + a_n^n e_n, \end{aligned}$$

выражающие векторы

$$e'_1 = Ae_1, \ \ldots, \ e'_n = Ae_n$$

через векторы базиса

$$e_1, \ \ldots, \ e_n,$$

лишь обозначениями отличаются от формул (1) п. 7 § 3, выражающих векторы базиса

$$e_{1'}, \ \ldots, \ e_{n'}$$

через векторы базиса

$$e_1, \ \ldots, \ e_n.$$

Таким образом, на эти формулы (при невырожденной матрице A) возможны две «сопряженные» точки зрения: их можно рассматривать либо как формулы перехода от одного («старого») базиса к другому («новому»), либо как формулы, описывающее действие некоторого обратимого линейного оператора на векторах данного базиса.

Аналогично, формулы

$$\begin{aligned} y^1 &= a_1^1 x^1 + \ldots + a_n^1 x^n, \\ &\cdots\cdots\cdots\cdots \\ y^n &= a_1^n x^1 + \ldots + a_n^n x^n \end{aligned}$$

также можно истолковывать двояко в зависимости от того, что понимается под величинами $x^1, \ldots, x^n$ и $y^1, \ldots, y^n$. При одном истолковании величины $x^1, \ldots, x^n$ являются координатами произвольного вектора $\boldsymbol{x}$ в базисе $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, а величины $y^1, \ldots, y^n$ — координатами в том же базисе преобразованного вектора $\boldsymbol{y} = \boldsymbol{Ax}$. При другом истолковании величины $x^1, \ldots, x^n$ являются координатами произвольного вектора $\boldsymbol{x}$ в «новом» базисе $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$, а величины $y^1, \ldots, y^n$ — координатами того же вектора $\boldsymbol{x}$ в «старом» базисе $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.

Замечание 1. Подчеркнем, что при первом истолковании величины $x^1, \ldots, x^n$ являются координатами вектора $\boldsymbol{x}$ в базисе $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, а при втором — в базисе $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$.

Таким образом, одна и та же алгебраическая конструкция (линейное преобразование неизвестных) имеет два различных (хотя и тесно между собой связанных) геометрических истолкования.

Объяснение этого факта можно получить, сравнив формулу

$$\boldsymbol{x} = x^1\boldsymbol{e}_1 + \ldots + x^n\boldsymbol{e}_n$$

с формулой

$$\boldsymbol{Ax} = x^1\boldsymbol{e}'_1 + \ldots + x^n\boldsymbol{e}'_n.$$

Мы видим, что

преобразованный вектор $\boldsymbol{Ax}$ имеет в преобразованном базисе $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ те же координаты, что и исходный вектор $\boldsymbol{x}$ в базисе $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.

Чтобы описать это явление, введем следующее

Определение 1. Пусть $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ и $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$ — два базиса в $\mathrm{Vect}(n)$. Говорят, что оператор $\boldsymbol{A}$ *действует по равенству координат* (в данных базисах), если произвольному вектору $\boldsymbol{x}$ он сопоставляет вектор $\boldsymbol{y} = \boldsymbol{Ax}$, имеющий в базисе $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$ те же координаты, что и вектор $\boldsymbol{x}$ в базисе $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.

В этой терминологии доказанное выше утверждение означает, что

любой обратимый линейный оператор действует по равенству координат в некоторых базисах.

Замечание 2. Легко видеть, что и, обратно,

любой оператор, действующий по равенству координат, является, во-первых, линейным, а во-вторых, обратимым оператором.

Задание. Докажите это утверждение.

Если мы теперь хотим найти координаты вектора $\boldsymbol{y} = \boldsymbol{Ax}$ не в базисе $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$, а в исходном базисе $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, нам достаточно воспользоваться формулами преобразования координат при переходе от базиса $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$ к базису $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. Это и объясняет, почему линейные преобразования координат

могут быть истолкованы и как преобразования, описывающие действие линейных операторов.

Замечание 3. Подчеркнем, что такого рода «двойное» истолкование возможно лишь для линейных преобразований неизвестных с невырожденной матрицей A. Линейные преобразования с вырожденной матрицей A описывают в координатах действия необратимых линейных операторов, но никакого отношения к замене базисов не имеют.

6. Обратимые линейные операторы и ориентации

Пусть

$$e_1, \ldots, e_n$$

и

$$f_1, \ldots, f_n$$

— два базиса в $\mathrm{Vect}(n)$ и пусть C — матрица перехода от первого базиса ко второму. По определению, это означает, что для векторных матриц-строк

$$e = (e_1, \ldots, e_n)$$
$$f = (f_1, \ldots, f_n)$$

имеет место равенство

$$f = eC.$$

Применив к этому равенству линейный оператор A, мы получим равенство

$$f' = e'C,$$

где

$$f' = (Af_1, \ldots, Af_n),$$
$$e' = (Ae_1, \ldots, Ae_n).$$

Если оператор A обратим, то семейства $Af_1, \ldots, Af_n$ и $Ae_1, \ldots, Ae_n$ также являются базисами, и доказанная формула означает, что эти базисы связаны той же самой матрицей перехода C.

Таким образом, мы доказали, что

если два базиса связаны матрицей перехода C, то базисы, получающиеся из данных воздействием некоторого обратимого линейного оператора A, также связаны матрицей перехода C.

Вспомним теперь, что два базиса называются одноименными (определяющими одну и ту же ориентацию), если определитель, связывающий их матрицы перехода, положителен. Отсюда и из только что доказанного утверждения непосредственно вытекает, что

обратимые линейные операторы переводят одноименные базисы в одноименные, а разноименные — в разноименные.

Это означает, что, сопоставив ориентации o, задаваемой базисом

$$e_1, \ldots, e_n,$$

ориентацию o', задаваемую базисом

$$e'_1 = Ae_1, \ldots, e'_n = Ae_n,$$

мы получим корректно определенное соответствие между ориентациями. Мы будем говорить, что это соответствие *индуцировано* обратимым линейным оператором A. Ориентацию o' мы будем обозначать символом $A(o)$.

Поскольку обратимый линейный оператор переводит разноименные базисы в разноименные, то противоположные ориентации линейный оператор переводит в противоположные:

$$A(-o) = -A(o).$$

Отсюда вытекает, что либо

$$A(o) = o$$

для любой ориентации o, либо

$$A(o) = -o$$

также для любой ориентации o.

В первом случае говорят, что обратимый линейный оператор *сохраняет ориентации*, а во втором — что он *обращает ориентации*.

Множество всех обратимых линейных операторов, сохраняющих ориентации, мы будем обозначать символом $\mathrm{Op}^+(n)$, а дополнительное множество обратимых линейных операторов, обращающих ориентации, — символом $\mathrm{Op}^-(n)$.

Ясно, что

1) *если обратимые линейные операторы* A *и* B *оба сохраняют или обращают ориентации, то их произведение* AB *сохраняет ориентации, а если один из операторов* A *или* B *сохраняет ориентации, а другой обращает, то оператор* AB *обращает ориентации*;

2) *если обратимый линейный оператор* A *сохраняет (обращает) ориентации, то обратный оператор* A^{-1} *также сохраняет (соответственно обращает) ориентации.*

На языке теории групп это означает, что

множество $\mathrm{Op}^+(n)$ *обратимых линейных операторов, сохраняющих ориентации, является подгруппой (даже нормальным делителем) группы* $\mathrm{Op}(n)$, *а множество* $\mathrm{Op}^-(n)$ *является смежным классом группы* $\mathrm{Op}(n)$ *по подгруппе* $\mathrm{Op}^+(n)$.

Тот факт, что обратимый линейный оператор A сохраняет (обращает) ориентации, означает, что произвольный базис

$$e_1, \ldots e_n$$

он переводит в одноименный (разноименный) базис

$$e'_1 = Ae_1, \ldots, e'_n = Ae_n.$$

Но, как мы знаем, матрицей перехода от базиса $e_1, \ldots, e_n$ к базису $e'_1, \ldots, e'_n$ служит матрица A оператора $\boldsymbol{A}$ в базисе $e_1, \ldots, e_n$. Следовательно,

обратимый линейный оператор тогда и только тогда сохраняет (обращает) ориентации, когда в произвольном базисе его матрица имеет положительный (соответственно отрицательный) определитель.

Пусть $\mathrm{GL}^+(n)$ — совокупность всех квадратных матриц порядка n с положительным определителем, а $\mathrm{GL}^-(x)$ — совокупность всех квадратных матриц порядка n с отрицательным определителем. Из того, что при умножении матриц их определители перемножаются, непосредственно вытекает, что

по отношению к умножению матриц множество $\mathrm{GL}^+(n)$ *является подгруппой (даже нормальным делителем) группы* $\mathrm{GL}(n)$, *а множество* $\mathrm{GL}^-(n)$ *— смежным классом по этой подгруппе.*

При этом согласно сказанному выше справедлива следующая

Теорема 1. *Соответствие*

«*линейный оператор*» $\longmapsto$ «*его матрица*»

определяет изоморфизм группы $\mathrm{Op}^+(n)$ *на группу* $\mathrm{GL}^+(n)$.

Семейство матриц $A(t)$, зависящих от числового параметра t, называется *непрерывным семейством*, если каждый элемент матрицы $A(t)$ является непрерывной функцией параметра t. В соответствии с этим семейство $\boldsymbol{A}(t)$ линейных операторов называется *непрерывным семейством*, если матрицы $A(t)$ операторов $\boldsymbol{A}(t)$ (в некотором базисе $e_1, \ldots, e_n$) образуют непрерывное семейство. Поскольку при изменении базиса матрицы $A(t)$ переходят (см. п. 2) в матрицы $C^{-1}A(t)C$ и поскольку семейство матриц $C^{-1}A(t)C$ непрерывно тогда и только тогда, когда непрерывно семейство $A(t)$, то свойство непрерывности семейства линейных операторов определено корректно (не зависит от выбора базиса).

Пусть, для определенности, числовой параметр t пробегает отрезок $[0, 1]$.

Определение 1. Если

$$\boldsymbol{A}(0) = \boldsymbol{A}, \quad \boldsymbol{A}(1) = \boldsymbol{B},$$

то непрерывное семейство обратимых линейных операторов $\boldsymbol{A}(t)$ мы будем называть *деформацией*, связывающей обратимый линейный оператор $\boldsymbol{A}$ с обратимым линейным оператором $\boldsymbol{B}$. Обратимые линейные операторы $\boldsymbol{A}$ и $\boldsymbol{B}$ мы будем называть

деформируемыми друг в друга, если существует хотя бы одна деформация, связывающая один оператор с другим.

Понятие деформации обратимых линейных операторов тесно связано с введенным в п. 1 § 4 понятием деформации базисов. Действительно, легко видеть, что

семейство $\boldsymbol{A}(t)$ обратимых линейных операторов тогда и только тогда является деформацией, связывающей оператор $\boldsymbol{A}$ с оператором $\boldsymbol{B}$, когда для любого базиса $e_1, \ldots, e_n$ вектор-функции

$$a_1(t) = \boldsymbol{A}(t)\, e_1, \ \ldots, \ a_n(t) = \boldsymbol{A}(t)\, e_n$$

составляют деформацию, связывающую базис

$$a_1 = \boldsymbol{A}e_1, \ \ldots, \ a_n = \boldsymbol{A}e_n$$

с базисом

$$b_1 = \boldsymbol{B}e_1, \ \ldots, \ b_n = \boldsymbol{B}e_n.$$

Но, как было показано в п. 1 § 4, два базиса тогда и только тогда деформируемы друг в друга, когда они одноименны. Следовательно,

обратимые линейные операторы $\boldsymbol{A}$ и $\boldsymbol{B}$ тогда и только тогда деформируемы друг в друга, когда для любого базиса

$$e_1, \ \ldots, \ e_n$$

базисы

$$\boldsymbol{A}e_1, \ \ldots, \ \boldsymbol{A}e_n$$

и

$$\boldsymbol{B}e_1, \ \ldots, \ \boldsymbol{B}e_n$$

одноименны.

Последнее условие, очевидно, означает, что операторы $\boldsymbol{A}$ и $\boldsymbol{B}$ одновременно либо сохраняют, либо обращают ориентации. Тем самым мы доказали следующее

Предложение 1. *Обратимые линейные операторы тогда и только тогда деформируемы друг в друга, когда они оба либо сохраняют, либо обращают ориентации.*

Другими словами,

обратимые линейные операторы, принадлежащие одновременно либо подгруппе $\mathrm{Or}^+(n)$, либо смежному классу $\mathrm{Or}^-(n)$, деформируемы друг в друга, а обратимые линейные операторы, один из которых принадлежит подгруппе $\mathrm{Or}^+(n)$, а другой — смежному классу $\mathrm{Or}^-(n)$, друг в друга не деформируемы.

В частности, мы видим, что

обратимый линейный оператор тогда и только тогда сохраняет ориентации (принадлежат подгруппе $\mathrm{Or}^+(n)$), когда он может быть продеформирован в тождественный оператор $\boldsymbol{E}$.

Для матриц доказанные утверждения означают, что

две квадратные невырожденные матрицы тогда и только тогда могут быть продеформированы друг в друга, когда их определители имеют один и тот же знак.

В частности,

квадратная невырожденная матрица тогда и только тогда может быть продеформирована в единичную матрицу, когда ее определитель положителен.

7. Изометричные операторы

Определение 1. Линейный оператор U называется *изометричным*, если он сохраняет скалярное произведение, т. е. для любых двух векторов x и y имеет место равенство

$$Ux \cdot Uy = xy.$$

Предложение 1. *Линейный оператор U тогда и только тогда изометричен, когда он сохраняет длины векторов, т. е.*

$$|Ux| = |x|$$

для любого вектора x.

Доказательство. Если оператор изометричен, то, в частности,

$$(Ux)^2 = x^2$$

для любого вектора x, т. е.

$$|Ux| = |x|.$$

Заметим, что линейностью оператора U мы здесь не пользовались.

Обратно, пусть оператор U сохраняет длины векторов. Тогда для любых векторов x и y

$$|Ux| = |x|, \quad |Uy| = |y|, \quad |U(x+y)| = |x+y|,$$

т. е.

$$(Ux)^2 = x^2, \quad (Uy)^2 = y^2, \quad (U(x+y))^2 = (x+y)^2.$$

Но в силу линейности

$$U(x+y) = Ux + Uy,$$

и потому

$$(U(x+y))^2 = (Ux)^2 + 2Ux \cdot Uy + (Uy)^2.$$

С другой стороны,

$$(U(x+y))^2 = (x+y)^2 = x^2 + 2xy + y^2 = (Ux)^2 + 2xy + (Uy)^2.$$

Поэтому

$$Ux \cdot Uy = xy,$$

так что оператор U изометричен.

Из этого предложения непосредственно вытекает, что *любой изометричный оператор U обратим.*

Действительно, если $Ux_0 = 0$, то $|x_0| = |Ux_0| = |0| = 0$, и потому $x_0 = 0$.

Далее, ясно, что

для любого изометричного оператора U *и любого ортонормированного базиса* $e_1, \ldots, e_n$ *векторы*

$$e'_1 = Ue_1, \ \ldots, \ e'_n = Ue_n$$

составляют ортонормированный базис.

Обратно,

если для линейного оператора U *и некоторого ортонормированного базиса* $e_1, \ldots, e_n$ *векторы*

$$e'_1 = Ue_1, \ \ldots, \ e'_n = Ue_n$$

составляют ортонормированный базис, то оператор U *изометричен.*

Действительно, для любого вектора

$$x = x^1 e_1 + \ \ldots \ + x^n e_n$$

вектор Ux выражается формулой

$$Ux = x^1 e'_1 + \ \ldots \ + x^n e'_n.$$

Поскольку векторы $e'_1, \ldots, e'_n$ составляют ортонормированный базис, отсюда следует, что

$$|Ux| = (x^1)^2 + \ \ldots \ + (x^n)^2 = |x|$$

и, следовательно, оператор U изометричен.

Оба последних утверждения мы можем объединить в следующем предложении:

Предложение 2. *Оператор* U *тогда и только тогда изометричен, когда он действует по равенству координат в двух ортонормированных базисах.*

Как мы знаем, матрица оператора U является матрицей перехода от базиса $e_1, \ldots, e_n$ к базису $e'_1, \ldots, e'_n$. Но (см. п. 6 § 5) матрица тогда и только тогда является матрицей перехода от одного ортонормированного базиса к другому, когда она ортогональна. Следовательно, имеет место следующее

Предложение 3. *Линейный оператор тогда и только тогда изометричен, когда в некотором (а потому и в любом) ортонормированном базисе его матрица ортогональна.*

Ясно, что произведение двух изометричных операторов, а также оператор, обратный к изометричному оператору, являются изометричными операторами. Это означает, что

совокупность $\mathrm{Iso}(n)$ *всех изометричных операторов является группой.*

Любой линейный изометричный оператор либо сохраняет ориентации, либо их обращает. Ясно, что

совокупность $\mathrm{Iso}^+(n)$ *всех линейных изометричных операторов, сохраняющих ориентации, является подгруппой (даже нормальным делителем) группы* $\mathrm{Iso}(n)$, *а совокупность*

$\mathrm{Iso}^-(n)$ *всех линейных изометричных операторов, обращающих ориентации, является смежным классом группы* $\mathrm{Iso}(n)$ *по этой подгруппе.*

Выбор ортонормированного базиса устанавливает изоморфизм группы $\mathrm{Iso}(n)$ на группу $\mathrm{O}(n)$ всех ортогональных матриц порядка n. Подгруппе $\mathrm{Iso}^+(n)$ соответствует при этом изоморфизме подгруппа $\mathrm{O}^+(n) = \mathrm{SO}(n)$ всех собственных ортогональных матриц.

8. Свойства изометричных операторов

В этом пункте мы докажем несколько лемм об изометричных операторах, которые нам понадобятся в гл. 7.

Сначала мы рассмотрим изометричные операторы на плоскости.

Лемма 1. *Для любого нетождественного сохраняющего ориентации изометричного оператора* $\boldsymbol{U}$ *на плоскости* (*т. е. в* $\mathrm{Vect}(2)$) *оператор* $\boldsymbol{E} - \boldsymbol{U}$ *обратим.*

Доказательство. Выбрав на плоскости ортонормированный базис, рассмотрим матрицу оператора $\boldsymbol{U}$. Как мы знаем, эта матрица является собственной ортогональной матрицей, и потому (см. п. 6 § 5) имеет вид

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}.$$

При этом $\alpha \neq 0$ (поскольку по условию $\boldsymbol{U} \neq \boldsymbol{E}$). Следовательно, матрица оператора $\boldsymbol{E} - \boldsymbol{U}$ имеет вид

$$\begin{pmatrix} 1-\cos\alpha & \sin\alpha \\ -\sin\alpha & 1-\cos\alpha \end{pmatrix}$$

и ее определитель $(1-\cos\alpha)^2 + \sin^2\alpha = 2(1-\cos\alpha)$ отличен от нуля. Таким образом, матрица оператора $\boldsymbol{E} - \boldsymbol{U}$ невырождена и потому этот оператор обратим.

Лемма 2. *Для любого обращающего ориентации изометричного оператора* $\boldsymbol{U}$ *на плоскости* (*в* $\mathrm{Vect}(2)$) *оператор* $\boldsymbol{E} - \boldsymbol{U}$ *необратим.*

Доказательство. В произвольном ортонормированном базисе матрица оператора $\boldsymbol{U}$, являясь несобственной ортогональной матрицей, имеет вид

$$\begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix},$$

и потому матрица оператора $\boldsymbol{E} - \boldsymbol{U}$ имеет вид

$$\begin{pmatrix} 1-\cos\alpha & -\sin\alpha \\ -\sin\alpha & 1+\cos\alpha \end{pmatrix}.$$

Поскольку

$$\begin{vmatrix} 1-\cos\alpha & -\sin\alpha \\ -\sin\alpha & 1+\cos\alpha \end{vmatrix} = 1-\cos^2\alpha-\sin^2\alpha = 0,$$

оператор $\boldsymbol{E}-\boldsymbol{U}$ необратим.

Замечание 1. Тот факт, что оператор $\boldsymbol{E}-\boldsymbol{U}$ необратим, означает (см. п. 4), что существует такой вектор $\boldsymbol{e}_1 \neq \boldsymbol{0}$, что

$$(\boldsymbol{E}-\boldsymbol{U})\,\boldsymbol{e}_1 = \boldsymbol{0},$$

т. е. такой, что

$$\boldsymbol{U}\boldsymbol{e}_1 = \boldsymbol{e}_1.$$

Ясно, что этот вектор мы можем считать единичным (при $|\boldsymbol{e}_1| \neq 1$ достаточно разделить его на $|\boldsymbol{e}_1|$).

Дополнив этот вектор ортогональным единичным вектором $\boldsymbol{e}_2$ до базиса, мы получим ортонормированный базис $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, в котором первый столбец матрицы оператора $\boldsymbol{U}$ имеет вид

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Следовательно, в этом базисе $\alpha = 0$, так что матрицей оператора $\boldsymbol{U}$ является матрица

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Тем самым мы доказали

Следствие. *Для любого обращающего ориентации изометричного оператора $\boldsymbol{U}$ на плоскости существует ортонормированный базис, в котором матрица оператора $\boldsymbol{U}$ имеет вид*

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Перейдем теперь к операторам в пространстве. Начнем с леммы, справедливой для любых (не обязательно изометричных) операторов.

Лемма 3. *Для любого линейного оператора $\boldsymbol{A}$ в пространстве* (в $\mathrm{Vect}(3)$) *существует отличный от нуля вектор $\boldsymbol{x}_0$, коллинеарный вектору $\boldsymbol{A}\boldsymbol{x}_0$, т. е. такой, что*

$$\boldsymbol{A}\boldsymbol{x}_0 = \lambda\boldsymbol{x}_0, \tag{1}$$

где λ — некоторое число (возможно, равное нулю).

Доказательство. Вводя оператор $\lambda\boldsymbol{E}$ умножения векторов на число λ, мы можем соотношение (1) переписать в следующем «операторном» виде:

$$(\boldsymbol{A}-\lambda\boldsymbol{E})\,\boldsymbol{x}_0 = \boldsymbol{0}. \tag{2}$$

Существование вектора $x_0 \neq 0$, удовлетворяющего этому уравнению, означает, что оператор $A - \lambda E$ необратим (см. п. 3), т. е. что его матрица (в некотором базисе e_1, e_2, e_3) вырождена. Но мы знаем, что матрицей этого оператора является матрица

$$A - \lambda E,$$

где

$$A = \begin{pmatrix} a_1^1 & a_2^1 & a_3^1 \\ a_1^2 & a_2^2 & a_3^2 \\ a_1^3 & a_2^3 & a_3^3 \end{pmatrix}$$

— матрица оператора A, а

$$\lambda E = \begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda \end{pmatrix}$$

— матрица оператора λE.

Поэтому векторное уравнение (2) тогда и только тогда имеет нетривиальное (отличное от нуля) решение x_0, когда число λ таково, что определитель $|A - \lambda E|$ матрицы $A - \lambda E$ равен нулю.

Равенство

$$|A - \lambda E| = 0,$$

т. е. равенство

$$\begin{vmatrix} a_1^1 - \lambda & a_2^1 & a_3^1 \\ a_1^2 & a_2^2 - \lambda & a_3^2 \\ a_1^3 & a_2^3 & a_3^3 - \lambda \end{vmatrix} = 0$$

является относительно λ уравнением третьей степени с вещественными коэффициентами и потому имеет хотя бы один вещественный корень.

Принимая в соотношении (2) за λ этот корень, мы получим, следовательно, уравнение, имеющее нетривиальное решение. Это решение и является искомым вектором x_0.

Следствие. *Для любого изометричного линейного оператора* U *в* Vect(3) *существует такой отличный от нуля вектор* x_0, *что*

$$Ux_0 = x_0 \quad \text{или} \quad Ux_0 = -x_0.$$

Доказательство. Достаточно заметить, что в силу изометричности оператора U векторы Ux_0 и x_0 должны иметь одну и ту же длину, и потому число λ в равенстве (1) должно быть равно ± 1.

Оказывается, что для изометричных операторов справедлива даже следующая более сильная

Лемма 4. *Для любого изометричного линейного оператора* $\boldsymbol{U}$ *в* $\mathrm{Vect}(3)$ *существует такой отличный от нуля вектор* $\boldsymbol{e}$, *что*

$$\boldsymbol{U}\boldsymbol{e} = \begin{cases} \boldsymbol{e}, & \text{если } \boldsymbol{U} \text{ сохраняет ориентации,} \\ -\boldsymbol{e}, & \text{если } \boldsymbol{U} \text{ обращает ориентации.} \end{cases}$$

Доказательство. Рассмотрим произвольную плоскость Π, перпендикулярную вектору $\boldsymbol{x}_0$. Линеал $\mathrm{Vect}_\Pi(2)$ векторов, параллельных этой плоскости, состоит, таким образом, из всех векторов, ортогональных вектору $\boldsymbol{x}_0$, т. е. из векторов $\boldsymbol{x} \in \mathrm{Vect}(3)$, удовлетворяющих соотношению

$$\boldsymbol{x}\boldsymbol{x}_0 = 0.$$

Поскольку оператор $\boldsymbol{U}$ изометричен и $\boldsymbol{U}\boldsymbol{x}_0 = \pm\boldsymbol{x}_0$, для любого такого вектора $\boldsymbol{x}$ будет иметь место равенство

$$\boldsymbol{U}\boldsymbol{x} \cdot \boldsymbol{x}_0 = \pm\, \boldsymbol{U}\boldsymbol{x} \cdot \boldsymbol{U}\boldsymbol{x}_0 = \pm\, \boldsymbol{x}\boldsymbol{x}_0 = 0,$$

показывающее, что

если $\boldsymbol{x} \in \mathrm{Vect}_\Pi(2)$, *то* $\boldsymbol{U}\boldsymbol{x} \in \mathrm{Vect}_\Pi(2)$.

Это позволяет нам построить в $\mathrm{Vect}_\Pi(2)$ *индуцированный оператор* $\boldsymbol{U}'$, определяющийся формулой

$$\boldsymbol{U}'\boldsymbol{x} = \boldsymbol{U}\boldsymbol{x}, \quad \boldsymbol{x} \in \mathrm{Vect}_\Pi(2)$$

(его отличие от оператора $\boldsymbol{U}$ состоит только в том, что он действует в $\mathrm{Vect}_\Pi(2)$, а не в $\mathrm{Vect}(3)$).

Ясно, что оператор $\boldsymbol{U}'$ также изометричен.

Рассмотрим теперь произвольный ортонормированный базис

$$\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \boldsymbol{e}_3,$$

последний вектор $\boldsymbol{e}_3$ которого коллинеарен вектору $\boldsymbol{x}_0$. Векторы

$$\boldsymbol{e}_1,\ \boldsymbol{e}_2$$

образуют ортонормированный базис линеала $\mathrm{Vect}_\Pi(2)$ и в этом базисе оператор $\boldsymbol{U}'$ имеет некоторую матрицу U'.

Что же касается вектора $\boldsymbol{e}_3$, то в силу линейности оператора $\boldsymbol{U}$ он будет обладать тем же свойством, что и коллинеарный ему вектор $\boldsymbol{x}_0$, т. е. для него будет справедливо равенство

$$\boldsymbol{U}\boldsymbol{e}_3 = \lambda\boldsymbol{e}_3,$$

где $\lambda = \pm 1$.

Отсюда непосредственно вытекает, что матрица U оператора $\boldsymbol{U}$ в базисе $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ имеет вид

$$\begin{pmatrix} & U' & & 0 \\ & & & 0 \\ 0 & 0 & & \lambda \end{pmatrix}.$$

Следовательно,

$$|U| = |U'| \cdot \lambda,$$

и потому

$$|U'| = |U| \cdot \lambda$$

(напомним, что $\lambda = \pm 1$ и, следовательно, $\lambda^{-1} = \lambda$).

Случай 1. $|U'| = 1$. В этом случае $\lambda = 1$, если оператор $\boldsymbol{U}$ сохраняет ориентации ($|U| = 1$), и $\lambda = -1$, если оператор $\boldsymbol{U}$ обращает ориентации ($|U| = -1$). Поэтому за вектор $\boldsymbol{e}$ мы можем принять вектор $\boldsymbol{x}_0$ (или любой вектор, ему коллинеарный, скажем, вектор $\boldsymbol{e}_3$).

Случай 2. $|U'| = -1$. В этом случае оператор $\boldsymbol{U}'$ является линейным изометричным оператором на плоскости, обращающим ориентации. Поэтому (см. выше следствие к лемме 2) в $\mathrm{Vect}_{\Pi}(2)$ существует ортонормированный базис $\boldsymbol{f}_1$, $\boldsymbol{f}_2$, в котором матрица этого оператора имеет вид

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

т. е. такой, что

$$\boldsymbol{U}'\boldsymbol{f}_1 = \boldsymbol{f}_1,$$
$$\boldsymbol{U}'\boldsymbol{f}_2 = -\boldsymbol{f}_2.$$

Для завершения доказательства остается за вектор $\boldsymbol{e}$ принять вектор $\boldsymbol{f}_1$, если оператор $\boldsymbol{U}$ сохраняет ориентации, и вектор $\boldsymbol{f}_2$, если оператор $\boldsymbol{U}$ обращает ориентации.

Тем самым лемма 4 полностью доказана.

Если мы теперь рассмотрим ортонормированный базис

$$\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3,$$

последний вектор $\boldsymbol{e}_3$ которого коллинеарен вектору $\boldsymbol{e}$, то, по сказанному выше, матрица оператора $\boldsymbol{U}$ в этом базисе будет иметь вид

$$\begin{pmatrix} & U' & & 0 \\ & & & 0 \\ 0 & 0 & & 1 \end{pmatrix},$$

если оператор $\boldsymbol{U}$ сохраняет ориентации, и вид

$$\begin{pmatrix} & U' & & 0 \\ & & & 0 \\ 0 & 0 & & -1 \end{pmatrix},$$

если оператор $\boldsymbol{U}$ обращает ориентации, где U' — матрица некоторого сохраняющего ориентации изометричного оператора в

$\mathrm{Vect}_{\pi}(2)$, т. е. некоторая собственная ортогональная матрица

$$U' = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}.$$

Тем самым мы доказали

Следствие. *Для любого линейного изометричного оператора* U *в* $\mathrm{Vect}(3)$ *существует ортонормированный базис, в котором матрица оператора имеет вид*

$$\begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad -\pi < \alpha \leqslant \pi,$$

если оператор сохраняет ориентации, и вид

$$\begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & \alpha & -1 \end{pmatrix}, \qquad -\pi < \alpha \leqslant \pi,$$

если оператор обращает ориентации.

Из этого следствия непосредственно вытекает

Лемма 5. *Для любого сохраняющего ориентации изометричного оператора* U *в пространстве оператор* $E - U$ *необратим.*

Доказательство. Согласно следствию существует ортонормированный базис, в котором матрица оператора U имеет вид

$$\begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Поэтому оператор $E - U$ будет иметь в этом базисе матрицу

$$\begin{pmatrix} 1-\cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & 1-\cos\alpha & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

и, следовательно, будет необратим.

Обратим внимание на аналогию между леммами 5 и 2: отличие состоит только в том, что лемма 2 относится к операторам, обращающим ориентации, а лемма 5 — к операторам, сохраняющим ориентации. Естественно ожидать, что в пространстве справедлива и лемма, аналогичная лемме 1 (и относящаяся, конечно, к изометричным операторам, обращающим ориентации). Тем не менее оказывается, что такая лемма неверна: в прост-

ранстве существуют обращающие ориентации изометричные операторы U, для которых оператор $E - U$ необратим. Примером может служить оператор, имеющий в некотором ортонормированном базисе матрицу вида

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \qquad (3)$$

Однако оказывается, что такого рода операторы являются единственными исключениями.

Лемма 6. *Для любого обращающего ориентации изометричного оператора U в* Vect(3), *обладающего тем свойством, что ни в одном ортонормированном базисе его матрица не имеет вида* (3), *оператор $E - U$ обратим.*

Доказательство. Согласно следствию к лемме 4 существует ортонормированный базис, в котором оператор U имеет матрицу

$$\begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

и, следовательно, оператор $E - U$ — матрицу

$$\begin{pmatrix} 1-\cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & 1-\cos\alpha & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Поскольку

$$\begin{vmatrix} 1-\cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & 1-\cos\alpha & 0 \\ 0 & 0 & 2 \end{vmatrix} = 4(1-\cos\alpha)$$

и поскольку, согласно условию, $\alpha \neq 0$, лемма 6 тем самым полностью доказана.

Глава 2

МЕТОД КООРДИНАТ

§ 1. КООРДИНАТЫ НА ПРЯМОЙ, В ПЛОСКОСТИ И В ПРОСТРАНСТВЕ

1. Аффинные координаты

Пусть на прямой, в плоскости или в пространстве задана некоторая точка O.

Определение 1. Для любой точки A вектор $\overrightarrow{OA}$ называется ее *радиус-вектором* (относительно точки O). Сама точка O называется при этом *начальной* (или *нулевой*) точкой.

Ясно, что, сопоставив каждой точке ее радиус-вектор, мы получим биективное соответствие

$$\text{точка } A \longmapsto \text{ее радиус-вектор } \boldsymbol{r}_A \tag{1}$$

между множеством всех точек (прямой, плоскости или пространства) и множеством всех векторов (т. е. соответствующим линеалом $\mathrm{Vect}(n)$, где, как всегда, $n = 1, 2$ или 3).

Подчеркнем, что соответствие (1) зависит от выбора точки O.

Это соответствие обладает, очевидно, тем свойством, что

произвольный вектор $\boldsymbol{x} = \overrightarrow{AB}$ является разностью радиус-вектора $\boldsymbol{r}_B$ его конца B и радиуса-вектора $\boldsymbol{r}_A$ его начала A:

$$\boldsymbol{x} = \boldsymbol{r}_B - \boldsymbol{r}_A. \tag{2}$$

Действительно,

$$\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA}.$$

Определение 2. Упорядоченная система $n+1$ точек $OE_1 \ldots E_n$ общего положения ($n = 1, 2, 3$) называется *аффинным координатным репером* (на прямой, на плоскости или в пространстве).

Таким образом,

на прямой репером является пара OE_1 различных точек,

на плоскости — тройка OE_1E_2 неколлинеарных точек,

в пространстве — четверка $OE_1E_2E_3$ некомпланарных точек.

Замечание об обозначениях. Реперы принято записывать, не разделяя точки никакими знаками препинания (и без скобок).

Так, например, хотя стандартным обозначением упорядоченной тройки точек является запись вида (O, E_1, E_2), но когда мы эту тройку рассматриваем как репер, она обозначается символом OE_1E_2.

Заметим, что тем же символом OE_1E_2 мы обозначаем плоскость, проходящую через точки O, E_1 и E_2. Эта двусмысленность обозначений ни к каким недоразумениям, как правило, не приводит.

Каждый репер $OE_1 \dots E_n$ однозначно определяет некоторый базис линеала Vect (n), а именно, базис, состоящий из векторов

$$\boldsymbol{e}_1 = \overrightarrow{OE}_1, \ \dots, \ \boldsymbol{e}_n = \overrightarrow{OE}_n. \tag{3}$$

Обратно, задание точки O и произвольного базиса $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ линеала Vect(n) однозначно определяет некоторый репер (а именно, репер $OE_1 \dots E_n$, связанный с векторами $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ формулами (3)).

На этом основании аффинными координатными реперами называются также системы вида $O\boldsymbol{e}_1 \dots \boldsymbol{e}_n$, состоящие из некоторой точки O и базиса $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ линеала Vect (n).

Понимаемые в этом смысле реперы имеют на прямой вид $O\boldsymbol{e}_1$, на плоскости — вид $O\boldsymbol{e}_1\boldsymbol{e}_2$ и в пространстве — вид $O\boldsymbol{e}_1\boldsymbol{e}_2\boldsymbol{e}_3$.

Пусть нам задан произвольный репер $O\boldsymbol{e}_1 \dots \boldsymbol{e}_n$. Тогда радиус-вектор $\overrightarrow{OA}$ каждой точки A мы можем разложить по базису $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$:

$$\overrightarrow{OA} = x^1\boldsymbol{e}_1 + \ \dots \ + x^n\boldsymbol{e}_n. \tag{4}$$

Определение 3. Определяемые формулой (4) числа $x^1, \dots, x^n$ называются *(аффинными) координатами* точки A в репере $O\boldsymbol{e}_1 \dots \boldsymbol{e}_n$. Для любого вектора $\boldsymbol{x}$ его координатами в репере $O\boldsymbol{e}_1 \dots \boldsymbol{e}_n$ называются его координаты в базисе $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$.

Согласно формуле (2)

координаты $x^1, \dots, x^n$ вектора $\boldsymbol{x} = \overrightarrow{AB}$ выражаются через координаты его концов по формулам

$$x^1 = x_2^1 - x_1^1, \ \dots, \ x^n = x_2^n - x_1^n, \tag{5}$$

где $x_1^1, \dots, x_1^n$ — координаты точки A, а $x_2^1, \dots, x_2^n$ — координаты точки B.

Строчка $(x^1, \dots, x^n)$ координат произвольной точки представляет собой элемент множества $\mathbb{R}^n$, состоящего из n-членных последовательностей вещественных чисел. Таким образом, задание репера $O\boldsymbol{e}_1 \dots \boldsymbol{e}_n$ определяет некоторое (очевидно, биективное) отображение

$$\text{«точка»} \longmapsto \text{«строчка ее координат»} \tag{6}$$

прямой, плоскости или пространства на множество $\mathbb{R}^n$ (с соответствующим $n = 1, 2, 3$).

Определение 4. Биективное отображение (6) называется *аффинной координатной системой*, определенной репером $Oe_1 \ldots e_n$. Эту систему принято обозначать символом $Ox_1 \ldots x_n$.

На прямой имеется только одна координата x^1, на плоскости — две координаты x^1, x^2 и в пространстве — три координаты x^1, x^2, x^3. Координата x^1 обычно обозначается (во всех трех случаях) символом x и называется *абсциссой* точки A. Аналогично, координата x^2 обозначается символом y и называется *ординатой*, а координата x^3 — символом z и называется *аппликатой*.

Когда координатная система фиксирована, точку A с координатами $x^1, \ldots, x^n$ бывает удобно обозначать символом $A(x^1, \ldots, x^n)$ (на прямой — символом $A(x)$, на плоскости — $A(x, y)$, в пространстве — $A(x, y, z)$).

Прямая, на которой задана аффинная координатная система Ox, часто обозначается тем же символом Ox. Аналогично, символ Oxy может обозначать как аффинную координатную систему на плоскости, так и саму эту плоскость, а символ $Oxyz$ — как аффинную координатную систему в пространстве, так и само это пространство (с заданной в нем аффинной координатной системой).

Обратим внимание на то, что

прямая, плоскость или пространство, для которых задана аффинная координатная система, автоматически ориентированы.

Именно, ориентация в них задается базисом $e_1, \ldots, e_n$.

В частности, мы видим, что прямая, на которой задана аффинная координатная система, является осью.

Определение 5. Две аффинные координатные системы называются *одноименными*, если одноименны (в смысле определения 4 п. 1 § 4 гл. 1) их реперы $OE_1 \ldots E_n$ и $O'E_{1'} \ldots E_n$ или (что равносильно) если одноименны базисы $e_1, \ldots, e_n$ и $e_{1'}, \ldots, e_{n'}$.

Ясно, что

отношение одноименности аффинных координатных систем является отношением эквивалентности, и соответствующие классы эквивалентности естественным образом отождествляются с ориентациями (прямой, плоскости или пространства).

Определение 6. Начальная точка O аффинной координатной системы Ox на прямой разбивает эту прямую на две полупрямые, называемые *полуосями*. Точки одной полуоси характеризуются условием $x > 0$ (и эта полуось называется *положительной*), а точки другой полуоси — условием $x < 0$ (и эта полуось называется *отрицательной*). Сама точка O характеризуется равенством $x = 0$.

Пусть на плоскости задана аффинная координатная система Oxy с репером e_1, e_2. Точка O и вектор e_1 составляют, очевидно, координатный репер Oe_1 прямой, проходящей через точку O параллельно вектору e_1. Эта прямая, снабженная соответствующей ориентацией, называется *первой координатной осью* или *осью абсцисс* рассматриваемой аффинной координатной системы. Аналогично определяется *вторая координатная ось* (или *ось ординат*) с координатным репером Oe_2.

Координата x является координатой на оси абсцисс в репере Oe_1, а координата y — координатой на оси ординат в репере Oe_2. Поэтому, в соответствии с введенными выше обозначениями, ось абсцисс обычно обозначается символом Ox, а ось ординат — символом Oy.

Ось абсцисс разбивает плоскость на две полуплоскости. Точки одной полуплоскости характеризуются условием $y > 0$, а другой — условием $y < 0$. Принято полуплоскость $y > 0$ называть *верхней* полуплоскостью, а полуплоскость $y < 0$ — *нижней* полуплоскостью. Для точек оси абсцисс $y = 0$.

Аналогично, ось ординат разбивает плоскость на *правую* полуплоскость, для которой $x > 0$, и *левую* полуплоскость, для которой $x < 0$. Для точек самой оси ординат $x = 0$.

Координатные оси вместе разбивают плоскость на четыре части, называемые *квадрантами*. Точки одного квадранта, называемого *первым*, характеризуются условиями $x > 0$, $y > 0$, точки следующего квадранта, называемого *вторым*, — условиями $x < 0$, $y > 0$, точки *третьего* квадранта — условиями $x < 0$, $y < 0$ и, наконец, точки *четвертого* квадранта — условиями $x > 0$, $y < 0$:

коорд. / квадр.	x	y
I	+	+
II	−	+
III	−	−
IV	+	−

Аналогично, в пространстве определяются (по отношению к заданной аффинной координатной системе $Oxyz$ с репером $Oe_1e_2e_3$) *ось абсцисс* Ox с репером Oe_1, *ось ординат* Oy с репером Oe_2 и *ось аппликат* (или *третья координатная ось*) Oz с репером Oe_3.

Координата x является координатой на оси Ox в репере Oe_1, координата y — координатой на оси Oy в репере Oe_2, а координата z — координатой на оси Oz в репере Oe_3.

Плоскости, проходящие через пары координатных осей, называются *координатными плоскостями.* Всего имеются три координатные плоскости Oxy, Oxz и Oyz. Репер Oe_1e_2 является координатным репером на плоскости Oxy, репер Oe_1e_3 — координатным репером на плоскости Oxz и репер Oe_2e_3 — координатным репером на плоскости Oyz.

Плоскость Oxy разбивает пространство на два полупространства. Точки одного полупространства характеризуются условием $z > 0$, а точки другого — условием $z < 0$. Точки самой плоскости Oxy характеризуются условием $z = 0$.

Аналогично, плоскость Oxz разбивает пространство на два полупространства $y > 0$ и $y < 0$, а плоскость Oyz — на полупространства $x > 0$ и $x < 0$. Для точек первой плоскости $y = 0$, а для точек второй плоскости $x = 0$.

Координатные плоскости вместе разбивают пространство на восемь частей, называемых *октантами.* Две точки тогда и только тогда принадлежат одному октанту, когда все три их координаты имеют одинаковые знаки. Октанты нумеруются в соответствии со следующим распределением знаков координат:

окт. \ координ.	x	y	z
I	+	+	+
II	−	+	+
III	−	−	+
IV	+	−	+
V	+	+	−
VI	−	+	−
VII	−	−	−
VIII	+	−	−

Замечание 1. Обратим внимание на то, что, например, координата x является координатой не только на оси Ox, но и на любой прямой, параллельной этой оси. Соответствующий репер имеет вид $O'e_1$, где O' — точка пересечения этой прямой с осью Oy (на плоскости) или с плоскостью Oyz (в пространстве).

Более того, легко видеть, что координата x является координатой и на любой прямой, не параллельной оси Oy (или плоскости Oyz). Соответствующий репер имеет вид $O'e_1'$, где e_1' — проекция вектора e_1 на рассматриваемую прямую параллельно оси Oy (плоскости Oyz).

Аналогичные утверждения справедливы, конечно, и для координат y и z.

В пространстве, скажем, координаты x, y являются координатами на любой плоскости, параллельной координатной плос-

кости Oxy, и вообще на любой плоскости, не параллельной оси Oz.

Введем еще одно полезное определение:

Определение 7. Точка E координатного репера OE на прямой называется его *единичной точкой* (она имеет координату $x = 1$). В соответствии с этим точки E_1, E_2 (и E_3) координатного репера OE_1E_2 (репера $OE_1E_2E_3$) на плоскости (в пространстве) называются *единичными точками его координатных осей.* Таким образом,

E_1 — единичная точка оси абсцисс,
E_2 — единичная точка оси ординат,
E_3 — единичная точка оси аппликат.

Точка O называется (на прямой, в плоскости и в пространстве) *нулевой* (или *начальной*) точкой координатного репера (ср. определение 1).

2. Замена аффинных координат

Пусть $Oe_1 \ldots e_n$ и $O'e_{1'} \ldots e_{n'}$ — два координатных репера ($n = 1, 2, 3$) и пусть $x^1, \ldots, x^n$ — координаты относительно репера $Oe_1 \ldots e_n$, а $x^{1'}, \ldots, x^{n'}$ — координаты относительно репера $O'e_{1'} \ldots e_{n'}$.

В этом пункте мы найдем формулы, выражающие координаты $x^{1'}, \ldots, x^{n'}$ через координаты $x^1, \ldots, x^n$.

Сначала мы рассмотрим частный случай, когда $O' = O$, т. е. когда рассматриваемые реперы имеют одну и ту же начальную точку и отличаются только векторными базисами.

В этом случае каждая точка A имеет в обоих реперах один и тот же радиус-вектор $\overrightarrow{OA}$ и потому ее координаты $x^1, \ldots, x^n$ и $x^{1'}, \ldots, x^{n'}$ являются не чем иным, как координатами одного и того же вектора $\boldsymbol{x} = \overrightarrow{OA}$ в двух различных базисах. Поэтому (см. п. 7 § 3 гл. 1) эти координаты связаны соотношениями вида

$$x^{1'} = c_1^{1'}x^1 + \ldots + c_n^{1'}x^n,$$
$$\ldots\ldots\ldots\ldots$$
$$x^{n'} = c_1^{n'}x^1 + \ldots + c_n^{n'}x^n$$

или в матричной записи

$$x' = Cx,$$

где $C = (c_i^{i'})$ — матрица перехода от «нового» базиса $\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}$ к «старому» базису $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, а x' и x — одностолбцовые матрицы, составленные из координат $x^{1'}, \ldots, x^{n'}$ и $x^1, \ldots, x^n$ соответственно. (Заметим, что в п. 7 § 3 гл. 1 символом C обозначалась не эта, а обратная матрица.)

Пусть теперь, наоборот, реперы $Oe_1 \dots e_n$ и $O'e_{1'} \dots e_{n'}$ отличаются лишь точками O и O' (так что $e_{1'} = e_1, \dots, e_{n'} = e_n$). Тогда для любой точки A

$$\overrightarrow{O'A} = \overrightarrow{OA} + \overrightarrow{O'O}$$

и потому ее координаты $x^1, \dots, x^n$ в репере $Oe_1 \dots e_n$ связаны с ее координатами $x^{1'}, \dots, x^{n'}$ в репере $O'e_1 \dots e_n$ соотношениями

$$\begin{aligned} x^{1'} &= x^1 + a^1, \\ &\dots\dots \\ x^{n'} &= x^n + a^n, \end{aligned}$$

где $a^1, \dots, a^n$ — координаты «старого» начала O в «новом» координатном репере $O'e_1 \dots e_n$. В матричной форме эти соотношения имеют вид

$$x' = x + a,$$

где a — одностолбцовая матрица

$$a = \begin{pmatrix} a^1 \\ \cdot \\ \cdot \\ \cdot \\ a^n \end{pmatrix}.$$

В общем случае переход от репера $Oe_1 \dots e_n$ к реперу $O'e_{1'} \dots e_{n'}$ может быть осуществлен в два этапа: сначала мы меняем базис, т. е. переходим от репера $Oe_1 \dots e_n$ к реперу $Oe_{1'} \dots e_{n'}$, а затем меняем точку O, т. е. переходим от репера $Oe_{1'} \dots e_{n'}$ к реперу $O'e_{1'} \dots e_{n'}$. Ясно, что соответствующие формулы перехода являются комбинациями полученных выше формул перехода и имеют вид

$$\begin{aligned} x^{1'} &= c_1^{1'}x^1 + \dots + c_n^{1'}x^n + a^{1'}, \\ &\dots\dots\dots\dots\dots\dots \\ x^{n'} &= c_1^{n'}x^1 + \dots + c_n^{n'}x^n + a^{n'}, \end{aligned} \tag{1}$$

или в матричной записи —

$$x' = Cx + a, \tag{1'}$$

где $C = (c_i^{i'})$ — матрица перехода от базиса $e_{1'}, \dots, e_{n'}$ к базису $e_1, \dots, e_n$, а $a = (a^{i'})$ — одностолбцовая матрица, составленная из координат «старого» начала O в «новом» репере $O'e_{1'} \dots e_{n'}$.

На языке алгебры это означает, что

переход от одной аффинной координатной системы к другой выражается неоднородным линейным преобразованием координат.

Замечание 1. Для конкретных значений n формулы (1) обычно пишутся в более наглядных обозначениях. Например, при $n = 2$ (случай плоскости) можно писать

$$x' = a_1 x + b_1 y + c_1,$$
$$y' = a_2 x + b_2 y + c_2$$

(здесь, естественно, x, y — координаты в «старом» репере, а x', y' — в «новом»).

Определение 1. Прямоугольную матрицу (C, a), получающуюся приписыванием к матрице C столбца a, мы будем называть *матрицей перехода* от аффинной координатной системы с репером $Oe_1 \ldots e_n$ к аффинной координатной системе с репером $O'e_{1'} \ldots e_{n'}$.

Ясно, что

прямоугольная матрица (размера $n \times (n+1)$) тогда и только тогда является матрицей перехода от одной аффинной координатной системы к другой, когда квадратная матрица, получающаяся вычеркиванием ее последнего столбца, невырождена.

Пусть нам даны три системы аффинных координат

(i) $x^1, \ldots, x^n$,

(ii) $x^{1'}, \ldots, x^{n'}$,

(iii) $x^{1''}, \ldots, x^{n''}$

и пусть

$$x' = Cx + a, \qquad x'' = C'x' + a'$$

— формулы перехода от первой системы ко второй и от второй к третьей. Тогда

$$x'' = C'(Cx + a) + a' = C'Cx + (C'a + a').$$

Тем самым доказано следующее

Предложение 1. *Если переход от координатной системы* (i) *к координатной системе* (ii) *описывается матрицей* (C, a), *а переход от координатной системы* (ii) *к координатной системе* (iii) *описывается матрицей* (C', a'), *то переход от координатной системы* (i) *к координатной системе* (iii) *описывается матрицей* $(C'C, C'a + a')$.

В частности, отсюда вытекает

Следствие. *Если переход от координатной системы $x^1, \ldots, x^n$ к координатной системе $x^{1'}, \ldots, x^{n'}$ описывается матрицей (C, a), то обратный переход от координатной системы $x^{1'}, \ldots, x^{n'}$ к координатной системе $x^1, \ldots, x^n$ описывается матрицей* $(C^{-1}, -C^{-1}a)$.

Действительно, пусть (C', a') — матрица, описывающая переход от координатной системы $x^{1'}, \ldots, x^{n'}$ к координатной

системе $x^1, \ldots, x^n$. Тогда согласно предложению 1 переход от системы $x^1, \ldots, x^n$ к самой себе будет описываться матрицей $(C'C, C'a + a')$. Но, с другой стороны, этот переход описывается матрицей $(E, 0)$. Следовательно,

$$C'C = E, \quad C'a + a' = 0.$$

Поэтому

$$C' = C^{-1}, \quad a' = -C'a = -C^{-1}a.$$

3. Деление отрезка в данном отношении

Используя координатные системы, мы можем задавать точки числами и любое геометрическое построение свести к вычислению над этими числами.

Рассмотрим, например, так называемую задачу о делении отрезка в данном отношении.

Пусть $\overline{A_0A_1}$ — произвольный отрезок прямой α. В элементарной геометрии говорят, что точка A отрезка $\overline{A_0A_1}$ делит его в отношении k, если отношение длин отрезков $\overline{A_0A}$ и $\overline{AA_1}$ равно k. Таким образом, $0 < k < \infty$.

Заметим, что фактически отрезок $\overline{A_0A_1}$ считается здесь направленным (ибо при перестановке точек A_0 и A_1 отношение k перейдет в отношение $1/k$). Таким образом, речь здесь на самом деле идет о делении в данном отношении направленного отрезка $\overrightarrow{A_0A_1}$ (предполагаемого невырожденным).

С другой стороны, поскольку векторы $\overrightarrow{A_0A}$ и $\overrightarrow{AA_1}$ одноименны (для точек A отрезка $\overline{A_0A_1}$), отношение их длин k совпадает с их отношением $\overrightarrow{A_0A} : \overrightarrow{AA_1}$. Поскольку это отношение определено для любой точки A прямой α (отличной от точки A_1), оказывается удобным ввести следующее общее

Определение 1. Говорят, что точка $A \neq A_1$ прямой α *делит* невырожденный направленный отрезок $\overrightarrow{A_0A_1}$ *в отношении* k, если

$$\overrightarrow{A_0A} : \overrightarrow{AA_1} = k. \tag{1}$$

При этом по-прежнему для точек A отрезка $\overline{A_0A_1}$ имеет место соотношение $0 < k < \infty$. В то же время соотношение $k < 0$ характеризует точки A, лежащие вне отрезка $\overline{A_0A_1}$; для точек A, предшествующих точке A_0 (в ориентации, в которой $A_0 < A_1$), имеют место неравенства $-1 < k < 0$, а для точек A, следующих за точкой A_1, — неравенства $-\infty < k < -1$.

При приближении точки A к точке A_1 по отрезку $\overline{A_0A_1}$ (т. е. так, что все время $A < A_1$) отношение k стремится к $+\infty$, а при приближении точки A к точке A_1 с другой стороны отношение k стремится к $-\infty$. Можно поэтому условно гово-

рить, что точка A_1 делит отрезок $\overrightarrow{A_0A_1}$ в отношении ∞ (без знака!).

При $A = A_0$ отношение k равно нулю.

Поставим задачу о вычислении аффинных координат точки A по известным координатам точек A_0 и A_1 (и числу k).

Пусть O — начальная точка координатного репера. Тогда

$$\overrightarrow{A_0A} = \overrightarrow{OA} - \overrightarrow{OA_0}, \quad \overrightarrow{AA_1} = \overrightarrow{OA_1} - \overrightarrow{OA},$$

т. е.

$$\overrightarrow{A_0A} = \boldsymbol{r} - \boldsymbol{r}_0, \quad \overrightarrow{AA_1} = \boldsymbol{r}_1 - \boldsymbol{r},$$

где

$$\boldsymbol{r} = \overrightarrow{OA}, \quad \boldsymbol{r}_0 = \overrightarrow{OA_0}, \quad \boldsymbol{r}_1 = \overrightarrow{OA_1}.$$

Поэтому соотношение (1), определяющее точку A, мы можем записать в следующем виде:

$$(\boldsymbol{r} - \boldsymbol{r}_0) : (\boldsymbol{r}_1 - \boldsymbol{r}) = k.$$

Отсюда

$$\boldsymbol{r} = \frac{\boldsymbol{r}_0 + k\boldsymbol{r}_1}{1 + k}.$$

Переходя от векторов к их координатам и считая, что $A = A(x, y, z)$, $A_0 = A_0(x_0, y_0, z_0)$, $A_1 = A_1(x_1, y_1, z_1)$, мы получаем (для случая пространства) следующие окончательные формулы, дающие решение нашей задачи:

$$x = \frac{x_0 + kx_1}{1 + k}, \quad y = \frac{y_0 + ky_1}{1 + k}, \quad z = \frac{z_0 + kz_1}{1 + k}. \tag{2}$$

В случае плоскости следует сохранить лишь две первые формулы (2), а для случая прямой — только первую.

По определению, для середины отрезка $\overline{A_0A_1}$ мы имеем $k = 1$. Поэтому

середина отрезка с концами $A_0(x_0, y_0, z_0)$ *и* $A_1(x_1, y_1, z_1)$ *имеет координаты*

$$x = \frac{x_0 + x_1}{2}, \quad y = \frac{y_0 + y_1}{2}, \quad z = \frac{z_0 + z_1}{2}.$$

На плоскости сохраняются лишь две первые формулы, а на прямой — только первая.

4. Прямоугольные координаты

Определение 1. Аффинный координатный репер $Oe_1 \ldots e_n$ называется *евклидовым*, если базис

$$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$$

ортонормирован. Аффинная координатная система, определенная прямоугольным репером, называется *евклидовой координ-*

натной системой, а соответствующие координаты $x^1, \ldots, x^n$ — *евклидовыми*.

При $n > 1$ евклидов репер обычно называется *прямоугольным репером*, евклидовы координаты — *прямоугольными координатами*, а евклидовы координатные системы — *системами прямоугольных координат*.

Евклидовы координаты должны быть известны читателю (по крайней мере на прямой и плоскости) из школьного курса. Поэтому мы на них подробно останавливаться не будем и ограничимся некоторыми замечаниями принципиального характера.

Поскольку евклидовы координаты являются частным случаем аффинных, все сказанное выше (в п. 1—3) автоматически применимо и к евклидовым координатам. Следует только отметить, что если, скажем, в пространстве заданы евклидовы координаты x, y, z, то, например, координаты x, y также будут евклидовыми координатами лишь на плоскостях, параллельных координатной плоскости Oxy (на любой другой плоскости, не перпендикулярной оси Oz, они будут лишь аффинными координатами).

Переход от одной евклидовой координатной системы к другой описывается формулами вида

$$x' = Cx + a,$$

где C — произвольная ортогональная матрица. Эта матрица тогда и только тогда является собственной ортогональной матрицей, когда рассматриваемые координатные системы одноименны.

При $n = 1$ (случай прямой) одноименные евклидовы координатные системы могут отличаться лишь выбором точки O, так что переход от одной такой системы к другой описывается формулой

$$x' = x + a.$$

При $n = 2$ (случай плоскости) переход от евклидовой координатной системы x, y к одноименной евклидовой системе x', y' описывается формулами

$$x' = x\cos\alpha - y\sin\alpha + x_0,$$
$$y' = x\sin\alpha + y\cos\alpha + y_0,$$

где α — угол от оси Ox к оси Ox'.

Прямоугольные координатные системы особенно удобны для решения метрических задач. Например, в любой такой системе расстояние d (в пространстве) между точками $A_1(x_1, y_1, z_1)$ и $A_2(x_2, y_2, z_2)$ выражается формулой

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

Действительно, как мы знаем, эта формула дает длину вектора $\overrightarrow{A_1A_2}$, равную, по определению, расстоянию между точками A_1 и A_2.

На плоскости соответствующая формула имеет вид

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

а на прямой — вид

$$d = \sqrt{(x_2 - x_1)^2} = |x_2 - x_1|.$$

В произвольных аффинных координатах формулы для расстояния имеют существенно более сложный вид. Например, на плоскости

$$d = \sqrt{g_{11}(x_2 - x_1)^2 + 2g_{12}(x_2 - x_1)(y_2 - y_1) + g_{22}(y_2 - y_1)^2}$$

(ср. формулы для длины вектора в п. 4 § 5 гл. 1).

5. Полярные, сферические и цилиндрические координаты

В ряде задач евклидовой геометрии (характеризующихся «круговой» или «сферической» симметрией) удобно вместо прямоугольных координат использовать так называемые «полярные координаты».

Определение 1. Пусть на плоскости заданы

а) некоторая ориентация,

б) некоторая точка O,

в) некоторая ориентированная прямая α, проходящая через точку O.

Тогда любой вектор $\boldsymbol{r}$ однозначно характеризуется его длиной r и (ориентированным) углом φ от прямой α до вектора $\boldsymbol{r}$. Числа r и φ, построенные для радиус-вектора $\boldsymbol{r} = \overrightarrow{OA}$ произвольной точки A плоскости, называются *полярными координатами* точки A (координата r — *полярным радиусом*, а координата φ — *полярным углом*). Точка O называется *полюсом* системы полярных координат, а прямая α — ее *полярной осью*.

Подчеркнем, что для задания полярных координат нужно, кроме полюса O и полярной оси α, задать также ориентацию плоскости.

Полярные координаты однозначно определены для любой точки $A \neq 0$. Для точки O, по определению, считается, что полярный радиус r равен нулю. Полярного угла φ точка O, по определению, не имеет (впрочем, иногда удобно считать, что этот угол существует, но может принимать любое значение).

Таким образом,

$$0 \leqslant r < \infty$$

и

$$-\pi < \varphi \leqslant \pi$$

(впрочем, часто удобно считать полярный угол φ принимающим любые вещественные значения, но определенным только с точностью до слагаемых, кратных 2π).

Систему прямоугольных координат x, y и систему полярных координат r, φ мы будем называть *согласованными*, если координатная система Oxy определяет данную ориентацию плоскости, точки O для обеих систем совпадают и полярная ось совпадает с осью абсцисс Ox. Ясно, что согласованные системы друг друга однозначно определяют. При этом

$$r = \sqrt{x^2 + y^2},$$

$$\cos\varphi = \frac{x}{\sqrt{x^2 + y^2}}, \quad \sin\varphi = \frac{y}{\sqrt{x^2 + y^2}},$$

и обратно,

$$x = r\cos\varphi,$$

$$y = r\sin\varphi.$$

Это и есть формулы перехода от прямоугольных координат к полярным и обратно.

Чтобы перейти от полярной системы координат к несогласованной с ней прямоугольной (или произвольной аффинной) координатной системе, следует сначала перейти к согласованной прямоугольной системе, а затем воспользоваться формулами перехода от одной прямоугольной системы к другой такой системе.

В пространстве аналогом полярных координат являются так называемые «сферические координаты».

Определение 2. Пусть в пространстве задана

а) некоторая ориентация,

б) некоторая ориентированная плоскость Π,

в) на плоскости Π — некоторая система полярных координат (т. е. полюс O и полярная ось α).

Сферическими координатами произвольной точки A называются

1) ее *полярный радиус* r, т. е. длина радиус-вектора $\overrightarrow{OA}$ (расстояние точки A от точки O); всегда $r \geqslant 0$ и $r = 0$ только при $A = 0$;

2) ее *долгота* φ, т. е. полярный угол ортогональной проекции A_0 точки A на плоскость Π относительно заданной в этой плоскости системы полярных координат; по определению, $-\pi < \varphi \leqslant \pi$;

3) ее *широта* ψ, т. е. угол между вектором $\overrightarrow{OA}$ и его проекцией $\overrightarrow{OA_0}$ на плоскость Π, считаемый положительным, $0 < \psi \leqslant \frac{\pi}{2}$, если вектор $\overrightarrow{OA}$ направлен в положительное полупространство (т. е. если произведение данной ориентации плоскости Π и определяемой этим вектором ориентации прямой OA совпадает с данной ориентацией пространства), и отрицательным, $-\frac{\pi}{2} \leqslant \psi < 0$, — в противном случае (здесь, конечно,

предполагается, что точка A не принадлежит плоскости П; если же это условие не выполнено, то угол ψ равен нулю и разговор о его знаке смысла не имеет). Таким образом, $-\frac{\pi}{2} \leqslant \psi \leqslant \frac{\pi}{2}$.

Сферические координаты однозначно определены для всех точек, не принадлежащих прямой Oz, проходящей через точку O перпендикулярно плоскости П. Для точек же этой прямой координата ψ смысла не имеет.

Система сферических координат определяет *согласованную* с ней систему прямоугольных координат x, y, z, для которой плоскость Oxy совпадает с плоскостью П, прямоугольные координаты x, y на этой плоскости согласованы с имеющимися на ней полярными координатами, и которая определяет данную ориентацию пространства. Легко видеть, что эти координаты x, y, z выражаются через координаты r, φ, ψ по формулам

$$x = r \cos\varphi \cos\psi,$$
$$y = r \sin\varphi \cos\psi,$$
$$z = r \sin\psi.$$

Очевидно, что эти формулы позволяют и, обратно, выразить координаты r, φ, ψ через координаты x, y, z.

Со сферическими координатами тесно связаны *цилиндрические координаты* ρ, φ, z, где φ и z имеют прежний смысл, а ρ представляет собой полярный радиус проекции A_0 точки A на плоскость П. С координатами x, y, z цилиндрические координаты связаны формулами

$$x = \rho \cos\varphi,$$
$$y = \rho \sin\varphi,$$
$$z = z,$$

а со сферическими координатами — формулами

$$r = \sqrt{\rho^2 + z^2},$$
$$\varphi = \varphi,$$
$$\psi = \operatorname{arctg} \frac{z}{\rho}.$$

6. Однородные координаты

Отношение k, в котором точка A прямой $\alpha = A_0A_1$ делит отрезок $\overrightarrow{A_0A_1}$ (см. п. 3), однозначно определяет эту точку. Поэтому его можно рассматривать как своеобразную координату этой точки. Эта координата называется *барицентрической координатой* точки A.

Это название объясняется тем, что, как показывается в механике, центром тяжести («барицентром») двух масс m_0 и m_1, помещенных в точках A_0 и A_1, является точка A, делящая отрезок $\overrightarrow{A_0A_1}$ в отношении $k = m_1/m_0$.

Если на прямой α введена аффинная координатная система Ox, то, как было показано в п. 3,

$$x = \frac{x_0 + kx_1}{1 + k}.$$

Это есть формула перехода от барицентрической координаты к аффинной координате x.

Подчеркнем, что в точке A_1 барицентрическая координата k не имеет определенного числового значения (в этой точке $k = \infty$; см. п. 3). Вместе с тем значению $k = -1$ не отвечает никакая точка прямой.

Чтобы избежать значения $k = \infty$ барицентрической координаты, удобно ввести следующее.

Определение 1. Числа X_0, X_1, связанные с барицентрической координатой k соотношением $k = X_1/X_0$, называются *однородными барицентрическими координатами* на прямой α.

Координаты X_0, X_1 определены только с точностью до пропорциональности. Другими словами, вместе с числами X_0, X_1 однородными координатами той же точки будут и любые числа вида ρX_0, ρX_1, где ρ — произвольное отличное от нуля число. Чтобы подчеркнуть последнее обстоятельство, тот факт, что числа X_0, X_1 являются однородными барицентрическими координатами точки A, обычно записывается формулой $A(X_0 : X_1)$. Так, например, $A_0(1 : 0)$.

Однородными барицентрическими координатами точки A_1, по определению, считаются числа 0, 1 (и любые числа, им пропорциональные). Таким образом, $A_1(0 : 1)$.

Для точки B, являющейся серединой отрезка $\overline{A_0A}$, мы имеем $B(1 : 1)$.

Заметим, что обе координаты X_0, X_1 нулю одновременно никогда не равны. Кроме того, на прямой нет точки с координатами $(-1 : 1)$ (поскольку нет точки с $k = -1$).

С аффинной координатой x однородные барицентрические координаты связаны, очевидно, формулой

$$x = \frac{x_0X_0 + x_1X_1}{X_0 + X_1}.$$

Для общего описания возникшей ситуации удобно ввести

Определение 2. Рассмотрим (для некоторого $n > 0$) множество

$$\mathbb{R}^{n+1} \setminus \{(0, \ldots, 0)\}$$

всех $n + 1$-членных последовательностей

$$(X_0, X_1, \ldots, X_n)$$

вещественных чисел, отличных от *нулевой последовательности*

$$(0, 0 \ldots 0).$$

Два элемента $(X_0, X_1, \ldots, X_n)$ и $(Y_0, Y_1, \ldots, Y_n)$ этого множества назы-

ваются *пропорциональными*, если существует такое (отличное от нуля) число ρ, что

$$Y_0 = \rho X_0, \quad Y_1 = \rho X_1, \ \ldots, \ Y_n = \rho X_n. \tag{1}$$

В этом случае мы будем писать

$$Y_0 : X_0 = Y_1 : X_1 = \ \ldots \ = Y_n : X_n$$

или даже

$$\frac{Y_0}{X_0} = \frac{Y_1}{X_1} = \ \ldots \ = \frac{Y_n}{X_n}. \tag{2}$$

Если все числа $X_0, \ldots, X_n$ отличны от нуля, то формулы (2) являются обычными равенствами между числами. Однако, если среди чисел $X_0, \ldots, X_n$ имеются равные нулю, то эти формулы уже не являются равенствами между числами и представляют собой лишь условную запись соотношений (1).

Заметим, что если в формулах (2) некоторый «знаменатель» X_i равен нулю, то обязательно равен нулю и соответствующий «числитель» Y_i. Обратно, если $Y_i = 0$, то $X_i = 0$.

Ясно, что отношение пропорциональности является на множестве $\mathbb{R}^{n+1} \setminus \{(0, \ldots, 0)\}$ отношением эквивалентности, так что это множество разбивается на классы пропорциональных последовательностей. Класс, содержащий последовательность $(X_0, X_1, \ldots, X_n)$, мы будем обозначать символом

$$(X_0 : X_1 : \ \ldots \ : X_n),$$

а множество всех таких классов (т. е. фактормножество множества $\mathbb{R}^{n+1} \setminus \{(0, \ldots, 0)\}$ по отношению пропорциональности) — символом $\mathbb{RP}^n$.

Нас будут интересовать в основном множества $\mathbb{RP}^1$, $\mathbb{RP}^2$ и $\mathbb{RP}^3$. По причинам, которые будут ясны в дальнейшем, множество $\mathbb{RP}^1$ называется *стандартной* (или *арифметической*) *проективной прямой*, множество $\mathbb{RP}^2$ — *стандартной* (*арифметической*) *проективной плоскостью*, а множество $\mathbb{RP}^3$ — *стандартным* (*арифметическим*) *проективным пространством*.

Множества $\mathbb{RP}^n$ устроены очень интересным и поучительным образом. Мы лишены здесь возможности исследовать их сколько-нибудь подробно. Скажем лишь несколько слов о множестве $\mathbb{RP}^1$ классов $(X_0 : X_1)$ пропорциональных пар чисел.

Все пары вида $(0, X_1)$, $X_1 \neq 0$, очевидно, пропорциональны (скажем, паре $(0, 1)$), т. е. определяют один элемент $(0 : 1)$ множества $\mathbb{RP}^1$. Аналогично, $(X_0 : 0) = (1 : 0)$.

Любой паре (X_0, X_1) с $X_0 \neq 0$ мы можем сопоставить число $x = \frac{X_1}{X_0}$. Ясно, что это число зависит только от класса $(X_0 : X_1)$ пары (X_0, X_1), так что соответствие $(X_0 : X_1) \longmapsto \frac{X_1}{X_0}$ корректно определено. Оно является биективным соответствием между множеством $\mathbb{RP}^1 \setminus \{(0 : 1)\}$ всех элементов множества $\mathbb{RP}^1$, отличных от элемента $(0 : 1)$, и множеством $\mathbb{R}$ всех вещественных чисел. Отождествив элемент $(X_0 : X_1)$ с числом $\frac{X_1}{X_0}$, мы можем, таким образом, считать, что множество $\mathbb{RP}^1$ получено из множества $\mathbb{R}$ присоединением еще одного элемента $(0 : 1)$ (играющего роль «бесконечности»).

Впрочем, элемент $(X_0 : X_1)$ можно отождествить и с числом $\frac{X_0}{X_1}$ (если, конечно, $X_1 \neq 0$). Тогда $\mathbb{R}P^1$ будет получаться из $\mathbb{R}$ присоединением элемента $(1 : 0)$. Вообще, удалив из $\mathbb{R}P^1$ произвольный элемент $(A_0 : A_1)$, мы можем множество всех остальных элементов отождествить с множеством $\mathbb{R}$, сопоставив элементу $(X_0 : X_1)$ число

$$\frac{A_0X_0 + A_1X_1}{A_0X_1 - A_1X_0}.$$

Таким образом, все элементы множества $\mathbb{R}P^1$ по существу совершенно равноправны.

Наглядно множество $\mathbb{R}P^1$ можно представлять себе в виде замкнутой линии (наподобие окружности), получающейся из прямой линии ее «замыканием в бесконечности».

Возвращаясь к барицентрическим координатам, мы теперь можем сказать, что

соответствие

«точка» $\longmapsto$ *«ее однородные барицентрические координаты»*

определяет биективное отображение прямой α *на множество* $\mathbb{R}P^1$, *из которого удален элемент* $(-1 : 1)$.

Тот факт, что элемент $(-1 : 1)$ играет исключительную роль, весьма затрудняет использование барицентрических координат, вынуждая к многочисленным оговоркам. Чтобы избежать этих оговорок, целесообразно «расширить» прямую, присоединив к ней некоторую фиктивную «несобственную точку» с барицентрическими координатами $(-1 : 1)$ (эту точку следует мыслить расположенной на прямой «в бесконечности», причем следует считать, что мы получаем одну и ту же несобственную точку при удалении по прямой как в одном, так и в другом направлении).

К вопросу о несобственной точке возможны три различных подхода, которые мы рассмотрим по степени их «решительности».

1°. Можно считать, что введение несобственной точки носит чисто «лингвистический» характер, способствуя лишь упрощению и облегчению формулировок. При этом, если при решении какой-нибудь задачи возникла в качестве ответа несобственная точка (например, при вычислении в барицентрических координатах получились координаты $(-1 : 1)$), то это попросту означает, что задача решения не имеет.

2°. Более решительная точка зрения состоит в том, что мы считаем несобственные точки столь же «законными», как и обыкновенные (собственные) точки. Конечно, это будет уже означать изменение предмета исследования: вместо обыкновенной прямой мы будем фактически изучать новый геометрический объект — *расширенную прямую*, существенно отличающийся по своим свойствам от привычной прямой. (Например, расширенная прямая, в отличие от обычной прямой, является замкнутой линией.) Другими словами, мы будем иметь дело уже с совсем другой «геометрией».

3°. Наконец, самая радикальная точка зрения состоит в том, что несобственная точка не только добавляется к обыкновенным точкам, но она считается совершенно равноправной с обыкновенными точками, так что сам термин «несобственная точка» теряет смысл. В этом случае расширенная прямая называется *проективной прямой*, а ее геометрия — *проективной геометрией*.

В дальнейшем мы еще раз будем возвращаться к затронутым здесь идеям и разовьем их более полно и систематично.

Наиболее общим образом «однородные» координаты $(X_0 : X_1)$ на прямой вводятся формулами вида

$$\begin{aligned} \rho X_0 &= a_0 x + b_0, \\ \rho X_1 &= a_1 x + b_1, \end{aligned} \tag{3}$$

где x — некоторая аффинная координата на прямой; a_0, b_0, a_1, b_1 — такие числа, что

$$\begin{vmatrix} a_0 & b_0 \\ a_1 & b_1 \end{vmatrix} \neq 0$$

(это условие необходимо, чтобы по координатам X_0, X_1 можно было обратно найти координату x); ρ — произвольный отличный от нуля множитель пропорциональности (мы его пишем, чтобы подчеркнуть однородный характер координат X_0, X_1, т. е. тот факт, что они определены только с точностью до пропорциональности).

Определение 3. Координаты X_0, X_1, связанные с аффинной координатой x формулами (3), называются (общими) *однородными проективными координатами* на прямой.

Аффинная координата x выражается, очевидно, через координаты $(X_0 : X_1)$ по формуле

$$x = \frac{-b_1 X_0 + b_0 X_1}{a_1 X_0 - a_0 X_1}.$$

В частности, мы видим, что в координатах (3) несобственная точка имеет координаты $(a_0 : a_1)$ (т. е. что координатам $a_0 : a_1$ не отвечает никакая собственная точка прямой).

Однородные барицентрические координаты являются частным случаем проективных координат, получающимся при

$$\begin{aligned} a_0 &= -1, \quad b_0 = x_1, \\ a_1 &= 1, \quad\;\; b_1 = -x_0. \end{aligned}$$

Другой важный частный случай проективных координат получается при

$$\begin{aligned} a_0 &= 0, \quad b_0 = 1, \\ a_1 &= 1, \quad b_1 = 1. \end{aligned}$$

Эти координаты называются *однородными аффинными координатами*. Они связаны с обыкновенной (неоднородной) аффинной координатой x формулой

$$x = \frac{X_1}{X_0}.$$

В этих координатах несобственная точка имеет координаты $(0:1)$.

Все сказанное выше непосредственно обобщается на случай плоскости и пространства.

Мы рассмотрим только случай плоскости, предоставляя случай пространства инициативе читателя.

Пусть x, y — аффинные координаты на плоскости. По произвольной невырожденной матрице

$$\begin{pmatrix} a_0 & b_0 & c_0 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{pmatrix}$$

составим числа

$$\begin{aligned} \rho X_0 &= a_0 x + b_0 y + c_0, \\ \rho X_1 &= a_1 x + b_1 y + c_1, \\ \rho X_2 &= a_2 x + b_2 y + c_2, \end{aligned} \tag{4}$$

где ρ — произвольный множитель пропорциональности.

Определение 4. Числа X_0, X_1, X_2, определенные (с точностью до пропорциональности) формулами (4), называются (общими) *однородными проективными координатами* на плоскости.

Аффинные координаты x, y выражаются через проективные координаты $(X_0 : X_1 : X_2)$ по формулам

$$x = \frac{\alpha_0 X_0 + \alpha_1 X_1 + \alpha_2 X_2}{\gamma_0 X_0 + \gamma_1 X_1 + \gamma_2 X_2}, \quad y = \frac{\beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2}{\gamma_0 X_0 + \gamma_1 X_1 + \gamma_2 X_2}, \tag{5}$$

где

$$\alpha_0 = \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}, \quad \alpha_1 = -\begin{vmatrix} b_0 & c_0 \\ b_2 & c_2 \end{vmatrix}, \quad \alpha_2 = \begin{vmatrix} b_0 & c_0 \\ b_1 & c_1 \end{vmatrix},$$

$$\beta_0 = -\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}, \quad \beta_1 = \begin{vmatrix} a_0 & c_0 \\ a_2 & c_2 \end{vmatrix}, \quad \beta_2 = -\begin{vmatrix} a_0 & c_0 \\ a_1 & c_1 \end{vmatrix},$$

$$\gamma_0 = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}, \quad \gamma_1 = -\begin{vmatrix} a_0 & b_0 \\ a_2 & b_2 \end{vmatrix}, \quad \gamma_2 = \begin{vmatrix} a_0 & b_0 \\ a_1 & b_1 \end{vmatrix}.$$

Задание. Докажите формулы (5). Указание: запишите эти формулы в следующем виде:

$$x=\frac{\begin{vmatrix} X_0 & b_0 & c_0 \\ X_1 & b_1 & c_1 \\ X_2 & b_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_0 & b_0 & X_0 \\ a_1 & b_1 & X_1 \\ a_2 & b_2 & X_2 \end{vmatrix}}, \quad y=\frac{\begin{vmatrix} a_0 & X_0 & c_0 \\ a_1 & X_1 & c_1 \\ a_2 & X_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_0 & b_0 & X_0 \\ a_1 & b_1 & X_1 \\ a_2 & b_2 & X_2 \end{vmatrix}}.$$

Однородные проективные координаты (4) осуществляют биективное отображение плоскости на множество $\mathbb{R}P^2$, из которого удалены все элементы $(X_0 : X_1 : X_2)$, для которых

$$\gamma_0 X_0 + \gamma_1 X_1 + \gamma_2 X_2 = 0$$

(этим элементам соответствуют «несобственные точки» плоскости).

Определение 5. Пусть

$$A_0(x_0, y_0), \quad A_1(x_1, y_1), \quad A_2(x_2, y_2)$$

— неколлинеарная тройка точек плоскости. Числа X_0, X_1, X_2, связанные с аффинными координатами x, y формулами

$$x=\frac{x_0X_0 + x_1X_1 + x_2X_2}{X_0 + X_1 + X_2}, \quad y=\frac{y_0X_0 + y_1X_1 + y_2X_2}{X_0 + X_1 + X_2},$$

называются *однородными барицентрическими координатами* на плоскости. Они являются частным случаем однородных проективных координат. Для них

$$\begin{aligned} &\alpha_0 = x_0, \quad \alpha_1 = x_1, \quad \alpha_2 = x_2, \\ &\beta_0 = y_0, \quad \beta_1 = y_1, \quad \beta_2 = y_2, \\ &\gamma_0 = 1, \quad \gamma_1 = 1, \quad \gamma_2 = 1. \end{aligned}$$

Механический смысл барицентрических координат (объясняющий их название) состоит в следующем: точка A с координатами $X_0 : X_1 : X_2$ является центром тяжести трех масс X_0, X_1, X_2, помещенных соответственно в точках A_0, A_1, A_2.

Несобственные точки плоскости характеризуются в барицентрических координатах соотношением

$$X_0 + X_1 + X_2 = 0.$$

Определение 6. *Однородными аффинными координатами* называются однородные проективные координаты, для которых

$$\begin{aligned} &\alpha_0 = 0, \quad \alpha_1 = 1, \quad \alpha_2 = 0, \\ &\beta_0 = 0, \quad \beta_1 = 0, \quad \beta_2 = 1, \\ &\gamma_0 = 1, \quad \gamma_1 = 0, \quad \gamma_0 = 0. \end{aligned}$$

Они связаны с аффинными координатами формулами

$$x = \frac{X_1}{X_0}, \quad y = \frac{X_2}{X_0}.$$

Несобственные точки характеризуются в этих координатах соотношением

$$X_0 = 0.$$

§ 2. УРАВНЕНИЯ ЛИНИЙ И ПОВЕРХНОСТЕЙ

1. Задание линий и поверхностей уравнениями

Интуитивно понятие линии знакомо каждому человеку, даже не изучавшему геометрию. Например, линиями на плоскости являются прямые и окружности, а в пространстве, скажем, — винтовые линии. С другой стороны, например, область, ограниченная окружностью («круг»), линией не является.

Однако задача дать строгое определение понятию линии (или, как говорят, дать его *экспликацию*) оказалась неожиданно очень трудной и, по-видимому, вполне удовлетворительного решения не имеющей.

Наиболее удачную экспликацию понятия линии предложил замечательный советский математик П. С. Урысон. Грубо говоря, множество точек является линией в смысле Урысона, если в окрестности любой его точки M все его точки, отличные от точки M, составляют несвязное множество (состоящее из нескольких «частей»). Однако под это определение подпадают некоторые весьма сложно устроенные множества, которые наша интуиция может считать линиями лишь с большой натяжкой.

Читатель, интересующийся этим вопросом, найдет подробности в книжке А. С. Пархоменко, Что такое линия, Гостехиздат, 1954.

Как бы то ни было, каждая линия на плоскости представляет собой множество («геометрическое место») точек, удовлетворяющих некоторому условию. В координатах x, y это условие может быть записано в виде равенства

$$f(x, y) = 0,$$

где $f(x, y)$ — некоторая функция двух переменных. Это равенство называется *уравнением* данной линии (в данной координатной системе Oxy).

Например, ось абсцисс определяется уравнением

$$y = 0,$$

а ось ординат — уравнением

$$x = 0.$$

В прямоугольной системе координат биссектриса первого и третьего квадрантов имеет уравнение $x = y$, т. е. уравнение

$$x - y = 0,$$

а биссектриса второго и четвертого квадрантов — уравнение

$$x + y = 0.$$

Окружность радиуса R с центром в точке (a, b) имеет (в прямоугольных координатах) уравнение

$$(x - a)^2 + (y - b)^2 = R^2.$$

В частности, окружность с центром в точке $O(0, 0)$ имеет уравнение

$$x^2 + y^2 = R^2.$$

Линии задаются уравнениями не только в аффинных, но и в любых других (например, в полярных) координатах. Так, в полярных координатах r, φ уравнение окружности с центром в точке O имеет особенно простой вид:

$$r = R.$$

Аналогично, поверхность в пространстве определяется одним уравнением вида

$$f(x, y, z) = 0.$$

Например, сфера радиуса R с центром в точке (a, b, c) имеет уравнение

$$(x - a)^2 + (y - b)^2 + (z - c)^2 = R^2.$$

Линия в пространстве задается двумя условиями, аналитически записываемыми в виде двух уравнений:

$$f(x, y, z) = 0,$$
$$g(x, y, z) = 0.$$

Иными словами, линия в пространстве задается как линия пересечения двух поверхностей $f(x, y, z) = 0$ и $g(x, y, z) = 0$.

Заметим, что (для данной линии) поверхности $f(x, y, z) = 0$ и $g(x, y, z) = 0$ можно выбирать с большой степенью произвола. Именно, за эти поверхности можно принять произвольные поверхности, содержащие данную линию и не имеющие более никаких общих точек.

Аналогом линий в пространстве являются на плоскости (неупорядоченные) наборы изолированных точек. Каждый такой набор может быть представлен как пересечение двух линий, т. е. может быть задан двумя уравнениями вида

$$f(x, y) = 0,$$
$$g(x, y) = 0.$$

Конечно, выбор этих уравнений также может быть осуществлен с большим произволом.

Замечание 1. Пусть нам дана некоторая поверхность

$$f(x, y, z) = 0$$

и некоторая плоскость П. Вообще говоря, плоскость П пересекает данную поверхность по некоторой линии. Поставим вопрос: как найти уравнение этой линии?

Конечно, чтобы этот вопрос имел смысл, необходимо задать на плоскости П некоторую систему координат x', y', а чтобы на этот вопрос можно было ответить, необходимо как-то связать координаты x, y, z в пространстве и координаты x', y' на плоскости.

Наиболее простой случай возникает, когда плоскость П является одной из координатных плоскостей, скажем, плоскостью

$$z = 0.$$

На этой плоскости величины x, y являются координатами и в этих координатах уравнение линии пересечения имеет, очевидно, вид

$$f(x, y, 0) = 0.$$

Общий случай сводится к этому простейшему преобразованием координат. Именно, нужно ввести в пространстве новые координаты x', y', z', в которых уравнение данной плоскости имеет вид

$$z' = 0,$$

и затем, пользуясь формулами преобразования координат, найти уравнение данной поверхности в координатах x', y', z'. После этого останется положить в этом уравнении $z' = 0$. Получится уравнение от x', y', являющееся уравнением искомой линии пересечения (в координатах x', y').

Замечание 2. Во многих курсах аналитической геометрии утверждается (иногда явно, иногда неявно), что не только любая линия на плоскости выражается (в аффинных координатах) некоторым уравнением вида $f(x, y) = 0$, но и обратно, каждое такое уравнение определяет некоторую линию. Несмотря на распространенность (и кажущуюся очевидность) этого утверждения, более внимательное рассмотрение показывает, что оно неверно.

Действительно, пусть X — произвольное множество точек плоскости (например, полуплоскость). Определим функцию $f(x, y)$ (так называемую *характеристическую функцию* множества X), считая, что $f(x, y) = 0$, если точка $A(x, y)$ принадлежит множеству X, и что $f(x, y) = 1$ в противном случае. Тогда множество X будет определяться уравнением $f(x, y) = 0$. Таким образом, уравнениями вида $f(x, y) = 0$ можно задать любое множество на плоскости (а не только линии).

Следовательно, чтобы сохранить утверждение о том, что любое уравнение $f(x, y) = 0$ определяет некоторую линию, мы должны как-то ограничить класс рассматриваемых функций $f(x, y)$, причем это мы должны сделать так, чтобы «вместе с

водой не выплеснуть и ребенка», т. е. так, чтобы любую линию по-прежнему можно было бы выразить уравнением $f(x, y) = 0$. Эта задача по существу равносильна задаче об экспликации понятия линии. Мы не будем здесь ею заниматься и ограничимся лишь указанием двух примеров (чтобы хотя бы намекнуть на характер возникающих здесь трудностей).

Пример 1. Рассмотрим функцию $f(x, y)$, определенную условиями:

$$f(x, y) = \begin{cases} 0, & \text{если } x \geqslant 0, \\ e^{-\frac{1}{x^2}}, & \text{если } x < 0. \end{cases}$$

Без труда проверяется, что эта функция непрерывна и, более того, имеет непрерывные частные производные всех порядков. Вместе с тем уравнение $f(x, y) = 0$ определяет, очевидно, полуплоскость $x \geqslant 0$.

Пример 2. Уравнение

$$|x| - x = 0$$

также определяет полуплоскость $x \geqslant 0$.

Таким образом, ни наложение условий гладкости (дифференцируемости), ни ограничение «элементарными функциями» не позволяет нам исключить «частей плоскости».

2. Алгебраические линии

Тем не менее, существуют, конечно, достаточно обширные классы функций $f(x, y)$, обладающие тем свойством, что уравнения $f(x, y) = 0$ определяют «настоящие» линии (не содержащие «плоскостных» кусков). Простейшим (и, по-видимому, важнейшим) таким классом является класс всех многочленов.

Напомним, что *многочленом* от переменных x, y называется сумма *одночленов* вида ax^py^q, где a — некоторое число (отличное от нуля), а p и q — целые неотрицательные числа. Число $n = p + q$ называется *степенью* одночлена ax^py^q. Наибольшая степень одночленов, из которых состоит данный многочлен, называется *степенью* этого многочлена. Многочлен называется *формой* (или *однородным многочленом*), если все его одночлены имеют одну и ту же степень. Ясно, что любой многочлен является суммой форм. Многочленами нулевой степени являются отличные от нуля числа и только они. Во избежание излишних оговорок, к многочленам причисляется также и число 0. Этот *нулевой многочлен* никакой степени не имеет (или, если угодно, имеет ту степень, которую мы желаем в данный момент ему приписать). Два многочлена f и g называются *пропорциональными*, если существует такое число $k \neq 0$, что $f = kg$. Ясно, что пропорциональные многочлены имеют одну и ту же степень.

Определение 1. Множество точек плоскости, аффинные координаты x, y которых удовлетворяют уравнению $f(x, y) = 0$, где

$f(x, y)$ — некоторый многочлен положительной степени, называется *алгебраической линией*. Степень многочлена $f(x, y)$ называется при этом *порядком* линии.

Изучение алгебраических линий является предметом так называемой алгебраической геометрии. В аналитической геометрии ограничиваются изучением линий лишь первого и второго порядков.

Чтобы оправдать введенную терминологию, следует, конечно, показать, что алгебраические линии соответствуют интуитивному понятию о линиях: например, не содержат «частей плоскости». Для линий первого и второго порядка мы сделаем это ниже, в явном виде построив все возможные такие линии. Для линий высших порядков соответствующее утверждение доказывается в курсах алгебраической геометрии.

Однако, понятие алгебраической линии все же не полностью соответствует интуитивному понятию линии. Например, уравнению

$$x^2 + y^2 + 1 = 0$$

не удовлетворяет ни одна точка плоскости, т. е. это уравнение определяет *пустое множество*. Аналогично, уравнению

$$x^2 + y^2 = 0$$

удовлетворяет только точка $O(0, 0)$, так что соответствующая алгебраическая линия состоит только из одной этой точки. Таким образом, вводя алгебраические линии, мы включаем в понятие линии пустое множество и множества, состоящие из отдельных точек. Это «насилие над интуицией» не столь велико, чтобы с ним нельзя было примириться, с другой стороны, включение в рассмотрение такого рода «особых» алгебраических линий существенно унифицирует и упрощает теорию.

В связи с понятием алгебраической линии, в первую очередь, встает вопрос о корректности его определения, т. е. об его независимости от выбора аффинной координатной системы. Соответствующее доказательство проводится без особого труда.

Действительно, пусть x', y' — другая аффинная координатная система. Чтобы получить в координатах x', y' уравнение [1]) $f'(x', y') = 0$ линии, имеющей в координатах x, y уравнение $f(x, y) = 0$, достаточно, очевидно, в функцию $f(x, y)$ подставить выражения координат x, y через координаты x', y'. Но, как мы знаем, координаты x, y линейно выражаются через координаты x', y', а при подстановке в многочлен $f(x, y)$ вместо x, y произвольных линейных функций переменных x', y' снова, очевидно, получится многочлен (уже от x', y'). Следовательно, линия, алгебраическая в координатах x, y, будет алгебраической

[1]) Здесь, конечно, символ f' никакого отношения к производным не имеет.

линией и в координатах x', y'. Тем самым корректность понятия алгебраической линии полностью доказана.

Правду сказать, в приведенном доказательстве допущена некоторая неточность (попытайтесь найти ее сами). Именно, многочлен, определяющий алгебраическую линию, должен иметь положительную степень, но априори не ясно, не перейдет ли при замене координат многочлен положительной степени в многочлен нулевой степени. Таким образом, нужно доказать, что это невозможно. Вместо этого мы докажем большее, а именно, что *при замене координат степень многочлена* $f(x, y)$ *вообще не меняется.*

С этой целью заметим, что при линейной замене переменных степень многочлена может только уменьшиться. Таким образом, если многочлен $f(x, y)$ имеет степень n, а получающийся из него в результате линейной замены переменных многочлен $f'(x', y')$ имеет степень n', то $n' \leqslant n$. Но многочлен $f'(x', y')$ в свою очередь получается из многочлена $f(x, y)$ линейной заменой переменных (обратной к предыдущей). Поэтому, по аналогичным соображениям, $n \leqslant n'$. Следовательно, $n = n'$.

Замечание 1. Доказательство равенства $n = n'$ можно рассматривать как доказательство корректности определения порядка алгебраической линии. Однако ситуация здесь несколько более сложная, чем это кажется на первый взгляд. Действительно, одна и та же алгебраическая линия (в одной и той же аффинной координатной системе) может, вообще говоря, выражаться различными уравнениями $f(x, y) = 0$ и $g(x, y) = 0$. Например, если многочлены $f(x, y)$ и $g(x, y)$ пропорциональны, то линии $f(x, y) = 0$ и $g(x, y) = 0$, очевидно, совпадают. Поскольку пропорциональные многочлены имеют одну и ту же степень, корректность определения порядка алгебраической линии будет установлена, если мы докажем, что справедливо и обратное утверждение, т. е. что справедлива следующая теорема единственности:

Алгебраические линии $f(x, y) = 0$ *и* $g(x, y) = 0$ *тогда и только тогда совпадают, когда многочлены* $f(x, y)$ *и* $g(x, y)$ *пропорциональны.*

Однако, к сожалению, в таком виде теорема неверна. Тому имеются две причины. Во-первых, если мы умножим многочлен $f(x, y)$ на произвольный многочлен положительной степени, ни при каких вещественных x, y не обращающийся в нуль (например, на многочлен $x^2 + y^2 + 1$), то мы получим многочлен, определяющий ту же самую алгебраическую линию, что и многочлен $f(x, y)$. Например, уравнения $x(x^2 + y^2 + 1) = 0$ и $x = 0$ определяют одну и ту же прямую (ось ординат). В то же время эти многочлены не пропорциональны (и имеют различные степени). Чтобы преодолеть эту трудность, следует соответствующим образом определить отношение «две алгебраические линии совпадают». Например, можно считать, что две алгебраи-

ческие линии $f(x, y) = 0$ и $g(x, y) = 0$ совпадают, если равенства $f(x, y) = 0$ и $g(x, y) = 0$ равносильны не только для вещественных переменных x и y, но и в случае, когда мы допускаем для этих переменных любые комплексные значения. Можно сказать, что, таким образом, мы вводим в рассмотрение «точки» с комплексными координатами (такие точки не имеют никакого геометрического смысла и представляют собой — при фиксированной аффинной координатной системе — просто пары (x, y) комплексных чисел) и определяем алгебраические линии как совокупности всех (в том числе и комплексных) точек, удовлетворяющих уравнениям вида $f(x, y) = 0$, где $f(x, y)$ — произвольный многочлен положительной степени. Из алгебры известно, что для любого многочлена положительной степени существуют комплексные точки (x, y), в которых он обращается в нуль. Поэтому после введения комплексных точек получить контрпримеры к теореме единственности описанным выше способом мы уже не можем.

Однако такие контрпримеры можно получить и по-другому, заметив, что уравнения $f(x, y) = 0$ и $f(x, y)^p = 0$, где p — произвольное натуральное число, определяют одну и ту же алгебраическую линию (безразлично, допускаются или нет точки с комплексными координатами). Например, оба уравнения $x = 0$ и $x^2 = 0$ определяют одну и ту же ось ординат. Эту трудность можно преодолеть, введя понятие «кратной точки» и считая, например, что уравнение $x^2 = 0$ определяет «дважды взятую» ось ординат. На этом пути возникают значительные трудности в связи с аккуратным обращением с кратными точками (даже строгое и «рабочее» определение понятия «кратности» дать не так-то просто). Тем не менее все эти трудности можно успешно преодолеть (что, кстати сказать, требует довольно изощренной алгебраической техники), после чего теорему единственности уже удается строго доказать.

Описанный путь обоснования теоремы единственности слишком труден, чтобы мы могли им здесь воспользоваться. Однако мы можем совершить «обходный маневр», заметив, что введение комплексных и кратных точек по существу лишает нашу теорию геометрической наглядности и переводит ее в область алгебры. Воспользовавшись принципом «снявши голову, по волосам не плачут», мы можем полностью «алгебраизироваться» и определить алгебраическую линию (в данной аффинной координатной системе x, y) просто как класс пропорциональных друг другу многочленов $f(x, y)$. Теорема единственности будет тогда автоматически справедлива. Мы, однако, не будем настаивать на этом определении, желая остаться в рамках наглядно геометрических понятий, хотя иногда оно и найдет отражение в нашей терминологии (аналогично тому, как мы позволим себе говорить, например, о «дважды взятых прямых», не давая общего определения кратных точек).

Мы можем надеяться сохранить (при первоначальном определении алгебраической линии) теорему единственности в несколько ослабленном виде, заметив, что во всех указанных выше контрпримерах многочлены $f(x, y) = 0$ и $g(x, y) = 0$, определяющие одну и ту же алгебраическую линию, имеют различные степени. Это наводит на мысль, не будет ли справедлива следующая теорема единственности для алгебраических линий фиксированного порядка:

Алгебраические линии $f(x, y) = 0$ и $g(x, y) = 0$, *определяемые многочленами одной и той же степени* n, *тогда и только тогда совпадают, когда эти многочлены пропорциональны.*

Однако и в этой ослабленной форме теорема единственности неверна, как показывает пример непропорциональных многочленов $x^2 + y^2 + 1$ и $x^2 + y^2 + 2$, определяющих одну и ту же алгебраическую линию (пустое множество). Тем не менее, эта теорема оказывается справедливой, если рассматриваемые линии имеют достаточно много точек. Например, как мы докажем ниже (см. п. 3 § 1 гл. 3 и п. 4 § 1 гл. 6), она справедлива *для любых линий первого порядка* (которые заведомо имеют много точек) и *для линий второго порядка, состоящих более чем из одной точки.*

Поскольку наша цель здесь состояла лишь в указании подводных камней в этих, казалось бы, спокойных и безопасных водах, мы ограничимся сказанным.

Конечно, аналогичные трудности (в осложненном виде) имеют место и для поверхностей (и линий) в пространстве.

3. Параметрические уравнения линий и поверхностей

Аналитическое задание линий уравнениями вида $f(x, y) = 0$ не является единственно возможным. Например, представляя данную линию как траекторию движущейся точки, мы получим (по-прежнему предполагая заданной систему аффинных координат x, y), что координаты x, y точек линии являются некоторыми функциями

$$\begin{aligned} x &= x(t), \\ y &= y(t) \end{aligned} \tag{1}$$

параметра t («времени»), меняющегося в некоторых определенных пределах $a \leqslant t \leqslant b$, где a и b — вещественные числа или символы $\pm\infty$. Такого рода задание линии называется заданием посредством *параметрических уравнений.*

Например, уравнения

$$\begin{aligned} x &= R\cos t, \\ y &= R\sin t, \end{aligned} \qquad -\pi < t \leqslant \pi, \tag{2}$$

являются (в прямоугольных координатах x, y) параметрическими уравнениями окружности радиуса R с центром в точке $O(0, 0)$.

Действительно, каждая точка $A(x, y)$ с координатами, выражающимися формулами (2), удовлетворяет уравнению окружности

$$x^2 + y^2 = R^2,$$

и обратно, если числа x, y удовлетворяют этому уравнению, то существует такое t, что выполнены уравнения (2).

Параметр t в уравнениях (2) имеет очевидный геометрический смысл: он равен углу от оси Ox к радиус-вектору $\overrightarrow{OA}$ точки $A(x, y)$.

Замечание 1. Одну и ту же линию можно задавать различными параметрическими уравнениями. Например, ту же окружность радиуса R с центром в точке O можно задать параметрическими уравнениями

$$\begin{aligned} x &= R\frac{1-t^2}{1+t^2}, \\ y &= R\frac{2t}{1+t^2}, \end{aligned} \qquad -\infty < t < +\infty$$

(строго говоря, мы здесь получим окружность без точки $(R, 0)$, соответствующей бесконечному значению параметра t). Конечно, здесь параметр t имеет уже совсем другое геометрическое истолкование (найдите его сами).

Каждому значению t_0 параметра t соответствует некоторая определенная точка A_0 рассматриваемой линии, имеющая координаты $x(t_0)$, $y(t_0)$. Однако одна и та же точка может, вообще говоря, соответствовать многим различным значениям параметра. Например, для уравнений (2) значения параметра $t = \pm\pi$ определяют одну и ту же точку окружности.

Чтобы от уравнения $f(x, y) = 0$ перейти к параметрическим уравнениям, следует в функцию $f(x, y)$ вместо x подставить некоторую функцию $x(t)$ и решить уравнение

$$f(x(t), y) = 0 \tag{3}$$

относительно y. При этом функцию $x(t)$ следует выбрать так, чтобы уравнение (3) имело однозначное решение $y = y(t)$. При неудачном выборе функции $x(t)$ уравнение (3) может вообще не иметь однозначного решения или иметь такое решение только на небольшом отрезке изменения параметра t (что даст нам только некоторую часть исследуемой линии). Детальное рассмотрение всех этих вопросов принадлежит, по существу, теории неявных функций, и мы здесь ими заниматься не будем.

Важный класс параметрических уравнений возникает, когда за параметр t принята координата x. В этом случае мы на самом деле имеем только одно уравнение вида

$$y = y(x). \tag{4}$$

Такое уравнение линии называется уравнением, *разрешенным относительно* x. В случае, когда координаты x, y прямоугольны,

наша линия представляет собой в этом случае не что иное, как *график* функции $y = y(x)$.

Для того чтобы линию можно было задать уравнением вида (4), необходимо и достаточно, чтобы каждая прямая, параллельная оси ординат, пересекала эту линию не более чем в одной точке. Например, окружность $x^2 + y^2 = R^2$ задать уравнением вида (4) нельзя. Однако ее «верхнюю» полуокружность $x^2 + y^2 = R^2$, $y \geqslant 0$, так задать можно:

$$y = \sqrt{R^2 - x^2}, \qquad -R \leqslant x \leqslant R.$$

Аналогично, уравнение

$$y = -\sqrt{R^2 - x^2}, \qquad -R \leqslant x \leqslant R,$$

определяет «нижнюю» полуокружность $x^2 + y^2 = R^2$, $y \leqslant 0$.

В пространстве параметрические уравнения линий имеют вид

$$\begin{aligned} x &= x(t), \\ y &= y(t), \qquad a \leqslant t \leqslant b. \\ z &= z(t), \end{aligned}$$

Все сказанное выше применимо, конечно (с соответствующими тривиальными изменениями), и в этом случае.

Вводя векторную функцию

$$\boldsymbol{r}(t) = x(t)\,\boldsymbol{e}_1 + y(t)\,\boldsymbol{e}_2 + z(t)\,\boldsymbol{e}_3$$

(на плоскости — векторную функцию $\boldsymbol{r}(t) = x(t)\boldsymbol{e}_1 + y(t)\boldsymbol{e}_2$), мы можем параметрические уравнения линии записать в виде одного *векторного параметрического уравнения*

$$\boldsymbol{r} = \boldsymbol{r}(t),$$

где $\boldsymbol{r}$ — радиус-вектор произвольной точки линии. Это уравнение имеет, формально, один и тот же вид как в плоскости, так и в пространстве.

Поверхности в пространстве также можно задавать параметрическими уравнениями (содержащими уже два независимых параметра). В векторной форме они имеют вид

$$\boldsymbol{r} = \boldsymbol{r}(u, v).$$

Поскольку общей теорией поверхностей мы по существу заниматься не будем, такого рода уравнения будут у нас возникать только мимоходом.

Обратим внимание на то, что интуитивное геометрическое различие между линиями и поверхностями, состоящее в том, что линии «одномерны», а поверхности «двумерны», находит свое аналитическое отражение в том, что линии определяются уравнениями, зависящими от одного параметра, а

поверхности — уравнениями, зависящими от д в у х параметров. Иначе говоря, линии представляют собой «однопараметрические семейства» точек, а поверхности — «двупараметрические семейства».

Владея параметрическими уравнениями линий, мы можем попытаться определить линию (на плоскости) как множество точек, координаты которых выражаются уравнениями вида (1), где функции $x(t)$ и $y(t)$ удовлетворяют определенным условиям (которые мы и должны найти). Само собой напрашивающееся условие состоит в требовании *непрерывности* этих функций. Такое определение линии предложил в прошлом веке французский математик Жордан, и потому линии в этом смысле называются «жордановыми кривыми». Однако вскоре после того, как Жордан предложил свое определение, итальянский математик Пеано построил пример жордановой кривой, проходящей через любую (!) точку некоторого квадрата. Пример Пеано показывает, таким образом, что понятие жордановой кривой оказывается слишком широким.

Это понятие можно сузить, потребовав, чтобы функции $x(t)$ и $y(t)$ были дифференцируемыми. Получающиеся так называемые «гладкие кривые» уже не могут заполнять никакой плоской области и вполне удовлетворяют интуитивному представлению о кривой линии. Однако вне рамок гладких кривых оказываются, например, все ломаные. Чтобы включить и их, следует допустить функции, обладающие конечным числом разрывов (первого рода). Соответствующие «кусочно гладкие кривые» уже охватывают по существу все интересные для математической практики кривые и, по-видимому, представляют (на современном уровне развития математики) наиболее удачный аналитический вариант строгого определения понятия «линий». Однако их определение содержит чуждое геометрии требование дифференцируемости, и потому с точки зрения «чистой геометрии» класс кусочно гладких кривых определен плохо.

Кроме того, все определения «по Жордану» (включающие или нет требования дифференцируемости) имеют еще и тот принципиальный недостаток, что они описывают линию не только как множество точек, но и включают в себя определенный «способ пробегания» этого множества. Это находит отражение в том, что две разные пары функций $x(t)$ и $y(t)$ могут определять одну и ту же линию (см. выше в п. 3 пример двух разных параметрических уравнений окружности). Строгое описание этого феномена возможно, но оказывается весьма длинным и утомительным (следует, например, учесть возможность того, что при изменении параметра t точка $(x(t), y(t))$ может «стоять на месте»).

Впрочем, все это имеет для нас чисто «академический» интерес, поскольку мы будем интересоваться лишь алгебраическими линиями первого и второго порядков, в связи с которыми

никаких трудностей с определением не возникает (мы не говорим сейчас о трудностях, о которых шла речь в конце п. 2). Мы коснулись этих вопросов лишь для того, чтобы читатель не думал, что здесь дело обстоит так просто, как это на первый взгляд кажется.

§ 3. КООРДИНАТНО-АКСИОМАТИЧЕСКОЕ ПОСТРОЕНИЕ ГЕОМЕТРИИ

1. Основные положения аксиоматического метода

Общеизвестен *дедуктивный характер* математики, т. е. тот факт, что истинность математических утверждений (в отличие от утверждений естественных наук) не устанавливается на основании опыта и наблюдения, а выводится (дедуцируется) из небольшого числа исходных утверждений (*аксиом*). Истинность аксиом принимается без доказательства; все остальные утверждения (*теоремы*) с помощью логических рассуждений выводятся из аксиом. Понятия, участвующие в утверждениях, также разделяются на два типа — *основные* (*исходные*) *понятия*, принимающиеся без определений, и *производные понятия*, которые должны быть определены через исходные.

Эти требования к дедуктивной науке были выработаны еще древнегреческими математиками и обычно связываются с именем Аристотеля. Основывающееся на них изложение геометрии, предпринятое Евклидом в его «Началах», в течение двух тысяч лет считалось основным и безупречным примером дедуктивного построения знания. Однако постепенно выяснилось, что изложение Евклида (на котором до сих пор во многом базируется преподавание математики в школе) не вполне удовлетворяет указанным требованиям. Например, Евклид делает попытки определения основных понятий (скажем, он пишет: «точка есть то, что не имеет частей») и, кроме того, его аксиомы недостаточны для доказательства всех теорем геометрии (необходимы дополнительные утверждения, Евклидом нигде явно не сформулированные).

Конечно, эти недостатки вполне исправимы и были действительно исправлены в конце XIX в., когда математическая мысль после многовекового перерыва опять обратилась к представлению геометрии (и других областей математики) в строгой форме, соответствующей древнегреческому идеалу аксиоматико-дедуктивной науки. Во многом это было стимулировано открытием Лобачевским его неевклидовой геометрии.

На первых порах геометры XIX в. исправляли Евклида, оставаясь в вопросе об истинности геометрии по существу на позициях Аристотеля, т. е. они считали, что:

1°. Исходные положения (аксиомы и основные понятия) имеют вполне определенный содержательный смысл; аксиомы истинны в силу своей очевидности.

2°. Теоремы истинны, так как они выводятся правильными логическими рассуждениями из истинных аксиом.

Эта точка зрения называется *содержательной*. Развитие математики в XIX и XX вв. привело к некоторому новому уровню дедуктивно-аксиоматических построений, который принято называть *формальным*. Этот уровень

стал необходимым в связи с созданием новых отделов математики, имеющих дело с очень отвлеченными и мало наглядными конструкциями. Наглядность и очевидность исходных утверждений перестала, тем самым, быть критерием правильности (и полезности) математических построений. Вперед выдвинулась логическая строгость математических теорий. Явная формулировка всех аксиом и посылок стала насущно необходимой.

При формально-аксиоматическом построении теории намеренно отвлекаются от содержательного понимания и вопрос об истинности или очевидности аксиом уже больше не ставится. Аксиоматически построенная математическая теория занимается изучением лишь логических следствий из принятых аксиом. Таким образом, формально-аксиоматический подход преодолевает аристотелевский тезис 1°.

Дальнейшее углубление дедуктивного метода, связанное с преодолением тезиса 2°, возникает при анализе понятия правильного логического вывода. Оказывается, что для логики также возможно чисто формальное построение, отвлекающееся от ее содержательного понимания. Последовательное проведение этой точки зрения увело бы нас слишком далеко и потому мы остановимся на полпути и будем рассматривать «формальные аксиоматики в неформальной логике».

Нашей основной целью будет дедуктивно-аксиоматическое построение геометрии (в первую очередь обыкновенной школьной, евклидовой геометрии, а затем аффинной и проективной). При этом для простоты мы ограничимся случаем плоскости. Переход к случаю пространства совершенно тривиален, и мы его оставим читателю.

Формально-аксиоматическое построение геометрии (как и любой другой математической теории) начинается с перечисления, во-первых, основных объектов, во-вторых, основных отношений между этими объектами и, в-третьих, аксиом.

Основные объекты и основные отношения никак не определяются, не поясняются и не описываются, а только называются. От них лишь требуется, чтобы они подчинялись аксиомам.

Все другие объекты и отношения, рассматриваемые в геометрии, должны определяться через основные. Именно потому, что цепь определений должна иметь начало, необходимы основные неопределяемые объекты и отношения.

Подчеркнем, что основными объектами данной аксиоматики можно считать объекты любой природы, а основным отношениям между ними можно придать любой конкретный смысл, лишь бы удовлетворялись все аксиомы. Тогда любая теорема, выведенная из аксиом, будет выражать определенный факт, относящийся к этим объектам, точнее, к тем их свойствам, которые фигурируют в аксиомах.

Любой конкретный выбор объектов и их отношений, удовлетворяющий аксиомам, называется *реализацией*, *интерпретацией* или *моделью* данной аксиоматики.

Впрочем, эта терминология постепенно выходит из употребления. Теперь предпочитают называть, скажем, модель аксиоматики плоской евклидовой геометрии «евклидовой плоскостью», а например, модель аксиоматики плоской аффинной геометрии — «аффинной плоскостью» и т. п.

Две модели называются *изоморфными*, если между ними может быть установлен изоморфизм, т. е. взаимно однозначное соответствие, сохраняющее все основные отношения (ср. п. 6 § 2 гл. 1).

Вообще говоря, основные объекты и отношения (а также связывающие их аксиомы) можно выбирать довольно свободно: их выбор диктуется теми целями, которые ставит перед собой автор аксиоматики. Нужно лишь следить, чтобы все рассматриваемые аксиоматики были друг другу равносильны, т. е. чтобы они описывали одну и ту же математическую теорию.

Что, собственно говоря, означает, что две аксиоматики «описывают одну и ту же математическую теорию»? Если обе аксиоматики имеют одни и те же основные объекты и основные отношения, ответ тривиален: каждая аксиома одной аксиоматики должна быть теоремой в другой и наоборот. Если же основные объекты и основные отношения двух рассматриваемых аксиоматик различны, то их равносильность означает, что в каждой из них можно соответствующими определениями ввести основные объекты и основные отношения другой аксиоматики так, чтобы для этих объектов и отношений были выполнены соответствующие аксиомы, т. е. чтобы эти аксиомы оказывались теоремами. Другими словами, это означает, что данные аксиоматики должны быть интерпретируемыми друг в друге.

Если мы, выбрав основные объекты и отношения, произвольно напишем ряд аксиом, то, вообще говоря, мы можем и не получить аксиоматики, описывающей какую-либо математическую теорию. Необходимо еще, чтобы эти аксиомы были *непротиворечивы*, т. е. чтобы из них нельзя было вывести некоторого утверждения вместе с его отрицанием. Противоречивые системы аксиом совершенно бесполезны, поскольку из любой такой системы можно по законам логики вывести произвольное утверждение.

Как доказать непротиворечивость данной системы аксиом? Для этого возможны два способа: «абсолютный» и «относительный».

«Абсолютный» способ основывается на тщательном исследовании формального строения всех предложений, выводимых в данной аксиоматике (ее «теорем»). Если такое исследование продвинуто достаточно далеко и, в частности, найдены формальные признаки, характеризующие выводимые предложения, то для доказательства непротиворечивости аксиоматики остается лишь доказать, что не существует предложения, обладающего этими признаками одновременно с его отрицанием.

Однако «абсолютный» способ для мало-мальски сложных систем аксиом (например, для аксиом арифметики, не говоря уже об аксиомах геометрии) наталкивается на принципиальные трудности и приводит к порочному кругу.

Поэтому нам остается лишь «относительный» способ, заключающийся в том, что для данной аксиоматики строится модель в некоторой другой аксиоматике, непротиворечивость которой уже известна. Возможность построения такой модели обеспечивает, очевидно, непротиворечивость и данной аксиоматики. Конечно, пользуясь «относительным» способом, мы рано или поздно придем к аксиоматике, которую «негде» интерпретировать и к которой, следовательно, необходимо применять «абсолютный» способ. Выяснением и решением возникающих здесь принципиальных проблем занимается математическая логика, лежащая вне рамок нашего изложения.

Мы будем предполагать, что непротиворечивость теории вещественных чисел нам уже известна, и будем доказывать непротиворечивость наших аксиоматик построением их «числовых моделей».

Кроме непротиворечивости, аксиоматики геометрии должны обладать также свойством «полноты». В наиболее сильной форме требование полноты означает, что любое утверждение, которое осмысленно может быть сформулировано в этой аксиоматике, может быть либо выведено из аксиом, либо опровергнуто (с помощью этих аксиом). Однако в математической логике доказывается, что никакая достаточно мощная аксиоматика (в частности, никакая аксиоматика геометрии) таким свойством обладать не может. Это утверждение известно как теорема Гёделя о неполноте.

Поэтому мы вынуждены принять следующее определение полноты, вполне достаточное для всех практических целей: аксиоматика называется *полной*, если любые две ее интерпретации изоморфны.

Конечно, в математике встречаются (и далее играют большую роль) и неполные аксиоматики. Примером неполной аксиоматики может служить система аксиом группы: она неполна, потому что существуют неизоморфные группы. Неполные аксиоматики служат для совершенно иных целей, чем полные, и мы ими здесь заниматься не будем.

На аксиоматику часто накладывается еще требование *минимальности*: в список аксиом не должны входить «лишние» аксиомы, которые могут быть выведены из остальных аксиом. Требование минимальности очень существенно, когда мы интересуемся внутренним логическим строением геометрии. Однако, если аксиомы нам нужны только для строгости изложения и мы интересуемся не столько логической структурой теории, сколько фактами, которые в ней могут быть доказаны, требование минимальности оказывается чрезвычайно стеснительным и порой просто вредным. Поэтому за минимальностью мы, как правило, следить не будем.

Наконец, обычно считается (точнее, считалось) необходимым, чтобы аксиоматика геометрии была *независимой*, т. е. чтобы в ней не участвовали понятия из других отделов математики. Это приводит к, по существу, неоправданному усложнению аксиоматики, поскольку, скажем, аксиомы, описывающие расположение точек на прямой, на самом деле ничем не отличаются от соответствующих аксиом теории вещественных чисел, так что в рамках независимой аксиоматики геометрии приходится заново строить большой фрагмент теории вещественных чисел. Кроме того, требование независимости запрещает использовать даже простейшие понятия теории множеств, что, естественно, также вносит определенные трудности.

По этим основаниям в последнее время требование независимости не стали считать так уж обязательным и все шире начали использовать аксиоматики геометрии, базирующиеся на теории вещественных чисел и теории множеств. Наша аксиоматика также будет основываться на этих теориях.

2. Аксиоматика евклидовой геометрии

В основу нашей аксиоматики геометрии мы положим понятие евклидовой координатной системы. Рассмотрим прежде всего, какими свойствами это понятие обладает с содержательной точки зрения.

По определению, каждая евклидова координатная система является некоторым биективным отображением α плоскости на множество $\mathbb{R}^2$ всех пар вещественных чисел: для любой точки A пара $\alpha(A)$ состоит из координат x, y этой точки.

Пусть нам даны две евклидовы координатные системы α и α'. Отображение α^{-1}, обратное координатной системе α, каждой паре $(x, y) \in \mathbb{R}^2$ сопоставляет некоторую точку A, а координатная система α' сопоставляет этой точке пару чисел (x', y'). Следовательно, отображение $\alpha' \circ \alpha^{-1}$ представляет собой отображение $\mathbb{R}^2 \to \mathbb{R}^2$, переводящее пару (x, y) в пару (x', y'). При этом (см. п. 4 § 1) числа x, y и x', y' связаны формулами перехода вида

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = C \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}, \tag{1}$$

где C — некоторая ортогональная матрица (второго порядка), а a и b — числа. Таким образом,

Для любых двух евклидовых координатных систем α и α' композиция $\alpha' \circ \alpha^{-1}$ представляет собой отображение $\mathbb{R}^2 \to \mathbb{R}^2$, задаваемое формулой вида (1).

Далее, пусть α — произвольная евклидова координатная система и пусть φ — произвольное отображение $\mathbb{R}^2 \to \mathbb{R}^2$, задаваемое формулой вида (1). Тогда величины x', y' также будут евклидовыми координатами на плоскости, т. е. отображение $A \mapsto (x', y')$ будет евклидовой координатной системой. Но, по определению, это отображение является не чем иным, как отображением $\varphi \circ \alpha$. Таким образом,

Для любой евклидовой координатной системы α и любого отображения φ: $\mathbb{R}^2 \to \mathbb{R}^2$, *задаваемого формулой вида* (1), *отображение $\varphi \circ \alpha$ также является евклидовой координатной системой.*

Как мы ниже покажем, из этих свойств евклидовых координатных систем можно вывести всю евклидову геометрию плоскости. Другими словами, их можно принять за аксиомы евклидовой геометрии плоскости.

Перейдем теперь к формальному описанию аксиом.

Мы будем рассматривать всевозможные преобразования

$$\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2,$$

задаваемые формулами вида (1).

Определение 1. Множество всех преобразований вида (1) мы будем обозначать символом $\mathrm{Ort}(\mathbb{R}^2)$.

Из содержательного смысла этих преобразований немедленно вытекает, что

множество $\mathrm{Ort}(\mathbb{R}^2)$ *является группой преобразований.*

Формальное доказательство сводится к тривиальным проверкам: если

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = C \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}$$

и

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = C' \begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} a' \\ b' \end{pmatrix},$$

то

$$\begin{pmatrix} x \\ y \end{pmatrix} = C^{-1} \begin{pmatrix} x' \\ y' \end{pmatrix} - C^{-1} \begin{pmatrix} a \\ b \end{pmatrix}$$

и

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = (C'C) \begin{pmatrix} x \\ y \end{pmatrix} + \left(C' \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} a' \\ b' \end{pmatrix} \right).$$

Поскольку для ортогональных матриц C и C' матрицы C^{-1} и $C'C$ ортогональны (п. 6 § 5 гл. 1), тем самым доказано, что если преобразования α и α' выражаются формулой вида (1), то преобразования α^{-1} и $\alpha' \circ \alpha$ также обладают этим свойством. Но это, по определению, и означает, что $\mathrm{Ort}(\mathbb{R}^2)$ является группой.

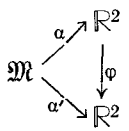
Теперь мы уже можем сформулировать наше основное определение.

Пусть нам дано некоторое множество $\mathfrak{M}$ вместе с непустым семейством $\mathrm{Coor}(\mathfrak{M})$ биективных отображений этого множества на множество $\mathbb{R}^2$ и пусть выполнены следующие две аксиомы:

Аксиома 1. *Для любого отображения* α: $\mathfrak{M} \to \mathbb{R}^2$, *принадлежащего семейству* $\mathrm{Coor}(\mathfrak{M})$, *и любого преобразования* φ: $\mathbb{R}^2 \to \mathbb{R}^2$, *принадлежащего группе* $\mathrm{Ort}(\mathbb{R}^2)$, *отображение* $\varphi \circ \alpha$: $\mathfrak{M} \to \mathbb{R}^2$ *принадлежит семейству* $\mathrm{Coor}(\mathfrak{M})$.

Аксиома 2. *Для любых двух отображений* α, α': $\mathfrak{M} \to \mathbb{R}^2$, *принадлежащих семейству* $\mathrm{Coor}(\mathfrak{M})$, *отображение* $\alpha' \circ \alpha^{-1}$: $\mathbb{R}^2 \to \mathbb{R}^2$ *принадлежит группе* $\mathrm{Ort}(\mathbb{R}^2)$.

Наглядно ситуацию, описываемую аксиомой 2, можно изобразить следующей диаграммой:



Определение 2. При выполнении аксиом 1 и 2 мы будем множество $\mathfrak{M}$ называть *евклидовой плоскостью*, его элементы — *точками*, отображения, принадлежащие семейству $\mathrm{Coor}(\mathfrak{M})$, — *евклидовыми координатными системами* на плоскости $\mathfrak{M}$, и для любой евклидовой координатной системы α: $\mathfrak{M} \to \mathbb{R}^2$ и любой точки $M \in \mathfrak{M}$ числа x, y, составляющие пару $\alpha(M) \in \mathfrak{M}$ — *евклидовыми координатами* точки M в координатной системе α (число x — *абсциссой*, а число y — *ординатой*).

Из сказанного в начале этого пункта непосредственно вытекает, что содержательно рассматриваемая обычная плоскость является в смысле этого определения евклидовой плоскостью.

Замечание 1. Совершенно ясно, как можно аналогичным образом формально определить *евклидово пространство* (достаточно во всем вышесказанном вместо множества $\mathbb{R}^2$ рассматривать множество $\mathbb{R}^3$ троек (x, y, z) вещественных чисел).

Более того, ясно, что мы можем на $n = 3$ не останавливаться: совершенно аналогичным образом можно для любого n определить *n-мерное евклидово пространство.*

Докажем теперь непротиворечивость и полноту нашей аксиоматики.

Чтобы доказать непротиворечивость, достаточно указать конкретный пример множества $\mathfrak{M}$ с семейством отображений $\mathrm{Coor}(\mathfrak{M})$, для которого справедливы аксиомы 1 и 2. Примем за $\mathfrak{M}$ множество $\mathbb{R}^2$, а за семейство $\mathrm{Coor}(\mathfrak{M})$ — группу $\mathrm{Ort}(\mathbb{R}^2)$. Тогда аксиомы 1 и 2 будут в точности означать, что семейство преобразований $\mathrm{Ort}(\mathbb{R}^2)$ является группой, и потому будут выполнены. Тем самым непротиворечивость наших аксиом полностью доказана. Вместе с тем доказано, что

множество $\mathbb{R}^2$, рассматриваемое вместе с группой $\mathrm{Ort}(\mathbb{R}^2)$ в качестве семейства евклидовых координатных систем, является евклидовой плоскостью.

Эту плоскость мы будем называть *стандартной* (или *арифметической*) *евклидовой плоскостью.*

Чтобы доказать полноту, мы сначала должны определить, что значит, что две евклидовы плоскости изоморфны.

Определение 3. Пусть $\mathfrak{M}$ и $\mathfrak{M}'$ — две евклидовы плоскости. Отображение

$$\mu\colon \mathfrak{M} \to \mathfrak{M}'$$

плоскости $\mathfrak{M}$ на плоскость $\mathfrak{M}'$ мы будем называть *изоморфным отображением* (или просто *изоморфизмом*), если

1) оно биективно;

2) для любой евклидовой координатной системы $\alpha'\colon \mathfrak{M}' \to \mathbb{R}^2$ на плоскости $\mathfrak{M}'$ отображение

$$\alpha' \circ \mu\colon \mathfrak{M} \to \mathbb{R}^2$$

является евклидовой координатной системой на плоскости $\mathfrak{M}$;

3) для любой евклидовой координатной системы $\alpha\colon \mathfrak{M} \to \mathbb{R}^2$ на плоскости $\mathfrak{M}$ существует такая (очевидно, единственная) евклидова координатная система $\alpha'\colon \mathfrak{M}' \to \mathbb{R}^2$ на плоскости $\mathfrak{M}'$, что

$$\alpha = \alpha' \circ \mu.$$

Короче можно сказать, что биективное отображение $\mu\colon \mathfrak{M} \to \mathfrak{M}'$ является изоморфизмом, если оно устанавливает

биективное соответствие между евклидовыми координатными системами на плоскостях $\mathfrak{M}$ и $\mathfrak{M}'$.

Заметим, что условие 1) определения изоморфизма вытекает из условия 2) или 3). Мы его включили в определение только для большей ясности.

Задание. Докажите, что тождественное отображение $1_{\mathfrak{M}}\colon \mathfrak{M} \to \mathfrak{M}$, а также композиция $\mu' \circ \mu\colon \mathfrak{M} \to \mathfrak{M}''$ двух изоморфизмов $\mu\colon \mathfrak{M} \to \mathfrak{M}'$ и $\mu'\colon \mathfrak{M}' \to \mathfrak{M}''$ являются изоморфизмами.

Две плоскости $\mathfrak{M}$ и $\mathfrak{M}'$ называются *изоморфными*, если между ними существует хотя бы один изоморфизм. Из только что сделанных замечаний непосредственно вытекает, что отношение изоморфности является отношением эквивалентности, так что совокупность всех евклидовых плоскостей распадается на классы изоморфных плоскостей.

Утверждение о полноте нашей аксиоматики означает, что на самом деле существует только один класс изоморфных плоскостей, т. е., иными словами, что

любые две евклидовы плоскости изоморфны.

Ясно, что для этого достаточно доказать, что

любая плоскость $\mathfrak{M}$ изоморфна стандартной плоскости $\mathbb{R}^2$.

В свою очередь это непосредственно вытекает из того, что

каждая евклидова координатная система $\alpha\colon \mathfrak{M} \to \mathbb{R}^2$ представляет собой изоморфизм плоскости $\mathfrak{M}$ на плоскость $\mathbb{R}^2$.

Для доказательства же последнего утверждения достаточно заметить, что условие 2) определения изоморфизма равносильно для отображения $\alpha\colon \mathfrak{M} \to \mathbb{R}^2$ аксиоме 1, а условие 3) — аксиоме 2.

Тем самым полнота аксиом 1 и 2 полностью доказана.

Заметим, что не только любая евклидова координатная система $\alpha\colon \mathfrak{M} \to \mathbb{R}^2$ является изоморфизмом, но и наоборот,

каждый изоморфизм

$$\mu\colon \mathfrak{M} \to \mathbb{R}^2$$

является евклидовой координатной системой на плоскости $\mathfrak{M}$.

Действительно, по определению, одной из евклидовых координатных систем на плоскости $\mathbb{R}^2$ является тождественное отображение $1_{\mathbb{R}^2}\colon \mathbb{R}^2 \to \mathbb{R}^2$. Поэтому согласно условию 2) определения изоморфизма отображение $1_{\mathbb{R}^2} \circ \mu = \mu$ является евклидовой координатной системой на плоскости $\mathfrak{M}$.

Замечание 2. Несмотря на то, что любые две евклидовы плоскости изоморфны, отождествлять их не очень удобно, поскольку никакого естественного однозначно определенного изоморфизма между ними установить, вообще говоря, нельзя.

Пусть $\mathfrak{M}$ и $\mathfrak{M}'$ — две евклидовы плоскости и пусть

$$\alpha\colon \mathfrak{M} \to \mathbb{R}^2, \quad \alpha'\colon \mathfrak{M}' \to \mathbb{R}^2$$

— произвольные евклидовы координатные системы на этих плоскостях. Поскольку отображение, обратное к изоморфизму, является изоморфизмом, и композиция двух изоморфизмов также представляет собой изоморфизм, отображение

$$\mu = \alpha^{-1} \circ \alpha'$$

плоскости $\mathfrak{M}'$ на плоскость $\mathfrak{M}$ является изоморфизмом.

Так как

$$\alpha \circ \mu = \alpha',$$

то для любой точки $M' \in \mathfrak{M}'$ равенство $M = \mu(M')$ равносильно равенству

$$\alpha(M) = \alpha'(M'). \tag{2}$$

Но, по определению, $\alpha(M)$ является парой $(x, y) \in \mathbb{R}^2$, состоящей из координат точки M в координатной системе α, а $\alpha'(M')$ — парой $(x', y') \in \mathbb{R}^2$, состоящей из координат точки M' в координатной системе α'. Следовательно, равенство (2) означает, что точки M и M' имеют одни и те же координаты (в координатных системах α и α' соответственно). На этом основании говорят, что изоморфизм μ *действует по равенству координат* в координатных системах α и α'.

Легко видеть, что

любой изоморфизм

$$\mu\colon \mathfrak{M}' \to \mathfrak{M}$$

действует по равенству координат в некоторых координатных системах α и α', причем одну из этих систем можно задать произвольно.

Действительно, пусть, например, нам задана координатная система

$$\alpha\colon \mathfrak{M} \to \mathbb{R}^2.$$

Рассмотрим отображение

$$\alpha' = \alpha \circ \mu\colon \mathfrak{M}' \to \mathbb{R}^2.$$

Будучи композицией изоморфизмов, это отображение представляет собой изоморфизм, т. е. является евклидовой координатной системой на плоскости $\mathfrak{M}'$. Для завершения доказательства остается заметить, что

$$\mu = \alpha^{-1} \circ \alpha'.$$

Собирая вместе доказанные утверждения, мы получаем следующее предложение, описывающее взаимоотношения между изоморфизмами и координатными системами.

Предложение 1. *Отображение*

$$\alpha\colon \mathfrak{M} \to \mathbb{R}^2$$

тогда и только тогда является евклидовой координатной системой, когда оно представляет собой изоморфизм.

Отображение

$$\mu\colon \mathfrak{M}' \to \mathfrak{M}$$

тогда и только тогда является изоморфизмом, когда существуют такие евклидовы координатные системы

$$\alpha\colon \mathfrak{M} \to \mathbb{R}^2, \quad \alpha'\colon \mathfrak{M}' \to \mathbb{R}^2,$$

что μ *действует по равенству координат в этих системах*:

$$\mu = \alpha^{-1} \circ \alpha'.$$

Особое значение имеют *автоморфизмы* евклидовой плоскости $\mathfrak{M}$, т. е. ее изоморфизмы на себя:

$$\Phi\colon \mathfrak{M} \to \mathfrak{M}.$$

Автоморфизмами стандартной плоскости $\mathbb{R}^2$ являются ее евклидовы координатные системы

$$\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$$

и только они (напомним, что, по определению, $\mathrm{Coor}(\mathbb{R}^2) = \mathrm{Ort}(\mathbb{R}^2)$). Другими словами,

преобразование плоскости $\mathbb{R}^2$ *тогда и только тогда является автоморфизмом, когда оно принадлежит группе* $\mathrm{Ort}(\mathbb{R}^2)$.

Автоморфизмы любой евклидовой плоскости $\mathfrak{M}$ составляют, очевидно, группу. Эту группу мы будем обозначать символом $\mathrm{Ort}(\mathfrak{M})$.

Легко видеть, что

для любых двух плоскостей $\mathfrak{M}$ *и* $\mathfrak{M}'$ *группы* $\mathrm{Ort}(\mathfrak{M})$ *и* $\mathrm{Ort}(\mathfrak{M}')$ *изоморфны.*

Действительно, выбрав некоторый изоморфизм

$$\mu\colon \mathfrak{M} \to \mathfrak{M}',$$

мы получим, очевидно, изоморфизм

$$\mathrm{Ort}(\mathfrak{M}) \to \mathrm{Ort}(\mathfrak{M}'),$$

сопоставив произвольному преобразованию $\Phi \in \mathrm{Ort}(\mathfrak{M})$ преобразование

$$\mu \circ \Phi \circ \mu^{-1} \in \mathrm{Ort}(\mathfrak{M}').$$

Об изоморфизме

$$\Phi \mapsto \mu \circ \Phi \circ \mu^{-1}$$

мы будем говорить, что он *индуцирован* изоморфизмом μ.

В частности, любая евклидова координатная система

$$\alpha\colon \mathfrak{M} \to \mathbb{R}^2$$

индуцирует изоморфизм

$$\Phi \longmapsto \alpha \circ \Phi \circ \alpha^{-1}$$

группы $\mathrm{Ort}(\mathfrak{M})$ на группу $\mathrm{Ort}(\mathbb{R}^2)$. Это означает, что координаты x, y произвольной точки $M \in \mathfrak{M}$ (в координатной системе α) связаны с координатами x', y' преобразованной точки $M' = \Phi(M)$ (в той же координатной системе α) формулой вида (1).

Принято говорить, что формула (1) *выражает* автоморфизм Φ в координатной системе α.

Мы видим, таким образом, что, рассматриваемая как абстрактная группа (т. е. «с точностью до изоморфизма»), группа $\mathrm{Ort}(\mathfrak{M})$ не зависит от выбора плоскости $\mathfrak{M}$.

Значение группы $\mathrm{Ort}(\mathfrak{M})$ определяется тем, что она позволяет обозреть всю совокупность изоморфизмов плоскости $\mathfrak{M}$ на любую другую плоскость $\mathfrak{M}'$. Точнее, это означает следующее.

Пусть нам задан некоторый изоморфизм

$$\mu_0\colon\ \mathfrak{M} \to \mathfrak{M}'.$$

Тогда для любого автоморфизма $\Phi \in \mathrm{Ort}(\mathfrak{M})$ отображение

$$\mu = \mu_0 \circ \Phi\colon\ \mathfrak{M} \to \mathfrak{M}'$$

также является изоморфизмом плоскости $\mathfrak{M}$ на плоскость $\mathfrak{M}'$. Тем самым мы получаем некоторое отображение

$$\Phi \longmapsto \mu \tag{3}$$

группы $\mathrm{Ort}(\mathfrak{M})$ в множество всех изоморфизмов $\mathfrak{M} \to \mathfrak{M}'$.

Задание. Докажите, что отображение (3) биективно.

Таким образом, грубо говоря, изоморфизмов $\mathfrak{M} \to \mathfrak{M}'$ существует «столько же», сколько автоморфизмов $\mathfrak{M} \to \mathfrak{M}$.

Определение 4. Пусть X и X' — некоторые множества точек евклидовой плоскости $\mathfrak{M}$ («фигуры»). Эти множества называются (*евклидово*) *эквивалентными*, если существует автоморфизм плоскости $\mathfrak{M}$, переводящий одно множество в другое.

Свойство $\mathscr{P}$ множеств плоскости $\mathfrak{M}$ называется (*евклидово*) *инвариантным* (или *геометрическим*), если одновременно с множеством X свойством $\mathscr{P}$ обладает и любое множество X', ему евклидово эквивалентное.

Класс (множество) фигур называется *инвариантно определенным*, если он характеризуется (см. п. 2 § 1 гл. 1) семейством инвариантных свойств.

Числовая функция $d(X)$, заданная на инвариантно определенном классе, называется *инвариантной функцией* (или просто *инвариантом*), если для любого числа d свойство «$d(X) = d$»

фигур X инвариантно, т. е. если для любых евклидово эквивалентных фигур X и X' имеет место равенство

$$d(X) = d(X').$$

Пусть $\mathfrak{M}$ и $\mathfrak{M}'$ — две евклидовы плоскости и пусть $\mathscr{P}$ — некоторое инвариантное свойство фигур на плоскости $\mathfrak{M}$. Тогда мы можем определить некоторое (тоже инвариантное) свойство $\mathscr{P}'$ фигур на плоскости $\mathfrak{M}'$. Именно, выбрав произвольный изоморфизм

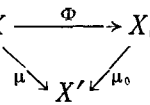
$$\mu_0\colon \mathfrak{M} \to \mathfrak{M}',$$

мы условимся, что фигура X' на плоскости $\mathfrak{M}'$ тогда и только тогда обладает свойством $\mathscr{P}'$, когда свойством $\mathscr{P}$ обладает фигура X_0 на плоскости $\mathfrak{M}$, переходящая при изоморфизме μ_0 в фигуру X'.

Это определение нуждается, конечно, в проверке корректности, т. е. в доказательстве того, что при другом выборе изоморфизма μ_0 получается то же самое свойство $\mathscr{P}'$. Но, как мы знаем, любой изоморфизм $\mu\colon \mathfrak{M} \to \mathfrak{M}'$ имеет вид

$$\mu = \mu_0 \circ \Phi,$$

где Φ — некоторый автоморфизм плоскости $\mathfrak{M}$. Следовательно, фигурой X, переходящей при изоморфизме μ в фигуру X', является фигура, переходящая в фигуру X_0 при автоморфизме Φ:



Для завершения доказательства остается заметить, что в силу инвариантности свойства $\mathscr{P}$ фигура X тогда и только тогда обладает этим свойством, когда им обладает фигура X_0.

Задание. Докажите, что свойство $\mathscr{P}'$ инвариантно.

Ясно, что изложенное построение устанавливает естественное биективное соответствие между инвариантными свойствами фигур на различных евклидовых плоскостях. Следовательно, изучение инвариантных свойств мы можем, как и следовало ожидать, производить на любой конкретной евклидовой плоскости. Результаты будут автоматически применимы к любой другой плоскости.

Определение 5. Предметом *евклидовой геометрии* (точнее, евклидовой планиметрии) является изучение евклидово инвариантных свойств фигур на евклидовых плоскостях.

Таким образом, согласно этому определению, «геометрический смысл» имеют только инвариантные свойства (и значит, в частности, инвариантные классы и инвариантные функции). По определению, все евклидово инвариантные свойства евклидово эквивалентных фигур одинаковы. Поэтому можно сказать

(ср. п. 2 § 1 гл. 1), что в евклидовой геометрии эквивалентные фигуры рассматриваются как «одинаковые».

Определение 5 является формальным уточнением несколько туманной идеи об евклидовой геометрии как науке, изучающей свойства «абстрактной евклидовой плоскости», конкретными воплощениями которой являются евклидовы плоскости в смысле определения 2. Оно согласуется с содержательным представлением об евклидовой геометрии как науке, изучающей свойства, общие у всех конгруэнтных фигур (переводящихся друг в друга некоторым движением), поскольку, как мы покажем в гл. 7, автоморфизмы евклидовой плоскости содержательно интерпретируются как движения (сопровождаемые, возможно, симметрией).

С описываемой точки зрения аналитическая геометрия может быть охарактеризована как наука, занимающаяся изучением евклидовой геометрии на ее стандартной модели $\mathbb{R}^2$. Переход от произвольной евклидовой плоскости $\mathfrak{M}$ к стандартной плоскости $\mathbb{R}^2$ осуществляется при этом выбором некоторой евклидовой координатной системы

$$\alpha\colon \mathfrak{M} \to \mathbb{R}^2.$$

Таким образом, вместо того чтобы изучать инвариантные свойства фигур на плоскости $\mathfrak{M}$, в аналитической геометрии изучают инвариантные свойства соответствующих фигур на плоскости $\mathbb{R}^2$. Можно сказать, что аналитическая геометрия основывается на отождествлении плоскости $\mathfrak{M}$ с плоскостью $\mathbb{R}^2$ посредством координатной системы α.

Само собой разумеется, что при этом нужно быть очень внимательным и рассматривать только инвариантные свойства фигур плоскости $\mathbb{R}^2$, т. е. (см. выше) свойства, перенесение которых на плоскость $\mathfrak{M}$ не зависит от выбора координатной системы α.

Пример 1. Свойство «координаты точки равны» неинвариантно. Например, при автоморфизме

$$x' = x,$$
$$y' = y + 1$$

точка (x, y), для которой $x = y$, переходит в точку (x', y'), для которой $x' \neq y'$. Таким образом, это свойство никакого «геометрического» (не зависящего от выбора координатной системы) смысла не имеет.

Можно показать, что вообще точка (т. е. фигура, состоящая из одной точки) никаких инвариантных свойств не имеет. Это связано с тем, что любую точку можно перевести в любую другую некоторым автоморфизмом (движением).

Пример 2. *Расстоянием* $d = d(A_1, A_2)$ между точками $A_1(x_1, y_1)$ и $A_2(x_2, y_2)$ называется число

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Покажем, что это число инвариантно (т. е. что расстояние d имеет «геометрический смысл»). Для этого мы должны доказать, что вычисленное в любой другой евклидовой координатной системе число d остается тем же самым.

Но, по определению, переход к другой координатной системе описывается формулами вида

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = C \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix},$$

где C — ортогональная матрица, а a и b — числа. В частности, если матрица C — собственная, т. е. если

$$C = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

то

$$x' = x\cos\theta - y\sin\theta + a,$$
$$y' = x\sin\theta + y\cos\theta + b,$$

и потому

$$x_2' - x_1' = (x_2 - x_1)\cos\theta - (y_2 - y_1)\sin\theta,$$
$$y_2' - y_1' = (x_2 - x_1)\sin\theta + (y_2 - y_1)\cos\theta.$$

Следовательно,

$$(x_2' - x_1')^2 + (y_2' - y_1')^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2.$$

Аналогично инвариантность числа d доказывается и для случая, когда матрица C несобственная.

Таким образом, *фигура, состоящая из двух точек, обладает инвариантом* d. Можно показать, что любой другой инвариант двух точек имеет вид $f(d)$, где f — некоторая функция. Это связано с тем, что пару точек A_1, A_2 можно автоморфизмом перевести в пару точек A_1', A_2' тогда и только тогда, когда $d(A_1', A_2') = d(A_1, A_2)$.

Замечание 3. Пример 2 иллюстрирует общий метод введения основных геометрических понятий при формально-аксиоматическом построении евклидовой геометрии: берется формула, известная из содержательных рассмотрений, и принимается за определение соответствующего понятия. Корректность обеспечивается содержательным смыслом, но, конечно, должна быть формально проверена. Дальнейшие примеры такого рода введения геометрических понятий мы рассмотрим после того, как аналогичным образом аксиоматизируем аффинную геометрию.

3. Аксиоматика аффинной геометрии

Мы пришли к аксиоматике евклидовой геометрии, дав формальное описание евклидовых координатных систем. Чтобы построить аффинную геометрию, достаточно аналогичным образом описать аффинные координатные системы.

В соответствии со сказанным в п. 2 § 1 о формулах перехода от одной аффинной координатной системы к другой, мы будем основываться на преобразованиях

$$\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2,$$

задаваемых формулами вида

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = C \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix},$$

где a, b — числа, а

$$C = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}$$

— произвольная невырожденная матрица, т. е. формулами

$$\begin{aligned} x' &= a_1 x + a_2 y + a, \\ y' &= b_1 x + b_2 y + b, \end{aligned} \qquad \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \neq 0. \tag{1}$$

Определение 1. Множество всех преобразований вида (1) мы будем обозначать символом $\mathrm{Aff}(\mathbb{R}^2)$.

Та же выкладка, что и для группы $\mathrm{Ort}(\mathbb{R}^2)$, показывает, что *множество* $\mathrm{Aff}(\mathbb{R}^2)$ *является группой преобразований.*

Пусть теперь нам дано некоторое множество $\mathfrak{M}$ вместе с непустым семейством $\mathrm{Coor}(\mathfrak{M})$ биективных отображений этого множества на множество $\mathbb{R}^2$ и пусть выполнены следующие две аксиомы:

Аксиома 1. *Для любого отображения* $\alpha\colon \mathfrak{M} \to \mathbb{R}^2$, *принадлежащего семейству* $\mathrm{Coor}(\mathfrak{M})$, *и любого преобразования* $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$, *принадлежащего группе* $\mathrm{Aff}(\mathbb{R}^2)$, *отображение* $\varphi \circ \alpha\colon \mathfrak{M} \to \mathbb{R}^2$ *принадлежит семейству* $\mathrm{Coor}(\mathfrak{M})$.

Аксиома 2. *Для любых двух отображений* α, $\alpha'\colon \mathfrak{M} \to \mathbb{R}^2$, *принадлежащих семейству* $\mathrm{Coor}(\mathfrak{M})$, *отображение* $\alpha' \circ \alpha^{-1}\colon \mathbb{R}^2 \to \mathbb{R}^2$ *принадлежит группе* $\mathrm{Aff}(\mathbb{R}^2)$.

Содержательно аксиома 1 означает, что для любых аффинных координат x, y величины x', y', выражающиеся формулами (1), также являются аффинными координатами, а аксиома 2 означает, что любые две системы аффинных координат x, y и x', y' связаны формулами (1).

Определение 2. При выполнении аксиом 1 и 2 мы будем множество $\mathfrak{M}$ называть *аффинной плоскостью*, его элементы — *точками*, отображения, принадлежащие семейству $\mathrm{Coor}(\mathfrak{M})$, — *аффинными координатными системами* на плоскости $\mathfrak{M}$, и для любой аффинной координатной системы $\alpha \in \mathrm{Coor}(\mathfrak{M})$ и любой точки $M \in \mathfrak{M}$ числа x, y, составляющие пару $\alpha(M) \in \mathbb{R}^2$, — *аффинными координатами* точки M в координатной системе α (число x — *абсциссой*, а число y — *ординатой*).

Замечание 1. Совершенно аналогично определяется *аффинное пространство.*

Мы видим, что аксиомы аффинной геометрии вполне аналогичны аксиомам евклидовой геометрии из п. 2. Поэтому все сказанное в п. 2 автоматически переносится (с незначительными изменениями) и на рассматриваемую сейчас ситуацию.

Например, для доказательства непротиворечивости наших аксиом достаточно заметить (ср. п. 2), что

множество $\mathbb{R}^2$ *будет аффинной плоскостью в смысле определения 2, если за семейство* $\mathrm{Coor}(\mathbb{R}^2)$ *принять группу* $\mathrm{Aff}(\mathbb{R}^2)$.

Эту плоскость мы будем называть *стандартной* (или *арифметической*) *аффинной плоскостью.*

Замечание 2. Таким образом, одно и то же множество $\mathbb{R}^2$ мы можем рассматривать и как аффинную и как евклидову плоскости — в зависимости от того, что мы принимаем за семейство $\mathrm{Coor}(\mathbb{R}^2)$. Вообще говоря, во избежание путаницы, целесообразно отразить это в обозначениях (например, можно писать $\mathbb{R}^2_{\text{орт}}$ и $\mathbb{R}^2_{\text{афф}}$).

Изоморфизм аффинных плоскостей определяется дословно так же, как изоморфизм евклидовых плоскостей (см. определение 3 п. 2), после чего мы немедленно получаем (ср. п. 2), что

каждая аффинная координатная система $\alpha\colon \mathfrak{M} \to \mathbb{R}^2$ *представляет собой изоморфизм плоскости* $\mathfrak{M}$ *на плоскость* $\mathbb{R}^2$.

Доказанное в п. 2 предложение 1 также, конечно, сохраняется (вместе с доказательством) и для аффинных плоскостей. В частности,

отображение $\mu\colon \mathfrak{M}' \to \mathfrak{M}$ *одной аффинной плоскости на другую тогда и только тогда является изоморфизмом, когда оно действует по равенству координат в двух аффинных координатных системах* α *и* α' *(т. е. когда* $\mu = \alpha^{-1} \circ \alpha'$).

В частности, это верно и для *автоморфизмов* произвольной аффинной плоскости (т. е. для ее изоморфизмов на себя).

Автоморфизмами стандартной плоскости $\mathbb{R}^2$ являются преобразования из группы $\mathrm{Aff}(\mathbb{R}^2)$ и только они.

Для любой аффинной плоскости $\mathfrak{M}$ группу всех ее автоморфизмов мы будем обозначать символом $\mathrm{Aff}(\mathfrak{M})$.

Так же, как в п. 2, немедленно доказывается, что

любой изоморфизм $\mu\colon \mathfrak{M} \to \mathfrak{M}'$ *индуцирует некоторый изоморфизм групп*

$$\mathrm{Aff}(\mathfrak{M}) \to \mathrm{Aff}(\mathfrak{M}').$$

В частности, любая формула вида (1) выражает некоторый автоморфизм Φ (в данной аффинной координатной системе $\alpha\colon \mathfrak{M} \to \mathbb{R}^2$).

Задание произвольного изоморфизма

$$\mu_0\colon \mathfrak{M} \to \mathfrak{M}'$$

двух аффинных плоскостей определяет по формуле

$$\Phi \longmapsto \mu_0 \circ \Phi$$

биективно отображение группы $\mathrm{Aff}(\mathfrak{M})$ на множество всех изоморфизмов $\mathfrak{M} \to \mathfrak{M}'$.

Определение 3. Две фигуры на аффинной плоскости $\mathfrak{M}$ называются *аффинно эквивалентными,* если одну можно получить из другой некоторым автоморфизмом $\Phi\colon \mathfrak{M} \to \mathfrak{M}$. Свойство фигур на аффинной плоскости называется *аффинно инвариантным,* если одновременно с фигурой X этим свойством обладает и любая фигура X', ей аффинно эквивалентная. Аналогично определяются *аффинно инвариантные* функции (ср. определение 4 п. 2).

Подобно евклидово инвариантным свойствам (см. п. 2) аффинно инвариантные свойства фигур на двух различных аффинных плоскостях находятся в естественном биективном соответствии. Это позволяет формально определить *аффинную геометрию* как науку о таких свойствах. Иначе можно сказать, что аффинная геометрия — это наука, в которой аффинно эквивалентные фигуры считаются «одинаковыми».

Мы не будем здесь детально развивать аффинную геометрию на основе этого определения, поскольку это без труда делается, исходя из ее содержательного представления на основе общего принципа, изложенного в замечании 4 п. 2. Мы ограничимся для примера только некоторыми замечаниями, касающимися понятия вектора.

Направленные отрезки определяются в формальной теории точно так же, как в содержательной, т. е. как пары, состоящие из двух точек рассматриваемой аффинной плоскости $\mathfrak{M}$. Два направленных отрезка $\overrightarrow{A_1A_2}$ и $\overrightarrow{A_1^*A_2^*}$ называются *эквиполлентными,* если в некоторой аффинной координатной системе $\alpha\colon \mathfrak{M} \to \mathbb{R}^2$ имеют место равенства

$$x_2^* - x_1^* = x_2 - x_1,$$
$$y_2^* - y_1^* = y_2 - y_1,$$

где (x_1, y_1) и (x_2, y_2) — координаты точек A_1 и A_2, а (x_1^*, y_1^*) и (x_2^*, y_2^*) — координаты точек A_1^* и A_2^*.

Конечно, здесь следует проверить корректность этого определения, т. е. независимость его от выбора аффинной координатной системы. Но это делается без труда.

Действительно, пусть

$$x_1' = a_1x_1 + a_2y_1 + a, \quad x_2' = a_1x_2 + a_2y_2 + a,$$
$$y_1' = b_1x_1 + b_2y_1 + b, \quad y_2' = b_1x_2 + b_2y_2 + b,$$
$$x_1^{*\prime} = a_1x_1^* + a_2y_1^* + a, \quad x_2^{*\prime} = a_1x_2^* + a_2y_2^* + a,$$
$$y_1^{*\prime} = b_1x_1^* + b_2y_1^* + b, \quad y_2^{*\prime} = b_1x_2^* + b_2y_2^* + b$$

— координаты точек A_1, A_2, A_1^*, A_2^* в некоторой другой аффинной координатной системе. Тогда

$$\begin{aligned} x_2' - x_1' &= a_1(x_2 - x_1) + a_2(y_2 - y_1), \\ y_2' - y_1' &= b_1(x_2 - x_1) + b_2(y_2 - y_1) \end{aligned} \tag{2}$$

и

$$\begin{aligned} x_2^{*\prime} - x_1^{*\prime} &= a_1(x_2^* - x_1^*) + a_2(y_2^* - y_1^*), \\ y_2^{*\prime} - y_1^{*\prime} &= b_1(x_2^* - x_1^*) + b_2(y_2^* - y_1^*). \end{aligned}$$

Следовательно, если $x_2 - x_1 = x_2^* - x_1^*$ и $y_2 - y_1 = y_2^* - y_1^*$, то $x_2' - x_1' = x_2^{*\prime} - x_1^{*\prime}$ и $y_2' - y_1' = y_2^{*\prime} - y_1^{*\prime}$.

Ясно, что отношение эквиполлентности направленных отрезков является отношением эквивалентности, так что определены классы эквиполлентных направленных отрезков. Эти классы и называются *векторами*.

Допуская вольность, мы будем (как и в содержательной теории) вектор, определенный направленным отрезком $\overrightarrow{A_1A_2}$, обозначать тем же символом $\overrightarrow{A_1A_2}$.

Поскольку, по определению, числа

$$x_2 - x_1, \quad y_2 - y_1$$

— одни и те же для всех эквиполлентных направленных отрезков $\overrightarrow{A_1A_2}$, они однозначно определяются данным вектором $\overrightarrow{A_1A_2}$. Эти числа называются *координатами* вектора $\overrightarrow{A_1A_2}$ в рассматриваемой аффинной координатной системе.

Ясно, что соответствие

«вектор» $\longmapsto$ «его координаты»

является биективным соответствием между множеством всех векторов на плоскости $\mathfrak{M}$ и множеством $\mathbb{R}^2$ всех пар (x, y).

Согласно полученным выше формулам (2) при переходе к другой аффинной координатной системе координаты векторов преобразуются по формулам

$$\begin{aligned} x' &= a_1x + a_2y, \\ y' &= b_1x + b_2y, \end{aligned} \quad \left|\begin{matrix} a_1 & a_2 \\ b_1 & b_2 \end{matrix}\right| \neq 0. \tag{2}$$

Подчеркнем, что это преобразование однородно (не содержит свободных членов), тогда как преобразования координат точек неоднородны.

Теперь уже ясно, что следует называть *суммой* векторов, *произведением* вектора на число и т. п. Легко видеть, что вся развитая в гл. 1 теория векторов (в ее аффинной части) полностью при этом сохраняется.

Репером произвольной аффинной координатной системы $\alpha\colon \mathfrak{M} \to \mathbb{R}^2$ называется тройка точек (O, E_1, E_2), переходящих

соответственно в пары (0, 0), (1, 0) и (0, 1). Репер однозначно определен, если задана точка O — *начало* координатной системы и векторы $\boldsymbol{e}_1 = \overrightarrow{OE}_1$ и $\boldsymbol{e}_2 = \overrightarrow{OE}_2$. Поэтому репером можно называть и тройку $O\boldsymbol{e}_1\boldsymbol{e}_2$.

Радиус-вектор $\overrightarrow{OA}$ произвольной точки $A(x, y)$ имеет координаты (x, y) и потому представляется в виде

$$\overrightarrow{OA} = x\boldsymbol{e}_1 + y\boldsymbol{e}_2.$$

Следовательно, зная репер аффинной координатной системы, мы можем найти координаты любой точки плоскости, т. е. полностью восстановить эту координатную систему. Иными словами,

аффинная координатная система однозначно определяется своим репером.

Покажем теперь, что

любая тройка $O\boldsymbol{e}_1\boldsymbol{e}_2$, *состоящая из произвольной точки* O *и двух линейно независимых векторов* $\boldsymbol{e}_1$ *и* $\boldsymbol{e}_2$, *является репером некоторой* (*однозначно определенной*) *аффинной координатной системы.*

Действительно, поскольку точка O дана, каждой точке A отвечает ее радиус-вектор $\overrightarrow{OA}$. Поскольку векторы $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$ линейно независимы, вектор $\overrightarrow{OA}$ однозначно по ним раскладывается:

$$\overrightarrow{OA} = x\boldsymbol{e}_1 + y\boldsymbol{e}_2. \tag{3}$$

Сопоставив точке A полученную таким образом пару (x, y), мы построим, следовательно, некоторое отображение α: $\mathfrak{M} \to \mathbb{R}^2$. Ясно, что при этом отображении точка O переходит в пару (0, 0), а концы E_1 и E_2 векторов $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$, отложенных от точки O, — в пары (1, 0) и (0, 1). Нам остается только доказать, что построенное отображение α: $\mathfrak{M} \to \mathbb{R}^2$ является аффинной координатной системой.

Рассмотрим с этой целью произвольную аффинную координатную систему

$$\alpha_0\colon\ \mathfrak{M} \to \mathbb{R}^2.$$

Пусть в этой системе векторы $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$ имеют координаты (a_1, b_1) и (a_2, b_2). Тогда ввиду равенства (3) вектор $\overrightarrow{OA}$ будет иметь координаты

$$(a_1x + a_2y,\quad b_1x + b_2y).$$

Далее, пусть (a, b) — координаты точки O в координатной системе α_0. Тогда координаты точки A будут иметь вид

$$(a_1x + a_2y + a,\quad b_1x + b_2y + b).$$

Это означает, что если мы введем в рассмотрение отображение

$$\varphi\colon\ \mathbb{R}^2 \to \mathbb{R}^2,$$

определенное формулами

$$x' = a_1 x + a_2 y + a,$$
$$y' = b_1 x + b_2 y + b,$$

то отображение $\varphi \circ \alpha$ будет совпадать с отображением α_0.

Но поскольку векторы e_1 и e_2 линейно независимы, определитель

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$

отличен от нуля, и потому отображение φ принадлежит группе $\mathrm{Aff}(\mathbb{R}^2)$. Следовательно, это отображение биективно и обратное к нему отображение φ^{-1} также принадлежит группе $\mathrm{Aff}(\mathbb{R}^2)$. Таким образом,

$$\alpha = \varphi^{-1} \circ \alpha_0,$$

где $\varphi^{-1} \in \mathrm{Aff}(\mathbb{R}^2)$. Но согласно аксиоме 1 отображение $\varphi^{-1} \circ \alpha_0$ является аффинной координатной системой. Следовательно, аффинной координатной системой является и отображение α.

Таким образом, мы видим, что и в формально-аксиоматической теории аффинные координатные системы могут быть описаны точно так же, как и в содержательной теории.

Рассмотрим теперь с формально-аксиоматической точки зрения вопрос о соотношении между евклидовой и аффинной геометриями.

Пусть $\mathfrak{M}$ — произвольная евклидова плоскость (чтобы подчеркнуть евклидовость, мы будем эту плоскость иногда обозначать символом $\mathfrak{M}_{\text{евкл}}$). Выбрав некоторую евклидову координатную систему

$$\alpha_0 \colon \mathfrak{M} \to \mathbb{R}^2,$$

рассмотрим всевозможные отображения $\alpha \colon \mathfrak{M} \to \mathbb{R}^2$, имеющие вид

$$\alpha = \varphi \circ \alpha_0, \tag{4}$$

где φ — произвольное преобразование из группы $\mathrm{Aff}(\mathbb{R}^2)$.

Задание. Докажите, что семейство всех преобразований вида (4) удовлетворяет аксиомам 1 и 2 определения аффинной плоскости.

Определение 4. Плоскость $\mathfrak{M}$, снабженную семейством аффинных координатных систем вида (4), мы будем называть аффинной плоскостью, *определенной* евклидовой плоскостью $\mathfrak{M}_{\text{евкл}}$, и будем обозначать ее символом $\mathfrak{M}_{\text{афф}}$.

Упражнение. Докажите корректность определения (4), т. е. тот факт, что семейство $\mathrm{Coor}(\mathfrak{M}_{\text{афф}})$ аффинных координатных систем вида (4) не зависит от выбора евклидовой координатной системы α_0. Указание: воспользуйтесь тем, что группа $\mathrm{Ort}(\mathbb{R}^2)$ является подгруппой группы $\mathrm{Aff}(\mathbb{R}^2)$:

$$\mathrm{Ort}(\mathbb{R}^2) \subset \mathrm{Aff}(\mathbb{R}^2), \tag{5}$$

и, следовательно, среди координатных систем вида (4) содержатся все евклидовы координатные системы на плоскости $\mathfrak{M}_{\text{евкл}}$.

Замечание 3. Обратим внимание на то, что возможен и «обратный» процесс построения по произвольной аффинной плоскости $\mathfrak{M}_{\text{афф}}$ некоторой евклидовой плоскости $\mathfrak{M}_{\text{евкл}}$. Для этого достаточно воспользоваться той же формулой (4), принимая за α_0 некоторую фиксированную аффинную координатную систему, а за φ — произвольное преобразование из группы $\mathrm{Ort}(\mathbb{R}^2)$. Однако получающаяся евклидова плоскость $\mathfrak{M}_{\text{евкл}}$ зависит от выбора координатной системы α_0. Чтобы получить вполне определенную евклидову плоскость, необходимо как-то ограничить выбор системы α_0. Обычно это делается посредством введения в аффинную плоскость некоторой *метрики*. Все эти вопросы подробно разбираются (сразу для любого n) в курсе линейной алгебры. Здесь же для нас достаточно иметь в виду, что одна и та же аффинная плоскость $\mathfrak{M}_{\text{афф}}$ может определяться различными (хотя, конечно, и изоморфными) евклидовыми плоскостями $\mathfrak{M}_{\text{евкл}}$.

Поскольку плоскость $\mathfrak{M}_{\text{афф}}$ однозначно определяется плоскостью $\mathfrak{M}_{\text{евкл}}$, можно (и удобно) несколько сократить терминологию, называя, например, автоморфизмы плоскости $\mathfrak{M}_{\text{афф}}$ *аффинными автоморфизмами* плоскости $\mathfrak{M}_{\text{евкл}} = \mathfrak{M}$ (и в соответствии с этим заменяя обозначение $\mathrm{Aff}(\mathfrak{M}_{\text{афф}})$ обозначением $\mathrm{Aff}(\mathfrak{M})$).

Из включения (5) немедленно вытекает соответствующее включение

$$\mathrm{Ort}(\mathfrak{M}) \subset \mathrm{Aff}(\mathfrak{M})$$

для автоморфизмов плоскости $\mathfrak{M}$, означающее, что

любой евклидов автоморфизм является аффинным автоморфизмом.

Поэтому любое аффинно инвариантное свойство фигур на плоскости $\mathfrak{M}$ (точнее, на плоскости $\mathfrak{M}_{\text{афф}}$) будет и евклидово инвариантным. Другими словами, *любое утверждение аффинной геометрии справедливо и в евклидовой геометрии.* В этом смысле аффинная геометрия является «частью» евклидовой геометрии.

Например, все сказанное выше о векторах в аффинной геометрии (на аффинной плоскости) автоматически справедливо и в евклидовой геометрии (на евклидовой плоскости). Но, конечно, на евклидовой плоскости можно пойти значительно дальше (построить метрическую теорию векторов).

Для этого в первую очередь необходимо определить *скалярное произведение* $\boldsymbol{a}_1\boldsymbol{a}_2$ двух векторов $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$.

Пусть в некоторой евклидовой координатной системе α вектор $\boldsymbol{a}_1$ имеет координаты (x_1, y_1), а вектор $\boldsymbol{a}_2$ — координаты (x_2, y_2). Тогда мы, по определению, положим

$$\boldsymbol{a}_1\boldsymbol{a}_2 = x_1x_2 + y_1y_2.$$

В любой другой евклидовой координатной системе α' координаты x', y' векторов связаны с их координатами x, y в системе α формулами вида (2), где теперь

$$\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}$$

— произвольная *ортогональная* матрица, т. е. такая (см. п. 6 § 5 гл. 1), что

$$a_1^2 + b_1^2 = 1, \quad a_1a_2 + b_1b_2 = 0, \quad a_2^2 + b_2^2 = 1.$$

Следовательно, вычисляя скалярное произведение в координатной системе α', мы получим

$$\begin{aligned} \boldsymbol{a}_1\boldsymbol{a}_2 = x_1'x_2' + y_1'y_2' = (a_1x_1 + a_2y_1)(a_1x_2 + a_2y_2) + \\ + (b_1x_1 + b_2y_1)(b_1x_2 + b_2y_2) = (a_1^2 + b_1^2)x_1x_2 + \\ + (a_1a_2 + b_1b_2)(x_1y_2 + x_2y_1) + (a_2^2 + b_2^2)y_1y_2 = x_1x_2 + y_1y_2, \end{aligned}$$

т. е. тот же результат, что и в координатной системе α.

Это означает, что наше определение скалярного произведения корректно.

Определив скалярное произведение, мы можем теперь определить *длину* $|\boldsymbol{a}|$ вектора $\boldsymbol{a}$ и *угол* $|\widehat{\boldsymbol{a}, \boldsymbol{b}}|$ между векторами $\boldsymbol{a}$ и $\boldsymbol{b}$ известными формулами

$$|\boldsymbol{a}| = \sqrt{\boldsymbol{a}^2}, \quad (\widehat{\boldsymbol{a}, \boldsymbol{b}}) = \arccos \frac{\boldsymbol{ab}}{|\boldsymbol{a}| \cdot |\boldsymbol{b}|}. \tag{6}$$

Задание. Докажите, что $\boldsymbol{a}^2 \geqslant 0$ и $\boldsymbol{ab} \leqslant |\boldsymbol{a}| \cdot |\boldsymbol{b}|$ (неравенство Коши — Буняковского), т. е. что формулы (6) действительно определяют некоторые числа.

Обратим внимание на то, что длина вектора $\boldsymbol{a} = \overrightarrow{A_1A_2}$ совпадает с расстоянием $d(A_1, A_2)$ между точками A_1 и A_2, как оно было определено в примере 2 п. 2.

Тот факт, что аффинная геометрия является частью евклидовой геометрии, позволяет доказывать теоремы аффинной геометрии «евклидовыми методами», т. е. с использованием метрических понятий. Формально это означает, что от аффинной плоскости $\mathfrak{M}_{\text{афф}}$ мы переходим к одной из евклидовых плоскостей $\mathfrak{M}_{\text{евкл}}$, определяющих плоскость $\mathfrak{M}_{\text{афф}}$. Конечно, мы должны при этом проследить, чтобы окончательный результат был аффинно инвариантен (не зависел от выбора евклидовой плоскости $\mathfrak{M}_{\text{евкл}}$).

Например, данное в начале п. 1 § 6 определение бивектора использует метрические понятия (площадь). Чтобы установить его аффинный характер, мы были вынуждены доказать его равносильность с другим, уже чисто аффинным определением (основывающимся на понятии элементарного преобразования).

4. Аффинная геометрия над полем комплексных чисел

При обсуждении в п. 2 § 2 теоремы единственности для алгебраических линий мы столкнулись с желательностью введения точек с комплексными координатами. Как мы увидим в гл. 6, уже в теории линий второго порядка такие точки оказываются совершенно необходимыми.

Конечно, введение в рассмотрение подобного рода «точек» означает окончательный переход к алгебре, лишь излагаемой в геометрических терминах. Все то же самое можно сделать, вообще не пользуясь геометрической терминологией. Однако опыт показывает, что даже чисто алгебраические вещи, изложенные на геометрическом языке, приобретают особую четкость и выразительность, и их «геометризация» оказывается не просто игрой в определения, а мощным орудием исследования, позволяющим привлечь к изучению алгебраических объектов геометрические методы и геометрическую интуицию.

Как же можно ввести «комплексные точки»? Если на плоскости выбрана аффинная система координат, то ее (обычные) точки находятся во взаимно однозначном соответствии с парами (x, y) вещественных чисел и могут быть с такими парами отождествлены. Поэтому мы можем ввести комплексные точки просто как пары (x, y) произвольных комплексных чисел. Так во многих учебниках аналитической геометрии и делается.

Однако эта конструкция комплексной плоскости зависит от выбора в обычной плоскости некоторой аффинной координатной системы, так что, строго говоря, мы получаем не одну комплексную плоскость, а столько, сколько существует различных аффинных координатных систем. Хотя, пользуясь формулами преобразования аффинных координат, эти комплексные плоскости можно друг с другом отождествить, все же значительно удобнее и изящнее дать определение комплексной плоскости, не зависящее от выбора в обычной плоскости аффинной системы координат.

С этой целью мы заметим, что в аксиомах аффинной геометрии, рассмотренных в предыдущем пункте, специфика поля $\mathbb{R}$ никак по существу не используется. Поэтому их можно переформулировать, заменяя всюду поле $\mathbb{R}$ полем комплексных чисел $\mathbb{C}$. В результате мы и получим аксиомы аффинной геометрии на комплексной плоскости.

Более того, эти аксиомы можно формулировать и для случая, когда поле $\mathbb{R}$ заменено произвольным полем $\mathbb{K}$. Тогда мы получим *аффинную геометрию плоскости над полем* $\mathbb{K}$.

Замечание 1. Очень любопытные «геометрии» получаются, когда поле $\mathbb{K}$ конечно. Например, для простейшего поля, состоящего из двух элементов, мы получаем геометрию, в которой имеется только четыре (!) точки. Мы такими геометриями здесь заниматься не будем (хотя они сами по себе очень

интересны и в последнее время получили совершенно неожиданные применения, например, в теории кодирования).

Поскольку аффинную геометрию, как выясняется, можно строить над произвольным полем $\mathbb{K}$, мы пока его специализировать не будем (только потом, когда мы перейдем к более глубоким вопросам, где специфика поля будет играть роль, мы будем считать, что им является интересующее нас сейчас в первую очередь поле $\mathbb{C}$).

Итак, мы будем считать данным некоторое (пока произвольное) поле $\mathbb{K}$. Элементы этого поля мы будем называть *числами*.

Определение 1. Пусть $\mathbb{K}^2$ — множество всех пар (x, y) чисел поля $\mathbb{K}$; обозначим через $\mathrm{Aff}(\mathbb{K}^2)$ совокупность всех преобразований

$$\varphi\colon \mathbb{K}^2 \to \mathbb{K}^2,$$

задаваемых формулами вида

$$\begin{array}{l} x' = a_1 x + a_2 y + a, \\ y' = b_1 x + b_2 y + b, \end{array} \quad \left| \begin{array}{cc} a_1 & a_2 \\ b_1 & b_2 \end{array} \right| \neq 0. \tag{1}$$

Ясно, что

множество $\mathrm{Aff}(\mathbb{K}^2)$ *является группой преобразований.*

Определение 2. Пусть $\mathfrak{M}$ — произвольное множество, для которого задано некоторое семейство $\mathrm{Coor}(\mathfrak{M})$ его биективных отображений

$$\alpha\colon \mathfrak{M} \to \mathbb{K}^2$$

на множество $\mathbb{K}^2$. Мы будем называть $\mathfrak{M}$ *аффинной плоскостью над полем* $\mathbb{K}$, а отображение $\alpha \in \mathrm{Coor}(\mathfrak{M})$ — *аффинными координатными системами* на этой плоскости, если (ср. п. 3) выполнены следующие две аксиомы:

Аксиома 1. *Если* $\alpha \in \mathrm{Coor}(\mathfrak{M})$ *и* $\varphi \in \mathrm{Aff}(\mathbb{K}^2)$, *то*

$$\varphi \circ \alpha \in \mathrm{Coor}(\mathfrak{M}).$$

Аксиома 2. *Если* α, $\alpha' \in \mathrm{Coor}(\mathfrak{M})$, *то*

$$\alpha' \circ \alpha^{-1} \in \mathrm{Aff}(\mathbb{K}^2).$$

При $\mathbb{K} = \mathbb{R}$ мы получаем уже известное нам определение обычной аффинной плоскости. Плоскость над полем $\mathbb{R}$ называется *вещественной плоскостью*. Аналогично, плоскость над полем $\mathbb{C}$ (случай, для нас сейчас наиболее интересный) называется *комплексной плоскостью*.

Ясно, что, положив $\mathrm{Coor}(\mathbb{K}^2) = \mathrm{Aff}(\mathbb{K}^2)$, мы определим $\mathbb{K}^2$ как аффинную плоскость. Эта плоскость называется *стандартной* (или *арифметической*) *плоскостью над полем* $\mathbb{K}$.

Следовательно, аксиомы 1 и 2 непротиворечивы (поскольку непротиворечивы аксиомы, определяющие поле $\mathbb{K}$).

Определение 3. Биективное отображение

$$\mu\colon \mathfrak{M} \to \mathfrak{M}'$$

одной плоскости над полем $\mathbb{K}$ на другую (или ту же самую) плоскость называется *изоморфизмом*, если отображение $\alpha'\colon \mathfrak{M}' \to \mathbb{K}^2$ тогда и только тогда является координатной системой на плоскости $\mathfrak{M}'$, когда отображение $\alpha' \circ \mu$ является координатной системой на плоскости $\mathfrak{M}$ (ср. определение 3 п. 2).

Ясно, что

отображение $\mu\colon \mathfrak{M} \to \mathbb{K}^2$ *тогда и только тогда является изоморфизмом, когда оно является аффинной координатной системой.*

Кроме того,

отображение $\mu\colon \mathfrak{M}' \to \mathfrak{M}$ *тогда и только тогда является изоморфизмом, когда оно действует по равенству координат в двух аффинных координатных системах* α *и* α', *т. е. когда*

$$\mu = \alpha^{-1} \circ \alpha'.$$

Как и в вещественном случае, группу всех автоморфизмов $\Phi\colon \mathfrak{M} \to \mathfrak{M}$ аффинной плоскости $\mathfrak{M}$ мы будем обозначать символом $\mathrm{Aff}(\mathfrak{M})$.

Ясно (ср. пп. 2 и 3), что

любой изоморфизм $\mu\colon \mathfrak{M} \to \mathfrak{M}'$ *индуцирует некоторый изоморфизм групп*

$$\mathrm{Aff}(\mathfrak{M}) \to \mathrm{Aff}(\mathfrak{M}').$$

Понятие *аффинно инвариантного свойства* фигур на плоскостях над полем $\mathbb{K}$ вводится точно так же, как и в случае поля $\mathbb{R}$. Аффинная геометрия над полем $\mathbb{K}$ является, по определению, наукой о таких свойствах.

В этой геометрии понятия «направленного отрезка», «вектора» и т. п. вводятся дословно так же, как в геометрии над полем $\mathbb{R}$ (см. п. 3). В частности, все сказанное в п. 3 о связи между аффинными координатными системами и реперами полностью остается в силе. Поэтому еще раз повторять это мы здесь не будем.

Задание. Проверьте, что вся аффинная теория векторов сохраняется и над любым полем $\mathbb{K}$.

Замечание 2 (очень важное!). Поскольку в произвольном поле нет понятий «положительное — отрицательное», все связанное с понятием ориентации не переносится на случай произвольного поля $\mathbb{K}$, и в частности, на случай поля $\mathbb{C}$. Например, не переносится понятие полуплоскости, так что говорить, скажем, о «полуплоскости комплексной плоскости» бессмысленно.

Понятие бивектора, тем не менее, имеет смысл над любым полем $\mathbb{K}$, поскольку сохраняется понятие элементарного преобразования.

Замечание 3. Поскольку понятие ортогональной матрицы (как матрицы, удовлетворяющей соотношению $C'C = E$) имеет смысл над любым полем $\mathbb{K}$, мы можем ввести группу $\mathrm{Ort}(\mathbb{K}^2)$ и уже известным нам способом определить *евклидову геометрию над полем* $\mathbb{K}$. Однако, в отличие от аффинной геометрии, эта геометрия даже формально резко отличается от евклидовой геометрии над полем $\mathbb{R}$. Например, в ней нельзя, вообще говоря, ввести понятие длины отрезка (поскольку в поле могут не существовать квадратные корни), не говоря уже об угле между отрезками. При $\mathbb{K} = \mathbb{C}$ длину и угол ввести, конечно, можно, но, например, длина будет комплексным числом и может быть равна нулю даже тогда, когда отрезок невырожден.

Все сказанное выше можно, конечно, повторить (с незначительными, само собой разумеющимися изменениями) и для любого n. В частности, при $n = 1$ и $\mathbb{K} = \mathbb{C}$ мы получаем понятие *комплексной прямой*. На такой прямой имеется одна комплексная координата z и переход от одной такой координаты к другой описывается формулой вида

$$z' = az + b,$$

где a и b — произвольные комплексные числа, причем $a \neq 0$.

Поскольку

$$z = x + iy,$$

точки комплексной прямой задаются двумя вещественными числами x и y, так что *геометрически комплексная прямая представляет собой плоскость.*

Однако отождествлять комплексную прямую с вещественной аффинной плоскостью нельзя, поскольку формулы преобразования координат в них различны. Действительно, положив

$$a = a_1 + ia_2, \quad b = b_1 + ib_2,$$

мы можем формулу (1) переписать в координатах x и y в следующем виде:

$$\begin{aligned} x' &= a_1x - a_2y + b_1, \\ y' &= a_2x + a_1y + b_2, \end{aligned} \tag{2}$$

— ясно, что не любое преобразование из группы $\mathrm{Aff}(\mathbb{R}^2)$ можно так записать.

Иначе говоря, группа $\mathrm{Aff}(\mathbb{C}^1)$ преобразований (1), записанных в виде (2), является лишь подгруппой группы $\mathrm{Aff}(\mathbb{R}^2)$.

Это показывает, что геометрия комплексной прямой (рассматриваемая как геометрия плоскости) богаче аффинной геометрии плоскости.

Ниже, в п. 2 § 4 гл. 7, мы идентифицируем эту геометрию с одной из плоских геометрий, читателю по существу известных.

5. Вещественно-комплексная геометрия

В построенной в предыдущем пункте (при $\mathbb{K} = \mathbb{C}$) аффинной геометрии нет места понятию «вещественная точка», поскольку точка, имеющая вещественные координаты в одной аффинной координатной системе, вполне может иметь невещественные координаты в другой. Это означает, что в нашем стремлении ввести комплексные точки мы зашли слишком далеко и потеряли по дороге контакт с обычной вещественной плоскостью. Хотелось бы остановиться «на полпути» и устроить комплексную плоскость так, чтобы она содержала обычную вещественную комплексную плоскость в качестве подмножества. Тогда мы могли бы работать, пока это возможно, в этой вещественной плоскости (сохраняя, тем самым, геометрическую наглядность) и выходить в объемлющую комплексную плоскость только тогда, когда это сделается совершенно необходимым.

Аксиоматическое построение такой «вещественно-комплексной» плоскости можно осуществить, заметив, что любое преобразование $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ из группы $\mathrm{Aff}(\mathbb{R}^2)$ однозначно определяет некоторое преобразование $\hat{\varphi}\colon \mathbb{C}^2 \to \mathbb{C}^2$ из группы $\mathrm{Aff}(\mathbb{C}^2)$, а именно, преобразование, задаваемое теми же формулами

$$x' = a_1 x + a_2 y + a,$$
$$y' = b_1 x + b_2 y + b.$$

(Таким образом, различие между преобразованиями φ и $\hat{\varphi}$ состоит лишь в том, что в случае преобразования φ величины x и y могут принимать лишь вещественные значения, тогда как для преобразования $\hat{\varphi}$ они могут иметь любые комплексные значения.)

Поскольку соответствие $\varphi \longmapsto \hat{\varphi}$ строится без какого-либо произвола (является «естественным» соответствием), мы можем преобразования φ и $\hat{\varphi}$ отождествлять или хотя бы обозначать одним и тем же символом. Ни к каким неудобствам это привести не может

Отождествив преобразования φ и $\hat{\varphi}$, мы можем, следовательно, считать, что *группа* $\mathrm{Aff}(\mathbb{R}^2)$ *вложена в группу* $\mathrm{Aff}(\mathbb{C}^2)$. По определению, преобразование $\varphi \in \mathrm{Aff}(\mathbb{C}^2)$ тогда и только тогда принадлежит $\mathrm{Aff}(\mathbb{R}^2)$, когда все коэффициенты этого преобразования являются вещественными числами.

Ясно, что если преобразование $\varphi \in \mathrm{Aff}(\mathbb{C}^2)$ принадлежит подмножеству $\mathrm{Aff}(\mathbb{R}^2)$, то преобразование φ^{-1} также принадлежит этому подмножеству, и если $\varphi_1 \in \mathrm{Aff}(\mathbb{R}^2)$ и $\varphi_2 \in \mathrm{Aff}(\mathbb{R}^2)$, то $\varphi_1 \circ \varphi_2 \in \mathrm{Aff}(\mathbb{R}^2)$. По определению, это означает, что *подмножество* $\mathrm{Aff}(\mathbb{R}^2)$ *группы* $\mathrm{Aff}(\mathbb{C}^2)$ *является ее подгруппой.*

Пусть теперь нам дано некоторое множество $\mathfrak{M}$ вместе с непустым семейством $\mathrm{Coor}(\mathfrak{M})$ его биективных отображений на множество $\mathbb{C}^2$ и пусть выполнены следующие две аксиомы:

Аксиома 1. *Для любого отображения* $\alpha\colon \mathfrak{M} \to \mathbb{C}^2$, *принадлежащего семейству* $\mathrm{Coor}(\mathfrak{M})$, *и любого преобразования* $\varphi\colon \mathbb{C}^2 \to \mathbb{C}^2$, *принадлежащего подгруппе* $\mathrm{Aff}(\mathbb{R}^2)$ *группы* $\mathrm{Aff}(\mathbb{C}^2)$, *отображение* $\varphi \circ \alpha$ *принадлежит семейству* $\mathrm{Coor}(\mathfrak{M})$.

Аксиома 2. *Для любых двух отображений* α, $\alpha'\colon \mathfrak{M} \to \mathbb{C}^2$, *принадлежащих семейству* $\mathrm{Coor}(\mathfrak{M})$, *отображение* $\alpha' \circ \alpha^{-1}\colon \mathbb{C}^2 \to \mathbb{C}^2$ *принадлежит подгруппе* $\mathrm{Aff}(\mathbb{R}^2)$.

Эти аксиомы полностью аналогичны аксиомам 1 и 2 для комплексной плоскости и отличаются от них только тем, что преобразования φ предполагаются теперь принадлежащими подгруппе $\mathrm{Aff}(\mathbb{R}^2)$, т. е. имеющими вещественные коэффициенты.

Определение 1. При выполнении аксиом 1 и 2 мы будем множество $\mathfrak{M}$ называть (*аффинной*) *вещественно-комплексной плоскостью*, его элементы — *точками*, отображения, принадлежащие семейству $\mathrm{Coor}(\mathfrak{M})$, — *аффинными координатными системами*, и для любой аффинной координатной системы $\alpha\colon \mathfrak{M} \to \mathbb{C}^2$ и любой точки $M \in \mathfrak{M}$ числа x, y, составляющие пару $\alpha(M) \in \mathbb{C}^2$, — *координатами* точки M в координатной системе α.

Таким образом, вещественно-комплексная плоскость содержит точки с произвольными комплексными координатами, но преобразования координат в ней допускаются только с вещественными коэффициентами.

Ясно, что в такой плоскости уже имеет смысл говорить о *вещественных точках*, поскольку при любом преобразовании координат вещественные координаты переходят в вещественные.

Непротиворечивость и полнота аксиом 1 и 2 доказываются по уже известному нам образцу: положив $\mathrm{Coor}(\mathbb{C}^2) = \mathrm{Aff}(\mathbb{R}^2)$, мы превращаем множество $\mathbb{C}^2$ в вещественно-комплексную плоскость (называемую *стандартной* или *арифметической вещественно-комплексной плоскостью*), и любая вещественно-комплексная плоскость $\mathfrak{M}$ изоморфна плоскости $\mathbb{C}^2$ (причем изоморфизмами $\mathfrak{M} \to \mathbb{C}^2$ являются координатные системы и только они).

Задание. Сформулируйте определение изоморфизма двух вещественно-комплексных плоскостей.

Замечание 1. Обратим внимание на то, что одно и то же множество $\mathbb{C}^2$ является областью действия различных геометрий: вещественно-комплексной и чисто комплексной. Все зависит от того, какие отображения мы считаем координатными системами.

Пусть $\mathfrak{M}^{\text{вещ}}$ — совокупность всех вещественных точек вещественно-комплексной плоскости $\mathfrak{M}$. По определению, каждая координатная система $\alpha\colon \mathfrak{M} \to \mathbb{C}^2$ отображает множество $\mathfrak{M}^{\text{вещ}}$ в подмножество $\mathbb{R}^2$ множества $\mathbb{C}^2$, состоящее из пар (x, y) вещественных чисел. Поэтому, если мы будем ее рассматривать

только на этом множестве, она будет некоторым отображением (очевидно, биективным) множества $\mathfrak{M}^{\text{вещ}}$ на множество $\mathbb{R}^2$. Очевидным образом проверяется, что полученное семейство $\mathrm{Coor}(\mathfrak{M}^{\text{вещ}})$ биективных отображений $\mathfrak{M}^{\text{вещ}} \to \mathbb{R}^2$ удовлетворяет аксиомам 1 и 2 определения плоскости над полем $\mathbb{R}$. Тем самым доказано, что

совокупность $\mathfrak{M}^{\text{вещ}}$ всех вещественных точек вещественно-комплексной плоскости $\mathfrak{M}$ естественным образом определяется как вещественная плоскость.

С другой стороны, пользуясь тем, что $\mathrm{Aff}(\mathbb{R}^2) \subset \mathrm{Aff}(\mathbb{C}^2)$, мы можем произвольной вещественно-комплексной плоскости $\mathfrak{M}$ сопоставить комплексную плоскость $\mathfrak{M}^{\text{компл}}$, точками которой являются точки плоскости $\mathfrak{M}$ (т. е. которая как множество совпадает с $\mathfrak{M}$), а семейство $\mathrm{Coor}(\mathfrak{M}^{\text{компл}})$ аффинных координатных систем которой состоит из всевозможных отображений $\mathfrak{M} \to \mathbb{C}^2$, имеющих вид $\varphi \circ \alpha$, где $\alpha \in \mathrm{Coor}(\mathfrak{M})$ и $\varphi \in \mathrm{Aff}(\mathbb{C}^2)$. Тот факт, что $\mathfrak{M}^{\text{компл}}$ действительно является комплексной плоскостью, т. е. что семейство $\mathrm{Coor}(\mathfrak{M}^{\text{компл}})$ удовлетворяет аксиомам 1 и 2 из п. 4 (при $\mathbb{K} = \mathbb{C}$), проверяется автоматически.

Таким образом,

любую вещественно-комплексную плоскость $\mathfrak{M}$ мы можем превратить в комплексную плоскость $\mathfrak{M}^{\text{компл}}$, приняв за ее аффинные координатные системы всевозможные отображения вида $\varphi \circ \alpha$, где $\alpha \in \mathrm{Coor}(\mathfrak{M})$ и $\varphi \in \mathrm{Aff}(\mathbb{C}^2)$.

Замечание 2. Переход от плоскости $\mathfrak{M}$ к плоскости $\mathfrak{M}^{\text{компл}}$ вполне аналогичен описанному в п. 3 переходу от евклидовой плоскости $\mathfrak{M}$ к аффинной плоскости $\mathfrak{M}_{\text{афф}}$, обязанному своей возможностью тому, что группа $\mathrm{Ort}(\mathbb{R}^2)$ является подгруппой группы $\mathrm{Aff}(\mathbb{R}^2)$.

Поскольку вещественно-комплексная плоскость получается из комплексной плоскости сужением семейства допустимых аффинных координатных систем, любое понятие, имеющее смысл в комплексной плоскости («вектор», «репер» и т. п.), будет иметь смысл и в вещественно-комплексной плоскости. При этом каждое утверждение, верное в комплексной плоскости, будет верно и в вещественно-комплексной плоскости.

Однако в вещественно-комплексной плоскости могут иметь смысл и понятия, бессмысленные в комплексной плоскости. Например, в вещественно-комплексной плоскости $\mathfrak{M}$ можно говорить о *вещественных векторах*, т. е. о векторах с вещественными координатами. Такие векторы естественным образом отождествляются с векторами на вещественной плоскости $\mathfrak{M}^{\text{вещ}}$.

Замечание 3. Строго говоря, векторы на вещественной плоскости и вещественные векторы представляют собой различные математические объекты: первые являются классами эквиполлентных направленных отрезков $\overrightarrow{AB}$ с вещественными концевыми точками A и B, тогда как вторые являются классами эквиполлентных направленных отрезков $\overrightarrow{AB}$, обладающих тем свойством, что координаты точек A и B отличаются на вещественные числа. Тем

не менее, поскольку эти классы находятся в естественном биективном соответствии, их можно безболезненно отождествлять.

Аналогично, для любой точки (вектора) на вещественно-комплексной плоскости можно говорить о *комплексно-сопряженной точке* (*векторе*). Точка (вектор) тогда и только тогда вещественна, когда она (он) совпадает с комплексно-сопряженной точкой.

Специфика вещественно-комплексной плоскости отражается также и в том, что для нее не любая тройка, состоящая из некоторой точки O и двух линейно независимых векторов $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$, является репером аффинной координатной системы (хотя попрежнему любая аффинная координатная система однозначно определяется своим репером). Действительно, ясно, что для этого необходимо, чтобы точка O и векторы $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$ были вещественны. Верно и обратное, т. е.

любая тройка $O\boldsymbol{e}_1\boldsymbol{e}_2$, состоящая из вещественной точки O и двух вещественных линейно независимых векторов $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$, является репером некоторой (однозначно определенной) аффинной координатной системы на вещественно-комплексной плоскости.

Доказательство по существу дословно совпадает с доказательством соответствующего утверждения из п. 3 и потому мы его опустим.

Поскольку мы не владеем формальным определением понятия линии (см. п. 4 § 2), определить, что такое линии на плоскости над произвольным полем $\mathbb{K}$, мы не можем. Однако понятие *алгебраической линии* (вполне достаточное для наших целей) имеет, очевидно, смысл над любым полем (поскольку понятие многочлена определено над любым полем и при линейной замене неизвестных многочлен переходит в многочлен). При этом понятие *порядка* алгебраической линии встречает те же трудности, что и в случае поля $\mathbb{R}$ (см. п. 2 § 2), усугубляющиеся для полей конечной характеристики еще и тем, что над таким полем многочлен может быть тождественно равен нулю даже тогда, когда его коэффициенты отличны от нуля.

Над полем $\mathbb{C}$ комплексных чисел по крайней мере часть из этих трудностей исчезает (см. п. 2 § 2). Это и является основной причиной, почему мы ввели в рассмотрение комплексные плоскости.

Определение 2. Линия (алгебраическая) на вещественно-комплексной плоскости $\mathfrak{M}$ называется *вещественной*, если в некоторой аффинной координатной системе она выражается уравнением с вещественными коэффициентами.

Ясно, что это определение корректно (линия, имеющая уравнение с вещественными коэффициентами в одной аффинной координатной системе, будет иметь уравнение с вещественными

коэффициентами и в любой другой аффинной координатной системе).

Каждая такая линия Г определяет некоторую линию γ (того же порядка) на вещественной плоскости $\mathfrak{M}^{\text{вещ}}$.

Подчеркнем, что как множества точек *линии* Г *и* γ *различны*: линия γ является пересечением линии Г с вещественной плоскостью $\mathfrak{M}^{\text{вещ}}$, так что в общепринятых теоретико-множественных обозначениях:

$$\gamma = \Gamma \cap \mathfrak{M}^{\text{вещ}}.$$

Тем не менее, по традиции принято называть точки линии Г, не принадлежащие линии γ, *мнимыми точками линии* γ.

Из того, что поле $\mathbb{C}$ комплексных чисел *алгебраически замкнуто* (любой многочлен положительной степени от одной переменной имеет в нем корень), непосредственно вытекает, что на вещественно-комплексной плоскости (так же, как и на комплексной плоскости) любая алгебраическая линия имеет точки (не является пустым множеством). Конечно, ее пересечение γ с вещественной плоскостью тем не менее может быть пустым (в этом случае она называется *нулевой линией*).

Подчеркнем, что хотя линии на вещественно-комплексной (и комплексной) плоскости определяются формально так же, как линии (алгебраические) на вещественной плоскости, с наглядно геометрической точки зрения они глубоко различны. Действительно, положение точки на линии в вещественной плоскости определяется, вообще говоря, значением одного вещественного параметра, и в этом смысле эти линии *одномерны*, тогда как положение точки на линии в комплексной (или вещественно-комплексной) плоскости определяется значением одного комплексного параметра, т. е. значениями двух вещественных параметров (действительной и мнимой частью комплексного параметра), так что эти линии *двумерны*. Иными словами, линии в комплексной (или вещественно-комплексной) плоскости представляют собой с наглядной точки зрения поверхности.

Сама комплексная (или вещественно-комплексная) плоскость описывается четырьмя вещественными параметрами и потому является четырехмерным образованием, наглядно геометрически не описываемым.

Мы видим, в частности, что основную массу точек каждой линии на вещественно-комплексной плоскости составляют мнимые точки (если представлять себе эту линию как поверхность, то ее вещественные точки будут изображаться на этой поверхности точками некоторой линии).

Очень важно иметь также в виду, что алгебраические линии на комплексной плоскости являются весьма специальным частным случаем двумерных образований, которые можно на этой плоскости рассматривать. Например, вещественная плоскость $\mathfrak{M}^{\text{вещ}}$ (также являющаяся двумерной поверхностью) алгебраической линией заведомо не является, т. е. условие вещественности координат x и y нельзя записать в виде равенства $f(x, y) = 0$, где $f(x, y)$ — некоторый многочлен (этот факт мы докажем ниже, в качестве простого следствия некоторых общих теорем; см. п. 2 § 1 гл. 3).

Так как

$$\mathrm{Ort}(\mathbb{R}^2) \subset \mathrm{Aff}(\mathbb{R}^2) \subset \mathrm{Aff}(\mathbb{C}^2),$$

то

группа $\mathrm{Ort}(\mathbb{R}^2)$ *является подгруппой группы* $\mathrm{Aff}(\mathbb{C}^2)$.

Это позволяет нам повторить определение вещественно-комплексной плоскости с заменой группы $\mathrm{Aff}(\mathbb{R}^2)$ группой $\mathrm{Ort}(\mathbb{R}^2)$. В результате мы получим определение *евклидовой вещественно-комплексной плоскости.*

Так же, как и выше, показывается, что

множество $\mathbb{C}^2$, *рассматриваемое вместе с семейством преобразований* $\mathrm{Ort}(\mathbb{R}^2)$ (*считаемых теперь преобразованиями* $\mathbb{C}^2 \to \mathbb{C}^2$), *является евклидовой вещественно-комплексной плоскостью,*

и что

любая евклидова вещественно-комплексная плоскость изоморфна стандартной евклидовой вещественно-комплексной плоскости $\mathbb{C}^2$.

Следовательно, аксиомы евклидовой вещественно-комплексной геометрии непротиворечивы и полны.

Ясно, что

вещественная плоскость $\mathfrak{M}^{\text{вещ}}$ *евклидовой вещественно-комплексной плоскости* $\mathfrak{M}$ *естественным образом определяется как евклидова плоскость в смысле* п. 2.

Это означает, что для вещественных точек евклидовой вещественно-комплексной плоскости сохраняется вся обычная евклидова геометрия.

Обратим еще раз внимание на тот замечательный факт, что разнообразные геометрии мы строили в этом параграфе по существу одним и тем же способом. Возможности этого способа отнюдь не исчерпываются построенными геометриями. Мы к этому вопросу еще вернемся в § 4 гл. 7.

Дополнение. **Аксиоматика Гильберта**

1. Формулировка аксиом

Аксиоматика Гильберта была первой формальной аксиоматикой геометрии, получившей широкое распространение. Мы приведем эту аксиоматику в чуть-чуть модифицированном виде.

В аксиоматике Гильберта имеется три рода основных объектов:

«точки», «прямые», «плоскости»,

и три основные отношения:

«принадлежать», «между», «конгруэнтность».

Отношение «принадлежать» связывает точки с прямыми и плоскостями: «точка принадлежит прямой», «точка принадлежит плоскости». Таким образом,

собственно говоря, имеется два отношения принадлежности: для точек и прямых и для точек и плоскостей. Вместо термина «принадлежит» употребляются также (исключительно для упрощения формулировок) такие термины, как «лежит на», «проходит через», «инцидентны» и т. п. Например, высказывания «точка A принадлежит прямой α», «прямая α проходит через точку A», «прямая α инцидентна точке A», «точка A инцидентна прямой α» означают, по определению, одно и то же. Вместо «точки A и B принадлежат прямой α» можно говорить «прямая α соединяет точки A и B», а вместо «точка A принадлежит прямым α и β» — «прямые α и β пересекаются в точке A» и т. д.

Вводится также неосновное отношение «прямая α принадлежит плоскости Π», по определению, означающее, что любая точка, принадлежащая прямой α, принадлежит также и плоскости Π.

Отношение «между» является отношением, связывающим три (различные) точки A, B, C, принадлежащие одной прямой. Если точки A, B, C связаны этим отношением, то говорят, что «точка B лежит между точками A и C».

Отношение «конгруэнтность» связывает некоторые неосновные объекты, называемые «отрезками» и «углами». Эти объекты мы определим ниже. Таким образом, на самом деле имеется два отношения конгруэнтности: одно — связывающее отрезки, а другое — углы.

Аксиомы Гильберта довольно сложны и их неудобно формулировать непосредственно для основных объектов и отношений (хотя конечно, это возможно). Целесообразнее вводить эти аксиомы постепенно (группами), перемежая их теоремами и определениями, необходимыми для формулировки следующих аксиом. Принято разбивать аксиомы Гильберта на пять групп.

Первая группа аксиом

(аксиомы принадлежности)

I_1. *Для любых двух точек A и B существует прямая, проходящая через эти точки.*

I_2. *Если точки A и B различны, то проходящая через них прямая единственна.*

I_3. *На каждой прямой существуют по крайней мере две точки.*

I_4. *Существуют по крайней мере три точки, не лежащие на одной прямой.*

I_5. *Для любых трех точек A, B, C существует проходящая через эти точки плоскость.*

I_6. *Если точки A, B, C не лежат на одной прямой, то проходящая через них плоскость единственна.*

I_7. *На каждой плоскости существует по крайней мере одна точка.*

I_8. *Если две различные точки A и B прямой α принадлежат плоскости Π, то прямая α принадлежит плоскости Π.*

I_9. *Если две плоскости имеют общую точку, то они имеют по крайней мере еще одну общую точку.*

I_{10}. *Существует по крайней мере четыре точки, не принадлежащие одной плоскости.*

С помощью этих аксиом могут быть уже доказаны некоторые теоремы. Например, можно доказать, что

1) *две (различные) прямые имеют не более одной общей точки*;

2) *две (различные) плоскости либо не имеют общих точек, либо имеют общую прямую, на которой лежат все общие точки этих плоскостей*;

3) *плоскость и не лежащая на ней прямая имеют не более одной общей точки*;

4) *через прямую и не лежащую на ней точку проходит одна и только одна плоскость*;

5) *через две пересекающиеся (но не совпадающие) прямые проходит одна и только одна плоскость*;

6) *каждая плоскость содержит по крайней мере три различные точки.*

Докажем, для примера, утверждение 6).

По аксиоме I_7 каждая плоскость П содержит по крайней мере одну точку A. По аксиоме I_{10} существует точка B, не принадлежащая плоскости П. По аксиомам I_1 и I_2 существует единственная прямая AB, проходящая через точки A и B. По аксиоме I_4 существует точка C, не принадлежащая прямой AB. По аксиомам I_5 и I_6 существует единственная плоскость ABC, проходящая через точки A, B, C. Плоскость ABC и данная плоскость П имеют общую точку A. Поэтому по аксиоме I_9 они имеют еще одну общую точку D. Точки A, B, D не принадлежат одной прямой (ибо в противном случае точка B принадлежала бы в силу аксиомы I_8 плоскости П) и потому через них проходит единственная плоскость ABD (аксиомы I_5 и I_6). По аксиоме I_{10} существует точка E, не принадлежащая плоскости ABD. Точки A, B, E не принадлежат одной прямой и потому содержатся в единственной плоскости ABE (аксиомы I_5 и I_6), отличной от плоскости ABD. Плоскости ABE и П имеют общую точку A и потому (аксиома I_9) они имеют еще одну общую точку F (не принадлежащую прямой AB). Так как точки D и F не принадлежат прямой AB, то (утверждение 2)) они не являются общими точками плоскостей ABD и ABF и, следовательно, различны. Таким образом, плоскость П содержит три различные точки A, D, F.

Задание. Докажите утверждения 1)—5).

Вторая группа аксиом

(аксиомы порядка)

II_1. *Каждая точка, лежащая между точками A и C, лежит также между точками C и A.*

II_2. *Для любых (различных) точек A и B на прямой AB существует такая точка C, что точка B лежит между точками A и C.*

II_3. *Среди любых трех точек произвольной прямой не более одной точки лежит между двумя другими.*

Эти аксиомы называются линейными аксиомами порядка. Кроме них группа II содержит еще так называемую аксиому Паша II_4, которую мы сформулируем ниже.

Пара (неупорядоченная) различных точек A и B называется *отрезком* прямой AB *с концами* A, B и обозначается символом $\overline{AB}$. Любая точка C,

лежащая между точками A и B, называется *внутренней точкой* отрезка $\overline{AB}$ (заметим, что существование таких точек аксиомами не утверждается). В силу аксиомы II_1 это определение корректно.

Точка C называется точкой, *принадлежащей* отрезку $\overline{AB}$, если она либо является его внутренней точкой, либо совпадает с одной из точек A или B. Все точки прямой AB, не принадлежащие отрезку $\overline{AB}$, называются *внешними* к отрезку $\overline{AB}$ (существование таких точек обеспечивается аксиомой II_2). Говорят, что прямая α *пересекает* отрезок $\overline{AB}$, если существует внутренняя точка отрезка $\overline{AB}$, принадлежащая прямой α.

II_4. *Пусть A, B, C — три точки, не лежащие на одной прямой, и пусть α — некоторая прямая, в плоскости ABC, не проходящая через точки A, B, C. Тогда, если прямая α пересекает отрезок $\overline{AB}$, то она пересекает также либо отрезок $\overline{AC}$, либо отрезок $\overline{BC}$.*

Из аксиом принадлежности и порядка уже можно вывести много важных фактов геометрии. Например, в порядке дополнения к аксиомам группы II можно показать, что

1) *любой отрезок $\overline{AB}$ содержит бесконечно много точек*;

2) *среди любых трех точек прямой одна и только одна точка лежит между двумя другими*;

3) *если в аксиоме Паша прямая α пересекает отрезок $\overline{AC}$, то она не пересекает отрезок $\overline{BC}$.*

Упражнение. Докажите утверждения 1), 2), 3).

Пусть O — точка прямой α. Введем на множестве $\alpha \setminus O$ всех точек прямой α, отличных от точки O, отношение $\sim$, считая, что $A \sim B$ тогда и только тогда, когда либо $A = B$, либо отрезок $\overline{AB}$ не содержит точки O.

Аналогично, пусть α — прямая плоскости Π. Введем на множестве $\Pi \setminus \alpha$ всех точек плоскости Π, не принадлежащих прямой α, отношение $\sim$, считая, что $A \sim B$ тогда и только тогда, когда либо $A = B$, либо прямая α не пересекает отрезок $\overline{AB}$.

Оказывается, что

в обоих случаях отношение $\sim$ является отношением эквивалентности, причем существует точно два класса эквивалентности.

В первом случае эти классы называются *лучами* (или *полупрямыми*), определенными точкой O (или имеющими начало в этой точке), а во втором случае — *полуплоскостями*, определенными прямой α.

Рассмотрим случай прямой на плоскости. По определению, отношение $\sim$ рефлексивно ($A \sim A$). Кроме того, оно, очевидно, симметрично (если $A \sim B$, то $B \sim A$). Покажем, что оно транзитивно (если $A \sim B$ и $B \sim C$, то $A \sim C$). Если среди точек A, B, C есть одинаковые, то транзитивность очевидна. Предположим, что все точки A, B, C различны и не принадлежат одной прямой. Если бы отношение $A \sim C$ места не имело, то прямая α пересекала бы отрезок $\overline{AC}$. Но тогда по аксиоме Паша II_4 она пересекала бы вопреки условию и один из отрезков $\overline{AB}$ или $\overline{BC}$. Следовательно, $A \sim C$. Случай, когда точки A, B, C принадлежат одной прямой, мы предоставим читателю разобрать самостоятельно.

Таким образом, отношение $\sim$ является отношением эквивалентности и потому все множество $\Pi \setminus \alpha$ разбивается на соответствующие классы эквивалентности. Осталось доказать, что этих классов ровно два.

Пусть A — произвольная точка из $\Pi \setminus \alpha$ и пусть B — произвольная точка прямой α. По аксиоме II_2 на прямой AB существует такая точка C, что точка B лежит между точками A и C, т. е. такая, что прямая α пересекает отрезок $\overline{AC}$ (в точке B). Следовательно, точки A и C не эквивалентны, так что существует по крайней мере два класса.

Покажем теперь, что более двух классов существовать не может, т. е. что любая точка M из $\Pi \setminus \alpha$ эквивалентна либо точке A, либо точке C. Но это немедленно вытекает из дополнения 3) к аксиоме Паша, поскольку прямая α, по условию, пересекает отрезок $\overline{AC}$ и, следовательно, пересекает один и только один из отрезков $\overline{AM}$ и $\overline{MC}$.

Задание. Дайте доказательство для случая точки на прямой.

Задание. Аналогично, вводя понятие *полупространства*, определенного данной плоскостью Π, докажите, что совокупность всех точек, не принадлежащих этой плоскости, разбивается точно на два полупространства.

О точках прямой, принадлежащих одному лучу с началом в точке O, говорят также, что они расположены *по одну сторону* от точки O. Аналогично о точках, принадлежащих одной полуплоскости (одному полупространству), говорят, что они расположены *по одну сторону* от соответствующей прямой (плоскости).

Два луча одной и той же прямой мы назовем *одинаково направленными*, если один из них целиком содержится в другом.

Задание. Докажите, что:

1) *отношение одинаковой направленности является на множестве всех лучей данной прямой отношением эквивалентности*;

2) *два луча с началом в произвольной точке O неодинаково направлены, но любой другой луч одинаково направлен с одним (и только одним) из этих лучей.*

Утверждение 2) означает, что на рассматриваемой прямой имеется точно два класса одинаково направленных лучей. Эти классы называются *ориентациями* прямой.

Пусть o — произвольная ориентация прямой α, A и B — две точки этой прямой, и α_A^+, α_B^+ — лучи прямой α, принадлежащие ориентации o и определенные соответственно точками A и B. Мы скажем, что $A \prec B$ в ориентации o, если луч α_B^+ целиком содержится в луче α_A^+. Таким образом, каждая ориентация прямой α определяет на множестве точек этой прямой некоторое отношение $A \prec B$.

Задание. Докажите, что это отношение является отношением предшествования, так что ориентации, как они сейчас определены, по существу совпадают с ориентациями в смысле п. 1 § 2 гл. 1.

Пусть Π — произвольная плоскость, α — прямая на плоскости Π и O — точка на прямой α. Фигура (O, α^+, Π^+), состоящая из точки O, некоторого луча α^+ прямой α с началом в точке O и некоторой полуплоскости Π^+, определенной прямой α, называется *флагом на плоскости* Π. Аналогично опреде-

ляется *флаг в пространстве* (добавляется еще одно из полупространств, определенных плоскостью Π) и *флаг* (O, α^+) *на прямой* (последнее понятие, впрочем, совпадает по существу с понятием луча).

Пара h, k лучей, выходящих из одной точки O и не принадлежащих одной прямой, называется *углом* и обозначается символом $\angle(h, k)$. Если A и B — точки лучей h и k, то угол $\angle(h, k)$ обозначается также символом $\angle AOB$.

Говорят, что угол $\angle(h, k)$ *примыкает* к флагу (O, α^+, Π^+), если луч h совпадает с лучом α^+, а луч k лежит в полуплоскости Π^+.

Аналогично, говорят, что отрезок $\overline{AB}$ *примыкает* к флагу (O, α^+), если точка A совпадает с точкой O, а точка B принадлежит лучу α^+.

Как уже было сказано, в аксиоматике Гильберта для отрезков и углов определено отношение «конгруэнтности».

Мы скажем, что угол $\angle(h, k)$ может быть *отложен* от флага (O, α^+, Π^+), если существует конгруэнтный ему угол, примыкающий к этому флагу. Аналогично, мы скажем, что отрезок $\overline{AB}$ может быть *отложен* от флага (O, α^+), если существует конгруэнтный ему отрезок, примыкающий к этому флагу.

Третья группа аксиом

(аксиомы конгруэнтности)

III_1. *Любой отрезок можно отложить от любого флага.*

III_2. *Если отрезки* $\overline{AB}$ *и* $\overline{A''B''}$ *конгруэнтны одному и тому же отрезку* $\overline{A'B'}$, *то отрезки* $\overline{AB}$ *и* $\overline{A''B''}$ *также конгруэнтны.*

III_3. *Для любой точки* B, *лежащей между точками* A *и* C, *и любой точки* B', *лежащей между точками* A' *и* C', *из того, что отрезок* $\overline{AB}$ *конгруэнтен отрезку* $\overline{A'B'}$, *а отрезок* $\overline{BC}$ *конгруэнтен отрезку* $\overline{B'C'}$, *следует, что отрезок* $\overline{AC}$ *конгруэнтен отрезку* $\overline{A'C'}$.

III_4. *Любой угол можно единственным образом отложить от любого флага.*

III_5. *Каждый угол конгруэнтен самому себе.*

III_6. *Пусть* A, B, C — *три точки, не лежащие на одной прямой, и* A', B', C' — *три точки, также не лежащие на одной прямой. Тогда, если отрезок* $\overline{AB}$ *конгруэнтен отрезку* $\overline{A'B'}$, *отрезок* $\overline{AC}$ *конгруэнтен отрезку* $\overline{A'C'}$, *а угол* $\angle BAC$ *конгруэнтен углу* $\angle B'A'C'$, *то угол* $\angle ABC$ *конгруэнтен углу* $\angle A'B'C'$.

Обратим внимание на то, что единственность откладывания углов требуется, а единственность откладывания отрезков — нет (она может быть доказана).

Из этих аксиом вытекает следующее утверждение, их существенно дополняющее:

отношения конгруэнтности отрезков и углов являются отношениями эквивалентности.

Действительно, пусть $\overline{AB}$ — произвольный отрезок. Отложив его от произвольного флага, мы получим некоторый отрезок $\overline{A'B'}$, которому конгруэнтен отрезок $\overline{AB}$ (аксиома III_1). Принимая теперь в аксиоме III_2 за отрезок

$\overline{A''B''}$ отрезок $\overline{AB}$, мы видим, что все условия этой аксиомы выполнены и, следовательно, отрезок $\overline{AB}$ конгруэнтен отрезку $\overline{A''B''}$, т. е. конгруэнтен самому себе. Следовательно, отношение конгруэнтности отрезков рефлексивно.

Его симметричность и транзитивность непосредственно вытекают теперь из аксиомы III_2.

Что же касается углов, то рефлексивность конгруэнтности для них требуется специальной аксиомой III_5, а симметричность и транзитивность без труда выводятся из симметричности и транзитивности конгруэнтности отрезков с помощью аксиомы III_6.

Упражнение. Докажите симметричность и транзитивность отношения конгруэнтности для углов.

Из аксиом групп I—III можно уже вывести довольно много теорем геометрии (все, для которых не нужна единственность параллельных и теория измерения). Например, можно доказать признаки равенства (конгруэнтности) треугольников, теорему о внешнем угле треугольника, теорему о перпендикуляре и наклонной и т. д. В частности, можно доказать, что через любую точку A, не принадлежащую прямой α, проходит прямая, *параллельная* этой прямой (т. е. содержащаяся в плоскости, определенной прямой α и точкой A, но не пересекающаяся с прямой α). Чтобы построить такую прямую, достаточно опустить из точки A на прямую α перпендикуляр (что это такое — определяется обычным образом) и затем восставить к нему перпендикуляр в точке A (конечно, нужно доказать, что все эти построения выполнимы).

Последние две группы аксиом состоят каждая из одной аксиомы.

Четвертая группа аксиом

(аксиома непрерывности)

IV. *Если все точки прямой распределены на два непустых класса так, что в некоторой ориентации каждая точка одного класса предшествует каждой точке другого класса, то либо в первом классе существует точка, которой предшествуют все остальные точки этого класса, либо во втором классе существует точка, которая предшествует остальным точкам этого класса.*

Аксиома IV известна как аксиома Дедекинда. Ее можно заменить двумя аксиомами: известной аксиомой Архимеда и аксиомой Кантора (принципом стягивающихся отрезков).

Пятая группа аксиом

(аксиома параллельности)

V. *Через любую точку A, не принадлежащую данной прямой α, проходит не более одной прямой, параллельной прямой α.*

Перечисленных аксиом оказывается достаточно для построения всей геометрии. В частности, они позволяют определить понятие координатной системы (евклидовой), что доказывает как их непротиворечивость, так и их полноту. Доказательство этого мы опустим.

Замечание 1. Аксиоматика Гильберта является аксиоматикой геометрии в пространстве (стереометрии). Чтобы получить аналогичную аксиоматику планиметрии, достаточно исключить плоскости из списка основных объектов и оставить в группе I аксиом принадлежности только первые четыре аксиомы. Ясно, что

аксиоматика планиметрии интерпретируется на любой плоскости П.

Это утверждение и является точной формулировкой того факта, что аксиомы планиметрии описывают геометрию на любой плоскости.

2. Обсуждение аксиом

Популярность аксиоматики Гильберта объясняется не только тем, что она была первой, получившей широкое признание аксиоматикой геометрии, но и ее близостью к традиционной школьной геометрии. Основные объекты и основные отношения этой аксиоматики известны из школы, а ее аксиомы либо очевидны, либо являются теоремами школьного курса (например, аксиома III_6 — это, по существу, первый признак равенства треугольников).

Аксиоматика Гильберта обладает также рядом принципиальных достоинств.

Во-первых, можно показать, что аксиоматика Гильберта *минимальна*: из нее не только нельзя удалить ни одной аксиомы, но и нельзя даже ослабить их формулировки. Впрочем, на практике это методологическое достоинство оборачивается недостатком: прежде чем дойти до действительно интересных геометрических фактов, приходится доказывать массу утверждений, тривиально дополняющих аксиомы (но доказательства которых часто совсем не тривиальны). Если не гнаться за минимальностью, то все эти утверждения естественно с самого начала включить в аксиомы.

Второе методологическое достоинство аксиоматики Гильберта состоит в ее *независимости* от каких-либо других математических теорий. Она в принципе независима даже от теории множеств. Действительно, по Гильберту, скажем, прямая отнюдь не является множеством точек, а отношение принадлежности не является теоретико-множественным отношением принадлежности элемента к множеству. Однако эта независимость от теории множеств на самом деле эфемерна. Уже полупрямая вводится по Гильберту по существу как множество точек. Еще хуже дело обстоит с аксиомой непрерывности Дедекинда, в которой понятие множества (класса) играет основную роль. (Правда, у самого Гильберта аксиомы Дедекинда нет: ее заменяет некая аксиома полноты, формально от теории множеств независимая.)

Как бы то ни было, в духе современных теоретико-множественных концепций, придавать большое значение этой особенности аксиоматики Гильберта не стоит. Более того, переходя на теоретико-множественную точку зрения, целесообразно видоизменить эту аксиоматику и считать прямую множеством принадлежащих ей точек. По существу мы выше так и поступали (по крайней мере в терминологическом аспекте). При таком видоизменении понятия прямой, мы, конечно, несколько сузим класс возможных интерпретаций: те интерпретации, в которых принадлежность не является теоретико-множественной, будут уже невозможны. Однако много мы при этом не потеряем, поскольку от любой такой интерпретации возможен по существу

автоматический переход к интерпретации с теоретико-множественным отношением принадлежности. Выиграем же мы не только в наглядности, но и в уменьшении списка основных отношений.

Если не заботиться о минимальности, то аксиомы принадлежности можно теперь сформулировать следующим образом (удобнее называть их теперь «аксиомами прямых и плоскостей»):

Аксиомы прямых и плоскостей

Пр. 1. *Любая прямая является бесконечным (несчетным) множеством точек.*

Пр. 2. *Любые две (различные) точки содержатся в одной и только одной прямой.*

Пр. 3. *Вне любой данной прямой существуют точки.*

Пл. 1. *Любая плоскость является бесконечным (несчетным) множеством точек.*

Пл. 2. *Любые три точки, не принадлежащие одной прямой, содержатся в одной и только одной плоскости.*

Пл. 3. *Если две различные точки прямой принадлежат плоскости, то вся прямая лежит в этой плоскости.*

Пл. 4. *Пересечение двух плоскостей либо пусто, либо является прямой.*

Пл. 5. *Вне любой плоскости существуют точки.*

Ступив раз на «теоретико-множественный» путь, мы можем пойти и дальше, убрав из списка основных отношений также и отношение коллинеарности и заменив его новым основным объектом — «движением». В рамках аксиоматики Гильберта *движения* определяются как (биективные) преобразования множества точек, переводящие конгруэнтные отрезки в конгруэнтные. С содержательной точки зрения «движением» является либо результат «перемещения» всего пространства по себе как твердого тела, либо результат такого перемещения, сопровождаемый некоторой симметрией (относительно плоскости).

Аксиомы движения

Дв. 1. *Любое движение является биективным отображением множества всех точек на себя (преобразованием), переводящим прямые в прямые.*

Дв. 2. *Для любых двух (пространственных) флагов существует единственное движение, переводящее первый флаг во второй.*

Упражнение. Выведите Дв. 1 и Дв. 2 из аксиоматики Гильберта. Обратно, исходя из движений, удовлетворяющих аксиомам Дв. 1 и Дв. 2, определите обычным образом конгруэнтность (как совмещаемость при некотором движении) и проверьте, что все аксиомы конгруэнтности будут выполнены.

Можно показать (см. ниже, § 1 гл. 7), что аксиома Дв. 1 влечет за собой сохранение при движениях отношения «между». Поэтому любое движение переводит луч в луч, полуплоскость в полуплоскость, полупространство в полупространство, а следовательно, флаг во флаг. Это делает осмысленной формулировку аксиомы Дв. 2.

Заметим еще, что из аксиомы Дв. 2 без труда вытекает, что все движения составляют, как и полагается, группу.

Читатель может сравнить сложные и в достаточной мере неуклюжие аксиомы Гильберта III_1—III_6 с простыми и наглядными аксиомами Дв. 1 и Дв. 2. Это не означает, конечно, что Гильберт не понимал преимуществ аксиом Дв. 1 и Дв. 2. Просто они для него были неприемлемы из-за его «антитеоретико-множественной» установки.

Можно считать, что в аксиоматике Гильберта неудачен и выбор в качестве одного из основных отношений отношения «между». Значительно удобнее (как по методическим, так и по принципиальным соображениям) принимать за основное отношение отношение одноименности направленных отрезков (вводимое обычным образом: см. п. 1 § 2 гл. 1), или (что равносильно, но означает переход на теоретико-множественные позиции) ввести новый основной объект — «ориентацию», подчинив его соответствующим аксиомам (перечисленным в п. 1 § 2 гл. 1). Введение в аксиоматику ориентаций устраняет последнее из оставшихся основных отношений (отношение «между»). Это, конечно, не означает, что в получающейся аксиоматике нет никаких отношений: просто этими отношениями будут стандартные отношения теории множеств.

Независимость аксиоматики Гильберта проявляется также и в том, что она не опирается на теорию вещественных чисел, из-за чего большой фрагмент теории чисел по существу заново строится в недрах этой аксиоматики (достаточно чуть-чуть переформулировать линейные аксиомы порядка и аксиому непрерывности, чтобы получить аксиоматику, описывающую множество вещественных чисел [1])). Но если мы уже решили отказаться от этой независимости, то естественно сделать следующий шаг и построить геометрию, явно опираясь на арифметику вещественных чисел. Проще всего это сделать, аксиоматизировав понятие «координатной системы на прямой».

Аксиомы координатных систем

КС1. *Любая координатная система x на прямой α представляет собой биективное отображение этой прямой на множество $\mathbb{R}$ вещественных чисел.*

КС2. *Для любого направленного отрезка $\overrightarrow{AB}$ прямой α существует на этой прямой единственная координатная система x, для которой*

$$x(A) = 0, \quad x(B) = 1.$$

КС3. *Координатные системы x и y, отвечающие направленным отрезкам $\overrightarrow{AB}$ и $\overrightarrow{CD}$ прямой α, связаны соотношением*

$$y = ax + b,$$

где

$$a = y(B) - y(A), \quad b = y(A).$$

Эти аксиомы равносильны аксиоме непрерывности IV и линейным аксиомам порядка II_1—II_3: достаточно определить отношение «между», считая, что точка B тогда и только тогда находится между точками A и C, когда в координатной системе x, отвечающей (в силу аксиомы КС2) направленному отрезку $\overrightarrow{AC}$, имеет место неравенство

$$0 < x(B) < 1.$$

[1]) Строго говоря, нужна еще аксиома III_1, позволяющая переносить отрезки по прямой.

Резюмируя все сказанное, мы получаем следующее

Предложение 1. *Евклидова геометрия может быть задана аксиоматикой, основными объектами которой являются*

«точки», «прямые», «плоскости»,

«движения», «координатные системы»,

удовлетворяющие аксиомам Пр. 1 — Пр. 3, Пл. 1 — Пл. 5, КС1 — КС3, Дв. 1 — Дв. 2, *аксиоме Паша* III_4 *и аксиоме параллельности* V.

Эта аксиоматика в достаточной степени наглядна и проста. Вместе с тем она не слишком отходит от традиционных «школьных» понятий (представление о прямой как о числовой оси в школе известно).

Чтобы получить из этой аксиоматики аксиоматику планиметрии, достаточно удалить аксиомы Пл. 1 — Пл. 5 (а аксиому Дв. 2 сформулировать для флагов на плоскости).

При желании можно, конечно, пойти и дальше, зааксиоматизировав понятие координатной системы в пространстве (для планиметрии — понятие координатной системы на плоскости). В результате мы придем к аксиоматике, рассмотренной в § 3 гл. 2 и содержащей только две аксиомы. Однако такая аксиоматика уже совершенно порывает со школьной традицией.

С другой стороны, можно несколько вернуться назад и заменить аксиомы КС1 — КС3 аксиомой ориентаций У (см. п. 1 § 2 гл. 1) и аксиомой Дедекинда IV. Как уже отмечалось, при этом придется заново построить практически всю теорию вещественных чисел (в части, не связанной с арифметическими операциями).

Особенностью аксиоматики Гильберта является также специальная роль, которую в ней играет аксиома параллельности V. Исключение этой аксиомы из группы I (где ей естественное место) и выделение ее в специальную (последнюю!) группу V позволяет отделить теоремы геометрии, не зависящие от аксиомы параллельности, от теорем, от этой аксиомы зависящих. Это играет существенную роль, когда мы хотим сравнить евклидову геометрию с *геометрией Лобачевского*, которая удовлетворяет всем аксиомам Гильберта, за исключением аксиомы параллельности, заменяемой следующей аксиомой:

Аксиома параллельности Лобачевского

V′. *В любой плоскости, содержащей произвольную прямую* α, *через любую точку, не принадлежащую прямой* α, *проходят по крайней мере две прямые, не пересекающие прямую* α.

Революционная идея Лобачевского о возможности многих геометрий, произведшая глубочайший переворот в понимании сути и значения не только геометрии, но и всей математики, была признана не сразу. Одной из целей Гильберта было показать, что геометрия Лобачевского с принципиальной, формально-логической, стороны ничем не хуже и не лучше евклидовой геометрии: каждая из этих геометрий может быть описана системами аксиом, отличающихся лишь одной аксиомой.

Однако, с современных позиций, геометрия Лобачевского является лишь одной из многих неевклидовых геометрий (хотя, возможно, и самой важной). Поэтому выделять аксиому параллельности нет сейчас никаких оснований.

С современной точки зрения представляется более важным отделить в аксиоматике не аксиому параллельности, а аксиомы, описывающие афинную геометрию.

В аксиоматике Гильберта аффинный характер имеют все аксиомы, за исключением аксиом конгруэнтности III. Однако из этих аксиом не вытекает существование параллельных прямых. Поэтому аксиому параллельности V целесообразно соответствующим образом усилить.

Усиленная аксиома параллельности

V*. *Через любую точку, не принадлежащую прямой* α, *можно провести единственную прямую, параллельную прямой* α.

Оказывается, что

аксиомы I_1—I_{10}, II_1—II_4, IV и V* (*или аксиомы* Пр. 1 — Пр. 3, Пл. 1 — Пл. 5, КС1 — КС3, II_4 *и* V*) *составляют полную систему аксиом афинной геометрии в пространстве.*

Доказательство мы приводить не будем.

Замечательно, что, опустив в этом списке пространственные аксиомы $I_1 — I_{10}$ (или соответствующие аксиомы Пл. 1 — Пл. 5), мы не получим полной системы аксиом афинной планиметрии: необходима некоторая дополнительная аксиома, связывающая точки различных прямых.

Пусть на плоскости заданы три прямые α_0, α и β, причем прямые α и β не параллельны прямой α_0. Тогда для любой точки A прямой β прямая, параллельная прямой α_0 и проходящая через точку A, пересекает прямую α в единственной точке B. Построенное таким образом отображение

$$\theta\colon A \longmapsto B$$

прямой β на прямую α называется, как мы знаем, *проектированием* параллельно прямой α_0.

Дополнительная аксиома координатных систем

КС4. *Для любой координатной системы* x *на прямой* α *и любого проектирования* θ *прямой* β *на прямую* α *композиция*

$$x \circ \theta\colon \beta \to \mathbb{R}$$

является координатной системой на прямой β.

Аксиома Паппа

Пп. *Если шесть различных точек* A, B, C, A', B', C' *одной плоскости таковы, что*

а) *точка* C *принадлежит прямой* AB, *а точка* C' *— прямой* $A'B'$;

б) *ни одна из точек* A, B, C, A', B', C' *не является общей точкой прямых* AB *и* $A'B'$;

в) *прямая* $B'C$ *параллельная прямой* BC', *а прямая* $A'C$ *параллельна прямой* CA',

то прямая $A'B$ *параллельна прямой* AC'.

Аксиома Дезарга

Дез. *Если шесть различных точек* A, B, C, A', B', C' *таковы, что*

а) *прямые* AA', BB' *и* CC' *либо параллельны, либо пересекаются в одной точке, отличной от точек* A, B, C, A', B', C',

б) *прямая* AB *параллельна прямой* $A'B'$, *а прямая* BC *параллельна прямой* $B'C'$,

то прямая AC *параллельна прямой* $A'C'$.

Можно показать (мы этого делать не будем), что

при добавлении к аксиомам $I_1 - I_4$, II, IV *и* V* (*или к аксиомам* Пр. 1—Пр. 3, КС1 — КС3, II_4 *и* V*) *любой из аксиом* КС4, Пп *или* Дез. *получается полная система аксиом аффинной геометрии на плоскости.*

Все эти аксиомы имеют «элементарно-геометрический» характер. С позиций аналитической геометрии наиболее приемлема, конечно, аксиоматика аффинной геометрии, изложенная в § 3 гл. 2 (и состоящая только из двух аксиом). Помимо всего прочего, последняя аксиоматика имеет и то преимущество, что она немедленно обобщается на случай любого поля $\mathbb{K}$ и на любое число измерений n.

Возможны, впрочем, и другие (часто более удобные) аксиоматики аффинной геометрии, обобщаемые на случай любых $\mathbb{K}$ и n. Наиболее известной из них является аксиоматика, предложенная немецким математиком Г. Вейлем.

В аксиоматике Вейля имеется два рода основных объектов:

«точки» и «векторы».

Предполагается, что для векторов определены операции сложения и умножения на вещественные числа, обладающие всеми восемью стандартными свойствами из п. 5 § 2 гл. 1 (т. е. превращающие множество всех векторов в линеал). Кроме того, предполагается, что любые четыре вектора линейно зависимы (если мы строим аффинную геометрию плоскости, то число «четыре» заменяется, естественно, числом «три»).

Далее, предполагается, что любым двум точкам A и B сопоставлен некоторый вектор $\overrightarrow{AB}$, причем

1) *для любой точки* A *и любого вектора* $\boldsymbol{a}$ *существует единственная точка* B *такая, что*

$$\boldsymbol{a} = \overrightarrow{AB};$$

2) *для любых трех точек* A, B, C *имеет место равенство*

$$\overrightarrow{AC} = \overrightarrow{AB} + \overrightarrow{BC}.$$

Оказывается, что этих аксиом уже достаточно для построения всей аффинной геометрии.

Эта аксиоматика немедленно обобщается на случай любого поля $\mathbb{K}$ (достаточно потребовать, чтобы векторы умножались не на вещественные числа, а на элементы поля $\mathbb{K}$), а также на случай любого числа измерений (достаточно число «4» заменить любым фиксированным числом «n»). Более того, ясно, что эти аксиомы имеют смысл и тогда, когда $\mathbb{K}$ является произвольным *телом* (полем с некоммутативным умножением). Они определяют в этом случае *аффинную геометрию над телом* $\mathbb{K}$.

Глава 3

ЛИНИИ И ПОВЕРХНОСТИ ПЕРВОГО ПОРЯДКА

§ 1. ПРЯМАЯ НА ПЛОСКОСТИ

1. Прямая как линия первого порядка

Рассмотрим плоскость с заданной на ней аффинной координатной системой Oxy и в ней произвольную прямую. Мы всегда можем найти (и даже многими способами) аффинную систему координат $O'x'y'$, для которой данная прямая является осью ординат и потому определяется уравнением $x' = 0$. Но как мы знаем (п. 2 § 1 гл. 2), координата x' выражается через исходные координаты x, y по формуле

$$x' = Ax + By + C,$$

где A, B и C — некоторые вещественные числа, причем хотя бы одно из чисел A или B отлично от нуля (в обозначениях п. 2 § 1 гл. 2, $A = c_1^{1\prime}$, $B = c_2^{1\prime}$, $C = a^1$). Тем самым доказано, что

любая прямая на плоскости имеет уравнение вида

$$Ax + By + C = 0, \tag{1}$$

где хотя бы одно из чисел A или B отлично от нуля.

Уравнение прямой вида (1) называется ее *общим уравнением.*

В терминологии, введенной в п. 2 § 2 гл. 2, доказанное утверждение означает, что

каждая прямая на плоскости является линией первого порядка.

Покажем, что верно и обратное, т. е. что

любая линия первого порядка является прямой.

Действительно, пусть

$$Ax + By + C = 0$$

— уравнение произвольной линии первого порядка. Подберем числа A_1, B_1 и C_1 так, чтобы определитель

$$\begin{vmatrix} A & B \\ A_1 & B_1 \end{vmatrix} \tag{2}$$

был отличен от нуля (например, можно положить $A_1 = -B$, $B_1 = A$). На число C_1 мы не накладываем никаких условий (например, можно считать, что $C_1 = 0$). Поскольку определитель (2) отличен от нуля, формулы

$$x' = Ax + By + C,$$
$$y' = A_1x + B_1y + C_1$$

описывают переход от координат x, y к некоторым новым координатам x', y'. В координатах x', y' данная линия выражается уравнением $x' = 0$ и, следовательно, является осью ординат, т. е. прямой. Поскольку тот факт, что данная линия является прямой, не зависит от системы координат, наше утверждение, тем самым, полностью доказано.

Собирая вместе оба доказанных утверждения, мы получаем следующую теорему:

Теорема 1. *Линия на плоскости тогда и только тогда представляет собой прямую, когда она является алгебраической линией первого порядка.*

Замечание 1. При формально-аксиоматическом построении геометрии (см. § 3 гл. 2) следует, как всегда, «обратить» полученный содержательный результат и принять его за определение.

Таким образом, на произвольной аффинной плоскости $\mathfrak{M}$ прямые определяются как алгебраические линии первого порядка, т. е. как множества точек, координаты x, y которых в произвольной аффинной координатной системе $\alpha\colon \mathfrak{M} \to \mathbb{R}^2$ удовлетворяют уравнению вида

$$Ax + By + C = 0,$$

где либо A, либо B не равно нулю. Это определение корректно, поскольку любое преобразование из группы Aff ($\mathbb{R}^2$) линейное уравнение переводит снова в линейное уравнение.

Заметим, что это определение применимо к аффинным (а потому и к евклидовым) плоскостям над л ю б ы м полем $\mathbb{K}$ (в частности, к комплексным и вещественно-комплексным плоскостям). При этом среди прямых на вещественно-комплексной плоскости выделяются *вещественные прямые*, имеющие уравнения с вещественными коэффициентами, а остальные прямые распадаются на пары *комплексно-сопряженных прямых*, уравнения которых имеют комплексно-сопряженные коэффициенты.

Соглашение об обозначениях. В дальнейшем символ λ будет обозначать прямую с уравнением (1). В случае, когда нам придется рассматривать несколько прямых, мы будем снабжать символ λ индексами. При этом мы всегда будем молчаливо подразумевать, что коэффициенты уравнения (1) снабжены соответствующими индексами. Так, например, прямая λ_1 будет иметь уравнение

$$A_1x + B_1y + C_1 = 0$$

и т. п.

Символ $\Leftrightarrow$ мы будем использовать для обозначения эквивалентности (равносильности) двух утверждений.

Символ $\parallel$ будет обозначать параллельность прямых, а символ $\perp$ — их перпендикулярность.

Равенство нулю одного (или двух) из коэффициентов A, B, C уравнения прямой указывает на специальное положение прямой относительно координатных осей.

Задание. Докажите, что

$$A = 0 \Leftrightarrow \lambda \parallel Ox;$$
$$B = 0 \Leftrightarrow \lambda \parallel Oy;$$
$$C = 0 \Leftrightarrow \lambda \text{ проходит через } O\,(0, 0).$$

Если $B \neq 0$, т. е. если прямая не параллельна оси ординат, ее уравнение может быть записано в виде

$$y = kx + b,$$

где $k = -\frac{A}{B}$, $b = -\frac{C}{B}$. Это уравнение называется *уравнением с угловым коэффициентом*, а число k называется *угловым коэффициентом* прямой. Число b называется *свободным членом*. Оно равно величине направленного отрезка $\overrightarrow{OB}$, отсекаемого прямой на оси ординат, т. е. равно ординате точки B пересечения прямой с осью ординат.

Замечание 2. Уравнение с угловым коэффициентом показывает (если мы пользуемся прямоугольными координатами), что *прямая (не параллельная оси ординат) является графиком линейной функции* $y = kx + b$.

Задание. Докажите, что
две прямые с угловыми коэффициентами k_1 и k_2 *тогда и только тогда параллельны, когда* $k_1 = k_2$.

При $k = 0$ (т. е. при $A = 0$) прямая имеет уравнение вида

$$y = b$$

и параллельна оси абсцисс.

При $B = 0$ уравнение с угловым коэффициентом написать невозможно. Вместо него можно рассматривать уравнение

$$x = a,$$

где $a = -\frac{C}{A}$ — величина направленного отрезка $\overrightarrow{OA}$, отсекаемого прямой на оси абсцисс, т. е. абсцисса точки A пересечения прямой с осью абсцисс.

В этом случае удобно условно считать, что $k = \infty$.

2. Параметрические и канонические уравнения прямой

Определение 1. *Направляющим вектором* прямой на плоскости (или в пространстве) называется произвольный отличный от нуля вектор, параллельный этой прямой, т. е. принадлежащий соответствующему линеалу Vect(1). Этот вектор определен с точностью до коллинеарности (и представляет собой не что иное, как базис линеала Vect(1)).

Предложение 1. *Отличный от нуля вектор* $\boldsymbol{a}(l, m)$ *тогда и только тогда является направляющим вектором прямой* λ, *когда*

$$Al + Bm = 0. \tag{1}$$

Доказательство. Пусть $M_0(x_0, y_0)$ — произвольная точка прямой λ и пусть $\boldsymbol{a} = \overrightarrow{M_0M_1}$. Ясно, что вектор $\boldsymbol{a}$ тогда и только тогда является направляющим вектором прямой λ, когда точка $M_1(x_0 + l, y_0 + m)$ принадлежит этой прямой, т. е. когда

$$A(x_0 + l) + B(y_0 + m) + C = 0.$$

Для завершения доказательства остается раскрыть скобки и учесть равенство

$$Ax_0 + By_0 + C = 0$$

(выражающее принадлежность точки M_0 прямой).

Условие (1), очевидно, выполнено при $l = B$ и $m = -A$. Таким образом, имеет место

Следствие. *Вектор* $\boldsymbol{a} = (B, -A)$ *является направляющим вектором прямой*

$$Ax + By + C = 0.$$

Замечание 1. Это следствие можно доказать и иначе, рассмотрев координаты

$$x' = Ax + By + C,$$
$$y' = A_1x + B_1y + C_1,$$

в которых прямая λ является осью ординат (см. предыдущий пункт). В этой координатной системе вектор $\boldsymbol{a}$ имеет координаты

$$A \cdot B + B \cdot (-A) = 0,$$
$$A_1 \cdot B + B_1 \cdot (-A) \neq 0.$$

Следовательно, он параллелен оси ординат $x' = 0$, т. е. прямой λ.

Пусть прямая с направляющим вектором $\boldsymbol{a}$ проходит через точку M_0. Ясно, что точка M тогда и только тогда принадлежит этой прямой, когда вектор $\overrightarrow{M_0M}$ коллинеарен вектору $\boldsymbol{a}$. По-

скольку $\boldsymbol{a} \neq \mathbf{0}$, это означает, что существует такое число t, что

$$\overrightarrow{M_0M} = t\boldsymbol{a}.$$

Пусть $\boldsymbol{r}_0$ — радиус-вектор точки M_0, а $\boldsymbol{r}$ — радиус-вектор точки M. Тогда предыдущее равенство может быть записано в виде

$$\boldsymbol{r} = \boldsymbol{r}_0 + t\boldsymbol{a}.$$

Это — *параметрическое уравнение прямой в векторной форме.* При любом t вектор $\boldsymbol{r}$, выражаемый этим уравнением, является радиус-вектором некоторой точки прямой, и обратно, радиус-вектор любой точки прямой имеет такой вид.

Заметим, что здесь мы нигде не пользовались тем, что прямая лежит в плоскости. Поэтому доказанное утверждение справедливо и для прямых в пространстве.

Записав векторное параметрическое уравнение прямой в координатах, мы получим параметрические уравнения прямой в координатной форме. На плоскости они имеют вид

$$\begin{aligned} x &= x_0 + lt, \\ y &= y_0 + mt, \end{aligned}$$

где x_0, y_0 — координаты точки M_0 и l, m — координаты направляющего вектора $\boldsymbol{a}$, а в пространстве — вид

$$\begin{aligned} x &= x_0 + lt, \\ y &= y_0 + mt, \\ z &= z_0 + nt. \end{aligned}$$

Точка M_0 и вектор $\boldsymbol{a}$ составляют некоторый аффинный координатный репер на рассматриваемой прямой. Ясно, что

параметр t является аффинной координатой на прямой относительно репера $M_0\boldsymbol{a}$.

Коллинеарность векторов $\overrightarrow{MM_0} = \boldsymbol{r} - \boldsymbol{r}_0$ и $\boldsymbol{a}$ означает также, что их координаты пропорциональны:

$$\frac{x - x_0}{l} = \frac{y - y_0}{m}$$

(мы снова рассматриваем лишь прямые на плоскости). Следовательно, это равенство является уравнением прямой, проходящей через точку $M_0(x_0, y_0)$ и имеющей направляющий вектор $\boldsymbol{a}(l, m)$. Уравнение прямой, записанное в таком виде, называется ее *каноническим уравнением.*

Чтобы по общему уравнению прямой на плоскости

$$Ax + By + C = 0$$

написать ее каноническое уравнение (или, что равносильно, ее параметрические уравнения), следует найти хотя бы одну ее точку $M_0(x_0, y_0)$ (например, при $A \neq 0$ можно положить $x_0 = -\frac{C}{A}$, $y_0 = 0$, а при $B \neq 0$ можно положить $x_0 = 0$, $y_0 = -\frac{C}{B}$) и хотя бы один направляющий вектор $\boldsymbol{a}$ (например, можно взять $\boldsymbol{a} = (-B, A)$).

Замечание 2. Каноническое уравнение является, собственно говоря, не равенством, а пропорцией и должно пониматься именно в этом смысле. Например, при $l = 0$ оно означает, что $x = x_0$.

Полагая $A = m$ и $B = -l$, мы можем каноническое уравнение записать в следующем виде:

$$A(x - x_0) + B(y - y_0) = 0.$$

Это — *общее уравнение прямой, проходящей через данную точку.*

Если на прямой даны две различные точки $M_0(x_0, y_0)$ и $M_1(x_1, y_1)$, то за направляющий вектор этой прямой мы можем принять вектор $\overrightarrow{M_0M_1}$ с координатами $x_1 - x_0$, $y_1 - y_0$. Следовательно,

прямая, проходящая через точки $M_0(x_0, y_0)$ *и* $M_1(x_1, y_1)$, *имеет уравнение*

$$\frac{x - x_0}{x_1 - x_0} = \frac{y - y_0}{y_1 - y_0}.$$

В детерминантной форме это уравнение имеет вид

$$\begin{vmatrix} x - x_0 & y - y_0 \\ x_1 - x_0 & y_1 - y_0 \end{vmatrix} = 0.$$

Иногда его удобно записывать в более симметричном виде:

$$\begin{vmatrix} x & y & 1 \\ x_0 & y_0 & 1 \\ x_1 & y_1 & 1 \end{vmatrix} = 0$$

(для доказательства равносильности этих уравнений достаточно вторую строку последнего определителя вычесть из его первой и третьей строк).

Параметрические уравнения рассматриваемой прямой имеют, очевидно, вид

$$\begin{aligned} x &= x_0 + (x_1 - x_0)t, \\ y &= y_0 + (y_1 - y_0)t, \end{aligned} \quad \text{или} \quad \begin{aligned} x &= (1 - t)x_0 + tx_1, \\ y &= (1 - t)y_0 + ty_1. \end{aligned}$$

Сравнивая эти уравнения с формулами

$$x = \frac{x_0 + kx_1}{1 + k}, \quad y = \frac{y_0 + ky_1}{1 + k},$$

определяющими координаты точки, делящей отрезок $\overrightarrow{M_0M_1}$ в данном отношении $k \neq -1$ (см. п. 3 § 1 гл. 2), мы немедленно получаем, что

точка M прямой, отвечающая значению параметра t, делит отрезок $\overrightarrow{M_0M_1}$ в отношении $t/(1-t)$.

Иными словами (см. п. 6 § 1 гл. 2),

числа $1-t$, t являются однородными барицентрическими координатами на прямой.

Полезно сравнить это истолкование параметра t с его истолкованием как аффинной координаты на прямой (см. выше).

Из обоих истолкований немедленно вытекает, в частности, что

точки отрезка $\overline{M_0M_1}$ характеризуются неравенствами

$$0 \leqslant t \leqslant 1$$

(*значению $t = 0$ отвечает точка M_0, а значению $t = 1$ — точка M_1*).

Заметим, что это утверждение остается справедливым и для прямой в пространстве.

Параметрические уравнения прямой

$$\begin{aligned} x &= x_0 + lt, \\ y &= y_0 + mt \end{aligned} \tag{2}$$

особенно полезны в задаче о вычислении точек пересечения этой прямой с произвольной линией

$$F(x, y) = 0. \tag{3}$$

Действительно, чтобы найти эти точки, достаточно решить (относительно t) уравнение

$$F(x_0 + lt, y_0 + mt) = 0. \tag{4}$$

Каждый корень t_1 этого уравнения дает нам точку с координатами

$$\begin{aligned} x_1 &= x_0 + lt_1, \\ y_1 &= y_0 + mt_1; \end{aligned} \tag{5}$$

принадлежащую одновременно прямой (2) и линии (3), т. е. являющуюся точкой пересечения прямой (2) и линии (3). Обратно, если точка (x_1, y_1) является точкой пересечения прямой

(2) и линии (3), то поскольку она принадлежит прямой (2), существует такое t_1, что имеют место равенства (5), а поскольку она принадлежит линии (3), это t_1 является корнем уравнения (4).

Замечание 3. Обратим внимание на то, что описанный метод имеет общий характер и применим к случаю, когда вместо прямой (2) мы рассматриваем произвольную (параметрически заданную) линию.

В частном случае, когда (3) является алгебраической линией порядка n, т. е. функция $F(x, y)$ — многочленом от x, y степени n, уравнение (4) имеет вид

$$\Phi_0 t^n + \Phi_1 t^{n-1} + \ldots + \Phi_{n-1} t + \Phi_n = 0, \qquad (6)$$

где Φ_0, Φ_1, ..., Φ_{n-1}, Φ_n — некоторые числа, т. е. является уравнением степени, не большей n (некоторые из чисел Φ_0, Φ_1, ... могут быть равны нулю).

Упражнение. Докажите, что

$$\Phi_0 = F_0(l, m), \quad \Phi_n = F(x_0, y_0),$$

где $F_0(x, y)$ — сумма всех старших (имеющих точно степень n) членов многочлена $F(x, y)$.

Поскольку уравнение степени, не большей n, имеет не более n корней (или удовлетворяется тождественно), тем самым доказано следующее

Предложение 2. *Произвольную алгебраическую линию порядка n каждая прямая (не принадлежащая целиком этой линии) пересекает не более чем в n точках.*

Замечание 4. Ясно, что все вышесказанное без каких-либо изменений справедливо и в (аксиоматически построенной) аффинной геометрии над любым полем $\mathbb{K}$ (конечно, в соответствии с общими правилами, направляющий вектор следует при этом определять как вектор, координаты которого удовлетворяют уравнению (1) [1]). В частности, предложение 2 также остается справедливым.

Заметим, однако, что не только в случае вещественной геометрии, но даже и в случае комплексной (или вещественно-комплексной) геометрии утверждать существование точно n точек пересечения было бы неосторожно, поскольку уравнение (6) может, во-первых, иметь кратные корни, а во-вторых, быть уравнением степени, меньшей n.

Поскольку каждая вещественная прямая в вещественно-комплексной плоскости $\mathfrak{M}$ пересекается с вещественной плоскостью $\mathfrak{M}^{\text{вещ}} \subset \mathfrak{M}$ по бесконечному числу точек (по всем ее вещественным точкам), а вместе с тем не содержится в этой плоскости (имеет невещественные точки), мы видим, что предположение о том, что $\mathfrak{M}^{\text{вещ}}$ является алгебраической линией, противоречит предложению 2 (см. п. 5 § 3 гл. 2, стр. 225).

---

[1]) Само собой разумеется, что при этом следует не забыть проверить корректность этого определения.

3. Взаимное расположение прямых на плоскости

Ясно, что две прямые тогда и только тогда параллельны [1], когда их направляющие векторы коллинеарны. Отсюда (и из следствия к предложению 1 п. 2) непосредственно вытекает следующее

Предложение 1.

$$\lambda \parallel \lambda_1 \Leftrightarrow \frac{A}{A_1} = \frac{B}{B_1} \Leftrightarrow \begin{vmatrix} A & B \\ A_1 & B_1 \end{vmatrix} = 0.$$

Параллельные прямые совпадают, если они имеют хотя бы одну общую точку $M_0(x_0, y_0)$, т. е. если существуют такие числа x_0, y_0, что

$$\begin{aligned} Ax_0 + By_0 + C &= 0, \\ A_1x_0 + B_1y_0 + C_1 &= 0. \end{aligned} \tag{1}$$

Пусть $A = \rho A_1$, $B = \rho B_1$. Вычитая из первого из равенств (1) второе, умноженное на ρ, мы немедленно получим, что $C - \rho C_1 = 0$, т. е. что $C = \rho C_1$. Ясно, что обратное также верно: если $C = \rho C_1$, то рассматриваемые прямые совпадают. Таким образом, справедливо следующее

Предложение 2.

$$\lambda = \lambda_1 \Leftrightarrow \frac{A}{A_1} = \frac{B}{B_1} = \frac{C}{C_1}.$$

Это — теорема единственности для линий первого порядка (см. п. 2 § 2 гл. 2).

Предложения 1 и 2 можно объединить в следующую теорему о взаимном расположении двух прямых на плоскости:

Теорема 1.

а) *прямые* λ *и* λ_1 *пересекаются* $\Leftrightarrow \frac{A}{A_1} \neq \frac{B}{B_1}$;

б) $\lambda \parallel \lambda_1$, *но* $\lambda \neq \lambda_1 \Leftrightarrow \frac{A}{A_1} = \frac{B}{B_1} \neq \frac{C}{C_1}$;

в) $\lambda = \lambda_1 \Leftrightarrow \frac{A}{A_1} = \frac{B}{B_1} = \frac{C}{C_1}$.

Если ввести в рассмотрение матрицы

$$\begin{pmatrix} A & B \\ A_1 & B_1 \end{pmatrix}, \quad \begin{pmatrix} A & B & C \\ A_1 & B_1 & C_1 \end{pmatrix}$$

и обозначить ранг первой матрицы символом r, а второй — символом R, то эту теорему можно будет переформулировать следующим образом:

---

[1] Напомним, что совпадающие прямые мы считаем параллельными.

Теорема 1'.

а) *прямые* λ *и* λ_1 *пересекаются* $\Leftrightarrow r=2$;

б) $\lambda \parallel \lambda_1 \Leftrightarrow r=1,\ R=2$;

в) $\lambda=\lambda_1 \Leftrightarrow r=R=1$.

В частности,

прямые λ *и* λ_1 *имеют хотя бы одну общую точку* $\Leftrightarrow r=R$.

Это — специальный случай теоремы Кронекера — Капелли о совместности систем линейных уравнений (см. дополнение к § 3 гл. 1).

Для трех прямых λ, λ_1, λ_2 на плоскости возможны семь существенно различных типов их взаимного расположения:

Тип 1. Прямые попарно не параллельны и не проходят через одну точку, т. е. образуют треугольник; это — *общий тип.*

Тип 2. Две прямые параллельны, но не совпадают, а третья их пересекает (в этом случае иногда говорят, что прямые образуют «открытый треугольник»).

Тип 3. Прямые не параллельны и проходят через одну точку.

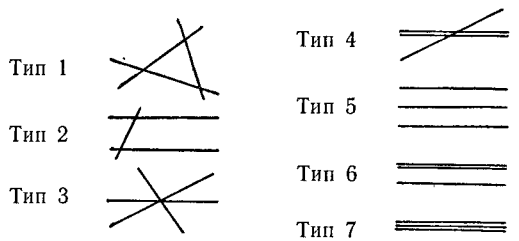
Тип 4. Две прямые совпадают, а третья их пересекает.

Тип 5. Все три прямые попарно параллельны, но попарно не совпадают.

Тип 6. Две прямые совпадают, а третья им параллельна, но от них отлична.

Тип 7. Все три прямые совпадают.

Эти типы расположения можно схематически изобразить следующим образом:



Для исследования этих типов удобно рассмотреть следующие четыре случая:

Случай 1 (типы 1 и 2): прямые не имеют ни одной общей точки и среди них имеются хотя бы две не параллельные (и не совпадающие) прямые.

Случай 2 (типы 3 и 4): прямые имеют единственную общую точку и среди них имеются хотя бы две не параллельные (и не совпадающие) прямые.

Случай 3 (типы 5 и 6): прямые не имеют ни одной общей точки и любые две из них параллельны или совпадают.

Случай 4 (тип 7): все три прямые совпадают.

Пусть r — ранг матрицы

$$\begin{pmatrix} A & B \\ A_1 & B_1 \\ C_1 & B_1 \end{pmatrix}, \tag{2}$$

а R — ранг матрицы

$$\begin{pmatrix} A & B & C \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{pmatrix}. \tag{3}$$

Тогда справедлива следующая теорема о взаимном расположении трех прямых на плоскости:

Теорема 2.

Случай 1 $\Leftrightarrow$ $r=2$, $R=3$;
случай 2 $\Leftrightarrow$ $r=2$, $R=2$;
случай 3 $\Leftrightarrow$ $r=1$, $R=2$;
случай 4 $\Leftrightarrow$ $r=1$, $R=1$.

Доказательство. Предположим сначала, что среди наших прямых имеется хотя бы одна пара непараллельных прямых (случаи 1 и 2). Пусть, для определенности, это — прямые λ и λ_1. Тогда

$$\begin{vmatrix} A & B \\ A_1 & B_1 \end{vmatrix} \neq 0$$

и потому $r=2$. Кроме того, поскольку прямые λ и λ_1 не параллельны (и в частности, не совпадают), они имеют единственную общую точку $M_0(x_0, y_0)$, т. е. существуют такие числа x_0, y_0, что

$$Ax_0 + By_0 + C = 0,$$
$$A_1x_0 + B_1y_0 + C_1 = 0.$$

Поэтому, прибавив к третьему столбцу матрицы (3) первый столбец, умноженный на x_0, и второй столбец, умноженный на y_0, мы получим матрицу

$$\begin{pmatrix} A & B & 0 \\ A_1 & B_1 & 0 \\ A_2 & B_2 & C_2' \end{pmatrix},$$

где $C_2' = A_2x_0 + B_2y_0 + C_2$. Поскольку при этом преобразовании определитель матрицы (3) не меняется, этим доказано, что он равен произведению

$$\begin{vmatrix} A & B \\ A_1 & B_1 \end{vmatrix} C_2'.$$

Следовательно, если $C_2' \neq 0$, т. е. если третья прямая не проходит через точку пересечения первых двух (случай 1), то определитель матрицы (3) отличен от нуля, и потому $R = 3$. Если же $C_2' = 0$ (случай 2), то определитель матрицы (3) равен нулю и потому $R \leqslant 2$, т. е. $R = 2$ (ибо всегда $r \leqslant R$, а как мы уже знаем, в рассматриваемом случае $r = 2$). Таким образом, *в случае* 1 *имеют место равенства* $R = 3$ *и* $r = 2$, *а в случае* 2 — *равенства* $R = 2$ *и* $r = 2$.

Пусть теперь любые две из прямых λ, λ_1 и λ_2 параллельны (случаи 3 и 4). Тогда

$$\frac{A}{A_1} = \frac{B}{B_1}, \quad \frac{A_1}{A_2} = \frac{B_1}{B_2}, \quad \frac{A}{A_2} = \frac{B}{B_2}$$

и потому $r = 1$. Следовательно, $R \leqslant 2$ (ибо всегда $R \leqslant r + 1$). В случае 3 существуют хотя бы две параллельные, но не совпадающие, прямые. Пусть, для определенности, это — прямые λ и λ_1. Тогда

$$\frac{B}{B_1} \neq \frac{C}{C_1}$$

и потому минор

$$\begin{vmatrix} B & C \\ B_1 & C_1 \end{vmatrix}$$

матрицы (3) отличен от нуля. Поэтому в этом случае $R = 2$. В случае 4 все миноры второго порядка матрицы (3) равны нулю и потому $R = 1$. Таким образом, *в случае* 3 *имеют место равенства* $R = 2$ *и* $r = 1$, *а в случае* 4 — *равенства* $R = 1$ *и* $r = 1$.

На первый взгляд кажется, что тем самым теорема уже полностью доказана. Однако это не совсем так, поскольку, кроме доказанных утверждений (выделенных выше курсивом), нужно также доказать и обратные утверждения. Для этого мы воспользуемся общим логическим принципом обращения, утверждающим, что если у нас имеется два ряда утверждений $\mathscr{A}_1$, $\mathscr{A}_2$, ... и $\mathscr{B}_1$, $\mathscr{B}_2$, ..., причем утверждения $\mathscr{A}_1$, $\mathscr{A}_2$, ... друг друга взаимно исключают, а утверждения $\mathscr{B}_1$, $\mathscr{B}_2$, ... исчерпывают все возможности, и если доказано, что из каждого утверждения $\mathscr{A}_i$ вытекает соответствующее утверждение $\mathscr{B}_i$, то и обратно, из каждого утверждения $\mathscr{B}_i$ вытекает соответствующее утверждение $\mathscr{A}_i$. В нашем случае утверждения $\mathscr{B}_i$ — это четыре комбинации чисел R и r, а утверждения $\mathscr{A}_i$ — это четыре случая взаимного расположения прямых. Поскольку различные случаи взаимного расположения прямых друг друга взаимно исключают, а рассмотренные комбинации значений R и r исчерпывают все возможности, этот принцип применим. Тем самым наша теорема полностью доказана.

Замечание 1. Чтобы в каждом из рассмотренных случаев различить соответствующие типы (например, при $R = 3$ и

$r = 2$ узнать, имеет ли место тип 1 или тип 2) проще всего воспользоваться теоремой о взаимном расположении двух прямых. Например, при $R = 3$ и $r = 2$ имеет место тип 1, если все миноры второго порядка матрицы (2) отличны от нуля, и тип 2, если один (и только один) из этих миноров равен нулю.

Из доказанной теоремы, в частности, вытекает, что

три прямые тогда и только тогда имеют хотя бы одну общую точку, когда $r = R$.

Это — снова частный случай теоремы Кронекера — Капелли.

Упражнение. Рассмотрите взаимное расположение четырех прямых на плоскости (должно получиться семнадцать типов, группирующихся в четыре случая) и проверьте для четырех прямых справедливость теоремы Кронекера — Капелли.

4. Полуплоскости, на которые прямая разбивает плоскость

Пусть

$$F(x, y) \equiv Ax + By + C = 0$$

— произвольная прямая на плоскости. Как известно, эта прямая разбивает плоскость на две полуплоскости: точки M_1 и M_2 (не принадлежащие прямой) тогда и только тогда принадлежат одной полуплоскости, когда отрезок $\overline{M_1M_2}$ не пересекает прямую.

Предложение 1. *Точки* $M_1(x_1, y_1)$ *и* $M_2(x_2, y_2)$ *тогда и только тогда принадлежат одной полуплоскости, определяемой данной прямой, когда величины* $F(x_1, y_1)$ *и* $F(x_2, y_2)$ *имеют один и тот же знак.*

Доказательство. Рассмотрим прямую, проходящую через точки M_1 и M_2. Как было показано в п. 2, ее параметрические уравнения имеют вид

$$\begin{aligned} x &= (1 - t)\, x_1 + t x_2, \\ y &= (1 - t)\, y_1 + t y_2, \end{aligned} \tag{1}$$

причем точкам отрезка $\overline{M_1M_2}$ отвечают значения параметра, удовлетворяющие неравенствам $0 \leqslant t \leqslant 1$.

Найдем точку пересечения этой прямой с прямой $F(x, y) = 0$ (если эта точка существует). Как мы знаем (см. п. 2), для этого нужно подставить выражения (1) в уравнение прямой:

$$A((1 - t)\, x_1 + t x_2) + B((1 - t)\, y_1 + t y_2) + C = 0 \tag{2}$$

и решить получившееся уравнение относительно t. Переписав уравнение (2) в виде

$$(1 - t) F(x_1, y_1) + t F(x_2, y_2) = 0,$$

мы немедленно получим, что оно имеет корень

$$t_0 = \frac{F(x_1, y_1)}{F(x_1, y_1) - F(x_2, y_2)}. \tag{3}$$

Но, по определению, точки M_1 и M_2 тогда и только тогда лежат в разных полуплоскостях, определяемых прямой $F(x, y) = 0$, когда точка пересечения существует и принадлежит отрезку $\overline{M_1M_2}$, т. е. когда t_0 определено $(F(x_1, y_1) \neq F(x_2, y_2))$ и $0 < t_0 < 1$. Для завершения доказательства остается заметить, что в силу формулы (3) условие $0 < t_0 < 1$ в точности равносильно тому, что числа $F(x_1, y_1)$ и $F(x_2, y_2)$ имеют разные знаки.

Таким образом, при данном уравнении прямой

$$\lambda: \; Ax + By + C = 0$$

точки одной полуплоскости, на которые эта прямая делит плоскость, характеризуются неравенством $Ax + By + C > 0$ (эту полуплоскость можно назвать *положительной*), а точки другой — неравенством $Ax + By + C < 0$ (эту полуплоскость можно назвать *отрицательной*). Заметим, однако, что при другом выборе уравнения прямой положительная полуплоскость может стать отрицательной, а отрицательная — положительной (достаточно данное уравнение умножить на отрицательное число). Таким образом, правильнее говорить о положительной (или отрицательной) полуплоскости *относительно данного уравнения* прямой.

В случае, когда прямая λ не проходит через начало координат, т. е. при $C \neq 0$, мы без ограничения общности можем считать, что $C < 0$. Тогда *положительная полуплоскость будет полуплоскостью, не содержащей точки O, а отрицательная — полуплоскостью, содержащей эту точку.*

Используя терминологию, введенную в п. 4 § 4 гл. 1, мы можем, следовательно, сказать, что *задание уравнения прямой определяет некоторую сторону этой прямой.* Но поскольку у нас на плоскости задана аффинная координатная система, эта плоскость ориентирована, а задание на ориентированной плоскости некоторой стороны прямой равносильно заданию ориентации этой прямой (п. 4 § 4 гл. 1). Таким образом,

каждое уравнение прямой задает некоторую ориентацию этой прямой; при умножении уравнения на положительное число эта ориентация не меняется, а при умножении на отрицательное число — переходит в противоположную.

Ориентация прямой, соответствующая уравнению $Ax + By + C = 0$, задается ее направляющим вектором $\boldsymbol{a}$, обладающим тем свойством, что базис $\boldsymbol{a}$, $\boldsymbol{n}$ плоскости, состоящий из вектора $\boldsymbol{a}$ и некоторого вектора $\boldsymbol{n}$, направленного в положительную сторону прямой, одноименен с координатным базисом.

Чтобы найти этот вектор $\boldsymbol{a}$, заметим сначала, что

вектор $\boldsymbol{n}(A, B)$ направлен в положительную сторону прямой λ.

Действительно, пусть $M_0(x_0, y_0)$ — произвольная точка прямой и пусть $\boldsymbol{n} = \overrightarrow{M_0M_1}$. Тогда точка M_1 имеет координаты $x_0 + A$ и $y_0 + B$. Остается заметить, что

$$A(x_0 + A) + B(y_0 + B) + C = A^2 + B^2 > 0.$$

Отсюда немедлено вытекает, что

описанная выше ориентация прямой λ задается направляющим вектором $\boldsymbol{a}(B, -A)$.

Действительно,

$$\begin{vmatrix} B & -A \\ A & B \end{vmatrix} = A^2 + B^2 > 0$$

и потому базис $\boldsymbol{a}(B, -A)$, $\boldsymbol{n}(A, B)$ одноименен с координатным базисом.

Замечание 1. Обратим внимание, что в то время как все сказанное в предыдущих пунктах справедливо в аффинной геометрии над любым полем $\mathbb{K}$, утверждения, доказанные в этом пункте, справедливы (и имеют смысл) только над полем вещественных чисел $\mathbb{R}$ (или, точнее, над любым полем, в котором определены понятия положительного и отрицательного числа, обладающие обычными свойствами [1])).

5. Прямая на евклидовой плоскости

До сих пор мы занимались изучением аффинных свойств прямых (принадлежащих аффинной геометрии). Теперь мы рассмотрим их метрические свойства (принадлежащие собственно евклидовой геометрии). В соответствии с этим мы будем впредь считать наши координаты *x, y прямоугольными.*

На евклидовой плоскости каждую прямую можно задать ее произвольной точкой $M_0(x_0, y_0)$ и некоторым (отличным от нуля) вектором $\boldsymbol{n}$, перпендикулярным этой прямой. Для любой точки $M(x, y)$ рассматриваемой прямой вектор $\overrightarrow{M_0M}$ ортогонален вектору $\boldsymbol{n}$ и потому скалярное произведение этих векторов равно нулю:

$$\boldsymbol{n}\overrightarrow{M_0M} = 0.$$

Обозначая радиус-вектор точки M_0 через $\boldsymbol{r}_0$, а радиус-вектор точки M — через $\boldsymbol{r}$, мы можем это равенство записать в следующем виде:

$$\boldsymbol{n}(\boldsymbol{r} - \boldsymbol{r}_0) = 0 \tag{1}$$

или, раскрывая скобки, в виде

$$\boldsymbol{n}\boldsymbol{r} - p = 0, \tag{2}$$

[1]) В алгебре такие поля называются *упорядоченными.* Кроме поля вещественных чисел, к ним принадлежит, например, поле рациональных чисел. Подчеркнем, что *поле* $\mathbb{C}$ *не является упорядоченным полем.*

где $p = \boldsymbol{n}r_0$. Таким образом, радиус-вектор каждой точки нашей прямой удовлетворяет уравнениям (1) или (2). Ясно, что и обратно, если радиус-вектор некоторой точки M удовлетворяет уравнению (1) или (2), то эта точка принадлежит рассматриваемой прямой. В соответствии с этим уравнения (1) и (2) называются *уравнениями прямой в векторной форме.*

Обозначая координаты вектора $\boldsymbol{n}$ через A, B и полагая $C = -p$, мы можем уравнение (2) переписать в уже известной нам *координатной форме*

$$Ax + By + C = 0.$$

Аналогично, уравнение (1) дает нам общее уравнение прямой, проходящей через данную точку:

$$A(x - x_0) + B(y - y_0) = 0.$$

Замечание 1. Изложенные соображения еще раз независимо показывают (правда, только для евклидовой плоскости), что любую прямую можно задать уравнением первого порядка.

Кроме того, мы видим, что
вектор $\boldsymbol{n}(A, B)$ *перпендикулярен прямой*

$$Ax + By + C = 0.$$

Уравнение (2) особо интересно в случае, когда вектор $\boldsymbol{n}$ является ортом (единичным вектором). Обозначая координаты этого орта через $\cos\alpha$, $\sin\alpha$ (это — его направляющие косинусы; см. п. 5 § 5 гл. 1), мы получаем в этом случае для рассматриваемой прямой уравнение

$$x\cos\alpha + y\sin\alpha - p = 0.$$

При $p \geqslant 0$ (а этого всегда можно добиться, заменив, в случае необходимости, вектор $\boldsymbol{n}$ противоположным вектором $-\boldsymbol{n}$) это уравнение называется *нормальным уравнением прямой* (конечно, лучше было бы говорить — «нормированным»). Таким образом, уравнение

$$Ax + By + C = 0$$

является нормальным уравнением, если

$$A^2 + B^2 = 1 \quad \text{и} \quad C \leqslant 0.$$

Чтобы из общего уравнения прямой получить нормальное уравнение, достаточно умножить его на «нормирующий множитель»

$$\pm \frac{1}{\sqrt{A^2 + B^2}}$$

(т. е. разделить на длину вектора $\boldsymbol{n}(A, B)$, взятую с некоторым знаком). Знак нормирующего множителя следует выбирать противоположным знаку свободного члена C (если $C \neq 0$; при $C = 0$ можно взять любой знак).

Заметим, что при $C \neq 0$ *нормальное уравнение однозначно определено прямой*, а при $C = 0$ (т. е. в случае, когда прямая проходит через начало координат) прямая имеет два нормальных уравнения, получающихся друг из друга умножением на -1.

Если в нормальном уравнении свободный член отличен от нуля (и следовательно, отрицателен), то начало координат O находится в отрицательной полуплоскости, определяемой данной прямой, т. е. вектор $\boldsymbol{n}(\cos\alpha, \sin\alpha)$ направлен в сторону, противоположную стороне, в которой находится точка O (см. п. 4).

Предложение 1. *Расстояние d произвольной точки $M(x, y)$ от прямой с нормальным уравнением*

$$x\cos\alpha + y\sin\alpha - p = 0$$

равно абсолютной величине числа, получающегося при подстановке координат этой точки в уравнение прямой:

$$d = |\, x\cos\alpha + y\sin\alpha - p \,| \tag{3}$$

(*как мы уже знаем, число $x\cos\alpha + y\sin\alpha - p$ отрицательно, если точка M расположена по ту же сторону от прямой, что и начало координат, и положительно в противном случае*).

Доказательство. В векторной форме формула (3) имеет вид

$$d = |\, \boldsymbol{nr} - p \,|,$$

где $\boldsymbol{r}$ — радиус-вектор точки M. В этом виде мы и будем ее доказывать.

Пусть M_0 — основание перпендикуляра, опущенного из точки M на нашу прямую. По определению, расстояние d точки M от прямой равно длине отрезка $\overline{M_0M}$, т. е. равно длине вектора $\overrightarrow{M_0M}$:

$$d = |\, \overrightarrow{M_0M} \,|.$$

Иными словами,

$$d = |\, \boldsymbol{r} - \boldsymbol{r}_0 \,|,$$

где $\boldsymbol{r}$ — радиус-вектор точки M, а $\boldsymbol{r}_0$ — радиус-вектор точки M_0. Но вектор $\boldsymbol{n}$ перпендикулярен прямой и потому коллинеарен вектору $\overrightarrow{M_0M} = \boldsymbol{r} - \boldsymbol{r}_0$. Следовательно, поскольку этот вектор является ортом, длина вектора $\boldsymbol{r} - \boldsymbol{r}_0$ равна абсолютной величине его скалярного произведения на вектор $\boldsymbol{n}$:

$$d = |\, \boldsymbol{n}(\boldsymbol{r} - \boldsymbol{r}_0) \,| = |\, \boldsymbol{nr} - \boldsymbol{nr}_0 \,|.$$

Для завершения доказательства остается заметить, что $\boldsymbol{nr}_0 = p$, ибо точка M_0 принадлежит прямой $\boldsymbol{nr} - p = 0$.

Следствие. *Число p равно расстоянию точки O от прямой, т. е. длине перпендикуляра, опущенного из точки O на прямую.*

Один из двух (элементарно-геометрических) углов, образованных пересекающимися прямыми λ_1 и λ_2, равен углу θ между их направляющими векторами $\boldsymbol{a}_1(B_1, -A_1)$ и $\boldsymbol{a}_2(B_2, -A_2)$ и потому определяется формулой

$$\cos\theta = \frac{A_1A_2 + B_1B_2}{\sqrt{A_1^2 + A_2^2}\sqrt{B_1^2 + B_2^2}}. \qquad (4)$$

При $A_1A_2 + B_1B_2 > 0$ этот угол — острый, а при $A_1A_2 + B_1B_2 < 0$ — тупой. При $A_1A_2 + B_1B_2 = 0$ этот угол равен $\pi/2$. Таким образом,

$$\lambda_1 \perp \lambda_2 \Leftrightarrow A_1A_2 + B_1B_2 = 0.$$

Чтобы получить острый (точнее, не тупой) угол между прямыми, следует правую часть формулы (4) взять по абсолютной величине.

Если прямые λ_1 и λ_2 заданы нормальными уравнениями

$$x\cos\alpha_1 + y\sin\alpha_1 - p_1 = 0$$

и

$$x\cos\alpha_2 + y\sin\alpha_2 - p_2 = 0,$$

то угол θ между ними определяется формулой

$$\cos\theta = \cos\alpha_1\cos\alpha_2 + \sin\alpha_1\sin\alpha_2 = \cos(\alpha_2 - \alpha_1).$$

Следовательно,

$$\theta = |\alpha_2 - \alpha_1|.$$

В частности,

$$\lambda_1 \perp \lambda_2 \Leftrightarrow |\alpha_2 - \alpha_1| = \frac{\pi}{2}.$$

Замечание 2. Согласно п. 4 прямые λ_1 и λ_2 естественным образом ориентированы. Ясно, что определенный формулой (4) угол θ является не чем иным, как углом между так ориентированными прямыми.

Это замечание можно развить, вспомнив, что наша плоскость также ориентирована (поскольку на ней фиксирована некоторая координатная система). Поэтому (см. дополнение к § 4 гл. I) определен угол φ от прямой λ_1 к прямой λ_2. По определению, $-\pi < \varphi \leqslant \pi$, причем $\varphi = \theta$, если $0 \leqslant \varphi \leqslant \pi$, и $\varphi = -\theta$, если $-\pi < \varphi \leqslant 0$. Угол φ однозначно определен, если известны его тригонометрические функции $\cos\varphi = \cos\theta$ и $\sin\varphi$.

Чтобы найти угол φ, мы рассмотрим наряду с направляющими векторами $\boldsymbol{a}_1(B_1, -A_1)$ и $\boldsymbol{a}_2(B_2, -A_2)$ ортогональные им векторы $\boldsymbol{n}_1(A_1, B_1)$ и $\boldsymbol{n}_2(A_2, B_2)$. Поскольку вектор $\boldsymbol{n}_2$ направлен в положительную сторону прямой λ_1, угол от вектора $\boldsymbol{a}_2$ к век-

тору n_2 равен $\pi/2$. Следовательно, угол φ' от вектора a_1 к вектору n_2 равен $\varphi+\frac{\pi}{2}$ и потому $\cos\varphi'=-\sin\varphi$. Но угол φ' и угол между векторами a_1 и n_2 имеют один и тот же косинус, так что косинус этого угла мы можем вычислить по известной формуле:

$$\cos\varphi'=\frac{a_1n_2}{|a_1|\cdot|n_2|}=\frac{-A_1B_2+A_2B_1}{\sqrt{A_1^2+B_1^2}\sqrt{A_2^2+B_2^2}}.$$

Таким образом, для угла φ имеют место формулы

$$\begin{aligned}\cos\varphi&=\frac{A_1A_2+B_1B_2}{\sqrt{A_1^2+A_2^2}\sqrt{B_1^2+B_2^2}},\\ \sin\varphi&=\frac{A_1B_2-A_2B_1}{\sqrt{A_1^2+A_2^2}\sqrt{B_1^2+B_2^2}},\end{aligned}\tag{5}$$

однозначно этот угол определяющие.

В частности, мы видим, что

$$\operatorname{tg}\varphi=\frac{A_1B_2-A_2B_1}{A_1A_2+B_1B_2}.$$

Вернемся теперь к элементарно-геометрическим углам между неориентированными прямыми. Легко видеть, что один из этих смежных углов (обозначим его через θ') имеет тот же тангенс, что и угол φ:

$$\operatorname{tg}\theta'=\operatorname{tg}\varphi,$$

так что

$$\operatorname{tg}\theta'=\frac{A_1B_2-A_2B_1}{A_1A_2+B_1B_2}$$

(угол θ' совпадает с углом θ, если $0\leqslant\varphi\leqslant\pi$, и равен $\pi-\theta$, если $-\pi<\varphi\leqslant 0$; в частности, угол θ' — острый, если числа $A_1B_2-A_2B_1$ и $A_1A_2+B_1B_2$ имеют одинаковые знаки, и тупой, если знаки этих чисел различны).

Через угловые коэффициенты $k_1=-\frac{A_1}{B_1}$ и $k_2=-\frac{A_2}{B_2}$ данных прямых формула для $\operatorname{tg}\theta'$ записывается следующим образом:

$$\operatorname{tg}\theta'=\frac{k_2-k_1}{1+k_1k_2}.\tag{6}$$

В частности, мы видим, что

прямые с угловыми коэффициентами k_1 и k_2 тогда и только тогда перпендикулярны, когда

$$k_1k_2=-1.$$

Применяя формулу (6) к оси абсцисс (имеющей угловой коэффициент $k_1 = 0$) и к произвольной прямой (с угловым коэффициентом $k_2 = k$), мы немедленно получаем, что

угловой коэффициент k произвольной прямой равен тангенсу угла θ', образованного этой прямой с осью абсцисс:

$$k = \operatorname{tg} \theta'.$$

Этот угол θ' часто называют поэтому *углом наклона* рассматриваемой прямой.

Предупреждение. Подчеркнем, что угол между прямыми мы рассматриваем по крайней мере в трех различных смыслах (θ, φ и θ'). Углы θ и θ' оба — элементарно-геометрические, но, вообще говоря, различные. Каждый раз, когда речь идет об угле между прямыми, необходимо четко отдавать себе отчет, в каком именно смысле этот угол понимается.

§ 2. ПЛОСКОСТЬ В ПРОСТРАНСТВЕ

Плоскость в пространстве рассматривается почти дословно так же, как прямая на плоскости.

1. Плоскость как поверхность первого порядка

Пусть в пространстве задана аффинная координатная система $Oxyz$. Для любой плоскости Π существует (отнюдь не единственная) аффинная координатная система $O'x'y'z'$, в которой эта плоскость определяется уравнением $x' = 0$. Но, как мы знаем (п. 2 § 1 гл. 2), координата x' выражается через координаты x, y, z по формуле

$$x' = Ax + By + Cz + D,$$

где A, B, C и D — некоторые вещественные числа, причем хотя бы одно из чисел A, B или C отлично от нуля (в обозначениях п. 2 § 1 гл. 2 $A = c_1^{1'}$, $B = c_2^{1'}$, $C = c_3^{1'}$ и $D = a^{1'}$). Тем самым доказано, что

любая плоскость определяется уравнением вида

$$Ax + By + Cz + D = 0, \tag{1}$$

где хотя бы одно из чисел A, B или C отлично от нуля.

Уравнение плоскости, имеющее вид (1), называется ее *общим уравнением*.

Доказанное утверждение означает, что

любая плоскость является поверхностью первого порядка.

Покажем, что верно и обратное, т. е. что

любая поверхность первого порядка является плоскостью.

Действительно, пусть

$$Ax + By + Cz + D = 0$$

— уравнение произвольной поверхности первого порядка. Подберем числа A_1, B_1, C_1 и A_2, B_2, C_2 так, чтобы определитель

$$\begin{vmatrix} A & B & C \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{vmatrix}$$

был отличен от нуля (например, при $A^2 + B^2 \neq 0$ можно положить $A_1 = -B$, $B_1 = A$, $C_1 = 0$ и $A_2 = B_2 = 0$, $C_2 = 1$, а при $A = B = 0$ можно положить $A_1 = C_1 = 0$, $B_1 = 1$ и $A_2 = 1$, $B_2 = C_2 = 0$). Тогда формулы

$$\begin{aligned} x' &= Ax + By + Cz + D, \\ y' &= A_1x + B_1y + C_1z + D_1, \\ z' &= A_2x + B_2y + C_2z + D_2, \end{aligned}$$

где D_1 и D_2 — произвольные числа (например, можно считать, что $D_1 = D_2 = 0$), описывают переход от координат x, y, z к некоторым новым координатам x', y', z'. В координатах x', y', z' данная поверхность выражается, очевидно, уравнением $x' = 0$ и, следовательно, является плоскостью. Поскольку тот факт, что данная поверхность является плоскостью, не зависит от системы координат, наше утверждение полностью доказано.

Собирая вместе доказанные утверждения, мы получаем следующую теорему:

Теорема 1. *Поверхность в пространстве тогда и только тогда представляет собой плоскость, когда она является алгебраической поверхностью первого порядка.*

Соглашение об обозначениях. В дальнейшем символом Π мы будем обозначать плоскость с уравнением (1). В случае, когда нам придется рассматривать несколько прямых, мы будем снабжать символ Π индексами и будем предполагать, что соответствующими индексами снабжены и коэффициенты уравнения (1) (ср. соглашение для прямых в п. 1 § 1).

Равенство нулю одного (или нескольких) коэффициентов A, B, C, D указывает на специальное положение плоскости относительно координатных осей.

Задание. Докажите, что

$A = 0 \Leftrightarrow \Pi \parallel Ox$,

$B = 0 \Leftrightarrow \Pi \parallel Oy$,

$C = 0 \Leftrightarrow \Pi \parallel Oz$,

$D = 0 \Leftrightarrow \Pi$ проходит через точку $O(0, 0, 0)$.

2. Параметрические уравнения плоскости

Определение 1. *Направляющим бивектором* плоскости Π называется произвольный отличный от нуля бивектор, параллельный этой плоскости, т. е. принадлежащий соответствующему линеалу $\mathrm{Biv}(2)$. Этот бивектор определен с точностью до коллинеарности (и представляет собой не что иное, как базис линеала $\mathrm{Biv}(2)$).

Предложение 1. *Вектор* $\boldsymbol{a}(l, m, n)$ *тогда и только тогда параллелен плоскости* Π (*принадлежит соответствующему линеалу* $\mathrm{Vect}_\Pi(2)$), *когда*

$$Al + Bm + Cn = 0. \tag{1}$$

Доказательство этого предложения аналогично доказательству предложения 1 из п. 1 § 1, и мы его опустим.

Задание. Докажите предложение 1.

Следствие. *Бивектор* $\mathfrak{a}(A, B, C)$ *является направляющим бивектором плоскости* Π.

Здесь запись $\mathfrak{a}(A, B, C)$ означает, что бивектор $\mathfrak{a}$ имеет координаты A, B, C в бивекторном базисе $\boldsymbol{e}_2 \wedge \boldsymbol{e}_3$, $\boldsymbol{e}_3 \wedge \boldsymbol{e}_1$, $\boldsymbol{e}_1 \wedge \boldsymbol{e}_2$, соответствующем данному базису $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ пространства.

Доказательство. В отличие от случая прямой (когда соответствующий вывод тривиален), для доказательства этого следствия требуется некоторая алгебраическая техника. Проще всего (хотя и несколько искусственно) это доказательство можно провести следующим образом.

Ясно, что уравнение (1) будет удовлетворено при

$$l = -B, \quad m = A, \quad n = 0.$$

Обозначим вектор с этими координатами через $\boldsymbol{a}_1$. С другой стороны, уравнение (1) удовлетворяется также и при

$$l = 0, \quad m = -C, \quad n = B.$$

Обозначим вектор с этими координатами через $\boldsymbol{a}_2$. Если $B \neq 0$, то векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ не коллинеарны, и потому их внешнее произведение $\boldsymbol{a}_1 \wedge \boldsymbol{a}_2$ отлично от нуля. Поскольку векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ параллельны согласно предложению 1 рассматриваемой плоскости, это внешнее произведение будет, поэтому, направляющим бивектором этой плоскости. Но тогда направляющим бивектором будет и бивектор

$$\mathfrak{a} = \frac{1}{B}(\boldsymbol{a}_1 \wedge \boldsymbol{a}_2).$$

Для завершения доказательства остается заметить, что согласно общей формуле для внешнего произведения двух векторов

(п. 3 § 6 гл. 1)

$$\mathfrak{a} = \frac{1}{B}\begin{vmatrix} e_2 \wedge e_3 & e_3 \wedge e_1 & e_1 \wedge e_2 \\ -B & A & 0 \\ 0 & -C & B \end{vmatrix} =$$

$$= A(e_2 \wedge e_3) + B(e_3 \wedge e_1) + C(e_1 \wedge e_2).$$

При $B = 0$, $A \neq 0$ доказательство проводится аналогично, только за вектор a_2 следует принять вектор с координатами

$$l = -C, \quad m = 0, \quad n = A.$$

При $A = B = 0$ (тогда обязательно $C \neq 0$) следует за вектор a_1 принять вектор с координатами

$$l = -C, \quad m = 0, \quad n = 0 \quad (= A).$$

Замечание 1. Это следствие можно доказать и по-иному, рассмотрев координаты

$$x' = Ax + By + Cz + D,$$
$$y' = A_1x + B_1y + C_1z + D_1,$$
$$z' = A_2x + B_2y + C_2z + D_2,$$

в которых данная плоскость П является плоскостью $x' = 0$ (см. выше). Согласно формулам преобразования координат бивекторов (п. 3 § 6 гл. 1) бивектор $\mathfrak{a}$ в этой координатной системе имеет координаты

$$\begin{vmatrix} A & A_1 & A_2 \\ B & B_1 & B_2 \\ C & C_1 & C_2 \end{vmatrix} \neq 0, \quad \begin{vmatrix} A & A & A_2 \\ B & B & B_2 \\ C & C & C_2 \end{vmatrix} = 0, \quad \begin{vmatrix} A & A_1 & A \\ B & B_1 & B \\ C & C_1 & C \end{vmatrix} = 0$$

и потому коллинеарен бивектору $e_2' \wedge e_3'$. Для завершения доказательства остается заметить, что последний бивектор, очевидно, параллелен рассматриваемой плоскости.

Пусть плоскость П с направляющим бивектором $\mathfrak{a}$ проходит через точку $M_0(x_0, y_0, z_0)$. Представим бивектор $\mathfrak{a}$ в виде внешнего произведения $\boldsymbol{a} \wedge \boldsymbol{b}$ каких-то двух векторов $\boldsymbol{a}$ и $\boldsymbol{b}$. Поскольку векторы $\boldsymbol{a}$, $\boldsymbol{b}$ составляют, очевидно, базис линеала $\mathrm{Vect}_{\Pi}(2)$, точка $M(x, y, z)$ тогда и только тогда принадлежит плоскости, когда вектор $\overrightarrow{M_0M}$ линейно выражается через векторы $\boldsymbol{a}$ и $\boldsymbol{b}$, т. е. когда существуют такие числа u и v, что

$$\overrightarrow{M_0M} = u\boldsymbol{a} + v\boldsymbol{b}.$$

Пусть $\boldsymbol{r}_0$ — радиус-вектор точки M_0, а $\boldsymbol{r}$ — радиус-вектор точки M. Тогда предыдущее равенство может быть записано в виде

$$\boldsymbol{r} = \boldsymbol{r}_0 + u\boldsymbol{a} + v\boldsymbol{b}.$$

Это — *параметрическое уравнение плоскости в векторной форме.* Записав его в координатах, мы получим *параметрические уравнения плоскости в координатной форме*:

$$x = x_0 + ua_1 + vb_1,$$
$$y = y_0 + ua_2 + vb_2,$$
$$z = z_0 + ua_3 + vb_3,$$

где a_1, a_2, a_3 — координаты вектора $\boldsymbol{a}$, а b_1, b_2, b_3 — координаты вектора $\boldsymbol{b}$.

Точка M_0 и векторы $\boldsymbol{a}$, $\boldsymbol{b}$ определяют на плоскости аффинную координатную систему с репером $M_0\boldsymbol{ab}$. Очевидно, что *параметры u и v являются аффинными координатами на плоскости относительно репера $M_0\boldsymbol{ab}$.*

Поскольку векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ не коллинеарны, тот факт, что вектор $\overrightarrow{M_0M_1}$ линейно через них выражается, равносилен тому, что векторы $\overrightarrow{M_0M}$, $\boldsymbol{a}$ и $\boldsymbol{b}$ линейно зависимы (компланарны), т. е. тому, что составленный из их координат определитель равен нулю:

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = 0. \qquad (2)$$

Это равенство выражает необходимое и достаточное условие того, что данная точка принадлежит рассматриваемой плоскости, т. е. является ее уравнением. Это уравнение плоскости иногда называется *уравнением в детерминантной форме.*

Обозначив алгебраические дополнения элементов первой строки определителя (2) через A, B и C и разложив определитель по этой строке, мы получим уравнение

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0.$$

Это — *общее уравнение плоскости, проходящей через данную точку.*

Плоскость полностью определена, если на ней заданы три точки $M_0(x_0, y_0, z_0)$, $M_1(x_1, y_1, z_1)$ и $M_2(x_2, y_2, z_2)$, не лежащие на одной прямой. Рассмотрим векторы

$$\boldsymbol{a} = \overrightarrow{M_0M_1}\,(x_1 - x_0,\ y_1 - y_0,\ z_1 - z_0),$$
$$\boldsymbol{b} = \overrightarrow{M_0M_2}\,(x_2 - x_0,\ y_2 - y_0,\ z_2 - z_0).$$

Эти векторы не коллинеарны, и потому их внешнее произведение $\boldsymbol{a} \wedge \boldsymbol{b}$ отлично от нуля. Оно, очевидно, является направляющим бивектором данной плоскости.

В соответствии со сказанным выше отсюда вытекает, что *плоскость, проходящая через точки* $M_0(x_0, y_0, z_0)$, $M_1(x_1, y_1, z_1)$ *и* $M_2(x_2, y_2, z_2)$, *имеет уравнение*

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \end{vmatrix} = 0. \tag{3}$$

Замечание 2. Уравнение (3) иногда удобно записывать в следующем более симметричном виде:

$$\begin{vmatrix} x & y & z & 1 \\ x_0 & y_0 & z_0 & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \end{vmatrix} = 0. \tag{4}$$

Для доказательства равносильности уравнений (3) и (4) достаточно вторую строку определителя (4) вычесть последовательно из всех остальных строк и после этого разложить определитель по последнему столбцу.

Параметрические уравнения рассматриваемой плоскости имеют, очевидно, вид

$$\begin{aligned} x &= x_0 + u(x_1 - x_0) + v(x_2 - x_0), \\ y &= y_0 + u(y_1 - y_0) + v(y_2 - y_0), \\ z &= z_0 + u(z_1 - z_0) + v(z_2 - z_0), \end{aligned}$$

т. е. вид

$$\begin{aligned} x &= (1 - u - v)\, x_0 + ux_1 + vx_2, \\ y &= (1 - u - v)\, y_0 + uy_1 + vy_2, \\ z &= (1 - u - v)\, z_0 + uz_1 + vz_2. \end{aligned}$$

Следовательно (см. п. 6 § 1 гл. 2),

числа $1 - u - v$, u, v *являются однородными барицентрическими координатами на плоскости.*

3. Взаимное расположение плоскостей в пространстве

Ясно, что две плоскости Π и Π_1 тогда и только тогда параллельны [1]), когда их направляющие бивекторы коллинеарны. Следовательно, справедливо следующее

Предложение 1.

$$\Pi \parallel \Pi_1 \Leftrightarrow \frac{A}{A_1} = \frac{B}{B_1} = \frac{C}{C_1} \Leftrightarrow r = 1.$$

[1]) Как и для случая прямых, совпадающие плоскости мы считаем параллельными.

Здесь r — ранг матрицы

$$\begin{pmatrix} A & B & C \\ A_1 & B_1 & C_1 \end{pmatrix}.$$

Замечание 1. Вместо направляющих бивекторов $\mathfrak{a}$ в этом рассуждении можно было бы рассматривать ассоциированные с ними векторы $\mathfrak{a}^{\perp}$, перпендикулярные данным плоскостям (см. п. 3 § 6 гл. 1). Однако это означает использование метрических понятий. Если оставаться в рамках аффинной геометрии, то использование бивекторов неизбежно (быть может, в завуалированной форме).

Параллельные плоскости совпадают, если они имеют хотя бы одну общую точку $M_0(x_0, y_0, z_0)$, т. е. если существуют такие числа x_0, y_0, z_0, что

$$\begin{aligned} Ax_0 + By_0 + Cz_0 + D = 0, \\ A_1x_0 + B_1y_0 + C_1z_0 + D_1 = 0. \end{aligned} \tag{1}$$

Пусть $A = \rho A_1$, $B = \rho B_1$, $C = \rho C_1$. Вычитая из первого из равенств (1) второе, умноженное на ρ, мы немедленно получим, что $D - \rho D_1 = 0$, т. е. что $D = \rho D_1$. Ясно, что обратное также верно: если $D = \rho D_1$, то рассматриваемые плоскости совпадают. Тем самым доказано следующее

Предложение 2.

$$\Pi = \Pi_1 \Leftrightarrow \frac{A}{A_1} = \frac{B}{B_1} = \frac{C}{C_1} = \frac{D}{D_1} \Leftrightarrow R = 1.$$

Здесь R — ранг матрицы

$$\begin{pmatrix} A & B & C & D \\ A_1 & B_1 & C_1 & D_1 \end{pmatrix}.$$

Это — теорема единственности для поверхностей первого порядка.

Оба доказанных утверждения можно объединить в следующую теорему о взаимном расположении двух плоскостей:

Теорема 1.

а) *плоскости* Π *и* Π_1 *пересекаются* (*по прямой*) $\Leftrightarrow r = 2$;

б) $\Pi \parallel \Pi_1$, *но* $\Pi \neq \Pi_1 \Leftrightarrow r = 1$, $R = 2$;

в) $\Pi = \Pi_1 \Leftrightarrow R = 1$.

В частности,

плоскости Π *и* Π_1 *тогда и только тогда имеют хотя бы одну общую точку, когда* $r = R$.

Это — снова специальный случай теоремы Кронекера — Капелли.

Для трех плоскостей Π, Π_1, Π_2 в пространстве возможны восемь существенно различных типов их взаимного расположения:

Тип 1. Плоскости попарно не параллельны и имеют одну и только одну общую точку; это — *общий тип.*

Тип 2. Плоскости попарно не параллельны, но прямые, по которым они попарно пересекаются, параллельны между собой и не совпадают (в этом случае плоскости не имеют ни одной общей точки и вырезают из пространства бесконечную треугольную призму).

Тип 3. Две из плоскостей параллельны, но не совпадают, а третья их пересекает.

Тип 4. Плоскости попарно не параллельны и имеют общую прямую.

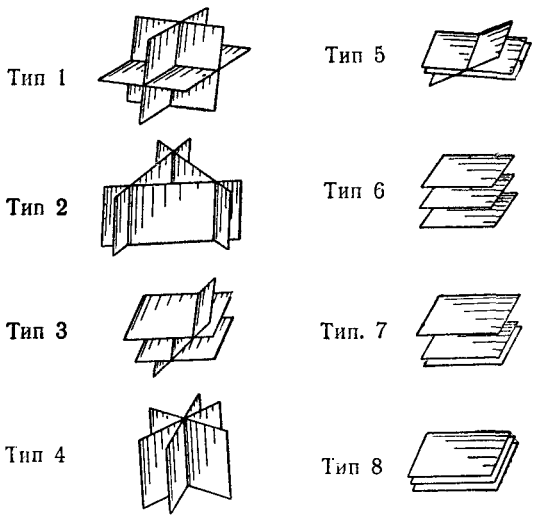
Тип 5. Две плоскости совпадают, а третья их пересекает.

Тип 6. Все три плоскости параллельны, но никакие две из них не совпадают.

Тип 7. Две плоскости совпадают, а третья им параллельна, но от них отлична.

Тип 8. Все три плоскости совпадают.

Схематически:



Для исследования этих типов взаимного расположения удобно рассмотреть следующие пять случаев:

Случай 1 (тип 1): имеется хотя бы одна пара непараллельных плоскостей и прямая их пересечения пересекает третью плоскость.

Случай 2 (типы 2 и 3): имеется хотя бы одна пара непараллельных плоскостей и прямая их пересечения параллельна третьей плоскости (но в ней не содержится).

Случай 3 (типы 4 и 5): имеется хотя бы одна пара непараллельных плоскостей и прямая их пересечения содержится в третьей плоскости.

Случай 4 (типы 6 и 7): все плоскости параллельны и среди них по крайней мере две не совпадают.

Случай 5 (тип 8): все плоскости совпадают.

Пусть r — ранг матрицы

$$\begin{pmatrix} A & B & C \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{pmatrix},$$

а R — ранг матрицы

$$\begin{pmatrix} A & B & C & D \\ A_1 & B_1 & C_1 & D_1 \\ A_2 & B_2 & C_2 & D_2 \end{pmatrix}.$$

Оказывается, что справедлива следующая теорема о взаимном расположении трех плоскостей в пространстве:

Теорема 2.

$$\begin{aligned} &\textit{Случай } 1 \Leftrightarrow r=3,\ R=3;\\ &\textit{случай } 2 \Leftrightarrow r=2,\ R=3;\\ &\textit{случай } 3 \Leftrightarrow r=2,\ R=2;\\ &\textit{случай } 4 \Leftrightarrow r=1,\ R=2;\\ &\textit{случай } 5 \Leftrightarrow r=1,\ R=1. \end{aligned}$$

Доказательство этой теоремы вполне аналогично доказательству теоремы 2 из п. 3 § 1, и мы его опустим.

Упражнение. Докажите теорему 2.

Из теоремы 2 снова вытекает соответствующий частный случай теоремы Кронекера — Капелли:

три плоскости тогда и только тогда имеют хотя бы одну общую точку, когда $r = R$.

4. Полупространства, на которые плоскость разбивает пространство

Подобно тому как прямая разбивает плоскость на две полуплоскости, плоскость разбивает пространство на два полупространства: точки M_1 и M_2 (не принадлежащие плоскости) тогда и только тогда принадлежат одному полупространству, когда отрезок $\overline{M_1M_2}$ не пересекает плоскость.

Пусть рассматриваемая плоскость задана уравнением

$$F(x, y, z) \equiv Ax + By + Cz + D = 0.$$

Предложение 1. *Точки* $M_1(x_1, y_1, z_1)$ *и* $M_2(x_2, y_2, z_2)$ *тогда и только тогда принадлежат одному полупространству, когда числа* $F(x_1, y_1, z_1)$ *и* $F(x_2, y_2, z_2)$ *имеют один и тот же знак.*

Доказательство (вполне аналогичное доказательству предложения 1 из п. 4 § 1). Рассмотрим прямую, проходящую через точки M_1 и M_2. Как было показано в п. 2 § 1, ее параметрические уравнения имеют вид

$$\begin{aligned} x &= (1-t)\,x_1 + tx_2, \\ y &= (1-t)\,y_1 + ty_2, \\ z &= (1-t)\,z_1 + tz_2, \end{aligned} \tag{1}$$

причем точкам отрезка $\overline{M_1M_2}$ отвечают значения параметра, удовлетворяющие неравенствам $0 \leqslant t \leqslant 1$.

Чтобы найти точку пересечения этой прямой с нашей плоскостью (когда эта точка существует), следует подставить выражения (1) в уравнение плоскости и решить получающееся уравнение относительно t. Сделав это, мы немедленно получим, что точке пересечения отвечает значение параметра

$$t_0 = \frac{F(x_1, y_1, z_1)}{F(x_1, y_1, z_1) - F(x_2, y_2, z_2)}. \tag{2}$$

Но, по определению, точки M_1 и M_2 тогда и только тогда лежат в разных полупространствах, когда точка пересечения существует и принадлежит отрезку $\overline{M_1M_2}$, т. е. когда t_0 определено и $0 < t_0 < 1$. Для завершения доказательства остается заметить, что в силу формулы (2) условие $0 < t_0 < 1$ равносильно тому, что числа $F(x_1, y_1, z_1)$ и $F(x_2, y_2, z_2)$ имеют разные знаки.

Таким образом, при данном уравнении плоскости П:

$$Ax + By + Cz + D = 0$$

точки одного полупространства, на которые эта плоскость делит пространство, характеризуются неравенством $Ax + By + Cz + D > 0$, а точки другого полупространства — неравенством $Ax + By + Cz + D < 0$. В соответствии с этим первое полупространство называется *положительным* (относительно данного уравнения плоскости), а второе — *отрицательным*.

Используя терминологию, введенную в п. 4 § 4 гл. 1, мы можем сказать, что *задание уравнения плоскости определяет некоторую сторону этой плоскости*. Но поскольку у нас в пространстве задана аффинная координатная система, пространство ориентировано, а задание в ориентированном пространстве некоторой стороны плоскости равносильно заданию ориентации этой плоскости (п. 4 § 4 гл. 1). Таким образом,

каждое уравнение плоскости задает некоторую ориентацию этой плоскости; при умножении уравнения на положительное число эта ориентация не меняется, а при умножении на отрицательное число — переходит в противоположную.

Ориентация плоскости, соответствующая уравнению $Ax + By + Cz + D = 0$, задается ее направляющим бивектором α,

обладающим тем свойством, что внешнее произведение $\mathfrak{a} \wedge \boldsymbol{n}$ бивектора $\mathfrak{a}$ и некоторого вектора $\boldsymbol{n}$, направленного в положительную сторону плоскости, определяет ту же ориентацию пространства, что и внешнее произведение $\boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \wedge \boldsymbol{e}_3$ векторов координатного базиса.

Чтобы найти этот бивектор, мы заметим сначала, что

вектор $\boldsymbol{n}(A, B, C)$ *направлен в положительную сторону плоскости* П.

Действительно, пусть $M_0(x_0, y_0, z_0)$ — произвольная точка плоскости и пусть $\boldsymbol{n} = \overrightarrow{M_0M_1}$. Тогда точка M_1 имеет координаты $x_0 + A$, $y_0 + B$ и $z_0 + C$. Остается заметить, что

$$A(x_0 + A) + B(y_0 + B) + C(z_0 + C) + D = A^2 + B^2 + C^2 > 0.$$

Отсюда немедленно вытекает, что

описанная выше ориентация плоскости П *задается направляющим бивектором* $\mathfrak{a}(A, B, C)$.

Действительно,

$$\mathfrak{a} \wedge \boldsymbol{n} = (A(\boldsymbol{e}_2 \wedge \boldsymbol{e}_3) + B(\boldsymbol{e}_3 \wedge \boldsymbol{e}_1) + C(\boldsymbol{e}_1 \wedge \boldsymbol{e}_2)) \wedge (A\boldsymbol{e}_1 + B\boldsymbol{e}_2 + C\boldsymbol{e}_3) = \\ = (A^2 + B^2 + C^2)(\boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \wedge \boldsymbol{e}_3).$$

Поскольку $A^2 + B^2 + C^2 > 0$, ориентации тривекторов $\mathfrak{a} \wedge \boldsymbol{n}$ и $\boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \wedge \boldsymbol{e}_3$ совпадают.

5. Плоскость в евклидовом пространстве

Мы переходим теперь к изучению евклидовых (метрических) свойств плоскостей в пространстве. В соответствии с этим мы будем впредь считать наши координаты x, y, z *прямоугольными*.

В евклидовом пространстве каждую плоскость можно задать ее произвольной точкой $M_0(x_0, y_0, z_0)$ и некоторым (отличным от нуля) вектором $\boldsymbol{n}$, перпендикулярным этой плоскости. Для любой точки $M(x, y, z)$ рассматриваемой плоскости вектор $\overrightarrow{M_0M}$ ортогонален вектору $\boldsymbol{n}$ и потому скалярное произведение этих векторов равно нулю:

$$\boldsymbol{n}\overrightarrow{M_0M} = 0.$$

Обозначая радиус-вектор точки M_0 через $\boldsymbol{r}_0$, а радиус-вектор точки M — через $\boldsymbol{r}$, мы можем это равенство записать в следующем виде:

$$\boldsymbol{n}(\boldsymbol{r} - \boldsymbol{r}_0) = 0 \qquad (1)$$

или, раскрывая скобки, — в виде

$$\boldsymbol{n}\boldsymbol{r} - p = 0, \qquad (2)$$

где $p = \boldsymbol{n}\boldsymbol{r}_0$. Таким образом, радиус-вектор каждой точки нашей плоскости удовлетворяет уравнениям (1) или (2). Ясно, что и

обратно, если радиус-вектор некоторой точки M удовлетворяет уравнениям (1) или (2), то эта точка принадлежит рассматриваемой плоскости. В соответствии с этим уравнения (1) и (2) называются *уравнениями плоскости в векторной форме.*

Обозначая координаты вектора $\boldsymbol{n}$ через A, B, C и полагая $D = -p$, мы можем уравнение (2) переписать в уже известной нам *координатной форме*

$$Ax + By + Cz + D = 0.$$

Аналогично, уравнение (1) дает нам *общее уравнение плоскости, проходящей через данную точку*:

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0.$$

Замечание 1. Изложенные соображения еще раз независимо показывают (правда, только для евклидова пространства), что любую плоскость можно задать уравнением первого порядка.

Кроме того, мы видим, что
вектор $\boldsymbol{n}(A, B, C)$ *перпендикулярен плоскости*

$$Ax + By + Cz + D = 0.$$

Замечание 2. Таким образом, вектор $\boldsymbol{a}(l, m, n)$ тогда и только тогда параллелен плоскости, когда он ортогонален вектору $\boldsymbol{n}(A, B, C)$, т. е. когда

$$Al + Bm + Cn = 0.$$

Это — уже известная нам формула (1) из п. 1. Подчеркнем, что согласно п. 1 она имеет место в любых аффинных координатах, когда вектор $\boldsymbol{n}(A, B, C)$ уже не перпендикулярен плоскости (а скалярное произведение двух векторов не является суммой произведений одноименных координат).

Рассмотрим случай, когда вектор $\boldsymbol{n}$ является ортом (единичным вектором). Обозначая координаты этого орта через $\cos\alpha$, $\cos\beta$, $\cos\gamma$ (это — его направляющие косинусы; см. п. 5 § 5 гл. 1) мы получаем в этом случае для рассматриваемой плоскости уравнение

$$x\cos\alpha + y\cos\beta + z\cos\gamma - p = 0.$$

При $p \geqslant 0$ (чего всегда можно добиться заменой, в случае необходимости, вектора $\boldsymbol{n}$ противоположным вектором $-\boldsymbol{n}$) это уравнение называется *нормальным уравнением* плоскости. Таким образом, уравнение

$$Ax + By + Cz + D = 0$$

является нормальным уравнением, если

$$A^2 + B^2 + C^2 = 1 \quad \text{и} \quad D \leqslant 0.$$

Чтобы из общего уравнения плоскости получить ее нормальное уравнение, достаточно умножить его на «нормирующий множитель»

$$\pm \frac{1}{\sqrt{A^2+B^2+C^2}}.$$

Знак нормирующего множителя следует выбирать противоположным знаку свободного члена D (если $D \neq 0$; при $D = 0$ можно взять любой знак).

Заметим, что *при* $D \neq 0$ *нормальное уравнение однозначно определено плоскостью*, а при $D = 0$ (т. е. в случае, когда плоскость проходит через начало координат) имеется два нормальных уравнения, получающихся друг из друга умножением на -1.

Предложение 1. *Расстояние* d *произвольной точки* $M(x, y, z)$ *от плоскости с нормальным уравнением*

$$x\cos\alpha + y\cos\beta + z\cos\gamma - p = 0$$

равно абсолютной величине числа, получающегося при подстановке координат этой точки в уравнение плоскости:

$$d = |\, x\cos\alpha + y\cos\beta + z\cos\gamma - p\,| \qquad (3)$$

(*как мы уже знаем, число* $x\cos\alpha + y\cos\beta + z\cos\gamma - p$ *отрицательно, если точка* M *расположена по ту же сторону плоскости, что и начало координат, и положительно в противном случае*).

Доказательство этого предложения по существу дословно совпадает с доказательством соответствующего предложения 1 из п. 5 § 1, и мы его опустим.

Задание. Докажите предложение 1.

В частности, мы видим, что

коэффициент p *равен расстоянию от начала координат до плоскости, т. е. равен длине перпендикуляра, опущенного из точки* O *на плоскость.*

Две (пересекающиеся) плоскости Π_1 и Π_2 образуют два смежных угла, сумма которых равна π. Если плоскости параллельны или совпадают, то угол между ними, по определению, считается равным 0 (или π).

Один из углов между плоскостями Π_1 и Π_2 (обозначим его символом θ) совпадает с углом между векторами $\boldsymbol{n}_1(A_1, B_1, C_1)$ и $\boldsymbol{n}_2(A_2, B_2, C_2)$, перпендикулярными этим плоскостям. Поэтому

$$\cos\theta = \frac{A_1A_2 + B_1B_2 + C_1C_2}{\sqrt{A_1^2+B_1^2+C_1^2}\sqrt{A_2^2+B_2^2+C_2^2}}. \qquad (4)$$

При $A_1A_2 + B_1B_2 + C_1C_2 > 0$ угол θ — острый, а при $A_1A_2 + B_1B_2 + C_1C_2 < 0$ — тупой. При $A_1A_2 + B_1B_2 + C_1C_2 = 0$ этот угол равен $\pi/2$. Таким образом,

$$\Pi_1 \perp \Pi_2 \Leftrightarrow A_1A_2 + B_1B_2 + C_1C_2 = 0.$$

Чтобы получить острый (точнее — не тупой) угол между плоскостями, следует правую часть формулы (4) взять по абсолютной величине.

В случае, когда плоскости Π_1 и Π_2 заданы нормальными уравнениями

$$x\cos\alpha_1 + y\cos\beta_1 + z\cos\gamma_1 - p_1 = 0$$

и

$$x\cos\alpha_2 + y\cos\beta_2 + z\cos\gamma_2 - p_2 = 0,$$

угол θ между ними определяется формулой

$$\cos\theta = \cos\alpha_1\cos\alpha_2 + \cos\beta_1\cos\beta_2 + \cos\gamma_1\cos\gamma_2.$$

В частности,

$$\Pi_1 \perp \Pi_2 \Leftrightarrow \cos\alpha_1\cos\alpha_2 + \cos\beta_1\cos\beta_2 + \cos\gamma_1\cos\gamma_2 = 0.$$

Далее можно продолжать так же, как и для прямых (ввести углы φ и θ' и получить для них соответствующие формулы). Без сомнения, читатель может теперь это сделать самостоятельно и потому мы этим заниматься здесь не будем.

§ 3. ПРЯМАЯ В ПРОСТРАНСТВЕ

1. Прямая в аффинном пространстве

Любая прямая в пространстве является пересечением двух непараллельных плоскостей и потому задается (в фиксированной аффинной координатной системе $Oxyz$) двумя уравнениями первой степени

$$A_1x + B_1y + C_1z + D_1 = 0,$$
$$A_2x + B_2y + C_2z + D_2 = 0.$$

Для того чтобы такие уравнения являлись уравнениями некоторой прямой, необходимо и достаточно, чтобы соответствующие плоскости были не параллельны, т. е. чтобы ранг матрицы

$$\begin{pmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{pmatrix} \tag{1}$$

был равен двум.

Прямую в пространстве можно задавать также ее произвольной точкой $M_0(x_0, y_0, z_0)$ и некоторым отличным от нуля

вектором $\boldsymbol{a}(l, m, n)$, параллельным этой прямой (т. е. ее *направляющим вектором*; ср. п. 2 § 1). На этом пути мы приходим к *параметрическим уравнениям*

$$\begin{aligned} x &= x_0 + tl, \\ y &= y_0 + tm, \\ z &= z_0 + tn \end{aligned}$$

прямой и к ее *каноническим уравнениям*

$$\frac{x - x_0}{l} = \frac{y - y_0}{m} = \frac{z - z_0}{n},$$

двумя разными способами выражающим тот факт, что точка $M(x, y, z)$ тогда и только тогда принадлежит прямой, когда вектор $\overrightarrow{M_0M}$ коллинеарен вектору $\boldsymbol{a}$ (ср. п. 2 § 1).

Заметим, что хотя канонических уравнений формально три:

$$\frac{x - x_0}{l} = \frac{y - y_0}{m}, \quad \frac{x - x_0}{l} = \frac{z - z_0}{n}, \quad \frac{y - y_0}{m} = \frac{z - z_0}{n},$$

но каждое из них является следствием двух других, так что на самом деле мы имеем, как и полагается, только два независимых уравнения.

Чтобы от общих уравнений

$$\begin{aligned} A_1x + B_1y + C_1z + D_1 &= 0, \\ A_2x + B_2y + C_2z + D_2 &= 0 \end{aligned} \tag{2}$$

перейти к каноническим уравнениям, нужно найти

1) хотя бы одну точку $M_0(x_0, y_0, z_0)$ прямой (2);

2) некоторый ее направляющий вектор $\boldsymbol{a}(l, m, n)$.

Для нахождения точки $M_0(x_0, y_0, z_0)$ следует одному из неизвестных x, y, z в уравнениях (2) придать произвольное значение (скажем, 0) и решить получающуюся систему двух линейных уравнений с двумя неизвестными. Чтобы обеспечить существование решения, нужно при этом «свободную» неизвестную (которой мы придаем произвольное значение) выбрать так, чтобы определитель, составленный из коэффициентов при других двух неизвестных, был отличен от нуля (это всегда можно сделать, поскольку ранг матрицы (1) по условию равен двум).

Чтобы найти направляющий вектор, можно, например, найти на данной прямой вторую точку $M_1(x_1, y_1, z_1)$, отличную от точки $M_0(x_0, y_0, z_0)$ (для этого достаточно дать свободной неизвестной иное значение) и принять за вектор $\boldsymbol{a}$ вектор $\overrightarrow{M_0M_1}$.

Можно поступить и более «инвариантным» образом, заметив, что вектор $\boldsymbol{a}$ тогда и только тогда является направляющим вектором прямой, когда он параллелен обеим плоскостям, пересе-

чением которых является эта прямая, т. е. (см. п. 2 § 2) когда его координаты l, m, n удовлетворяют уравнениям

$$A_1 l + B_1 m + C_1 n = 0,$$
$$A_2 l + B_2 m + C_2 n = 0.$$

Тривиальное вычисление показывает, что этим уравнениям удовлетворяют, например, числа

$$l = \begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix}, \quad m = -\begin{vmatrix} A_1 & C_1 \\ A_2 & C_2 \end{vmatrix}, \quad n = \begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}$$

(среди которых хотя бы одно не равно нулю). Таким образом, эти числа можно принять за координаты вектора $\boldsymbol{a}$.

Этот направляющий вектор можно записать в следующем виде:

$$\boldsymbol{a} = \begin{vmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{vmatrix},$$

удобном для запоминания.

Замечание 1. В случае прямоугольных координат (и только в этом случае!) — т. е. в случае, когда базис $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ ортонормирован — вектор $\boldsymbol{a}$ является векторным произведением векторов $\boldsymbol{n}_1(A_1, B_1, C_1)$ и $\boldsymbol{n}_2(A_2, B_2, C_2)$, перпендикулярных рассматриваемым плоскостям.

Из полученных формул для направляющего вектора $\boldsymbol{a}$ немедленно вытекает, в частности, что

для прямой

$$\lambda: \begin{aligned} A_1 x + B_1 y + C_1 z + D_1 &= 0, \\ A_2 x + B_2 y + C_2 z + D_2 &= 0 \end{aligned}$$

имеют место следующие эквивалентности:

а) $\lambda \parallel Oxy \Leftrightarrow \begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} = 0,$

б) $\lambda \parallel Oxz \Leftrightarrow \begin{vmatrix} A_1 & C_1 \\ A_2 & C_2 \end{vmatrix} = 0,$

в) $\lambda \parallel Oyz \Leftrightarrow \begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix} = 0.$

Замечание 2. Мы видим, таким образом, что равенство нулю какого-либо минора матрицы

$$\begin{pmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{pmatrix}$$

имеет (при фиксированной координатной системе) *инвариантный геометрический смысл*, т. е. сохраняется при любом

другом выборе пары плоскостей, определяющих данную прямую. В дополнении к § 2 гл. 4 мы выясним причины этого.

В случае, когда $\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} \neq 0$, т. е. когда рассматриваемая прямая не параллельна координатной плоскости Oxy, мы можем решить уравнения этой прямой относительно x и y, т. е. привести их к виду

$$\begin{aligned} x &= az + p, \\ y &= bz + q. \end{aligned} \tag{3}$$

Такие уравнения прямой называются ее *приведенными уравнениями*. Задание прямой приведенными уравнениями означает ее представление в виде пересечения двух плоскостей, одна из которых параллельна оси ординат Oy (плоскость $x = az + p$), а другая — оси абсцисс Ox (плоскость $y = bz + q$). Первое из уравнений (3) (уравнение $x = az + p$) является одновременно уравнением проекции данной прямой (параллельно оси Oy) на плоскость Oxz (в имеющихся на этой плоскости координатах x, z), а второе (уравнение $y = bz + q$) — уравнением проекции данной прямой (параллельно оси Ox) на плоскость Oyz (в имеющихся на этой плоскости координатах y, z).

Чтобы помочь читателю освоиться с уравнениями прямой, решим для примера две простые задачи на прямую и плоскость.

Задача 1. *Найти плоскость, проходящую через прямую*

$$\frac{x - x_0}{l} = \frac{y - y_0}{m} = \frac{z - z_0}{n}$$

и точку $M_1(x_1, y_1, z_1)$, *не лежащую на этой прямой.*

Ясно, что точка $M(x, y, z)$ тогда и только тогда принадлежит искомой плоскости, когда векторы

$$\overrightarrow{M_0M}(x - x_0,\ y - y_0,\ z - z_0), \quad \overrightarrow{M_0M_1}(x_1 - x_0,\ y_1 - y_0,\ z_1 - z_0)$$

и

$$\boldsymbol{a}(l, m, n)$$

компланарны. Поэтому уравнение этой плоскости имеет вид

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ l & m & n \end{vmatrix} = 0.$$

Это действительно уравнение первой степени (т. е. хотя бы один из коэффициентов при x, y, z отличен от нуля), поскольку векторы $\overrightarrow{M_0M_1}(x_1 - x_0, y_1 - y_0, z_1 - z_0)$ и $\boldsymbol{a}(l, m, n)$ не коллинеарны.

Задача 2. *Найти плоскость, проходящую через прямую*

$$\frac{x-x_0}{l}=\frac{y-y_0}{m}=\frac{z-z_0}{n}$$

параллельно прямой

$$\frac{x-x_1}{l_1}=\frac{y-y_1}{m_1}=\frac{z-z_1}{n_1},$$

в предположении, что эти прямые не параллельны (иначе задача была бы неопределенной).

Искомая плоскость может быть рассматриваема как плоскость, проходящая через точку $M_0(x_0, y_0, z_0)$ параллельно неколлинеарным векторам $\boldsymbol{a}(l, m, n)$ и $\boldsymbol{a}_1(l_1, m_1, n_1)$. Следовательно (п. 2 § 2), ее уравнение имеет вид

$$\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ l & m & n \\ l_1 & m_1 & n_1 \end{vmatrix}=0.$$

2. Взаимное расположение прямых и плоскостей

Две прямые λ_1 и λ_2 в пространстве могут

а) скрещиваться (т. е. не принадлежать одной плоскости); это — *общий тип* расположения двух прямых в пространстве;

б) пересекаться (и потому лежать в одной плоскости);

в) быть параллельными (и потому также лежать в одной плоскости), но не совпадать;

г) совпадать.

Пусть прямые λ_1 и λ_2 заданы их каноническими уравнениями

$$\frac{x-x_1}{l_1}=\frac{y-y_1}{m_1}=\frac{z-z_1}{n_1}$$

и

$$\frac{x-x_2}{l_2}=\frac{y-y_2}{m_2}=\frac{z-z_2}{n_2}.$$

Рассмотрим направляющие векторы $\boldsymbol{a}_1(l_1, m_1, n_1)$, $\boldsymbol{a}_2(l_2, m_2, n_2)$ этих прямых и вектор $\boldsymbol{r}=\overrightarrow{M_1M_2}(x_2-x_1,\ y_2-y_1,\ z_2-z_1)$. Очевидно, что прямые тогда и только тогда скрещиваются, когда векторы $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ и $\boldsymbol{r}$ не компланарны. Если эти векторы компланарны, но векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ не коллинеарны, то прямые пересекаются. Если векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ коллинеарны, но не коллинеарны вектору $\boldsymbol{r}$, то прямые параллельны, но не совпадают. Если все три вектора $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ и $\boldsymbol{r}$ попарно коллинеарны, то прямые совпадают. Вспоминая условия компланарности и коллинеарности векторов и обозначая через r ранг матрицы

$$\begin{pmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{pmatrix},$$

а через R — ранг матрицы

$$\begin{pmatrix} x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{pmatrix}, \qquad (1)$$

мы немедленно получаем отсюда, что

прямые λ_1 *и* λ_2

а) *скрещиваются* $\Leftrightarrow R = 3$;

б) *пересекаются* $\Leftrightarrow r = 2,\ R = 2$;

в) *параллельны* $\Leftrightarrow r = 1,\ R = 2$;

г) *совпадают* $\Leftrightarrow r = 1,\ R = 1$.

В частности,

прямые λ_1 *и* λ_2 *тогда и только тогда компланарны* (*лежат в одной плоскости*), *когда определитель матрицы* (1) *равен нулю.*

Прямая λ и плоскость Π в пространстве могут

а) пересекаться в единственной точке; это — *общий тип* расположения прямой и плоскости в пространстве;

б) не иметь ни одной общей точки (прямая параллельна плоскости, но в ней не содержится);

в) быть инцидентными (по определению, это означает, что прямая содержится в плоскости).

Пусть прямая λ задана каноническими уравнениями

$$\frac{x - x_0}{l} = \frac{y - y_0}{m} = \frac{z - z_0}{n},$$

а плоскость Π — общим уравнением

$$Ax + By + Cz + D = 0.$$

Прямая тогда и только тогда пересекает плоскость в единственной точке, когда ее направляющий вектор $\boldsymbol{a}(l, m, n)$ не параллелен этой плоскости, т. е. (п. 2 § 2) когда

$$Al + Bm + Cn \neq 0.$$

Если $Al + Bm + Cn = 0$, но ни одна точка прямой, и, в частности, точка $M_0(x_0, y_0, z_0)$, не принадлежит плоскости, то прямая и плоскость параллельны, но не инцидентны. Наконец, если $Al + Bm + Cn = 0$ и точка M_0 принадлежит плоскости ($Ax_0 + By_0 + Cz_0 + D = 0$), то прямая инцидентна плоскости. Таким образом,

прямая λ *и плоскость* Π

а) *пересекаются* $\Leftrightarrow Al + Bm + Cn \neq 0$;

б) *параллельны, но не инцидентны* $\Leftrightarrow Al + Bm + Cn = 0$, *но* $Ax_0 + By_0 + Cz_0 + D \neq 0$;

в) *инцидентны* $\Leftrightarrow Al + Bm + Cn = 0$ *и* $Ax_0 + By_0 + Cz_0 + D = 0$.

Это утверждение можно доказать и по-иному, если поставить задачу о разыскании точек пересечения (общих точек) прямой и плоскости. Параметрические уравнения нашей прямой имеют вид

$$x = x_0 + lt,$$
$$y = y_0 + mt,$$
$$z = z_0 + nt,$$

и потому значения параметра t, соответствующие точкам пересечения, определяются из уравнения

$$A(x_0 + lt) + B(y_0 + mt) + C(z_0 + nt) + D = 0,$$

т. е. из уравнения

$$(Al + Bm + Cn)t + (Ax_0 + By_0 + Cz_0 + D) = 0.$$

При $Al + Bm + Cn \neq 0$ это уравнение имеет единственное решение (и, следовательно, имеет место случай а)), при $Al + Bm + Cn = 0$ и $Ax_0 + By_0 + Cz_0 + D \neq 0$ это уравнение не имеет решений (случай б)), а при $Al + Bm + Cn = 0$ и $Ax_0 + By_0 + Cz_0 + D = 0$ оно удовлетворяется тождественно (случай в)).

Вопрос о взаимном расположении прямой и плоскости можно ставить и тогда, когда прямая λ задана своими общими уравнениями

$$A_1x + B_1y + C_1z + D_1 = 0,$$
$$A_2x + B_2y + C_2z + D_2 = 0.$$

В этом случае все сводится к теореме 2 п. 3 § 2 о взаимном расположении трех плоскостей.

Пусть r — ранг матрицы

$$\begin{pmatrix} A & B & C \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{pmatrix},$$

а R — ранг матрицы

$$\begin{pmatrix} A & B & C & D \\ A_1 & B_1 & C_1 & D_1 \\ A_2 & B_2 & C_2 & D_2 \end{pmatrix}.$$

Согласно теореме 2 п. 3 § 2

прямая λ *и плоскость* Π

а) *пересекаются в одной точке* $\Leftrightarrow$ $r = 3$, $R = 3$;

б) *параллельны, но не инцидентны* $\Leftrightarrow$ $r = 2$, $R = 3$;

в) *инцидентны* $\Leftrightarrow$ $r = 2$, $R = 2$.

Действительно, случай а) имеет место тогда и только тогда, когда все три рассматриваемые плоскости расположены по типу 1, случай б) — когда они расположены по типу 2 или 3, а случай в) — когда они расположены по типу 4 или 5. (Заметим, что типы расположения 6, 7 и 8 в рассматриваемом случае невозможны.)

3. Прямая в евклидовом пространстве

Напомним, что расстоянием точки от прямой называется длина перпендикуляра, опущенного из этой точки на прямую.

Пусть

$$\frac{x-x_0}{l}=\frac{y-y_0}{m}=\frac{z-z_0}{n} \tag{1}$$

— канонические уравнения произвольной прямой. Для любой точки $M(x, y, z)$ (с радиус-вектором $\boldsymbol{r}$), не принадлежащей этой прямой, рассмотрим параллелограмм, построенный на векторах $\overrightarrow{M_0M}=\boldsymbol{r}-\boldsymbol{r}_0$ и $\boldsymbol{a}(l, m, n)$ (имеется в виду, что эти векторы отложены от точки $M_0(x_0, y_0, z_0)$. Высота этого параллелограмма равна, по определению, расстоянию d точки M от прямой (1). Поэтому его площадь равна произведению расстояния d на длину $|\boldsymbol{a}|$ вектора $\boldsymbol{a}$. Но эта площадь равна площади бивектора $(\boldsymbol{r}-\boldsymbol{r}_0)\wedge\boldsymbol{a}$ (т. е. длине $|\boldsymbol{r}-\boldsymbol{r}_0)\times\boldsymbol{a}|$ векторного произведения [1]) $(\boldsymbol{r}-\boldsymbol{r}_0)\times\boldsymbol{a}$). Следовательно,

расстояние d точки $M(x, y, z)$ от прямой (1) *выражается формулой*

$$d=\frac{|(\boldsymbol{r}-\boldsymbol{r}_0)\times\boldsymbol{a}|}{|\boldsymbol{a}|}. \tag{2}$$

В координатах эта формула имеет вид

$$d=\frac{\begin{vmatrix} x-x_0 & y-y_0 \\ l & m \end{vmatrix}^2+\begin{vmatrix} x-x_0 & z-z_0 \\ l & n \end{vmatrix}^2+\begin{vmatrix} y-y_0 & z-z_0 \\ m & n \end{vmatrix}^2}{l^2+m^2+n^2}.$$

Формулу (2) можно получить и иначе, заметив, что по известной теореме о перпендикуляре и наклонной, расстояние d точки M от прямой является минимумом расстояний от этой точки до произвольной точки рассматриваемой прямой. Поскольку каждая точка нашей прямой имеет радиус-вектор $\boldsymbol{r}_0+\boldsymbol{a}t$, где t — некоторое число, мы, обозначая расстояние от точки M до этой точки символом $\omega(t)$, можем поэтому написать, что

$$d=\min_t \omega(t),$$

где минимум берется по всем вещественным t. Для вычислений удобнее эту формулу возвести в квадрат:

$$d^2=\min_t \omega^2(t).$$

[1]) Поскольку наши рассуждения имеют сугубо метрический характер, переход от бивекторов к ассоциированным векторам вполне оправдан.

Согласно формуле для расстояния между двумя точками

$$\omega^2(t) = (r - r_0 - at)^2,$$

где r, как и выше, — радиус-вектор точки M. Дифференцируя эту формулу по t (легко видеть, что обычные правила дифференцирования суммы и произведения остаются справедливыми и для векторов[1])), мы немедленно находим, что

$$\frac{d\omega^2(t)}{dt} = -2(r - r_0 - at)\,a, \qquad \frac{d^2\omega^2(t)}{dt^2} = 2a^2 > 0.$$

Поскольку $\frac{d^2\omega^2(t)}{dt^2} > 0$, значение t, при котором $\frac{d\omega^2(t)}{dt} = 0$, соответствует минимуму функции $\omega^2(t)$. Это значение, очевидно, равно

$$t_0 = \frac{(r - r_0)\,a}{a^2}.$$

Следовательно,

$$d^2 = \omega^2(t_0) = \left(r - r_0 - \frac{(r - r_0)\,a}{a^2}\,a\right)^2 = (r - r_0)^2 - 2\frac{((r - r_0)\,a)^2}{a^2} + \\ + \frac{((r - r_0)\,a)^2}{a^2} = (r - r_0)^2 - \frac{((r - r_0)\,a)^2}{a^2}.$$

Полученную формулу можно переписать в более симметричном виде:

$$d^2 = \frac{\begin{vmatrix} (r - r_0)^2 & a\,(r - r_0) \\ (r - r_0)\,a & a^2 \end{vmatrix}}{a^2}.$$

Определитель в числителе является не чем иным, как определителем Грама $\Gamma(r - r_0,\ a)$ векторов $r - r_0$ и a (см. п. 4 § 6 гл. 1), и потому он равен квадрату площади их внешнего произведения. Следовательно, окончательно,

$$d = \frac{|(r - r_0) \wedge a|}{|a|},$$

что совпадает с полученным выше результатом.

Заметим, что мы не только вычислили минимум, но и доказали его существование.

Кроме того, мы теперь без особого труда можем написать параметрические (или, что равносильно, канонические) уравнения перпендикуляра, опущенного из произвольной точки $M_1(x_1, y_1, z_1)$ на прямую (1).

Действительно, как мы видели, основанию этого перпендикуляра соответствует значение

$$t_0 = \frac{(r_1 - r_0)\,a}{a^2}$$

параметра t. Поэтому его радиус-вектор равен

$$r_0 + \frac{(r_1 - r_0)\,a}{a^2}\,a,$$

откуда после тривиальных алгебраических преобразований получается, что *перпендикуляр, опущенный из точки $M_1(r_1)$ на прямую*

$$r = r_0 + at,$$

---

[1]) Проще всего это проверить, написав соответствующие формулы в координатах.

задается параметрическим уравнением

$$\boldsymbol{r} = \frac{\begin{vmatrix} (1-t)\,\boldsymbol{r}_0 + t\boldsymbol{r}_1 & (1-t)\,\boldsymbol{a} \\ (\boldsymbol{r}_1 - \boldsymbol{r}_0)\,\boldsymbol{a} & \boldsymbol{a}^2 \end{vmatrix}}{\boldsymbol{a}^2}. \tag{3}$$

Напомним, что *углом* между двумя прямыми в пространстве называется угол между параллельными им прямыми, проходящими через произвольную точку (и потому лежащими в одной плоскости). Очевидно, что это определение корректно (не зависит от выбора точки). Оно дает для угла между прямыми два значения, дополняющие друг друга до π.

Один из этих дополнительных углов равен, очевидно, углу θ между направляющими векторами рассматриваемых прямых. Поэтому

для угла θ между прямыми

$$\lambda_1\colon \frac{x-x_1}{l_1} = \frac{y-y_1}{m_1} = \frac{z-z_1}{n_1}$$

и

$$\lambda_2\colon \frac{x-x_2}{l_2} = \frac{y-y_2}{m_2} = \frac{z-z_2}{n_2}$$

справедлива формула

$$\cos\theta = \frac{l_1l_2 + m_1m_2 + n_1n_2}{\sqrt{l_1^2+m_1^2+n_1^2}\,\sqrt{l_2^2+m_2^2+n_2^2}}. \tag{4}$$

Угол θ — тупой, если $l_1l_2 + m_1m_2 + n_1n_2 < 0$, острый, — если $l_1l_2 + m_1m_2 + n_1n_2 > 0$, и равен $\pi/2$, если $l_1l_2 + m_1m_2 + n_1n_2 = 0$.

Чтобы получить острый (точнее — не тупой) угол между прямыми, следует правую часть формулы (4) взять по абсолютной величине.

В частности, мы видим, что

$$\lambda_1 \perp \lambda_2 \iff l_1l_2 + m_1m_2 + n_1n_2 = 0.$$

Угол *между прямой и плоскостью*, подобно углу между двумя прямыми, принимает, вообще говоря, два значения, дополняющие друг друга до π. Мы будем интересоваться только острым (точнее — не тупым) из этих двух углов. Обозначим его через θ. Этот угол можно определить как дополнение до $\pi/2$ нетупого угла φ между данной прямой и перпендикуляром к данной плоскости:

$$\theta = \frac{\pi}{2} - \varphi.$$

При $\varphi \neq 0$, т. е. в случае когда прямая и плоскость не перпендикулярны, угол θ можно также определить как острый угол между прямой и ее ортогональной проекцией на плоскость.

Пусть прямая λ задана ее каноническими уравнениями

$$\frac{x-x_0}{l}=\frac{y-y_0}{m}=\frac{z-z_0}{n},$$

а плоскость Π — ее общим уравнением

$$Ax+By+Cz+D=0.$$

По определению, угол φ является либо углом между векторами $\boldsymbol{a}(l, m, n)$ и $\boldsymbol{n}(A, B, C)$, либо представляет собой дополнение этого угла до π. Поэтому

$$\cos\varphi=\left|\frac{Al+Bm+Cn}{\sqrt{A^2+B^2+C^2}\sqrt{l^2+m^2+n^2}}\right|.$$

Но $\cos\varphi=\sin\theta$, ибо $\theta=\frac{\pi}{2}-\varphi$. Следовательно,

$$\sin\theta=\left|\frac{Al+Bm+Cn}{\sqrt{A^2+B^2+C^2}\sqrt{l^2+m^2+n^2}}\right|. \tag{5}$$

В частности, мы снова получаем, что

$$\lambda\parallel\Pi \Leftrightarrow Al+Bm+Cn=0.$$

Из формулы (5) довольно трудно получить условие, когда $\theta=\pi/2$, т. е. когда $\lambda\perp\Pi$. Проще, не пользуясь этой формулой, непосредственно воспользоваться тем, что вектор $\boldsymbol{a}(l, m, n)$ параллелен прямой λ, а вектор $\boldsymbol{n}(A, B, C)$ перпендикулярен плоскости Π. Поэтому

$$\lambda\perp\Pi \Leftrightarrow \frac{l}{A}=\frac{m}{B}=\frac{n}{C}.$$

Отсюда, в частности, непосредственно вытекает, что

уравнения прямой, проходящей через точку $M_0(x_0, y_0, z_0)$ *перпендикулярно плоскости*

$$Ax+By+Cz+D=0,$$

имеют вид

$$\frac{x-x_0}{A}=\frac{y-y_0}{B}=\frac{z-z_0}{C},$$

а уравнение плоскости, проходящей через точку $M_0(x_0, y_0, z_0)$ *перпендикулярно прямой*

$$\frac{x-x_1}{l}=\frac{y-y_1}{m}=\frac{z-z_1}{n},$$

имеет вид

$$l(x-x_0)+m(y-y_0)+n(z-z_0)=0.$$

Плоскость, проходящая через прямую λ перпендикулярно плоскости Π, может быть охарактеризована как плоскость, проходящая через точку $M_0(x_0, y_0, z_0)$ параллельно двум векторам $\boldsymbol{a}(l, m, n)$ и $\boldsymbol{n}(A, B, C)$ (мы предполагаем, что эти векторы

не коллинеарны, т. е. что прямая λ и плоскость Π не перпендикулярны, ибо, в противном случае, искомая плоскость не была бы однозначно определена). Поэтому

уравнение плоскости, проходящей через прямую λ *перпендикулярно плоскости* Π, *имеет вид*

$$\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ l & m & n \\ A & B & C \end{vmatrix} = 0.$$

Аналогично, перпендикуляр, опущенный из точки $M_1(x_1, y_1, z_1)$ на прямую λ, является пересечением двух плоскостей: одной, проходящей через точку M_1 перпендикулярно прямой λ, и другой, проходящей через точку M_1 и прямую λ (см. задачу 1 в п. 1). Поэтому

перпендикуляр, опущенный из точки $M_1(x_1, y_1, z_1)$ *на прямую* λ, *может быть задан уравнениями*

$$\begin{gathered} l(x-x_1)+m(y-y_1)+n(z-z_1)=0 \\ \begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ x_1-x_0 & y_1-y_0 & z_1-z_0 \\ l & m & n \end{vmatrix} = 0. \end{gathered} \tag{6}$$

Упражнение. Сравните уравнения (6) с полученными выше параметрическими уравнениями этого перпендикуляра (см. формулы (3)).

4. Расстояние между двумя прямыми в пространстве

По определению, расстояние между двумя прямыми равно минимуму расстояний точек одной прямой до точек другой.

Если прямые пересекаются (или совпадают), то этот минимум равен нулю. Если прямые параллельны, то этот минимум равен расстоянию о произвольной точки одной прямой до другой прямой и может быть вычислен по формуле (2) предыдущего пункта. Поэтому вопрос о вычислении расстояния между двумя прямыми интересен лишь для скрещивающихся прямых.

Пусть

$$\frac{x-x_1}{l_1} = \frac{y-y_1}{m_1} = \frac{z-z_1}{n_1},$$

$$\frac{x-x_2}{l_2} = \frac{y-y_2}{m_2} = \frac{z-z_2}{n_2}$$

— канонические уравнения двух скрещивающихся прямых λ_1 и λ_2. Радиус-векторы точек прямой λ_1 имеют вид $\boldsymbol{r}_1 + \boldsymbol{a}_1 t_1$, а прямой λ_2 — вид $\boldsymbol{r}_2 + \boldsymbol{a}_2 t_2$, где $\boldsymbol{r}_1$ и $\boldsymbol{r}_2$ — радиус-векторы точек $M_1(x_1, y_1, z_1)$ и $M_2(x_2, y_2, z_2)$, $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ — направляющие векторы (l_1, m_1, n_1) и (l_2, m_2, n_2), а t_1 и t_2 — числовые параметры. По-

этому расстояние между точкой прямой λ_1, отвечающей значению параметра t_1, и точкой прямой λ_2, отвечающей значению параметра t_2, равно

$$\omega(t_1, t_2) = |\boldsymbol{r}_2 - \boldsymbol{r}_1 + \boldsymbol{a}_2 t_2 - \boldsymbol{a}_1 t_1|,$$

а квадрат этого расстояния равен

$$\omega^2(t_1, t_2) = (\boldsymbol{r}_2 - \boldsymbol{r}_1 + \boldsymbol{a}_2 t_2 - \boldsymbol{a}_1 t_1)^2.$$

По определению, расстояние d между рассматриваемыми прямыми выражается формулой

$$d = \min_{t_1, t_2} \omega(t_1, t_2),$$

а квадрат этого расстояния — формулой

$$d^2 = \min_{t_1, t_2} \omega^2(t_1, t_2),$$

где минимум берется по всем вещественным t_1 и t_2.

Чтобы найти этот минимум, вычислим [1]) частные производные функции $\omega^2(t_1, t_2)$ по t_1 и t_2. Имеем:

$$\frac{\partial \omega^2(t_1, t_2)}{\partial t_1} = -2(\boldsymbol{r}_2 - \boldsymbol{r}_1 + \boldsymbol{a}_2 t_2 - \boldsymbol{a}_1 t_1)\boldsymbol{a}_1,$$

$$\frac{\partial \omega^2(t_1, t_2)}{\partial t_2} = 2(\boldsymbol{r}_2 - \boldsymbol{r}_1 + \boldsymbol{a}_2 t_2 - \boldsymbol{a}_1 t_1)\boldsymbol{a}_2$$

и

$$\frac{\partial^2 \omega^2}{\partial t_1^2} = 2\boldsymbol{a}_1^2, \quad \frac{\partial^2 \omega^2}{\partial t_1 \partial t_2} = -2\boldsymbol{a}_1\boldsymbol{a}_2, \quad \frac{\partial^2 \omega^2}{\partial t_2^2} = 2\boldsymbol{a}_2^2.$$

Таким образом,

$$\begin{vmatrix} \frac{\partial^2 \omega^2}{\partial t_1^2} & \frac{\partial^2 \omega^2}{\partial t_1 \partial t_2} \\ \frac{\partial^2 \omega^2}{\partial t_1 \partial t_2} & \frac{\partial^2 \omega}{\partial t_2^2} \end{vmatrix} = 4\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2) > 0,$$

где

$$\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2) = \begin{vmatrix} \boldsymbol{a}_1^2 & \boldsymbol{a}_1\boldsymbol{a}_2 \\ \boldsymbol{a}_2\boldsymbol{a}_1 & \boldsymbol{a}_2^2 \end{vmatrix}$$

— определитель Грама векторов $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$. Поэтому значения t_1 и t_2, при которых

$$\frac{\partial^2 \omega^2(t_1, t_2)}{\partial t_1} = \frac{\partial^2 \omega^2(t_1, t_2)}{\partial t_2} = 0,$$

[1]) Читатель, не знакомый с аналитическими методами отыскания экстремумов функций двух переменных, найдет геометрический вывод формулы для расстояния d в конце этого пункта.

соответствуют минимуму функции $\omega^2(t_1, t_2)$, т. е. при этих значениях функция $\omega^2(t_1, t_2)$ равна квадрату d^2 искомого расстояния d.

Эти значения являются решениями системы двух уравнений с двумя неизвестными

$$\begin{aligned}(\boldsymbol{r}_2 - \boldsymbol{r}_1 + \boldsymbol{a}_2 t_2 - \boldsymbol{a}_1 t_1)\,\boldsymbol{a}_1 = 0,\\ (\boldsymbol{r}_2 - \boldsymbol{r}_1 + \boldsymbol{a}_2 t_2 - \boldsymbol{a}_1 t_1)\,\boldsymbol{a}_2 = 0,\end{aligned} \tag{1}$$

т. е. системы

$$\begin{aligned}\boldsymbol{a}_1^2 t_1 - \boldsymbol{a}_1\boldsymbol{a}_2 t_2 = (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_1,\\ \boldsymbol{a}_2\boldsymbol{a}_1 t_1 - \boldsymbol{a}_2^2 t_2 = (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_2\end{aligned}$$

с определителем $-\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2) \neq 0$, и потому выражаются формулами

$$t_1 = \frac{\begin{vmatrix}(\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_1 & \boldsymbol{a}_1\boldsymbol{a}_2\\ (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_2 & \boldsymbol{a}_2^2\end{vmatrix}}{\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2)},$$

$$t_2 = \frac{\begin{vmatrix}\boldsymbol{a}_1^2 & (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_1\\ \boldsymbol{a}_2\boldsymbol{a}_1 & (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_2\end{vmatrix}}{\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2)}.$$

Подставив эти значения в формулу для $\omega^2(t_1, t_2)$, мы получим, что

$$d^2 = \left(\boldsymbol{r}_2 - \boldsymbol{r}_1 - \frac{\begin{vmatrix}\boldsymbol{a}_1^2 & (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_1\\ \boldsymbol{a}_2\boldsymbol{a}_1 & (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_2\end{vmatrix}}{\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2)}\boldsymbol{a}_2 - \frac{\begin{vmatrix}(\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_1 & \boldsymbol{a}_1\boldsymbol{a}_2\\ (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_2 & \boldsymbol{a}_2^2\end{vmatrix}}{\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2)}\boldsymbol{a}_1\right)^2.$$

Используя формулу разложения определителя третьего порядка по элементам первой строки и учитывая, что $\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2) = -\begin{vmatrix}\boldsymbol{a}_1^2 & \boldsymbol{a}_1\boldsymbol{a}_2\\ \boldsymbol{a}_2\boldsymbol{a}_1 & \boldsymbol{a}_2^2\end{vmatrix}$, эту формулу можно записать в более симметричном виде:

$$d^2 = \frac{\begin{vmatrix}\boldsymbol{r}_2 - \boldsymbol{r}_1 & \boldsymbol{a}_1 & \boldsymbol{a}_2\\ (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_1 & \boldsymbol{a}_1^2 & \boldsymbol{a}_1\boldsymbol{a}_2\\ (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_2 & \boldsymbol{a}_1\boldsymbol{a}_2 & \boldsymbol{a}_2^2\end{vmatrix}^2}{\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2)^2}. \tag{2}$$

Формулу (2) можно привести к более простому виду. Мы не будем это делать прямым вычислением, а предпочтем обходной путь, не требующий почти никаких выкладок.

Пусть N_1 и N_2 — точки наших прямых, соответствующие найденным значениям параметров t_1 и t_2. Тогда

$$d^2 = (\overrightarrow{N_1N_2})^2,$$

причем, по доказанному,

$$\overrightarrow{N_1N_2} = \boldsymbol{r}_2 - \boldsymbol{r}_1 + \boldsymbol{a}_2 t_2 - \boldsymbol{a}_1 t_1 = \frac{\begin{vmatrix} \boldsymbol{r}_2 - \boldsymbol{r}_1 & \boldsymbol{a}_1 & \boldsymbol{a}_2 \\ (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_1 & \boldsymbol{a}_1^2 & \boldsymbol{a}_1\boldsymbol{a}_2 \\ (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_2 & \boldsymbol{a}_1\boldsymbol{a}_2 & \boldsymbol{a}_2^2 \end{vmatrix}}{\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2)}.$$

Заметим теперь, что уравнения (1) могут быть переписаны в следующем виде:

$$\begin{aligned} \overrightarrow{N_1N_2}\,\boldsymbol{a}_1 &= 0, \\ \overrightarrow{N_1N_2}\,\boldsymbol{a}_2 &= 0, \end{aligned} \tag{3}$$

означающем, что вектор $\overrightarrow{N_1N_2}$ ортогонален векторам $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$. Поэтому

$$d^2 = (\overrightarrow{N_1N_2})^2 = \overrightarrow{N_1N_2}\,(\boldsymbol{r}_2 - \boldsymbol{r}_1 + \boldsymbol{a}_2 t_2 - \boldsymbol{a}_1 t_1) = \overrightarrow{N_1N_2}\,(\boldsymbol{r}_2 - \boldsymbol{r}_1).$$

Но умножение определителя, одна строка которого состоит из векторов, на некоторый вектор можно, очевидно [1]), произвести, умножая эту строку на данный вектор. Поэтому

$$d^2 = \frac{\begin{vmatrix} (\boldsymbol{r}_2 - \boldsymbol{r}_1)^2 & \boldsymbol{a}_1\,(\boldsymbol{r}_2 - \boldsymbol{r}_1) & \boldsymbol{a}_2\,(\boldsymbol{r}_2 - \boldsymbol{r}_1) \\ (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_1 & \boldsymbol{a}_1^2 & \boldsymbol{a}_2\boldsymbol{a}_1 \\ (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_2 & \boldsymbol{a}_1\boldsymbol{a}_2 & \boldsymbol{a}_2^2 \end{vmatrix}}{\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2)}.$$

Числитель этой формулы является определителем Грама $\Gamma(\boldsymbol{r}_2 - \boldsymbol{r}_1, \boldsymbol{a}_1, \boldsymbol{a}_2)$ векторов $\boldsymbol{r}_2 - \boldsymbol{r}_1$, $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$. Тем самым нами доказано, что

квадрат d^2 расстояния d между прямыми $\boldsymbol{r} = \boldsymbol{r}_1 + \boldsymbol{a}_1 t$ и $\boldsymbol{r} = \boldsymbol{r}_2 + \boldsymbol{a}_2 t$ выражается формулой

$$d^2 = \frac{\Gamma(\boldsymbol{r}_2 - \boldsymbol{r}_1, \boldsymbol{a}_1, \boldsymbol{a}_2)}{\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2)}. \tag{4}$$

Поскольку определитель Грама трех векторов равен квадрату их смешанного произведения, а определитель Грама двух векторов равен квадрату длины их векторного произведения, формулу (4) можно переписать (извлекая квадратный корень) в следующем виде:

$$d = \frac{|(\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_1\boldsymbol{a}_2|}{|\boldsymbol{a}_1 \times \boldsymbol{a}_2|}. \tag{5}$$

[1]) Достаточно разложить определитель по элементам «векторной» строки.

В координатах формула (5) (до извлечения корня) имеет вид

$$d^2 = \frac{\begin{vmatrix} x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{vmatrix}^2}{\begin{vmatrix} l_1 & m_1 \\ l_2 & m_2 \end{vmatrix}^2 + \begin{vmatrix} l_1 & n_1 \\ l_2 & n_2 \end{vmatrix}^2 + \begin{vmatrix} m_1 & n_1 \\ m_2 & n_2 \end{vmatrix}^2}.$$

Подчеркнем, что мы не только нашли расстояние d, но и доказали, что оно существует, т. е. что соответствующий минимум достигается.

Замечание 1. Поскольку векторы $\boldsymbol{r}_2 - \boldsymbol{r}_1$, $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ могут быть произвольными (не компланарными) векторами пространства, сравнение формул (2) и (4) показывает, что для любых трех векторов $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ справедлива формула

$$\begin{vmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \\ \boldsymbol{ab} & \boldsymbol{b}^2 & \boldsymbol{cb} \\ \boldsymbol{ac} & \boldsymbol{bc} & \boldsymbol{c}^2 \end{vmatrix}^2 = \Gamma(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})\,\Gamma(\boldsymbol{b}, \boldsymbol{c}).$$

Упражнение. Докажите эту формулу прямым вычислением и выясните ее геометрический смысл. Напишите (и докажите) аналог этой формулы для двух векторов.

Как известно, *общим перпендикуляром* двух прямых λ_1 и λ_2 называется прямая, пересекающая каждую из прямых и им перпендикулярная. Из только что полученных результатов непосредственно вытекает, что

для любых двух скрещивающихся прямых существует единственный общий перпендикуляр.

Действительно, пусть N_1 и N_2 — построенные выше точки прямых λ_1 и λ_2. Рассмотрим прямую N_1N_2. Согласно формулам (3) направляющий вектор $\overrightarrow{N_1N_2}$ этой прямой ортогонален направляющим векторам $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ прямых λ_1 и λ_2, т. е. прямая N_1N_2 перпендикулярна этим прямым. Кроме того, она пересекает обе прямые λ_1 и λ_2. Следовательно, прямая N_1N_2 является их общим перпендикуляром.

Рассмотрим теперь произвольный общий перпендикуляр $N_1'N_2'$, проходящий через некоторые точки N_1' и N_2' прямых λ_1 и λ_2. Единственность общего перпендикуляра будет доказана, если мы покажем, что $N_1 = N_1'$ и $N_2 = N_2'$. С этой целью заметим, что поскольку прямая $N_1'N_2'$ перпендикулярна прямым λ_1 и λ_2, ее направляющий вектор $\overrightarrow{N_1'N_2'}$ ортогонален направляющим векторам $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$, т. е. для него имеют место равенства (3). Но эти равенства равносильны уравнениям (1), т. е. тому, что при соответствующих значениях параметров t_1 и t_2 обращаются в нуль частные производные $\frac{\partial \omega^2(t_1, t_2)}{\partial t_1}$ и $\frac{\partial \omega^2(t_1, t_2)}{\partial t^2}$. Но мы уже видели, что эти частные производные обращаются в нуль только

для одних вполне определенных значений t_1 и t_2, которым соответствуют точки N_1 и N_2. Следовательно, $N_1' = N_1$ и $N_2' = N_2$. Тем самым существование и единственность общего перпендикуляра полностью доказаны.

Для пересекающихся прямых также существует единственный общий перпендикуляр (проходящий через точку их пересечения перпендикулярно плоскости, в которой эти прямые содержатся). Для параллельных (или совпадающих) прямых существует бесчисленное множество общих перпендикуляров.

Пользуясь полученными выше формулами, можно также без труда написать и уравнения общего перпендикуляра. Поскольку нам известна его точка N_1 с радиус-вектором

$$\boldsymbol{r}_1 + \frac{\begin{vmatrix} (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_1 & \boldsymbol{a}_1\boldsymbol{a}_2 \\ (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_2 & \boldsymbol{a}_2^2 \end{vmatrix}}{\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2)}\,\boldsymbol{a}_1$$

и направляющий вектор

$$\overrightarrow{N_1N_2} = \frac{\begin{vmatrix} \boldsymbol{r}_2 - \boldsymbol{r}_1 & \boldsymbol{a}_1 & \boldsymbol{a}_2 \\ (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_1 & \boldsymbol{a}_1^2 & \boldsymbol{a}_2\boldsymbol{a}_1 \\ (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_2 & \boldsymbol{a}_1\boldsymbol{a}_2 & \boldsymbol{a}_2^2 \end{vmatrix}}{\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2)},$$

проще всего написать его параметрические (или канонические) уравнения. Сделав это, мы после тривиальных алгебраических преобразований получим, что

векторное параметрическое уравнение общего перпендикуляра к двум скрещивающимся прямым $\boldsymbol{r} = \boldsymbol{r}_1 + \boldsymbol{a}_1 t$ *и* $\boldsymbol{r} = \boldsymbol{r}_2 + \boldsymbol{a}_2 t$ *может быть записано в следующем виде*:

$$\boldsymbol{r} = \frac{\begin{vmatrix} (1-t)\,\boldsymbol{r}_1 + t\boldsymbol{r}_2 & -(1-t)\,\boldsymbol{a}_1 & t\boldsymbol{a}_2 \\ (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_1 & \boldsymbol{a}_1^2 & \boldsymbol{a}_2\boldsymbol{a}_1 \\ (\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_2 & \boldsymbol{a}_1\boldsymbol{a}_2 & \boldsymbol{a}_2^2 \end{vmatrix}}{\Gamma(\boldsymbol{a}_1, \boldsymbol{a}_2)}. \qquad (6)$$

Чтобы получить *общие уравнения* общего перпендикуляра, т. е. представить этот перпендикуляр в виде пересечения двух плоскостей, удобнее всего рассмотреть плоскости, содержащие этот перпендикуляр и одну из данных прямых. Первую из этих плоскостей можно определить как плоскость, проходящую через точку M_1 параллельно неколлинеарным векторам $\overrightarrow{N_1N_2}$ и $\boldsymbol{a}_1$, а вторую — как плоскость, проходящую через точку M_2 параллельно неколлинеарным векторам $\overrightarrow{N_1N_2}$ и $\boldsymbol{a}_2$. Получающиеся формулы можно упростить, заметив, что вектор $\overrightarrow{N_1N_2}$, будучи ортогональным неколлинеарным векторам $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, коллинеарен их

векторному произведению $\boldsymbol{a}_1 \times \boldsymbol{a}_2$ и потому может быть им заменен. Переходя к координатам, мы получаем, что

общий перпендикуляр к двум скрещивающимся прямым

$$\frac{x-x_1}{l_1}=\frac{y-y_1}{m_1}=\frac{z-z_1}{n_1} \quad \text{и} \quad \frac{x-x_2}{l_2}=\frac{y-y_2}{m_2}=\frac{z-z_2}{n_2}$$

может быть задан уравнениями

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ l_1 & m_1 & n_1 \\ \begin{vmatrix} m_1 & n_1 \\ m_2 & n_2 \end{vmatrix} & -\begin{vmatrix} l_1 & n_1 \\ l_2 & n_2 \end{vmatrix} & \begin{vmatrix} l_1 & m_1 \\ l_2 & m_2 \end{vmatrix} \end{vmatrix}=0,$$

$$\begin{vmatrix} x-x_2 & y-y_2 & z-z_2 \\ l_2 & m_2 & n_2 \\ \begin{vmatrix} m_1 & n_1 \\ m_2 & n_2 \end{vmatrix} & -\begin{vmatrix} l_1 & n_1 \\ l_2 & n_2 \end{vmatrix} & \begin{vmatrix} l_1 & m_1 \\ l_2 & m_2 \end{vmatrix} \end{vmatrix}=0. \tag{7}$$

Замечание 2. Тот факт, что векторное произведение $\boldsymbol{a}_1 \times \boldsymbol{a}_2$ коллинеарно вектору $\overrightarrow{N_1N_2}$, означает, что для любых трех векторов $\boldsymbol{a}=\boldsymbol{a}_1$, $\boldsymbol{b}=\boldsymbol{b}_1$ и $\boldsymbol{r}=\boldsymbol{r}_2-\boldsymbol{r}_1$ вектор

$$\begin{vmatrix} \boldsymbol{r} & \boldsymbol{a} & \boldsymbol{b} \\ \boldsymbol{ra} & \boldsymbol{a}^2 & \boldsymbol{ba} \\ \boldsymbol{rb} & \boldsymbol{ab} & \boldsymbol{b}^2 \end{vmatrix}$$

с точностью до коллинеарности не зависит от вектора $\boldsymbol{r}$. Это показывает, что должна иметь место формула

$$\begin{vmatrix} \boldsymbol{r} & \boldsymbol{a} & \boldsymbol{b} \\ \boldsymbol{ra} & \boldsymbol{a}^2 & \boldsymbol{ba} \\ \boldsymbol{rb} & \boldsymbol{ab} & \boldsymbol{b}^2 \end{vmatrix}=C(\boldsymbol{a}\times\boldsymbol{b}), \tag{8}$$

где C — некоторый числовой коэффициент (зависящий от векторов $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{r}$).

Упражнение. Докажите формулу (8) прямым вычислением и найдите коэффициент C.

Допуская определенную вольность речи, принято называть *длиной общего перпендикуляра* двух скрещивающихся прямых длину его отрезка $\overline{N_1N_2}$. Но, как мы знаем, длина этого отрезка равна расстоянию d между прямыми. Следовательно,

длина общего перпендикуляра двух скрещивающихся прямых равна расстоянию между этими прямыми.

Это открывает возможности для воспроизведения полученных выше результатов более геометрическим образом. Для этого следует, в первую очередь, определить расстояние между двумя скрещивающимися прямыми λ_1 и λ_2 как длину их общего перпендикуляра, предварительно доказав, конечно, его существование и единственность (что без труда делается на основе стандартных элементарно-геометрических соображений). Затем можно рассуждать, например, следующим образом.

Рассмотрим параллелепипед, построенный на векторах $\overrightarrow{M_1M_2}$, $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ (отложенных от точки M_1). Высота этого параллелепипеда равна, очевидно, длине d общего перпендикуляра прямых λ_1 и λ_2, а его основание является параллелограммом, построенным на векторах $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$. Поэтому длина d равна отношению объема параллелепипеда (т. е. абсолютного значения величины смешанного произведения $(\boldsymbol{r}_2 - \boldsymbol{r}_1)\boldsymbol{a}_1\boldsymbol{a}_2$ векторов $\boldsymbol{r}_2 - \boldsymbol{r}_1$, $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$) к площади его основания (т. е. к длине векторного произведения $\boldsymbol{a}_1 \times \boldsymbol{a}_2$ векторов $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$):

$$d = \frac{|(\boldsymbol{r}_2 - \boldsymbol{r}_1)\,\boldsymbol{a}_1\boldsymbol{a}_2|}{|\boldsymbol{a}_1 \times \boldsymbol{a}_2|}.$$

Тем самым мы заново получили формулу (5).

Можно поступить и иначе, заметив, что для точек пересечения N_1 и N_2 общего перпендикуляра с прямыми λ_1 и λ_2 должны иметь место формулы (3) (уже не как раньше — в силу равенства нулю частных производных, а потому, что прямая N_1N_2 перпендикуляриа прямым λ_1 и λ_2). Из этих уравнений (написанных в форме (1)) мы, как и раньше, найдем соответствующие значения параметров t_1 и t_2, т. е. найдем радиус-векторы точек N_1 и N_2, а значит, и величину d^2.

Имея выражения для радиус-векторов точек N_1 и N_2, мы точно так же, как и выше, можем написать параметрические уравнения (6) общего перпендикуляра N_1N_2. Что же касается его общих уравнений (в форме (7)), то для их вывода предварительное вычисление точек N_1 и N_2 даже и не требуется.

Глава 4

ГЕОМЕТРИИ ПРЯМЫХ, ПЛОСКОСТЕЙ И ОКРУЖНОСТЕЙ

§ 1. ГЕОМЕТРИЯ ПРЯМЫХ НА ПЛОСКОСТИ

В обычной «школьной» геометрии основными геометрическими элементами считаются точки, а линии и поверхности рассматриваются как семейства («геометрические места») этих точек. Однако можно развить «геометрию», в которой основными элементами («точками») являются, скажем, прямые и в которой изучаются семейства прямых, зависящие от одного или нескольких параметров (семейства, зависящие от одного параметра, аналогичны линиям, а зависящие от двух параметров — поверхностям). Отдел геометрии, занимающийся этими вопросами, называют часто *высшей геометрией прямых*. Точно так же можно говорить о *высшей геометрии плоскостей*, *окружностей*, *сфер* и т. п. Впрочем, в сочинениях последнего времени эпитет «высшая» принято опускать.

Наиболее простой из этих «неточечных» геометрий является геометрия прямых на плоскости. Поэтому мы в первую очередь ею и займемся.

1. Пучки прямых

Для применения аналитических методов необходимо научиться задавать прямые некоторыми наборами чисел (координатами). Проще всего принять за эти «координаты» коэффициенты A, B, C общего уравнения

$$Ax + By + C = 0 \qquad (1)$$

прямой в данной аффинной системе координат. Тем самым мы приходим к следующему определению:

Определение 1. *Координатами* прямой (1) в данной аффинной координатной системе Oxy называются коэффициенты A, B, C ее общего уравнения.

Координаты вполне определяют прямую. Вместе с тем прямая определяет координаты только с точностью до пропорциональности. Это означает, что координаты A, B, C являются *однородными координатами* (ср. п. 9 § 1 гл. 2). В соответствии с этим мы будем обозначать прямую с координатами A, B, C символом $(A : B : C)$.

Обратим внимание на то, что координаты прямых не могут быть выбраны произвольно: чтобы тройка (A, B, C) являлась тройкой координат некоторой прямой, необходимо и достаточно, чтобы хотя бы одно из чисел A, B было отлично от нуля.

Замечание 1. Можно ввести и неоднородные координаты прямых. Такими координатами являются, например, числа $k = -\frac{A}{B}$ и $b = -\frac{C}{B}$. Однако они определены лишь для прямых, не параллельных оси ординат. Можно показать (см. ниже замечание 2 п. 6), что неоднородных координат, пригодных для всех прямых плоскости, существовать не может. Этим геометрия прямых принципиально отличается от геометрии точек.

Аналогом прямой линии в геометрии прямых является семейство прямых $(A : B : C)$, задаваемое формулами [1])

$$\begin{aligned} A &= \mu A_0 + \nu A_1, \\ B &= \mu B_0 + \nu B_1, \\ C &= \mu C_0 + \nu C_1. \end{aligned} \tag{2}$$

Здесь (A_0, B_0, C_0) и (A_1, B_1, C_1) — произвольные непропорциональные тройки чисел, а μ и ν — параметры, одновременно не равные нулю.

Определение 2. Семейство прямых, задаваемое уравнениями вида (2), называется *пучком.*

Предположим сначала, что в уравнениях (2) хотя бы один из коэффициентов A_0, B_0 отличен от нуля и хотя бы один из коэффициентов A_1, B_1 также отличен от нуля. Тогда определены прямые

$$\begin{aligned} A_0 x + B_0 y + C_0 &= 0, \\ A_1 x + B_1 y + C_1 &= 0. \end{aligned} \tag{3}$$

Говорят, что пучок (2) *проходит* через эти прямые (или им *инцидентен*).

Обозначив левые части уравнений (3) символами f и g, мы можем сказать, что пучок (2) состоит из всех прямых, имеющих уравнения вида

$$\mu f + \nu g = 0.$$

Поскольку тройки (A_0, B_0, C_0) и (A_1, B_1, C_1), по условию, не пропорциональны, прямые (3) различны.

---

[1]) Действительно, перейдя в параметрических уравнениях прямой $x = x_0 + lt$, $y = y_0 + mt$ к однородным координатам (например, приняв $x = \frac{X_1}{X_0}$, $y = \frac{X_2}{X_0}$) и положив $t = \frac{\nu}{\mu}$, мы получим уравнения

$$\begin{aligned} X_0 &= \mu X_0^{(0)} + \nu X_0^{(1)}, \\ X_1 &= \mu X_1^{(0)} + \nu X_1^{(1)}, \\ X_2 &= \mu X_2^{(0)} + \nu X_2^{(1)}, \end{aligned}$$

лишь обозначениями отличающиеся от уравнений (2).

Определение 3. Если прямые (3) не параллельны, пучок (2) называется *собственным*, а если они параллельны, — то *несобственным*.

Пучок, для которого одна из пар (A_0, B_0) или (A_1, B_1) равна нулю, мы также будем считать *несобственным* (заметим, что обе эти пары равны нулю быть не могут, потому что тогда тройки (A_0, B_0, C_0) и (A_1, B_1, C_1) были бы пропорциональны).

Собственные пучки характеризуются тем, что для них определитель

$$\begin{vmatrix} A_0 & B_0 \\ A_1 & B_1 \end{vmatrix}$$

отличен от нуля, а несобственные — тем, что этот определитель равен нулю.

Пусть пучок (2) — собственный. Поскольку прямые (3) не параллельны, они пересекаются в единственной точке $M_0(x_0, y_0)$, координаты которой удовлетворяют системе двух линейных уравнений

$$\begin{aligned} A_0x_0 + B_0y_0 + C_0 &= 0, \\ A_1x_0 + B_1y_0 + C_1 &= 0. \end{aligned} \tag{4}$$

Определение 4. Точка $M_0(x_0, y_0)$ называется *центром* собственного пучка.

Предложение 1. *Собственный пучок состоит из всех прямых, проходящих через его центр.*

Доказательство. Любая прямая пучка имеет вид

$$\mu(A_0x + B_0y + C_0) + \nu(A_1x + B_1y + C_1) = 0.$$

Так как ввиду соотношений (4) координаты x_0, y_0 удовлетворяют этому уравнению, то следовательно, любая прямая пучка проходит через его центр.

Обратно, пусть

$$Ax + By + C = 0 \tag{5}$$

— произвольная прямая, проходящая через точку $M_0(x_0, y_0)$. Выбрав на этой прямой произвольную точку $M_1(x_1, y_1)$, отличную от точки $M_0(x_0, y_0)$, рассмотрим числа

$$\begin{aligned} \mu_0 &= A_1x_1 + B_1y_1 + C_1, \\ -\nu_0 &= A_0x_1 + B_0y_1 + C_0. \end{aligned}$$

Ясно, что хотя бы одно из этих чисел отлично от нуля. Поэтому определена прямая

$$\mu_0 f + \nu_0 g = 0. \tag{6}$$

Являясь прямой рассматриваемого пучка, эта прямая проходит через точку $M_0(x_0, y_0)$. Кроме того, она проходит и через точку $M_1(x_1, y_1)$, ибо

$$\mu_0(A_0x_1 + B_0y_1 + C_0) + \nu_0(A_1x_1 + B_1y_1 + C_1) = \mu_0(-\nu_0) + \nu_0\mu_0 = 0.$$

Поскольку через две различные точки плоскости проходит только одна прямая, тем самым доказано, что прямая (5) совпадает с прямой (6), т. е. принадлежит нашему пучку.

Тем самым предложение 1 полностью доказано.

Замечание 2. Тот факт, что прямая (5) принадлежит пучку (2), непосредственно вытекает также из теоремы 2 п. 3 § 1 гл. 3. Действительно, так как прямые (3) и (5) проходят через одну и ту ж точку, ранг R матрицы

$$\begin{pmatrix} A_0 & B_0 & C_0 \\ A_1 & B_1 & C_1 \\ A & B & C \end{pmatrix}$$

равен двум. Поэтому (здесь используется теорема о ранге матрицы) ее строки линейно зависимы. С другой стороны, первые две строки этой матрицы, по условию, линейно независимы (не пропорциональны). Это возможно тогда и только тогда, когда последняя строка этой матрицы является линейной комбинацией ее первых строк, т. е. когда прямая (5) имеет вид (6).

Задание. Докажите, что значения параметров μ и ν, дающие прямую (5), определены этой прямой с точностью до пропорциональности однозначно.

Заметим, что

любая точка $M_0(x_0, y_0)$ плоскости является центром (единственного) собственного пучка.

Действительно, достаточно взять две прямые (3), проходящие через точку M_0, и построить по ним пучок (2).

Это простое наблюдение дальше будет играть весьма важную роль.

Пусть теперь пучок (2) — несобственный, т. е. пусть либо прямые (3) параллельны ($A_1 : A_0 = B_1 : B_0$), либо одна из пар (A_0, B_0) или (A_1, B_1) равна нулю (пусть, для определенности, равна нулю пара (A_1, B_1) и, значит, отлична от нуля пара (A_0, B_0)). Рассмотрим отношение $A_1 : A_0$ (при $A_0 = 0$ — отношение $B_1 : B_0$). Если числа μ и ν обладают тем свойством, что $\mu : \nu = -A_1 : A_0$, то

$$\mu A_0 + \nu A_1 = 0,$$
$$\mu B_0 + \nu B_1 = 0,$$

и потому при этих μ и ν мы не получаем никакой прямой. Напротив, если $\mu : \nu \neq -A_1 : A_0$, то, во-первых, хотя бы одно из чисел $\mu A_0 + \nu A_1$ и $\mu B_0 + \nu B_1$ отлично от нуля, и, во-вторых, имеет место пропорция

$$\frac{\mu A_0 + \nu A_1}{A_0} = \frac{\mu B_0 + \nu B_1}{B_0}.$$

Следовательно, при этих μ и ν мы получаем прямую, параллельную прямой

$$A_0 x + B_0 y + C_0 = 0. \qquad (7)$$

Обратно, пусть

$$Ax + By + C = 0 \qquad (8)$$

— произвольная прямая, параллельная прямой (7). Выбрав на ней произвольную точку $M_1(x_1, y_1)$, рассмотрим числа

$$\mu_0 = A_1x_1 + B_1y_1 + C_1,$$
$$-\nu_0 = A_0x_1 + B_0y_1 + C_0$$

(см. выше). Хотя бы одно из этих чисел отлично от нуля и потому определена прямая

$$\mu_0 f + \nu_0 g = 0.$$

Эта прямая параллельна прямой (7) (поскольку она принадлежит рассматриваемому пучку) и проходит через точку $M_1(x_1, y_1)$. Следовательно, она совпадает с прямой (8).

Тем самым доказано следующее

Предложение 2. *Несобственный пучок состоит из всех прямых, параллельных некоторой фиксированной прямой.*

При этом ясно, что

совокупность всех прямых, параллельных произвольной фиксированной прямой, является несобственным пучком.

Таким образом, термин «несобственный пучок» является синонимом термина «направление» (см. сноску на стр. 17).

2. Расширенная плоскость

Как было объяснено в п. 6 § 1 гл. 2, удобно считать, что кроме обыкновенных («собственных») точек каждая прямая содержит некоторую фиктивную «несобственную» точку, лежащую, так сказать, «в бесконечности». Мы теперь дополнительно условимся, что несобственные точки параллельных прямых совпадают (а непараллельных — различны). Это согласуется с интуитивным представлением параллельных прямых как прямых, «пересекающихся в бесконечности».

Таким образом, все прямые каждого несобственного пучка будут иметь теперь некоторую общую несобственную точку — его *центр*, а сам пучок будет состоять из всех прямых, проходящих через эту точку. Другими словами, мы теперь можем сказать, что

любой пучок является множеством всех прямых, проходящих через некоторую фиксированную точку.

Эта точка (центр пучка) — собственная, если пучок — собственный, и несобственная, если пучок — несобственный.

При этом

любая точка (собственная или несобственная) является центром некоторого (единственного) пучка.

Введение несобственных точек позволяет, таким образом, давать единые формулировки, пригодные как для собственных, так и для несобственных пучков.

При этом следует отчетливо понимать, что это усовершенствование носит пока чисто «лингвистический» характер, поскольку несобственные точки на аффинной плоскости не существуют и участвуют лишь в выражениях типа «прямые пересекаются в несобственной точке», а такое выражение является лишь синонимом выражения «прямые параллельны».

Однако можно сделать и следующий шаг: *расширить* нашу плоскость Π и считать ее подмножеством некоторого большего множества — *расширенной плоскости* Π^+. Элементы множества Π^+ мы по-прежнему будем называть *точками* (собственными, если они принадлежат Π, и несобственными — в противном случае). При этом мы будем считать, что нам задано некоторое соответствие, сопоставляющее каждой прямой α на плоскости Π некоторую несобственную точку плоскости Π^+ и обладающее тем свойством, что двум прямым тогда и только тогда сопоставлена одна и та же несобственная точка, когда эти прямые параллельны, и что любая несобственная точка соответствует хотя бы одной прямой. Несобственную точку, сопоставленную прямой, мы будем называть несобственной точкой *этой прямой*. Множество α^+, получающееся из прямой α присоединением ее несобственной точки, мы будем называть *прямой на расширенной плоскости*. Таким образом, любые две (различные) прямые α^+ и β^+ на расширенной плоскости будут обязательно пересекаться в единственной точке (являющейся точкой пересечения прямых α и β, если эти прямые не параллельны, и являющейся их общей несобственной точкой в противном случае).

Переход от аффинной плоскости Π к расширенной плоскости Π^+ означает, что от аффинной геометрии мы переходим к принципиально новой геометрии — *геометрии расширенной плоскости*.

Подчеркнем, что одно только использование несобственных точек еще не означает переход от аффинной геометрии к геометрии расширенной плоскости: если мы используем несобственные точки только как удобное средство для унификации формулировок, мы остаемся в рамках аффинной геометрии. Геометрия расширенной плоскости возникает тогда, когда мы фактически присоединяем к плоскости несобственные точки и считаем их равноправными с собственными точками.

Множество всех несобственных точек расширенной плоскости мы также будем считать прямой на этой плоскости и будем называть его *несобственной прямой*. Это оправдывается тем, что с любой «настоящей» прямой оно пересекается в единственной точке (а именно, в несобственной точке этой прямой). Таким образом, можно сказать, что *расширенная плоскость*

получается из аффинной присоединением одной дополнительной прямой.

Пучком прямых на расширенной плоскости мы будем называть совокупность всех прямых, проходящих через некоторую фиксированную точку — *центр пучка.* Пучок называется *собственным,* если его центр является собственной точкой. Такому пучку соответствует собственный пучок на аффинной плоскости, состоящий из тех же прямых (а точнее, из тех же прямых, но рассматриваемых без их несобственных точек). Пучок называется *несобственным,* если его центр является несобственной точкой. Такой пучок получается из несобственного пучка на аффинной плоскости присоединением несобственной прямой (и, конечно, добавлением к каждой прямой пучка ее несобственной точки). Так же как и на аффинной плоскости,

любые две различные прямые содержатся в единственном пучке.

С другой стороны, на расширенной плоскости

любые два (различных) пучка содержат единственную общую прямую.

Последнее утверждение на аффинной плоскости уже места не имеет.

Основное отличие геометрии расширенной плоскости от аффинной геометрии состоит в том, что, как уже отмечалось,

любые две различные прямые расширенной плоскости пересекаются в единственной точке.

Вместе с тем, так же как в аффинной геометрии плоскости,

через любые две различные точки расширенной плоскости проходит единственная прямая.

Заметим, что хотя последнее утверждение формально совпадает с соответствующим утверждением в аффинной геометрии, его геометрическое содержание шире: оно содержит не только утверждение «через любые две различные точки аффинной плоскости проходит единственная прямая», но и утверждение «через любую точку аффинной плоскости проходит единственная прямая, параллельная данной прямой». Более того, оно не исчерпывается даже этими двумя утверждениями, поскольку в нем не исключен случай, когда обе данные точки несобственные — случай, «аффинными формулировками» не охватываемый. Все же, если отвлечься от такого рода исключительных случаев, можно сказать, что каждое утверждение геометрии расширенной плоскости эквивалентно нескольким (вообще говоря, многим) утверждениям аффинной геометрии. Внутреннее родство этих утверждений, проявляющееся в геометрии расширенной плоскости, остается в аффинной геометрии скрытым.

Резюмируя, мы можем сказать, что по сравнению с аффинной геометрией геометрия расширенной плоскости логически бо-

лее стройна: ее утверждения имеют более общий характер и не сопровождаются исключениями.

Сказанное выше является содержательным (не формальным) обсуждением понятия «расширенная плоскость». Формально-аксиоматически, это понятие вводится следующим образом.

Определение 1. Множество $\mathfrak{M}^+$ называется *расширенной плоскостью*, если

1) оно содержит некоторую аффинную плоскость $\mathfrak{M}$ (в смысле п. 3 § 3 гл. 2);

2) каждой прямой α в плоскости $\mathfrak{M}$ (см. замечание 1 в п. 1 § 1 гл. 3) сопоставлен некоторый элемент

$$\alpha_\infty \in \mathfrak{M}^+ \setminus \mathfrak{M};$$

3) элементы α_∞ и β_∞, сопоставленные двум прямым α и β, тогда и только тогда совпадают, когда прямые α и β параллельны;

4) любой элемент из $\mathfrak{M}^+ \setminus \mathfrak{M}$ имеет вид α_∞ для некоторой прямой α.

Элементы множества $\mathfrak{M}^+ \setminus \mathfrak{M}$ называются *несобственными точками* плоскости $\mathfrak{M}^+$ (или плоскости $\mathfrak{M}$). Несобственная точка α_∞ называется при этом *несобственной точкой прямой* α (говорят также, что точка α_∞ *инцидентна* прямой α). Точки плоскости $\mathfrak{M}$ называются *собственными точками плоскости* $\mathfrak{M}^+$.

Замечание 1 (очень важное!). Данное определение полностью пригодно и в случае, когда $\mathfrak{M}$ является плоскостью над произвольным полем $\mathbb{K}$. В этом случае $\mathfrak{M}^+$ называется расширенной плоскостью *над полем* $\mathbb{K}$. В частности, при $\mathbb{K} = \mathbb{C}$ мы получаем *комплексную расширенную плоскость*. Если $\mathfrak{M}$ является вещественно-комплексной плоскостью (см. п. 5 § 3 гл. 2), то $\mathfrak{M}^+$ называется *вещественно-комплексной расширенной плоскостью.*

Более того, в определении 1 нет нужды обязательно считать плоскость $\mathfrak{M}$ аффинной, она вполне может быть евклидовой. В этом случае мы получаем *евклидову расширенную плоскость* (вещественную, вещественно-комплексную или плоскость над любым полем $\mathbb{K}$).

Аффинная расширенная плоскость называется часто *аффинно-проективной плоскостью*, а евклидова расширенная плоскость — *евклидово-проективной плоскостью.*

Таким образом, мы получаем целый букет различных плоскостей (который в дальнейшем мы еще пополним), и следовательно, различных «геометрий». Относительно каждой геометрической конструкции или теоремы (безразлично, в содержательном или формально-аксиоматическом построении) следует отчетливо представлять себе, к какой собственно геометрии она относится, т. е. какой «плоскости» выполняется. Соответствующие указания мы будем часто опускать, поскольку

принадлежность теоремы или конструкции к той или иной геометрии, как правило, ясна из контекста. Например, если упоминаются несобственные точки, то мы находимся в расширенной плоскости, а если упоминаются расстояния или углы — то в евклидовой[1]). В каждом случае, когда возможны недоразумения, мы, конечно, будем явно указывать, какой именно геометрией мы в данный момент занимаемся.

3. Полнота и непротиворечивость аксиом геометрии расширенной плоскости

Определение 1. Отображение

$$\eta^+\colon \mathfrak{M}_1^+ \to \mathfrak{M}_2^+$$

одной расширенной плоскости на другую называется *изоморфизмом*, если

а) отображение η^+ переводит собственные точки в собственные, а несобственные — в несобственные;

б) получающееся отображение

$$\eta\colon \mathfrak{M}_1 \to \mathfrak{M}_2$$

нерасширенных плоскостей является изоморфизмом;

в) отображение η^+ сохраняет отношение «несобственная точка инцидентна собственной прямой», т. е. для любой прямой α плоскости $\mathfrak{M}_1$ ее несобственная точка α_∞ переходит при отображении η^+ в несобственную точку β_∞ прямой β плоскости $\mathfrak{M}_2$, являющейся образом при изоморфизме η прямой α:

$$\begin{array}{ccc} \alpha & \to & \alpha_\infty \\ {\scriptstyle\eta}\downarrow & & \downarrow{\scriptstyle\eta^+} \\ \beta & \to & \beta_\infty \end{array}$$

Замечание 1. На самом деле здесь дано не одно определение изоморфизма, а несколько (в зависимости от того, какими считаются нерасширенные плоскости — аффинными или евклидовыми, вещественными или комплексными и т. п.). Формально это находит отражение в возможности различного понимания термина «изоморфизм» в условии б).

Согласно определению 1 каждый изоморфизм

$$\eta^+\colon \mathfrak{M}_1^+ \to \mathfrak{M}_2^+$$

индуцирует некоторый изоморфизм

$$\eta\colon \mathfrak{M}_1 \to \mathfrak{M}_2.$$

[1]) Следует иметь в виду, что на евклидово-проективной плоскости углы и расстояния определяются только для собственных элементов (например, об угле между собственной и несобственной прямой говорить бессмысленно).

Оказывается, что это устанавливает биективное соответствие между изоморфизмами, т. е.

любой изоморфизм

$$\eta\colon \mathfrak{M}_1 \to \mathfrak{M}_2$$

индуцируется некоторым (единственным) изоморфизмом

$$\eta^+\colon \mathfrak{M}_1^+ \to \mathfrak{M}_2^+.$$

Действительно, пусть α — произвольная прямая плоскости $\mathfrak{M}_1$ и α_∞ — ее несобственная точка. Изоморфизм η переводит прямую α в некоторую прямую β плоскости $\mathfrak{M}_1$. Пусть β_∞ — несобственная точка прямой β. Мы положим, по определению,

$$\eta^+(\alpha_\infty) = \beta_\infty.$$

Очевидно, что это определение корректно (точка β_∞ зависит только от точки α_∞, но не от прямой α). Для собственных точек $M \in \mathfrak{M}_1$ мы, естественно, положим

$$\eta^+(M) = \eta(M).$$

Ясно, что построенное отображение

$$\eta^+\colon \mathfrak{M}_1^+ \to \mathfrak{M}_2^+$$

и является искомым изоморфизмом.

Мы будем говорить, что изоморфизм η^+ является *продолжением* изоморфизма η.

Поскольку любые две нерасширенные плоскости изоморфны, из доказанного утверждения непосредственно вытекает, что

любые две расширенные плоскости изоморфны.

Другими словами, аксиоматика геометрии расширенной плоскости полна.

Чтобы доказать непротиворечивость этой аксиоматики, нам нужно построить некоторую ее модель, т. е. проинтерпретировать ее в рамках другой аксиоматики, непротиворечивость которой уже известна. Такая модель не только докажет непротиворечивость, но и даст более наглядное представление о (несколько абстрактной) геометрии расширенной плоскости. Поэтому мы укажем не одну, а четыре различных модели, по-разному интерпретирующие расширенную плоскость.

Пусть $\mathfrak{M}'$ — произвольная нерасширенная плоскость (скажем, для определенности, аффинная и вещественная). Рассмотрим множество $\mathfrak{M}^+$ *всех пучков прямых* на плоскости $\mathfrak{M}'$ и его подмножество $\mathfrak{M}$, состоящее из собственных пучков. Как мы знаем, соответствие

«собственный пучок» $\longmapsto$ «его центр» (1)

является биективным соответствием между всевозможными собственными пучками и точками плоскости $\mathfrak{M}'$. Поэтому это соответствие позволяет определить множество $\mathfrak{M}$ как аффинную плоскость. «Точками» этой плоскости являются собственные пучки плоскости $\mathfrak{M}'$, а, например, «прямыми» — семейства пучков, обладающие тем свойством, что их центры лежат на одной прямой плоскости $\mathfrak{M}'$. Можно сказать, что плоскость $\mathfrak{M}'$ мы отождествляем с множеством $\mathfrak{M}$.

В силу этого отождествления для множества $\mathfrak{M}^+$ оказывается выполненным условие 1) определения 1 п. 2.

Пусть теперь α — некоторая прямая на плоскости $\mathfrak{M}$. Как только что было сказано, эта прямая состоит из всех пучков, центры которых принадлежат некоторой прямой α' плоскости $\mathfrak{M}'$. Пусть α_∞ — несобственный пучок прямых плоскости $\mathfrak{M}'$, состоящий из всех прямых, параллельных прямой α'. Тем самым мы построили отображение $\alpha \longmapsto \alpha_\infty$, предусмотренное условием 2) определения 1.

Ясно, что это отображение обладает свойствами 3) и 4). Таким образом,

множество $\mathfrak{M}^+$ всех пучков аффинной плоскости $\mathfrak{M}'$ является аффинной расширенной плоскостью.

Если мы за $\mathfrak{M}'$ возьмем плоскость над полем $\mathbb{K}$, то $\mathfrak{M}^+$ будет расширенной плоскостью над полем $\mathbb{K}$, а если $\mathfrak{M}'$ будет евклидовой плоскостью, то $\mathfrak{M}^+$ будет евклидовой расширенной плоскостью.

Построенная «пучковая» модель позволяет любое утверждение геометрии расширенной плоскости истолковать как некоторое утверждение (плоской) аффинной геометрии. Однако, как правило, это истолкование не очень наглядно. Более наглядную интерпретацию мы получим, «выйдя в пространство». Для простоты мы станем при этом на содержательную, наглядно геометрическую, точку зрения (и, в частности, ограничимся интерпретацией вещественной расширенной плоскости). Более формальное построение читатель, без сомнения, может провести самостоятельно.

Пусть в пространстве задана некоторая аффинная координатная система $Oxyz$. Рассмотрим координатную плоскость Oxy и координаты x, y на этой плоскости. Каждой прямой

$$Ax + By + C = 0$$

на плоскости Oxy мы сопоставим вектор в пространстве, имеющий координаты A, B, C. Тем самым каждому вектору $\boldsymbol{a}(A, B, C)$, для которого $A \neq 0$ или $B \neq 0$, т. е. каждому вектору, не параллельному оси Oz, будет соответствовать некоторая прямая на плоскости, а каждой прямой будет соответствовать семейство отличных от нуля коллинеарных векторов (не параллельных оси Oz), т. е., другими словами, семейство от-

личных от нуля векторов, параллельных некоторой прямой (которую, для определенности, мы можем считать проходящей через точку O).

Таким образом, мы получаем биективное соответствие между прямыми на плоскости и прямыми в пространстве, проходящими через точку O и отличными от оси Oz. В этом соответствии прямой $(A : B : C)$ отвечает прямая

$$x = At,$$
$$y = Bt,$$
$$z = Ct.$$

Пусть теперь

$$A = \mu A_0 + \nu A_1,$$
$$B = \mu B_0 + \nu B_1,$$
$$C = \mu C_0 + \nu C_1$$

— произвольный пучок прямых на плоскости Oxy. Прямым этого пучка отвечают отличные от нуля векторы $\boldsymbol{a}(A, B, C)$, имеющие вид

$$\boldsymbol{a} = \mu \boldsymbol{a}_0 + \nu \boldsymbol{a}_1, \qquad (1)$$

где $\boldsymbol{a}_0(A_0, B_0, C_0)$ и $\boldsymbol{a}_1(A_1, B_1, C_1)$, т. е. отличные от нуля векторы, компланарные векторам $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$. Поскольку векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$, по условию, линейно независимы, векторы (1) могут быть охарактеризованы как отличные от нуля векторы, параллельные некоторой плоскости (проходящей через точку O). Тем самым мы получили биективное соответствие между пучками прямых на плоскости Oxy и всевозможными плоскостями в пространстве, проходящими через точку O.

Рассматриваемый пучок — собственный тогда и только тогда, когда среди векторов (1) не содержится ни одного вектора, параллельного оси Oz, т. е. когда соответствующая плоскость не параллельна этой оси.

Комбинируя построенное биективное соответствие

«пучок» $\longmapsto$ «плоскость»

с биективным соответствием

«собственный пучок» $\longmapsto$ «его центр»,

мы получим биективное соответствие между точками плоскости Oxy и плоскостями в пространстве, проходящими через точку O и не параллельными оси Oz.

Следовательно, объявив последние плоскости «точками», мы снова получим модель аффинной плоскости. В этой модели «прямыми» являются множества плоскостей, проходящих через фиксированную прямую пространства (проходящую через точку O и отличную от оси Oz).

Как получить из этой модели модель расширенной плоскости, теперь уже совершенно ясно: нужно откинуть условие, что

плоскости не параллельны оси Oz. Таким образом, «точками» этой модели являются всевозможные плоскости пространства, проходящие через точку O, а «прямыми» — множества таких плоскостей, проходящих через фиксированную прямую пространства (проходящую через точку O, но уже может быть совпадающую с осью Oz). Несобственными точками этой модели являются плоскости, проходящие через ось Oz.

«Прямые» построенной модели естественным образом отождествляются с проходящими через точку O прямыми пространства. В частности, несобственная прямая отождествляется с осью Oz.

Построенная модель позволяет любое утверждение геометрии расширенной плоскости переформулировать в виде некоторого утверждения о прямых и плоскостях в пространстве. Например, утверждение, что через любые две различные точки расширенной плоскости проходит единственная прямая, переходит при таком переформулировании в утверждение, что любые две различные плоскости, проходящие через точку O, пересекаются по единственной прямой.

Эта модель нагляднее предыдущей «пучковой» модели. Однако, последняя модель обладает свойством естественности, т. е. лежащее в ее основе соответствие «пучок»$\longmapsto$«точка» строится без всякого произвола. В то же время соответствие «пучок»$\longmapsto$«плоскость», на котором основана рассматриваемая сейчас модель, существенно зависит от выбора системы координат $Oxyz$ и естественным инвариантным способом построено быть не может.

4. Координаты на расширенной плоскости

Еще одну полезную модель расширенной плоскости мы получим, привлекая вещественные числа. Построение такой «числовой модели» с содержательной точки зрения равносильно введению в расширенную плоскость координат. Поэтому мы в первую очередь рассмотрим (пока с содержательных позиций), как в геометрии расширенной плоскости могут быть введены координаты.

Ясно, что распространить на несобственные точки аффинные координаты x, y невозможно, поскольку каждая пара (x, y) определяет некоторую собственную точку и для несобственных точек «не остается координат».

По-иному дело обстоит с однородными координатами, поскольку для них имеются системы значений, не отвечающие никакой собственной точке, и можно надеяться, что эти «запрещенные» значения можно сопоставить несобственным точкам.

Однородные координаты в расширенной плоскости проще всего ввести, воспользовавшись установленным выше биективным соответствием между точками расширенной плоскости и

плоскостями в пространстве, проходящими через начало координат O.

Уравнение произвольной такой плоскости имеет вид

$$Xx + Yy + Zz = 0,$$

где хотя бы один из коэффициентов X, Y, Z отличен от нуля. При этом плоскость однозначно определяется этими коэффициентами и, обратно, однозначно с точностью до пропорциональности их определяет.

Следовательно, комбинируя биективные соответствия

«точка расширенной плоскости» $\longmapsto$ «плоскость в пространстве»

и

«плоскость в пространстве» $\longmapsto$ «класс $(X : Y : Z)$ пропорциональности коэффициентов ее уравнения»,

мы получим биективное соответствие между точками расширенной плоскости и классами $(X : Y : Z)$ пропорциональных троек $(X, Y, Z) \neq (0, 0, 0)$.

Числа X, Y, Z мы и будем называть (*однородными*) *координатами* на расширенной плоскости. Они определены с точностью до пропорциональности.

Чтобы найти для этих координат явные формулы, мы должны выразить в координатах построенное соответствие.

Пусть $M_0(x_0, y_0)$ — произвольная собственная точка плоскости. Любая прямая, проходящая через точку $M_0(x_0, y_0)$, имеет уравнение вида

$$A(x - x_0) + B(y - y_0) = 0,$$

где A и B — некоторые числа, одновременно не равные нулю (а в остальном — произвольные). Это показывает, что прямым собственного пучка с центром в точке $M_0(x_0, y_0)$ отвечают в пространстве векторы вида

$$(A, B, -Ax_0 - By_0).$$

Но ясно, что вектор в пространстве тогда и только тогда имеет такой вид, когда он параллелен плоскости

$$x_0x + y_0y + z = 0. \qquad (1)$$

Таким образом,

Собственная точка $M_0(x_0, y_0)$ изображается в пространстве плоскостью (1).

По определению, это означает, что

точка $M_0(x_0, y_0)$ имеет однородные координаты

$$X_0 : Y_0 : Z_0 = x_0 : y_0 : 1.$$

Замечание 1. Таким образом, построенные координаты $X_0 : Y_0 : Z_0$ являются (с точностью до перестановки) *однородными аффинными координатами* в смысле п. 6 § 1 гл. 2.

Рассмотрим теперь произвольную несобственную точку. Пусть эта точка является несобственной точкой прямой

$$A_0x + B_0y + C_0 = 0. \tag{2}$$

Тогда все прямые соответствующего несобственного пучка будут иметь вид

$$A_0x + B_0y + C = 0,$$

где C — произвольное число, т. е. этим прямым будут отвечать в пространстве векторы вида

$$(A_0,\ B_0,\ C)$$

(и все векторы, им коллинеарные). Но ясно, что вектор в пространстве тогда и только тогда имеет такой вид, когда он параллелен плоскости

$$-B_0x + A_0y = 0. \tag{3}$$

Таким образом,

несобственная точка прямой (2) *изображается в пространстве плоскостью* (3).

По определению, это означает, что

несобственная точка прямой (2) *имеет однородные координаты*

$$X_0 : Y_0 : Z_0 = -B_0 : A_0 : 0.$$

Мы видим, в частности, что координаты $X : Y : Z$ собственных точек характеризуются условием $Z \neq 0$, а координаты несобственных точек — условием $Z = 0$. Другими словами,

уравнение

$$Z = 0$$

является в координатах $X : Y : Z$ *уравнением несобственной прямой.*

Аффинные координаты собственных точек выражаются через координаты $X : Y : Z$ по формулам

$$x = \frac{X}{Z},$$

$$y = \frac{Y}{Z}.$$

Подставив эти выражения в уравнение

$$Ax + By + C = 0 \tag{4}$$

произвольной прямой, мы получим (после умножения на Z) уравнение

$$AX + BY + CZ = 0. \tag{5}$$

Координаты $X : Y : Z$ собственной точки $M(x, y)$ тогда и только тогда удовлетворяют этому уравнению, когда эта точка принадлежит прямой (4). Что же касается несобственных точек, то,

полагая в уравнении (5) $Z = 0$, мы для координат X и Y получим пропорцию

$$X : Y = -B : A.$$

Следовательно, уравнению (5) удовлетворяют координаты единственной несобственной точки, а именно, несобственной точки $(-B:A:0)$ прямой (4).

Тем самым доказано, что

уравнение (5) *выражает прямую* (4), *пополненную несобственной точкой.*

Резюмируя, мы получаем следующее

Предложение 1. *Любое уравнение вида*

$$AX + BY + CZ = 0$$

(*хотя бы один из коэффициентов* A, B, C *которого отличен от нуля*) *выражает в координатах* $X : Y : Z$ *некоторую прямую расширенной плоскости, собственную, если* $A \neq 0$ *или* $B \neq 0$, *и несобственную, если* $A = B = 0$.

Теперь ясно, как построить числовую модель расширенной (аффинно-проективной) плоскости. *Точками* этой модели являются классы $(X : Y : Z)$, т. е. элементы множества $\mathbb{R}P^2$ (см. п. 6 § 1 гл. 2). Точка называется *несобственной*, если $Z = 0$. Формулы

$$x = \frac{X}{Z}, \quad y = \frac{Y}{Z}$$

определяют естественное биективное соответствие между множеством всех собственных точек из $\mathbb{R}P^2$ и стандартной аффинной плоскостью $\mathbb{R}^2$ пар (x, y). Тем самым это множество определяется как аффинная плоскость. Прямыми на этой плоскости являются множества собственных точек $(X : Y : Z)$, удовлетворяющих уравнениям вида

$$AX + BY + CZ = 0,$$

где либо $A \neq 0$, либо $B \neq 0$. Ясно, что все условия определения 1 из п. 2 выполнены, так что $\mathbb{R}P^2$ действительно является расширенной плоскостью.

Определение 1. Построенная плоскость $\mathbb{R}P^2$ называется *стандартной* (или *арифметической*) *аффинно-проективной плоскостью.*

Замечание 2. Ясно, что все сказанное выше сохраняется над любым полем $\mathbb{K}$ (конечно, вместо $\mathbb{R}P^2$ нужно рассматривать множество $\mathbb{K}P^2$ классов пропорциональных троек $(X, Y, Z) \neq (0, 0, 0)$ элементов поля $\mathbb{K}$).

Различных систем однородных аффинных координат $X : Y : Z$ на расширенной плоскости может быть много (а именно, «столько же», сколько имеется различных систем аффинных

координат x, y на нерасширенной плоскости). При этом переход от одних координат $X:Y:Z$ к другим координатам $X':Y':Z'$ описывается, как легко видеть, формулами вида

$$\begin{aligned} X' &= a_1X + a_2Y + a_3Z, \\ Y' &= b_1X + b_2Y + b_3Z, \\ Z' &= Z, \end{aligned} \tag{6}$$

где

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \neq 0$$

(см. формулы перехода от координат x, y к координатам x', y' в п. 2 § 1 гл. 2).

Задание. Докажите формулы (6).

Это наводит на мысль о возможности независимого определения понятия аффинно-проективной плоскости, не опирающегося как определение 1 из п. 2 на понятие аффинной плоскости, а повторяющее (с соответствующими изменениями) определения из § 3 гл. 2.

Введем с этой целью группу $\mathrm{Aff}(\mathbb{R}P^2)$ всех преобразований стандартной плоскости $\mathbb{R}P^2$, выражающихся формулами (6) (обратим внимание на то, что эта группа естественно изоморфна группе $\mathrm{Aff}(\mathbb{R}^2)$), и затем определим *аффинно-проективную плоскость* как произвольное множество $\mathfrak{M}^+$, для которого задано семейство $\mathrm{Coor}(\mathfrak{M}^+)$ биективных отображений

$$\alpha\colon \mathfrak{M}^+ \to \mathbb{R}P^2,$$

удовлетворяющее (по отношению к группе $\mathrm{Aff}(\mathbb{R}P^2)$) известным нам аксиомам 1 и 2 (см. § 3).

Сказанное выше об однородных аффинных координатах показывает, что любая расширенная (аффинно-проективная) плоскость в смысле определения 1 из п. 3 является аффинно-проективной плоскостью и в смысле этого нового определения.

Обратно, аффинно-проективная плоскость в новом смысле будет, очевидно, аффинно-проективной плоскостью в смысле определения 1 п. 3, если мы определим ее несобственные точки как точки, координата Z которых в любой координатной системе $\alpha\colon \mathfrak{M}^+ \to \mathbb{R}P^2$ равна нулю (ясно, что это определение корректно, т. е. не зависит от выбора системы α).

Чтобы получить аналогичное определение *евклидовой расширенной* (в другой терминологии — *евклидово-проективной*) *плоскости*, достаточно потребовать, чтобы для преобразований (6) матрица

$$\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}$$

была ортогональной матрицей (такие преобразования составляют группу $\mathrm{Ort}(\mathbb{R}P^2)$, изоморфную группе $\mathrm{Ort}(\mathbb{R}^2)$). С содержательной точки зрения координатами $X:Y:Z$ на евклидово-проективной плоскости являются *однородные евклидовы координаты*, получающиеся из евклидовых (прямоугольных) координат x, y точно так же, как однородные аффинные координаты получались выше из аффинных.

Поскольку группа $\mathrm{Ort}(\mathbb{R}P^2)$ является подгруппой группы $\mathrm{Aff}(\mathbb{R}P^2)$, мы можем любую евклидово-проективную плоскость единственным образом превратить в аффинно-проективную плоскость; ср. п. 3 § 3 гл. 2, где подробно обсужден аналогичный вопрос для евклидовых и аффинных плоскостей (см., в частности, определение 5 п. 3 § 3 гл. 2).

Аналогично, чтобы получить определение аффинно-проективной плоскости над произвольным полем $\mathbb{K}$, достаточно заменить $\mathbb{R}P^2$ на $\mathbb{K}P^2$ (и, конечно, считать, что все коэффициенты преобразования (6) принадлежат полю $\mathbb{K}$).

Например, *комплексная аффинно-проективная плоскость* получается, когда преобразования (6) имеют комплексные коэффициенты и являются преобразованиями в $\mathbb{C}P^2$.

Вещественно-комплексная аффинно-проективная плоскость получается, когда преобразования (6) с вещественными коэффициентами рассматриваются как преобразования в комплексной стандартной плоскости $\mathbb{C}P^2$.

5. Проективная плоскость

В расширенной плоскости мы сохраняем определенное различие между собственными и несобственными точками. Можно, однако, сделать следующий шаг и считать собственные и несобственные точки абсолютно равноправными и неотличимыми друг от друга. Расширенная плоскость, в которой различие между собственными и несобственными точками игнорируется, называется *проективной плоскостью*, а ее геометрия — *проективной геометрией* (более точное, формальное определение см. ниже).

Чтобы получить наглядную интерпретацию проективной геометрии, мы рассмотрим еще одну модель расширенной (аффинно-проективной) плоскости.

Пусть O — произвольная точка пространства и Π — некоторая плоскость, не проходящая через точку O. Тогда любая точка M плоскости Π будет однозначно определять в пространстве прямую OM, причем соответствие

«точка O» $\longmapsto$ «прямая OM»

будет биективным соответствием между точками плоскости Π и прямыми в пространстве, проходящими через точку O и не параллельными плоскости Π. Тем самым получили некоторую

интерпретацию аффинной плоскости, «точками» которой являются прямые в пространстве.

Ясно, как эту интерпретацию можно распространить до интерпретации всей расширенной плоскости: достаточно включить в рассмотрение и прямые, параллельные плоскости Π (проходящие через точку O), считая, что каждая такая прямая изображает несобственную точку всех прямых плоскости Π, параллельных этой прямой.

В этой интерпретации прямые плоскости Π будут, очевидно, изображаться плоскостями в пространстве (проходящими через точку O) или, точнее, множествами всех прямых (проходящих через точку O), содержащихся в таких плоскостях. Если плоскость пересекает плоскость Π, то она изображает собственную прямую (линию пересечения). Плоскость, параллельная плоскости Π (и проходящая через точку O), изображает несобственную прямую плоскости Π.

Теперь ясно, как от этой интерпретации аффинно-проективной плоскости перейти к интерпретации проективной плоскости: нужно «забыть» о плоскости Π и об особом положении, которое имеют прямые, параллельные этой плоскости.

Построенную таким образом модель проективной плоскости мы будем обозначать символом $\mathfrak{M}_O$. По определению, точками этой модели являются прямые в пространстве, проходящие через точку O, а прямыми — плоскости в пространстве (также проходящие через точку O).

Пусть $Oe_1e_2e_3$ — произвольный аффинный репер в пространстве (с началом в точке O) и пусть A — произвольная точка проективной плоскости $\mathfrak{M}_O$, т. е. прямая в пространстве, проходящая через точку O. Параметрические уравнения этой прямой имеют вид

$$x = Xt,$$
$$y = Yt,$$
$$z = Zt,$$

где X, Y, Z — координаты (в репере $Oe_1e_2e_3$) некоторой точки прямой A (отличной от точки O). Ясно, что числа X, Y, Z однозначно определяют прямую A и, с точностью до пропорциональности, однозначно этой прямой определены.

Определение 1. Числа X, Y, Z мы будем называть *проективными координатами* рассматриваемой точки проективной плоскости $\mathfrak{M}_O$, *ассоциированными* с данным репером $Oe_1e_2e_3$.

Вообще говоря, разные реперы $Oe_1e_2e_3$ и $Oe_{1'}e_{2'}e_{3'}$ могут приводить к одним и тем же координатам $X : Y : Z$. Например, это заведомо так, если реперы *гомотетичны*, т. е. если существует такое число $h \neq 0$, что

$$e_{1'} = he_1, \quad e_{2'} = he_2, \quad e_{3'} = he_3.$$

Покажем, что этим исчерпываются все возможности, т. е. *если для любой точки проективной плоскости $\mathfrak{M}_O$ координаты $X : Y : Z$, ассоциированные с реперами $Oe_1e_2e_3$ и $Oe_{1'}e_{2'}e_{3'}$, совпадают, то эти реперы гомотетичны.*

Действительно, рассмотрим, например, прямую E_1, проходящую через точку O параллельно вектору e_1. Эта точка плоскости $\mathfrak{M}_O$ имеет в координатной системе, ассоциированной с репером $Oe_1e_2e_3$, проективные координаты $1 : 0 : 0$. По условию те же координаты она имеет и в координатной системе, ассоциированной с репером $Oe_{1'}e_{2'}e_{3'}$, а это означает, что вектор $e_{1'}$ параллелен прямой E_1. Следовательно, векторы e_1 и $e_{1'}$ коллинеарны, т. е. существует такое число h_1, что

$$e_{1'} = h_1 e_1.$$

Аналогично доказывается, что

$$e_{2'} = h_2 e_2, \quad e_{3'} = h_3 e_3.$$

Рассмотрим теперь прямую E_0, проходящую через точку O параллельно вектору

$$e_0 = e_1 + e_2 + e_3 = \frac{e_{1'}}{h_1} + \frac{e_{2'}}{h_2} + \frac{e_{3'}}{h_3}.$$

Ассоциированные с репером $Oe_1e_2e_3$ координаты этой точки проективной плоскости $\mathfrak{M}_O$ равны $1 : 1 : 1$, а ассоциированные с репером $Oe_{1'}e_{2'}e_{3'}$ равны $\frac{1}{h_1} : \frac{1}{h_2} : \frac{1}{h_3}$. Так как эти координаты должны совпадать, то, следовательно,

$$h_1 = h_2 = h_3.$$

Таким образом, реперы $Oe_1e_2e_3$ и $Oe_{1'}e_{2'}e_{3'}$ гомотетичны.

Чтобы указать удобные условия гомотетичности реперов, мы произвольный репер $Oe_1e_2e_3$ пополним вектором

$$e_0 = e_1 + e_2 + e_3.$$

Легко видеть, что

реперы $Oe_1e_2e_3$ и $Oe_{1'}e_{2'}e_{3'}$ тогда и только тогда гомотетичны, когда для любого $i = 0, 1, 2, 3$ вектор $e_{i'}$ коллинеарен вектору e_i.

Действительно, если реперы гомотетичны:

$$e_{1'} = he_1, \quad e_{2'} = he_2, \quad e_{3'} = he_3,$$

то

$$e_{0'} = e_{1'} + e_{2'} + e_{3'} = h(e_1 + e_2 + e_3) = he_0,$$

так что для любого $i = 0, 1, 2, 3$ вектор $e_{i'}$ коллинеарен вектору e_i.

Обратно, пусть для любого $i = 0, 1, 2, 3$ вектор $e_{i'}$ коллинеарен вектору e_i, т. е. пусть существуют такие числа h_0, h_1, h_2 и h_3, что

$$e_{0'} = h_0 e_0, \quad e_{1'} = h_1 e_1, \quad e_{2'} = h_2 e_2, \quad e_{3'} = h_3 e_3.$$

Тогда

$$h_0(\boldsymbol{e}_1+\boldsymbol{e}_2+\boldsymbol{e}_3)=h_0\boldsymbol{e}_0=\boldsymbol{e}_{0'}=\boldsymbol{e}_{1'}+\boldsymbol{e}_{2'}+\boldsymbol{e}_{3'}=h_1\boldsymbol{e}_1+h_2\boldsymbol{e}_2+h_3\boldsymbol{e}_3,$$

и потому

$$h_0=h_1=h_2=h_3.$$

Следовательно, реперы $O\boldsymbol{e}_1\boldsymbol{e}_2\boldsymbol{e}_3$ и $O\boldsymbol{e}_{1'}\boldsymbol{e}_{2'}\boldsymbol{e}_{3'}$ гомотетичны.

Ясно, что любые три из четырех векторов

$$\boldsymbol{e}_0=\boldsymbol{e}_1+\boldsymbol{e}_2+\boldsymbol{e}_3,\ \boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \boldsymbol{e}_3$$

линейно независимы. Это означает, что проходящие через точку O прямые E_0, E_1, E_2, E_3 с направляющими векторами $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ составляют в проективной плоскости $\mathfrak{M}_O$ четверку точек, обладающую тем свойством, что никакие три из них не принадлежат одной прямой.

Обратно, пусть в проективной плоскости $\mathfrak{M}_O$ нам дана четверка E_0, E_1, E_2, E_3 точек, никакие три из которых не лежат на одной прямой. Мы всегда можем выбрать направляющие векторы $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ этих прямых так, чтобы имело место равенство

$$\boldsymbol{e}_0=\boldsymbol{e}_1+\boldsymbol{e}_2+\boldsymbol{e}_3.$$

Задание. Докажите возможность такого выбора векторов $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$.

Согласно доказанному выше утверждению, этим условием векторы $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ однозначно определены при данных прямых E_0, E_1, E_2, E_3 с точностью до общего множителя, т. е. репер $O\boldsymbol{e}_1\boldsymbol{e}_2\boldsymbol{e}_3$ однозначно определен с точностью до гомотетии. Следовательно, однозначно определены и ассоциированные проективные координаты.

Определение 2. Четыре точки проективной плоскости, из которых никакие три не лежат на одной прямой, называются *точками общего положения.*

Определение 3. О точках, имеющих в данной проективной координатной системе координаты

$$1:1:1,\quad 1:0:0,\quad 0:1:0,\quad 0:0:1,$$

говорят, что они составляют *проективный координатный репер* этой координатной системы.

В силу этих определений доказанные выше утверждения мы можем суммировать в следующем предложении:

Предложение 1. *Проективная координатная система однозначно определяется своим проективным координатным репером. Точки E_0, E_1, E_2, E_3 тогда и только тогда составляют проективный координатный репер, когда они являются точками общего положения.*

Определение 4. Прямые

$$E_2E_3,\quad E_1E_3,\quad E_1E_2 \tag{1}$$

называются *координатными прямыми* данного репера (или соответствующей проективной координатной системы), а точка E_0 — его *единичной точкой*. Тройка прямых (1) называется *координатным треугольником* репера (проективной координатной системы). Поскольку этот треугольник однозначно определяет точки E_1, E_2, E_3,

проективная координатная система однозначно определяется ее координатным треугольником и единичной точкой.

Поскольку прямая E_2E_3 изображается в пространстве плоскостью, проходящей через прямые E_2 и E_3 с направляющими векторами $\boldsymbol{e}_2$ и $\boldsymbol{e}_3$, то для всех точек этой плоскости их абсцисса (в репере $Oe_1e_2e_3$) равна нулю. Это показывает, что

в проективных координатах $X : Y : Z$ с реперем E_0, E_1, E_2, E_3 точки координатной прямой E_2E_3 характеризуются условием

$$X = 0.$$

Аналогично показывается, что

точки координатной прямой E_1E_3 характеризуются условием

$$Y = 0,$$

а точки координатной прямой E_1E_2 — условием

$$Z = 0.$$

Это позволяет переформулировать предложение 1 в следующем виде:

Предложение 2. *Проективная координатная система однозначно определяется координатными прямыми $X = 0$, $Y = 0$, $Z = 0$ и единичной точкой $E_0(1 : 1 : 1)$. Чтобы тройка прямых была тройкой координатных прямых некоторой проективной координатной системы, необходимо и достаточно, чтобы эти прямые не проходили через одну точку. Единичной точкой может служить любая точка, не принадлежащая этим прямым.*

Рассмотрим теперь вопрос о формулах перехода от одной проективной координатной системы к другой.

Поскольку проективные координаты (на проективной плоскости $\mathfrak{M}_O$) являются не чем иным, как аффинными координатами в пространстве, переход от одних проективных координат к другим сводится к переходу от одних аффинных координат в пространстве к другим аффинным координатам (с тем же началом O), т. е. к переходу, описывать который мы умеем (см. п. 2 § 1 гл. 2).

Таким образом,

переход от одних проективных координат $X : Y : Z$ к другим проективным координатам $X' : Y' : Z'$ описывается формулами вида

$$\begin{aligned} \rho X' &= a_1X + a_2Y + a_3Z, \\ \rho Y' &= b_1X + b_2Y + b_3Z, \\ \rho Z' &= c_1X + c_2Y + c_3Z, \end{aligned} \tag{2}$$

где

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \neq 0.$$

Здесь ρ — произвольный множитель (подчеркивающий, что проективные координаты определены только с точностью до пропорциональности).

Сказанное выше представляет собой содержательное обсуждение понятий проективной плоскости и проективной координатной системы. Теперь мы можем (по уже известному образцу) дать этим понятиям и формально-аксиоматическое определение.

С этой целью мы введем в рассмотрение всевозможные преобразования множества $\mathbb{R}P^2$, задаваемые формулами вида (2). Ясно, что

совокупность $\mathrm{Proj}(\mathbb{R}P^2)$ *всех таких преобразований образует группу.*

Определение 5. Множество $\mathfrak{M}$ называется *проективной плоскостью*, если задано семейство $\mathrm{Coor}(\mathfrak{M})$ биективных отображений

$$\alpha\colon \mathfrak{M} \to \mathbb{R}P^2$$

(называемых *проективными координатными системами*) и выполнены следующие (уже привычные для нас) аксиомы:

Аксиома 1. *Если* $\alpha \in \mathrm{Coor}(\mathfrak{M})$ *и* $\varphi \in \mathrm{Proj}(\mathbb{R}P^2)$, *то*

$$\varphi \circ \alpha \in \mathrm{Coor}(\mathfrak{M}).$$

Аксиома 2. *Если* α, $\alpha' \in \mathrm{Coor}(\mathfrak{M})$, *то*

$$\alpha' \circ \alpha^{-1} \in \mathrm{Proj}(\mathbb{R}P^2).$$

Примером проективной плоскости (доказывающим непротиворечивость аксиом) является *стандартная* (или *арифметическая*) *проективная плоскость* $\mathbb{R}P^2$, для которой, по определению, $\mathrm{Coor}(\mathbb{R}P^2) = \mathrm{Proj}(\mathbb{R}P^2)$ (а также, конечно, рассмотренная выше плоскость $\mathfrak{M}_0$).

Понятие *изоморфизма* проективных плоскостей вводится точно так же, как, скажем, для евклидовых плоскостей (см. п. 2 § 3 гл. 2), причем все перечисленные в п. 2 § 3 гл. 2 свойства изоморфизмов полностью сохраняются (например, любой изоморфизм является отображением по равенству координат и отображение $\alpha\colon \mathfrak{M} \to \mathbb{R}P^2$ тогда и только тогда является изоморфизмом, когда оно представляет собой проективную координатную систему из $\mathrm{Coor}(\mathfrak{M})$). В частности, мы видим, что

любая проективная плоскость $\mathfrak{M}$ *изоморфна стандартной проективной плоскости* $\mathbb{R}P^2$.

Следовательно, аксиомы 1 и 2 полны.

Замечание 1. Совершенно аналогичным образом определяется проективная плоскость *над любым полем* $\mathbb{K}$, *в частности, комплексная проективная плоскость*. Чтобы получить *вещественно-комплексную проективную плоскость*, следует группу $\mathrm{Proj}(\mathbb{R}P^2)$ рассматривать как группу преобразований множества $\mathbb{C}P^2$.

Конечно, можно, не ограничиваясь случаем $n = 2$, аналогично определить *проективное пространство* (над любым полем).

Покажем, что и с формально-аксиоматической точки зрения проективные плоскости можно рассматривать как расширенные плоскости, в которых игнорировано различие между собственными и несобственными точками.

С этой целью заметим, что группа $\mathrm{Aff}(\mathbb{R}P^2)$ (см. предыдущий пункт) является, очевидно, подгруппой группы $\mathrm{Proj}(\mathbb{R}P^2)$. Поэтому (ср. п. 3 § 3 гл. 2) каждой аффинно-проективной плоскости $\mathfrak{M}^+$ мы можем сопоставить вполне определенную проективную плоскость $\mathfrak{M}_{\text{проект}}$, состоящую из тех же точек, проективными координатными системами которой являются отображения вида $\varphi \circ \alpha$, где α — произвольная однородная аффинная координатная система на плоскости $\mathfrak{M}^+$, а φ — произвольное преобразование из группы $\mathrm{Proj}(\mathbb{R}P^2)$.

Это означает, что проективная геометрия находится по отношению к аффинно-проективной геометрии в положении, сходном с положением аффинной геометрии по отношению к евклидовой (первая является «частью» второй или, что то же самое, вторая «богаче» первой).

Замечание 2. Тот факт, что проективные координатные системы имеют вид $\varphi \circ \alpha$, где $\alpha \in \mathrm{Coor}(\mathfrak{M}^+)$, $\varphi \in \mathrm{Proj}(\mathbb{R}P^2)$, означает, что соответствующие координаты $X' : Y' : Z'$ связаны с некоторыми однородными аффинными координатами $X : Y : Z$ формулами вида (1). Заменив в этих формулах однородные координаты соответствующими неоднородными аффинными координатами x, y, мы получим формулы, только обозначениями отличающиеся от формул (4) п. 6 § 1 гл. 2. Это показывает, что теперешняя концепция «проективных координат» совпадает с концепцией «однородных проективных координат» в смысле п. 6 § 1, гл. 2.

Чтобы, обратно, получить из проективной плоскости $\mathfrak{M} = \mathfrak{M}_{\text{проект}}$ аффинно-проективную плоскость, нужно выбрать некоторую проективную координатную систему

$$\alpha_0 \colon \mathfrak{M} \to \mathbb{R}P^2$$

и принять за однородные аффинные координатные системы отображения вида $\varphi \circ \alpha_0$, где $\varphi \in \mathrm{Aff}(\mathbb{R}P^2)$. Получающаяся аффинно-проективная плоскость $\mathfrak{M}_{\alpha_0}$ зависит от выбора системы α_0.

Задание. Докажите, что

плоскости $\mathfrak{M}_{\alpha_0}$ *и* $\mathfrak{M}_{\alpha_1}$, *отвечающие двум координатным системам* $\alpha_0 \colon \mathfrak{M} \to \mathbb{R}P^2$ *и* $\alpha_1 \colon \mathfrak{M} \to \mathbb{R}P^2$, *тогда и только тогда совпадают, когда*

$$\alpha_1 \circ \alpha_0^{-1} \in \mathrm{Aff}(\mathbb{R}P^2). \tag{3}$$

Пусть теперь λ_0 — прямая на плоскости $\mathfrak{M}$, определяемая уравнением

$$Z = 0 \tag{4}$$

в координатной системе α_0, а λ_1 — прямая, определяемая тем же уравнением, но в координатной системе α_1.

Покажем, что

соотношение (3) *имеет место тогда и только тогда, когда прямые* λ_0 *и* λ_1 *совпадают.*

Действительно, пусть $X : Y : Z$ — координаты в системе α_0, а $X' : Y' : Z'$ — координаты в системе α_1. Если выполнено (3), то эти координаты связаны формулами (6) п. 4, и, в частности, $Z' = Z$. Поэтому $Z' = 0$ тогда и только тогда, когда $Z = 0$.

Обратно, пусть прямые λ_0 и λ_1 совпадают, т. е. $Z = 0$ тогда и только тогда, когда $Z' = 0$. Согласно аксиоме 2, $\alpha_1 \circ \alpha_0^{-1} \in \mathrm{Proj}(\mathbb{R}P^2)$, т. е. координаты $X : Y : Z$ и $X' : Y' : Z'$ связаны формулами (2), и, в частности,

$$Z' = c_1 X + c_2 Y + c_3 Z.$$

Рассмотрим точку, для которой $X : Y : Z = 1 : 0 : 0$. Для этой точки $Z' = c_1$, а, с другой стороны, по условию $Z = 0$. Следовательно, $c_1 = 0$. Рассмотрев точку $X : Y : Z = 0 : 1 : 0$, мы аналогично получим, что $c_2 = 0$. Кроме того, поскольку все наши координаты заданы только с точностью до пропорциональности, мы без ограничения общности можем предполагать, что $c_3 = 1$ (вопрос: почему $c_3 \neq 0$?). Таким образом, формулы (2) совпадают в рассматриваемом случае с формулами (6) п. 4, т. е. имеет место соотношение (3).

Поскольку прямая (4) является в плоскости $\mathfrak{M}_{\alpha_0}$ несобственной прямой, мы видим, таким образом, что

переход от проективной плоскости к аффинно-проективной сводится к выбору некоторой (совершенно произвольной) прямой в качестве несобственной прямой.

Другими словами,

две аффинно-проективные плоскости, определяющие одну и ту же проективную плоскость, отличаются друг от друга только тем, какая прямая является несобственной прямой.

Это и является точным формальным выражением утверждения, что проективная плоскость получается из аффинно-проективной плоскости игнорированием различия между собственными и несобственными точками.

Упражнение. Докажите, что, аналогичным образом, вещественно-комплексные плоскости $\mathfrak{M}$, определяющие одну и ту же комплексную плоскость $\mathfrak{M}^{\text{компл}}$ (см. п. 5 § 3 гл. 2), отличаются только тем, какое подмножество является вещественной плоскостью $\mathfrak{M}^{\text{вещ}}$.

Мы ввели, таким образом, целую иерархию различных плоскостей и соответствующих геометрий. В некотором смысле наиболее общей (наиболее «бедной») является проективная гео-

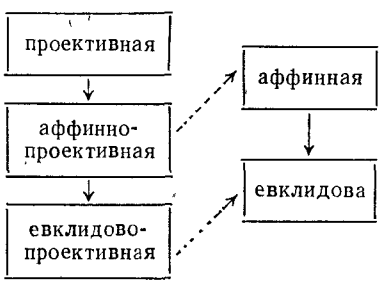
метрия. Чтобы получить более «богатую» аффинно-проективную геометрию, нужно в проективной плоскости фиксировать некоторую прямую. Это отражается в том, что от группы $\mathrm{Proj}(\mathbb{R}P^2)$ мы переходим к «меньшей» группе $\mathrm{Aff}(\mathbb{R}P^2)$.

Чтобы затем получить аффинную геометрию, надо эту фиксированную прямую удалить. Группа $\mathrm{Aff}(\mathbb{R}P^2)$ при этом не меняется (точнее, заменяется изоморфной группой $\mathrm{Aff}(\mathbb{R}^2)$), а геометрия несколько «обедняется» (ее «поле действия» уменьшается).

Переход от аффинной геометрии к евклидовой, связанный с уменьшением группы $\mathrm{Aff}(\mathbb{R}^2)$ до ее подгруппы $\mathrm{Ort}(\mathbb{R}^2)$, снова существенно обогащает геометрию.

Вместо того чтобы переходить от аффинно-проективной геометрии к аффинной, мы можем перейти к евклидово-проективной геометрии, сузив группу $\mathrm{Proj}(\mathbb{R}P^2)$ до ее подгруппы $\mathrm{Ort}(\mathbb{R}P^2)$, и лишь затем получить евклидову геометрию, перейдя к изоморфной группе $\mathrm{Ort}(\mathbb{R}^2)$ (и удалив несобственную прямую).

Эти взаимоотношения между геометриями можно изобразить следующей схемой:



Сплошные стрелки здесь означают переход к меньшей группе, а пунктирные — удаление несобственной прямой.

Конечно, аналогичная схема имеет место для любого поля $\mathbb{K}$, а для случая поля $\mathbb{C}$ даже в двух вариантах — «чисто комплексном» и «вещественно-комплексном».

Кроме того, все эти геометрии можно строить, например, при $n = 1$. В результате мы получаем понятия *проективной прямой*, *аффинно-проективной прямой* и т. д.

Задание. Опишите все соответствующие группы преобразований $(\mathrm{Proj}(\mathbb{R}P^1)$, $\mathrm{Aff}(\mathbb{R}P^1)$ и т. д.).

Конечно, такого рода «одномерные» геометрии весьма просты и потому большого интереса не представляют (см., впрочем, конец п. 4 § 3 гл. 2).

Замечание 3 (очень важное!). Ясно, что любая прямая евклидовой или аффинной плоскости автоматически является евклидовой или, соответственно, аффинной прямой. То же самое верно и для прямых на проективной плоскости. Интересно, что для прямых на аффинно-проективной плоскости это, вообще

говоря, неверно. Действительно, для несобственной прямой у нас нет никакого естественного способа определить ее как аффинно-проективную прямую, т. е. выделить на ней некоторую особую («несобственную») точку. Напротив, собственные прямые аффинно-проективной плоскости естественным образом определяются как аффинно-проективные прямые.

Таким образом, *несобственную прямую аффинно-проективной плоскости мы можем определить только как проективную прямую.* Аналитически это отражается в том, что на ней могут быть заданы однородные координаты, осуществляющие ее изоморфное отображение на стандартную проективную прямую $\mathbb{R}P^1$. Такие координаты мы, например, получим, сопоставив каждой несобственной точке плоскости координаты $l : m$ направляющего вектора произвольной собственной прямой, проходящей через эту несобственную точку.

Как уже отмечалось, нужно всегда отчетливо понимать, в какой именно геометрии справедливо то или иное утверждение. Помощь в этом может оказать тот факт, что для каждой геометрии существуют *предпочитаемые координатные системы*, наиболее приспособленные для ее изучения. Для проективной геометрии это — проективные координатные системы, для аффинно-проективной — однородные аффинные координатные системы и т. д. Поэтому использование, скажем, проективных координат, как правило, указывает, что соответствующее утверждение относится к проективной геометрии.

Конечно, любую геометрию можно изучать и в «чужих» координатах. Однако это всегда приводит к дополнительным осложнениям. Например, формулы евклидовой геометрии в аффинных координатах получаются существенно более сложными, чем в евклидовых (прямоугольных) координатах (появляются метрические коэффициенты базиса) и т. д.

6. Интерпретации проективной геометрии и их применения

Поскольку проективную плоскость можно получить из аффинно-проективной плоскости игнорированием различия между собственными и несобственными точками, построенные в п. 3 модели аффинно-проективной геометрии («пучковая» и «пространственная») являются одновременно и моделями проективной геометрии (только, конечно, в первой модели следует не различать собственные и несобственные пучки, а во второй — не обращать внимания на расположение прямых и плоскостей по отношению к оси Oz).

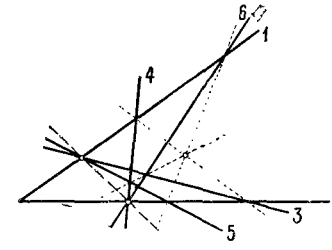
Это позволяет проинтерпретировать любое утверждение проективной геометрии как некоторое высказывание, скажем, о прямых и плоскостях в пространстве и потому применить к его доказательству стереометрические соображения.

Впрочем, на практике более удобно интерпретировать, скажем, прямые на проективной плоскости не прямыми в пространстве, а их направляющими векторами (как, по существу, мы и делали в п. 3), что позволяет применить к доказательству теорем проективной геометрии разработанный аппарат векторного исчисления.

Предупреждение. Следует всегда помнить, что прямой соответствует не один вектор, а целое семейство коллинеарных векторов. Нужно также не забывать, что нулевой вектор не соответствует никакой прямой.

Проиллюстрируем эти общие соображения на примере доказательства знаменитой теоремы Паппа — Брианшона.

Пусть на проективной плоскости даны шесть попарно различных прямых. Будем обозначать эти прямые цифрами **1**, **2**, **3**, **4**, **5**, **6**. Точку пересечения прямых n и m мы обозначим символом nm, а прямую, проходящую через точки nm и n_1m_1 (при условии, что $nm \neq n_1m_1$), — символом $nm \cdot n_1m_1$.

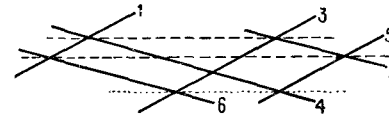


Теорема Паппа — Брианшона. *Если прямые* **1**, **3**, **5** *пересекаются в одной точке* $(13 = 15 = 35)$, *а прямые* **2**, **4**, **6** *также пересекаются в одной точке* $(24 = 26 = 46)$, *причем* $13 \neq 24$ *и ни одна из данных прямых не совпадает с прямой* **13·24**, *то прямые*

$$14 \cdot 23,\ 25 \cdot 16,\ 36 \cdot 45$$

также пересекаются в одной точке.

Замечание 1. Являясь теоремой проективной геометрии, эта теорема справедлива и в аффинно-проективной геометрии. Однако на аффинно-проективной плоскости она распадается на несколько утверждений, в зависимости от того, сколько точек, явно или не явно участвующих в теореме, являются собственными, а сколько — несобственными. При соответствующей переформулировке эти утверждения становятся теоремами аффинной геометрии. Одним из таких частных случаев теоремы Паппа—Брианшона на аффинной плоскости является, например, следующее утверждение: если прямые **1**, **3**, **5** параллельны и прямые **2**, **4**, **6** также параллельны (но не параллельны прямым **1**, **3**, **5**) и если, кроме того, прямые **14·23** и **25·16** параллельны, то прямая **36·45** параллельна прямым **14·23** и **25·16**.



Задание. Составьте полный список утверждений аффинной геометрии, являющихся частными случаями теоремы Паппа — Брианшона (и сделайте соответствующие чертежи).

Для доказательства теоремы Паппа — Брианшона мы переформулируем ее на языке векторного исчисления.

Пусть a_1, a_2, a_3, a_4, a_5, a_6 — векторы, соответствующие прямым **1**, **2**, **3**, **4**, **5**, **6**, а a_7, a_8, a_9 — векторы, соответствующие прямым **14·23**, **25·16**, **36·45**. Каждой прямой, проходящей через точку **14**, соответствует вектор вида $\mu_1 a_1 + \mu_4 a_4$, а каждой прямой, проходящей через точку **23**, — вектор вида $\nu_2 a_2 + \nu_3 a_3$. Следовательно, вектор a_7 обладает тем свойством, что он может быть представлен как в виде $\mu_1 a_1 + \mu_4 a_4$, так и в виде $\nu_2 a_2 + \nu_3 a_3$. Аналогично, вектор a_8 может быть представлен как в виде $\mu_2 a_2 + \mu_5 a_5$, так и в виде $\nu_1 a_1 + \nu_6 a_6$, а вектор a_9 — как в виде $\mu_3 a_3 + \mu_6 a_6$, так и в виде $\nu_4 a_4 + \nu_5 a_5$. Далее, тот факт, что прямые **1**, **3**, **5** проходят через одну точку (принадлежат одному пучку), означает, что векторы a_1, a_3, a_5 компланарны. Аналогично, тот факт, что прямые **2**, **4**, **6** проходят через одну точку, означает, что векторы a_2, a_4, a_6 компланарны. Наконец, условие, что $\mathbf{13} \neq \mathbf{24}$ и что ни одна из прямых **1**, **2**, **3**, **4**, **5**, **6** не совпадает с прямой **13·24**, означает, что ни один из векторов a_1, a_3, a_5 не компланарен векторам a_2, a_4, a_6 и, обратно, ни один из векторов a_2, a_4, a_6 не компланарен векторам a_1, a_3, a_5.

Это означает, что теорема Паппа — Брианшона может быть переформулирована следующим образом:

если

а) *векторы* a_1, a_2, a_3, a_4, a_5, a_6 *попарно не коллинеарны*;

б) *векторы* a_1, a_3, a_5, *а также векторы* a_2, a_4, a_6, *компланарны*;

в) *ни один из векторов* a_1, a_3, a_5 *не компланарен векторам* a_2, a_4, a_6 *и ни один из векторов* a_2, a_4, a_6 *не компланарен векторам* a_1, a_3, a_5,

то векторы a_7, a_8, a_9, *удовлетворяющие соотношениям*

$$\begin{aligned} a_7 &= \mu_1 a_1 + \mu_4 a_4 = \nu_2 a_2 + \nu_3 a_3, \\ a_8 &= \mu_2 a_2 + \mu_5 a_5 = \nu_1 a_1 + \nu_6 a_6, \\ a_9 &= \mu_3 a_3 + \mu_6 a_6 = \nu_4 a_4 + \nu_5 a_5, \end{aligned} \tag{1}$$

компланарны.

В этом виде мы и будем ее доказывать.

Поскольку векторы a_1, a_3, a_5 компланарны, они параллельны некоторой плоскости Π_1. Аналогично, векторы a_2, a_4, a_6 параллельны некоторой плоскости Π_2. Плоскости Π_1 и Π_2 не параллельны, и потому существует единственная прямая, параллельная обеим этим плоскостям. Пусть e_0 — некоторый направляющий вектор этой прямой. Пусть, далее, e_1 — произвольный вектор, параллельный плоскости Π_1, но не коллинеарный вектору e_0, а e_2 — произвольный вектор, параллельный плоскости Π_2 и также не коллинеарный вектору e_0. Из условий, наложенных на векторы $a_1, \ldots, a_6$, немедленно следует, что векторы e_0, e_1, e_2 линейно независимы (составляют базис пространства) и что ни один из векторов $a_1, \ldots, a_6$ не коллинеарен вектору e_0.

При этом векторы a_1, a_3, a_5 выражаются через векторы e_0, e_1, а векторы a_2, a_4, a_6 — через векторы e_0, e_2.

Заметим теперь, что векторы $a_1, \ldots, a_6$ нам заданы лишь с точностью до коллинеарности. Поэтому, без потери общности, мы можем предполагать, что

$$a_1 = k_1 e_0 + e_1, \quad a_2 = k_2 e_0 + e_2,$$
$$a_3 = k_3 e_0 + e_1, \quad a_4 = k_4 e_0 + e_2,$$
$$a_5 = k_5 e_0 + e_1, \quad a_6 = k_6 e_0 + e_2,$$

где k_1, k_2, k_3, k_4, k_5, k_6 — некоторые числа.

Подставив эти выражения в первое из равенств (1), мы немедленно получим, что

$$a_7 = (\mu_1 k_1 + \mu_4 k_4) e_0 + \mu_1 e_1 + \mu_4 e_2 = (\nu_2 k_2 + \nu_3 k_3) e_0 + \nu_3 e_1 + \nu_2 e_2.$$

Так как векторы e_0, e_1, e_2 линейно независимы, это равенство может иметь место только тогда, когда

$$\mu_1 = \nu_3, \quad \mu_4 = \nu_2$$

и

$$\mu_1 k_1 + \mu_4 k_4 = \nu_2 k_2 + \nu_3 k_3.$$

Следовательно,

$$\mu_1 (k_1 - k_3) = \mu_4 (k_2 - k_4).$$

Поскольку вектор a_7 нам задан также лишь с точностью до коллинеарности, мы без ограничения общности можем считать, что $\mu_1 = k_2 - k_4$ (заметим, что $k_2 \neq k_4$, ибо в противном случае векторы a_2 и a_4 совпадают). Тогда $\mu_4 = k_1 - k_3$ и, следовательно,

$$a_7 = [(k_2 - k_4) k_1 + (k_1 - k_3) k_4] e_0 + (k_2 - k_4) e_1 + (k_1 - k_3) e_2 =$$
$$= (k_1 k_2 - k_3 k_4) e_0 + (k_2 - k_4) e_1 + (k_1 - k_3) e_2.$$

Аналогично показывается, что

$$a_8 = (k_1 k_2 - k_5 k_6) e_0 + (k_2 - k_6) e_1 + (k_1 - k_5) e_2,$$
$$a_9 = (k_3 k_4 - k_5 k_6) e_0 + (k_4 - k_6) e_1 + (k_3 - k_5) e_2.$$

Но теперь ясно, что

$$a_7 + a_9 = a_8$$

и, следовательно, векторы a_7, a_8, a_9 компланарны. Тем самым теорема Паппа — Брианшона полностью доказана.

Вспомним теперь, что наряду с «пространственной» интерпретацией, которой мы воспользовались для доказательства теоремы Паппа — Брианшона (и в которой точки изображаются

плоскостями, а прямые — прямыми), у нас есть также и рассмотренная в п. 4 интерпретация $\mathfrak{M}_O$ (в которой точки изображаются прямыми, а прямые — плоскостями), которую мы с равным правом можем использовать для доказательства теорем проективной геометрии. Поэтому одно и то же стереометрическое рассуждение даст нам две (как правило, различные) теоремы проективной геометрии: одну, когда мы воспользуемся первой интерпретацией, и другую, когда мы воспользуемся второй интерпретацией. Чтобы сравнить друг с другом эти теоремы, нам удобно усовершенствовать терминологию (впрочем, этой усовершенствованной терминологией мы уже иногда пользовались, так что читателю она должна быть знакома). Именно, вместо того чтобы говорить «прямая проходит через точку», т. е. что «точка лежит на прямой», мы будем говорить, что прямая и точка *инцидентны*. Аналогично, мы будем говорить, что прямая и плоскость инцидентны, если прямая содержится в плоскости.

Теорему, в которой говорится только о точках и прямых (или прямых и плоскостях) и об отношении инцидентности между ними, мы будем называть *конфигурационной теоремой*.

Обратим теперь внимание на то, что обе построенные выше интерпретации *сохраняют отношение инцидентности*, т. е. если точка и прямая инцидентны на проективной плоскости, то соответствующие прямая и плоскость в пространстве также инцидентны (и наоборот). Поэтому любая верная конфигурационная теорема $\mathscr{A}$ на проективной плоскости переходит при каждой интерпретации в верную конфигурационную теорему в пространстве. Пусть это будет теорема $\mathscr{A}_1$ для первой интерпретации и теорема $\mathscr{A}_2$ для второй интерпретации. Формально переход от теоремы $\mathscr{A}$, скажем, к теореме $\mathscr{A}_1$, состоит в том, что слово «точка» всюду заменяется словом «плоскость». Аналогично, чтобы из теоремы $\mathscr{A}$ получить теорему $\mathscr{A}_2$, нужно слово «точка» заменить словом «прямая», а слово «прямая» — словом «плоскость». Поэтому теоремы $\mathscr{A}_1$ и $\mathscr{A}_2$ (вообще говоря, различные) переходят друг в друга при замене слова «плоскость» словом «прямая», а слова «прямая» словом «плоскость».

Например, если мы примем за $\mathscr{A}$ теорему «любые две различные точки инцидентны единственной прямой», то теоремой $\mathscr{A}_1$ будет высказывание «любые две различные плоскости инцидентны единственной прямой», а теоремой $\mathscr{A}_2$ — высказывание «любые две различные прямые инцидентны единственной плоскости» (напомним, что мы рассматриваем лишь прямые и плоскости пространства, проходящие через фиксированную точку O).

Определение 6. Две конфигурационные теоремы $\mathscr{A}$ и $\mathscr{A}'$ на проективной плоскости называются *двойственными*, если одна из них получается из второй заменой (всюду) слова «точка» словом «прямая» и слова «прямая» — словом «точка».

Из сказанного выше непосредственно вытекает, что

теоремы $\mathcal{A}$ и $\mathcal{A}'$ тогда и только тогда двойственны, когда теорема $\mathcal{A}_1$, отвечающая теореме $\mathcal{A}$ в первой интерпретации, совпадает с теоремой $\mathcal{A}'_2$, отвечающей теореме $\mathcal{A}'$ во второй интерпретации:

$$\mathcal{A}_1 = \mathcal{A}'_2.$$

Но мы знаем, что теоремы $\mathcal{A}$ и $\mathcal{A}_1$ (а также теоремы $\mathcal{A}'$ и $\mathcal{A}'_2$) одновременно верны или неверны. Поэтому справедлив следующий общий принцип:

Принцип двойственности. *Конфигурационная теорема, двойственная верной теореме, также верна.*

Например, верной теореме «любые две различные точки инцидентны единственной прямой» двойственна верная (на проективной плоскости!) теорема «любые две различные прямые инцидентны единственной точке».

Замечание 2. Подчеркнем, что принцип двойственности не является теоремой проективной геометрии, поскольку он говорит не о точках и прямых, а о теоремах. Он принадлежит науке, изучающей теоремы проективной геометрии (эту науку можно, если угодно, назвать *метагеометрией*).

Принцип двойственности позволяет «бесплатно» получать из доказанных теорем новые верные утверждения (правда, иногда теорема может оказаться «самодвойственной» и тогда никакой новой теоремы мы не получим).

Например, теореме Паппа — Брианшона двойственна следующая

Теорема Паппа—Паскаля. *Если из шести точек* **1**, **2**, **3**, **4**, **5**, **6** *точки* **1**, **3**, **5** *расположены на одной прямой и точки* **2**, **4**, **6** *также расположены на одной прямой (отличной от первой), причем ни одна из точек* **1**, ..., **6** *не совпадает с точкой пересечения этих прямых, то точка пересечения прямой* **14** *с прямой* **23**, *точка пересечения прямой* **25** *с прямой* **16** *и точка пересечения прямой* **36** *с прямой* **45** *расположены на одной прямой*[1]).

Чтобы полнее уяснить содержание этой теоремы, мы заметим, что в ней участвуют девять точек: данные точки **1**, **2**, **3**, **4**, **5**, **6** и три точки пересечения (точка **7** пересечения прямых **14** и **23**, точка **8** пересечения прямых **25** и **16**, точка **9** пересечения прямых **36** и **45**), а также девять прямых

13, **14**, **25**, 27, 39, 46, 59, 68, 79.

Обозначив эти прямые (в указанном порядке) цифрами

5, 7, 8, 3, 6, 2, 4, 1, 9,

мы можем теорему Паппа — Паскаля переформулировать следующим образом:

[1]) Обратим внимание на то, что аксиома Паппа аффинной планиметрии (см. дополнение к § 3 гл. 2) является одним из аффинных частных случаев этой теоремы.

(*) *если каждая из точек* **1**—**8** *инцидентна с прямой, обозначенной той же цифрой, то точка* **9** *также инцидентна с прямой* **9**.

Конфигурация, состоящая из точек **1**—**9** и прямых **1**—**9**, называется *конфигурацией Паппа*. В ней каждая прямая инцидентна трем точкам и каждая точка инцидентна трем прямым.

Эта конфигурация обладает замечательным свойством симметрии, состоящим в том, что если мы выберем в ней две тройки коллинеарных точек и будем по ним строить (указанным выше образом) конфигурацию Паппа, то в результате получим ту же конфигурацию. Другими словами, конфигурация Паппа не зависит от того, какие составляющие ее точки мы обозначаем цифрами **1**—**6**. Еще иначе, можно сказать, что *все точки конфигурации Паппа совершенно равноправны.*

Двойственная теорема Паппа — Брианшона аналогичным образом может рассматриваться как теорема, обеспечивающая возможность построения некоторой конфигурации точек и прямых, исходя из двух троек прямых, проходящих через одну точку. Замечательно, что при этом получается т а ж е конфигурация Паппа. Действительно, переформулировав теорему Паппа—Брианшона, подобно тому как мы выше переформулировали теорему Паппа—Паскаля, как легко видеть, получим ту же формулировку (*) (достаточно заметить, что формулировка (*) сама себе двойственна). Таким образом, *обе теоремы Паппа—Паскаля и Паппа — Брианшона выражают в сущности, один и тот же математический факт* — существование конфигурации Паппа; только теорема Паппа — Паскаля обеспечивает возможность построения этой конфигурации, «исходя из точек», а теорема Паппа — Брианшона — «исходя из прямых».

Заметим, что теорема Паппа — Брианшона обеспечивает также полную симметричность конфигурации Паппа и в отношении прямых.

7. Конфигурационная геометрия

При исследовании отношения инцидентности точек и прямых на проективной плоскости целесообразно несколько изменить точку зрения и рассматривать прямые не как множества точек, а как некие новые объекты, связанные с точками отношением инцидентности. Формализуя этот подход, мы приходим к следующему определению:

Определение 1. Пусть $\mathfrak{P}$ — множество элементов двух сортов, одни из которых называются *точками*, а другие — *прямыми*. Предположим, что на множестве $\mathfrak{P}$ задано некоторое отношение, связывающее точки и прямые и называемое отношением *инцидентности*. Такое множество $\mathfrak{P}$ называется *конфигурационной плоскостью*, если выполнены следующие аксиомы:

Инц. 1. *Любые две различные точки инцидентны одной и только одной прямой.*

Инц. 2. *Любые две различные прямые инцидентны одной и только одной точке.*

Инц. 3. *Любая прямая инцидентна по крайней мере трем различным точкам.*

Инц. 4. *Любая точка инцидентна по крайней мере трем различным прямым.*

Замечание 1. Термин «конфигурационная плоскость» мало употребителен. Обычно предпочитают говорить о «проективной плоскости». Однако это для нас неприемлемо, поскольку термин «проективная плоскость» мы уже употребляем в другом смысле (см. определение 1 п. 5). Конечно, каждая проективная плоскость является конфигурационной плоскостью (точнее, множество ее точек и прямых составляют конфигурационную плоскость), но (как показывают простые примеры) обратное неверно.

Трудно ожидать, чтобы из тощих аксиом Инц. 1—4 можно было извлечь интересные следствия. Поэтому обычно к этим аксиомам добавляют еще другие дополнительные аксиомы. Наиболее важной из них является следующая аксиома Паппа:

Инц. 5. *Возможно построение конфигураций Паппа, т. е. выполнены теоремы Паппа — Паскаля и Паппа — Брианшона.*

Определение 2. Конфигурационная плоскость, в которой выполнена аксиома Паппа, называется *папповой плоскостью.*

Поскольку доказательство теоремы Паппа — Брианшона, изложенное в предыдущем пункте, проходит, очевидно, в проективной геометрии над любым полем $\mathbb{K}$, то

проективная плоскость над любым полем $\mathbb{K}$ *является папповой плоскостью.*

Замечательно, что обратное утверждение также справедливо.

Теорема 1. *Любая паппова плоскость изоморфна проективной плоскости над некоторым полем* $\mathbb{K}$.

Термин «изоморфна» понимается здесь в естественном смысле: существует биективное отображение, переводящее точки в точки, прямые в прямые и сохраняющее отношение инцидентности.

Доказательство теоремы 1 в достаточной мере трудно и мы его опустим [1]).

Из теоремы 1 следует, что аксиомы Инц. 1—5 *полны*, в том (новом для нас) смысле, что каждая конфигурационная теорема, верная в проективной геометрии над любым полем $\mathbb{K}$, может быть выведена из этих аксиом.

Это не мешает, конечно, существованию конфигурационных теорем, не вытекающих из аксиом Инц. 1—5. Примером является следующая теорема:

---

[1]) Доказательство можно найти в книжке Р. Хартсхорна, Основы проективной геометрии, Мир, 1970 г. (а также в книжке Э. Артина, Геометрическая алгебра, Наука, 1969 г.).

Теорема Фано. *Для любой четверки точек* **1, 2, 3, 4**, *любые три из которых неколлинеарны, точки*

12·34, 13·24, 14·23

не коллинеарны.

Упражнение. Докажите, что *в проективной геометрии над полем* $\mathbb{K}$ *тогда и только тогда справедлива теорема Фано, когда характеристика поля* $\mathbb{K}$ *отлична от двух.*

Указание: рассмотрите проективную координатную систему, в которой точки 1, 2, 3, 4 имеют координаты (1 : 0 : 0), (0 : 1 : 0), (0 : 0 : 1) и (1 : 1 : 1).

Примером теоремы, вытекающей из аксиом Инц. 1—5, является следующая

Теорема Дезарга. *Пусть* **1, 2, 3, 4, 5, 6** — *такие* (*попарно различные*) *точки, что прямые* **14, 25** *и* **36** *инцидентны одной точке* (*отличной от точек* **1—6**). *Тогда точки*

12·45, 13·46, 23·56

коллинеарны [1]).

Упражнение. Докажите, что теорема Дезарга справедлива в проективной геометрии над любым полем $\mathbb{K}$. Затем выведите ее из аксиом Инц. 1—5, т. е. покажите, что она справедлива в любой папповой плоскости.

Задание. Сформулируйте теорему, обратную к теореме Дезарга, и теорему, двойственную теореме Дезарга. Установите, что обе эти теоремы равносильны теореме Дезарга (обеспечивают возможность построения одной и той же конфигурации).

К вопросу о роли теоремы Дезарга в геометрии плоскости мы еще вернемся в п. 4 § 2.

«Конфигурационная» точка зрения проливает также новый свет на принцип двойственности (см. п. 6). Действительно, очевидно, что аксиомы Инц. 1 — 4 попарно двойственны. Кроме того, аксиома Инц. 5, как мы знаем, сама себе двойственна.

Поэтому, если мы возьмем произвольную конфигурационную теорему, вытекающую из аксиом Инц. 1—5, и в ее выводе всюду поменяем местами слова «точка» и «прямая», то получим правильный вывод двойственной теоремы из двойственных (т. е. тех же) аксиом. Это доказывает, что

в любой папповой плоскости справедлив принцип двойственности.

Поскольку любая проективная плоскость (над произвольным полем $\mathbb{K}$) является папповой плоскостью, отсюда, казалось бы, вытекает принцип двойственности и для проективных плоскостей (и притом над произвольным полем $\mathbb{K}$). Однако это не совсем так: таким способом принцип двойственности обосновывается лишь для теорем, вытекающих из аксиом Инц. 1—5 (т. е. справедливых над любым полем $\mathbb{K}$). Применим ли принцип

[1]) Аффинным частным случаем этой теоремы является аксиома Дезарга из дополнения к § 3 гл. 2.

двойственности, например, к теореме Фано, остается при этом совершенно неясным.

Чтобы «конфигурационными» рассуждениями доказать принцип двойственности в полном объеме (и над любым полем $\mathbb{K}$), можно поступить, например, следующим образом.

Определение 3. *Антиавтоморфизмом* (употребителен также термин *коррелятивное преобразование*) конфигурационной плоскости $\mathfrak{P}$ называется ее биективное отображение на себя (преобразование), переводящее точки в прямые, прямые в точки и сохраняющее отношение инцидентности (т. е. обладающее тем свойством, что точка A и прямая α тогда и только тогда инцидентны, когда прямая, соответствующая точке A, инцидентна точке, соответствующей прямой α).

Ясно, что

если конфигурационная плоскость обладает хотя бы одним антиавтоморфизмом, то на ней выполнен принцип двойственности.

Поэтому для доказательства принципа двойственности в проективной геометрии над полем $\mathbb{K}$ достаточно доказать, что

проективная плоскость над полем $\mathbb{K}$ *обладает хотя бы одним антиавтоморфизмом.*

Но какой-нибудь антиавтоморфизм проективной плоскости строится очень легко: достаточно выбрать на плоскости произвольную проективную координатную систему и сопоставить каждой точке $(X:Y:Z)$ прямую с теми же координатами (т. е. имеющую числа X, Y, Z коэффициентами ее общего уравнения). Ясно, что это соответствие биективно. Покажем, что оно сохраняет отношение инцидентности.

Пусть точка $(X:Y:Z)$ и прямая $(A:B:C)$ инцидентны. Это означает, что

$$AX + BY + CZ = 0. \tag{1}$$

При нашем соответствии точка $(X:Y:Z)$ переходит в прямую $(X:Y:Z)$, а прямая $(A:B:C)$ — в точку $(A:B:C)$. Поскольку инцидентность прямой $(X:Y:Z)$ и точки $(A:B:C)$ выражается тем же уравнением (1), наше утверждение, тем самым, полностью доказано.

Факт существования антиавтоморфизма означает, что

проективную плоскость можно проинтерпретировать в себе так, чтобы точкам соответствовали прямые, а прямым — точки.

В этом смысле геометрия прямых на проективной плоскости полностью эквивалентна геометрии точек.

Именно это обстоятельство и определяет красоту и изящество проективной геометрии (по сравнению с аффинной).

Замечание 2. Теперь мы можем объяснить, почему прямые на аффинной плоскости нельзя описывать неоднородными (всюду определенными) координатами (см. замечание 1 п. 1). Действительно, множество всех прямых аффинной плоскости

получается из множества всех прямых проективной плоскости удалением одной прямой. Поэтому любой антиавтоморфизм проективной плоскости переводит это множество в множество всех точек проективной плоскости, за исключением одной точки. Если бы в множество всех прямых аффинной плоскости можно было ввести неоднородные координаты, то неоднородные координаты можно было бы ввести и в множество точек проективной плоскости, получающееся удалением одной точки. Но это, как мы знаем, сделать нельзя (нужно удалить целую прямую).

Дополнение. **Трилинейные координаты**

Определение 1. Пусть на евклидовой плоскости дана некоторая прямая λ и точка M_0, не принадлежащая прямой λ. Для любой точки M плоскости отношение r расстояния от точки M до прямой λ к расстоянию от M_0 до λ, взятое со знаком «плюс», если точки M и M_0 находятся по одну сторону от λ, и со знаком «минус» в противном случае, мы будем называть *расстоянием точки M до прямой λ по отношению к точке M_0.*

Пусть x, y — прямоугольные координаты на плоскости,

$$Ax + By + C = 0$$

— уравнение прямой λ в этих координатах и (x_0, y_0) — координаты точки M_0. В силу результатов п. 5 § 1 гл. 3 расстояние r произвольной точки $M(x, y)$ до прямой λ по отношению к точке M_0 выражается формулой

$$r = \frac{Ax + By + C}{Ax_0 + By_0 + C}. \qquad (1)$$

Несмотря на то, что в определении расстояния r участвуют метрические понятия, оказывается, что *это расстояние аффинно-инвариантно* (т. е. принадлежит аффинной геометрии).

Упражнение. Покажите, что формула (1) остается верной в любых аффинных координатах x, y, дайте также геометрическое определение величины r в терминах, имеющих аффинно-инвариантный смысл.

Предположим теперь, что на (аффинной) плоскости нам даны три прямые общего положения

$$\lambda_1, \quad \lambda_2, \quad \lambda_3$$

и некоторая точка E, не принадлежащая ни одной из этих прямых. Тогда для любой точки M будут определены три числа:

x_1 — расстояние точки M до прямой λ_1 по отношению к точке E;
x_2 — расстояние точки M до прямой λ_2 по отношению к точке E;
x_3 — расстояние точки M до прямой λ_3 по отношению к точке E.

Определение 2. Числа x_1, x_2, x_3 называются *трилинейными координатами* точки M, определенными прямыми λ_1, λ_2, λ_3 и точкой E.

Основной целью этого дополнения является доказательство следующего предложения:

Предложение 1. *Пусть $\mathfrak{M}^+$ — аффинно-проективная плоскость, полученная расширением аффинной плоскости $\mathfrak{M}$. Рассмотрим на плоскости $\mathfrak{M}^+$ проектив-*

ную координатную систему с репером E_0, E_1, E_2, E_3, состоящим из собственных точек (т. е. из точек плоскости $\mathfrak{M}$). Пусть

$$\lambda_1 = E_2E_3, \quad \lambda_2 = E_1E_3, \quad \lambda_3 = E_1E_2$$

— соответствующие координатные прямые. Тогда для любой собственной точки M ее проективные координаты $X : Y : Z$ пропорциональны ее трилинейным координатам x_1, x_2, x_3, определенным прямыми λ_1, λ_2, λ_3 и точкой E_0.

Короче:

проективные координаты пропорциональны трилинейным.

Это предложение дает вполне удовлетворительное наглядное описание проективных координат любых собственных точек.

Доказательство. Поскольку все аффинно-проективные плоскости изоморфны, мы без ограничения общности можем считать, что плоскостью $\mathfrak{M}^+$ является рассмотренная в начале п. 5 аффинно-проективная плоскость, точками которой являются прямые в пространстве, проходящие через фиксированную точку O. По определению, точка этой плоскости является собственной точкой, если как прямая в пространстве она не параллельна некоторой фиксированной плоскости П, не проходящей через точку O. Следовательно, сопоставив такой прямой ее точку пересечения с плоскостью П, мы можем аффинную плоскость $\mathfrak{M} \subset \mathfrak{M}^+$ отождествить с плоскостью П. В частности, точки E_0, E_1, E_2, E_3 мы можем считать точками плоскости П.

Согласно п. 5 для построения проективной координатной системы с репером E_0, E_1, E_2, E_3 мы должны рассмотреть в пространстве прямые

$$OE_0, \quad OE_1, \quad OE_2, \quad OE_3$$

и найти такие направляющие векторы e_1, e_2, e_3 прямых OE_1, OE_2, OE_3, чтобы вектор

$$e_0 = e_1 + e_2 + e_3$$

был направляющим вектором прямой OE_0. Тогда для любой собственной точки плоскости $\mathfrak{M}^+$ (т. е. некоторой точки M плоскости П) ее проективные координаты $X : Y : Z$ в координатной системе с репером E_0, E_1, E_2, E_3 будут пропорциональны ее аффинным координатам x, y, z (как точки пространства) в реперe $Oe_1e_2e_3$. Таким образом, для доказательства предложения 1 нам нужно только доказать, что

$$x : y : z = x_1 : x_2 : x_3$$

для любой точки плоскости П.

Пусть

$$\overrightarrow{OE_0} = k_0e_0, \quad \overrightarrow{OE_1} = k_1e_1, \quad \overrightarrow{OE_2} = k_2e_2, \quad \overrightarrow{OE_3} = k_3e_3.$$

Тогда точки E_0, E_1, E_2, E_3 будут иметь в пространстве (относительно репера $Oe_1e_2e_3$) аффинные координаты

$$E_0(k_0, k_0, k_0), \quad E_1(k_1, 0, 0), \quad E_2(0, k_2, 0), \quad E_3(0, 0, k_3),$$

причем будет иметь место равенство

$$\frac{1}{k_0} = \frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_3}.$$

Рассмотрим теперь в пространстве новый аффинный координатный репер $O'e_1'e_2'e_3'$, для которого

$$O' = E_1, \quad e_1' = \overrightarrow{OE_1}, \quad e_2' = \overrightarrow{E_1E_2}, \quad e_3' = \overrightarrow{E_1E_3}.$$

Так как

$$e_1' = \overrightarrow{OE_1} = k_1e_1,$$

$$e_2' = -\overrightarrow{OE_1} + \overrightarrow{OE_2} = -k_1e_1 + k_2e_2,$$

$$e_3' = -\overrightarrow{OE_1} + \overrightarrow{OE_3} = -k_1e_1 + k_3e_3$$

и

$$\overrightarrow{OO'} = \overrightarrow{OE_1} = k_1e_1,$$

то формулы перехода от координат x', y', z' в репере $O'e_1'e_2'e_3'$ к координатам x, y, z в репере $Oe_1e_2e_3$ имеют вид

$$\begin{aligned} x &= k_1x' - k_1y' - k_1z' + k_1, \\ y &= \quad k_2y', \\ z &= \quad\quad k_3z', \end{aligned}$$

а потому формулы перехода от координат x, y, z к координатам x', y', z' — вид

$$\begin{aligned} x' &= \frac{x}{k_1} + \frac{y}{k_2} + \frac{z}{k_3} - 1, \\ y' &= \quad \frac{y}{k_2}, \\ z' &= \quad\quad \frac{z}{k_3}. \end{aligned}$$

При этом уравнение плоскости Π в координатах x', y', z' будет иметь вид

$$x' = 0,$$

а величины

$$\xi = y', \quad \eta = z'$$

будут аффинными координатами на плоскости Π, определенными репером $O'e_2'e_3'$.

В координатах ξ, η точки E_0, E_1, E_2, E_3 плоскости Π имеют, очевидно, координаты:

$$E_0\left(\frac{k_0}{k_2}, \frac{k_0}{k_3}\right), \quad E_1(0, 0), \quad E_2(1, 0), \quad E_3(0, 1).$$

Поэтому координатная прямая $\lambda_1 = E_2E_3$ будет иметь уравнение

$$\xi + \eta - 1 = 0,$$

координатная прямая $\lambda_2 = E_1E_3$ — уравнение

$$\xi = 0$$

(т. е. будет осью ординат), а координатная прямая $\lambda_3 = E_1E_2$ будет осью абсцисс

$$\eta = 0.$$

Следовательно, для любой точки $M(\xi, \eta)$ плоскости Π ее трилинейные координаты x_1, x_2, x_3, определенные прямыми λ_1, λ_2, λ_3 и точкой E_0, будут выражаться формулами

$$x_1 = \frac{\xi + \eta - 1}{\frac{k_0}{k_2} + \frac{k_0}{k_3} - 1} = \frac{k_1}{k_0}(1 - \xi - \eta),$$

$$x_2 = \frac{k_2}{k_0}\xi,$$

$$x_3 = \frac{k_3}{k_0}\eta,$$

т. е. формулами

$$x_1 = \frac{k_1}{k_0}(1 - y' - z') = \frac{x - k_1 x'}{k_0} = \frac{x}{k_0},$$

$$x_2 = \frac{k_2}{k_0}y' = \frac{y}{k_0},$$

$$x_3 = \frac{k_3}{k_0}z' = \frac{z}{k_0}$$

(в первом преобразовании мы воспользовались тем, что $x' = 0$ для точек плоскости Π).

Таким образом, действительно $x : y : z = x_1 : x_2 : x_3$.

Пусть x, y — произвольные аффинные координаты на плоскости Π и пусть

$$(x_0, y_0), \quad (x_1, y_1), \quad (x_2, y_2), \quad (x_3, y_3)$$

— координаты точек E_0, E_1, E_2, E_3 в этой координатной системе. Тогда уравнения прямых λ_1, λ_2, λ_3 будут иметь вид

$$(y_3 - y_2)(x - x_2) - (x_3 - x_2)(y - y_2) = 0,$$
$$(y_3 - y_1)(x - x_1) - (x_3 - x_1)(y - y_1) = 0,$$
$$(y_2 - y_1)(x - x_1) - (x_2 - x_1)(y - y_1) = 0,$$

и, следовательно, трилинейные координаты x_1, x_2, x_3 произвольной точки $M(x, y)$ будут выражаться формулами

$$x_1 = \frac{(y_3 - y_2)(x - x_2) - (x_3 - x_2)(y - y_2)}{(y_3 - y_2)(x_0 - x_2) - (x_3 - x_2)(y_0 - y_2)},$$

$$x_2 = \frac{(y_3 - y_1)(x - x_1) - (x_3 - x_1)(y - y_1)}{(y_3 - y_1)(x_0 - x_1) - (x_3 - x_1)(y_0 - y_1)},$$

$$x_3 = \frac{(y_2 - y_1)(x - x_1) - (x_2 - x_1)(y - y_1)}{(y_2 - y_1)(x_0 - x_1) - (x_2 - x_1)(y_0 - y_1)}.$$

Тем самым доказано, что

проективные координаты $X : Y : Z$ произвольной (собственной) точки $M(x, y)$ в проективной системе координат с репером

$$E_0(x_0, y_0), \quad E_1(x_1, y_1), \quad E_2(x_2, y_2), \quad E_3(x_3, y_3)$$

выражаются формулами

$$\begin{aligned}
\rho X &= \frac{(y_3 - y_2)(x - x_2) - (x_3 - x_2)(y - y_2)}{(y_3 - y_2)(x_0 - x_2) - (x_3 - x_2)(y_0 - y_2)}, \\
\rho Y &= \frac{(y_3 - y_1)(x - x_1) - (x_3 - x_1)(y - y_1)}{(y_3 - y_1)(x_0 - x_1) - (x_3 - x_1)(y_0 - y_1)}, \\
\rho Z &= \frac{(y_2 - y_1)(x - x_1) - (x_2 - x_1)(y - y_1)}{(y_2 - y_1)(x_0 - x_1) - (x_2 - x_1)(y_0 - y_1)},
\end{aligned} \tag{1}$$

где ρ *— произвольный множитель пропорциональности.*

Формулы (1) представляют собой *формулы перехода* от аффинных координат x, y к проективным координатам $X : Y : Z$.

Формулы (1) можно разрешить относительно x, y, т. е. выразить координаты x, y через координаты $X : Y : Z$. В общем случае мы этого делать не будем, а ограничимся только случаем, когда точки E_1, E_2, E_3 имеют координаты

$$(1, 0),\ (0, 1),\ (0, 0)$$

(заметим, что для любых трех неколлинеарных точек E_1, E_2, E_3 всегда существует, и притом единственная, аффинная координатная система, в которой точки E_1, E_2, E_3 имеют эти координаты).

В этом случае формулы (1) приобретают вид

$$\begin{aligned}
\rho X &= \frac{x}{x_0}, \\
\rho Y &= \frac{y}{y_0}, \\
\rho Z &= \frac{x + y - 1}{x_0 + y_0 - 1}.
\end{aligned}$$

Находя из этих формул x, y, мы получаем

$$\begin{aligned}
x &= \frac{Xx_0}{Xx_0 + Yy_0 - Z(x_0 + y_0 - 1)}, \\
y &= \frac{Yy_0}{Xx_0 + Yy_0 - Z(x_0 + y_0 - 1)}.
\end{aligned}$$

Если

$$x_0 = \frac{1}{3},\quad y_0 = \frac{1}{3},$$

т. е. если точка E_0 является центром тяжести точек E_1, E_2, E_3, эти формулы упрощаются:

$$\begin{aligned}
x &= \frac{X}{X + Y + Z}, \\
y &= \frac{Y}{X + Y + Z}.
\end{aligned}$$

Полученные формулы означают, что числа x, y являются координатами центра тяжести масс X, Y, Z, помещенных в точках E_1, E_2, E_3. Следовательно, $X : Y : Z$ являются не чем иным, как барицентрическими координатами

(см. п. 6 § 1 гл. 2). Таким образом, *барицентрические координаты являются частным случаем проективных координат, получающимся, когда точка E_0 является центром тяжести точек E_1, E_2, E_3.*

Аналогично можно найти формулы перехода от аффинных координат x, y к проективным координатам $X : Y : Z$, определенным репером, одна или две точки которого являются несобственными точками. Мы не будем разбирать этот вопрос в полной общности (хотя и настоятельно рекомендуем сделать это читателю), а ограничимся случаем, когда точки E_0 и E_3 — собственные, а точки E_1 и E_2 — несобственные. При этом мы дополнительно будем предполагать, что точка E_0 имеет координаты (1, 1), точка E_3 — координаты (0, 0) (является началом координат), точка E_1 является несобственной точкой оси абсцисс, а точка E_2 — несобственной точкой оси ординат.

Согласно общему определению из п. 5 проективные координаты с таким репером являются аффинными координатами в пространстве, отвечающими аффинному координатному реперу $Oe_1e_2e_3$, для которого векторы e_1, e_2 параллельны осям координат данной аффинной координатной системы Oxy на плоскости П, вектор e_3 параллелен прямой OE_3, а вектор

$$e_0 = e_1 + e_2 + e_3$$

обладает тем свойством, что он параллелен прямой OE_0. Ясно, что все эти условия будут выполнены, если мы выберем репер $Oe_1e_2e_3$ так, чтобы в соответствующих аффинных координатах X, Y, Z плоскость П определялась уравнением

$$Z = 1,$$

а координаты X, Y совпадали на плоскости П с данными координатами x, y. (Заметим, что эти условия однозначно определяют репер $Oe_1e_2e_3$.)

Поскольку проективными координатами точки $M(x, y)$ плоскости являются, по определению, ее координаты X, Y, Z в этом репере, тем самым доказано, что

проективные координаты $X : Y : Z$ произвольной (собственной) точки $M(x, y)$ в проективной координатной системе с репером E_0, E_1, E_2, E_3, для которого

а) *точка E_0 имеет координаты* (1, 1);

б) *точка E_1 является несобственной точкой оси абсцисс Ox;*

в) *точка E_2 является несобственной точкой оси ординат Oy;*

г) *точка E_3 имеет координаты* (0,0) (*является началом координат*), *определяются соотношением*

$$X : Y : Z = x : y : 1,$$

т. е. являются однородными аффинными координатами.

§ 2. ГЕОМЕТРИЯ ПЛОСКОСТЕЙ В ПРОСТРАНСТВЕ

Геометрия плоскостей в пространстве во многом аналогична геометрии прямых на плоскости. Поэтому некоторые утверждения мы будем оставлять читателю в качестве упражнений и заданий.

1. Пучки плоскостей

Пусть в пространстве фиксирована некоторая аффинная система координат $Oxyz$.

Определение 1. *Координатами* произвольной плоскости

$$Ax + By + Cz + D = 0$$

в этой координатной системе мы будем называть коэффициенты A, B, C, D ее уравнения.

Координаты плоскости — это однородные координаты, т. е. по плоскости они определены только с точностью до пропорциональности. Плоскость с координатами A, B, C, D мы будем обозначать символом $(A : B : C : D)$. Числа A, B, C, D тогда и только тогда являются координатами некоторой плоскости, когда хотя бы одно из чисел A, B, C отлично от нуля.

Определение 2. *Пучком плоскостей* называется множество плоскостей $(A : B : C : D)$, координаты которых выражаются формулами

$$\begin{aligned} A &= \mu A_0 + \nu A_1, \\ B &= \mu B_0 + \nu B_1, \\ C &= \mu C_0 + \nu C_1, \\ D &= \mu D_0 + \nu D_1, \end{aligned} \tag{1}$$

где A_0, A_1, B_0, B_1, C_0, C_1, D_0, D_1 — постоянные числа, обладающие тем свойством, что ранг матрицы

$$\begin{pmatrix} A_0 & B_0 & C_0 & D_0 \\ A_1 & B_1 & C_1 & D_1 \end{pmatrix} \tag{2}$$

равен двум, а μ и ν — переменные параметры, одновременно не равные нулю. (Если ранг матрицы (2) равен единице, т. е. если четверки (A_0, B_0, C_0, D_0) и (A_1, B_1, C_1, D_1) пропорциональны, но пучок вырождается в единственную плоскость.)

Если каждая из троек (A_0, B_0, C_0) и (A_1, B_1, C_1) отлична от нулевой тройки (содержит отличное от нуля число) и если эти тройки не пропорциональны, то пучок (1) называется *собственным*. В противном случае, этот пучок называется *несобственным*. Таким образом, пучок тогда и только тогда — собственный, когда ранг матрицы

$$\begin{pmatrix} A_0 & B_0 & C_0 \\ A_1 & B_1 & C_1 \end{pmatrix}$$

равен двум. Для собственного пучка определены две плоскости с уравнениями

$$\begin{aligned} A_0x + B_0y + C_0z + D_0 &= 0, \\ A_1x + B_1y + C_1z + D_1 &= 0, \end{aligned}$$

и эти плоскости не параллельны (и не совпадают). Прямую пересечения этих плоскостей мы будем называть *центральной прямой* пучка.

Предложение 1. *Каждый собственный пучок состоит из всех плоскостей, проходящих через центральную прямую пучка.*

Предложение 2. *Каждый несобственный пучок состоит из всех плоскостей, параллельных некоторой плоскости.*

При этом

любая прямая является центральной прямой некоторого собственного пучка

и

совокупность всех плоскостей, параллельных произвольной фиксированной плоскости, является несобственным пучком.

Кроме того,

для собственных пучков плоскостей любые значения параметров μ *и* ν *(конечно, одновременно отличные от нуля) определяют некоторую плоскость, тогда как для каждого несобственного пучка имеются значения этих параметров (с точностью до пропорциональности, однозначно определенные), которым никакая плоскость не соответствует.*

Задание. Докажите предложения 1 и 2 и следующие за ними утверждения (ср. п. 1 § 1).

2. Связки плоскостей

Пучки плоскостей (подобно пучкам прямым) аналогичны прямым «точечной» геометрии. Рассмотрим теперь аналоги плоскостей.

Определение 1. *Связкой плоскостей* называется множество плоскостей $(A : B : C : D)$, координаты которых выражаются формулами

$$\begin{aligned} A &= \mu A_0 + \nu A_1 + \xi A_2, \\ B &= \mu B_0 + \nu B_1 + \xi B_2, \\ C &= \mu C_0 + \nu C_1 + \xi C_2, \\ D &= \mu D_0 + \nu D_1 + \xi D_2, \end{aligned} \tag{1}$$

где $A_0, A_1, \ldots, D_1, D_2$ — постоянные числа, обладающие тем свойством, что ранг матрицы

$$\begin{pmatrix} A_0 & A_1 & A_2 \\ B_0 & B_1 & B_2 \\ C_0 & C_1 & C_2 \\ D_0 & D_1 & D_2 \end{pmatrix} \tag{2}$$

равен трем, а μ, ν, ξ — переменные параметры, одновременно не равные нулю. (Если ранг матрицы (2) меньше трех, связка вырождается в пучок.)

Связка называется *собственной*, если определитель

$$\begin{vmatrix} A_0 & A_1 & A_2 \\ B_0 & B_1 & B_2 \\ C_0 & C_1 & C_2 \end{vmatrix} \tag{3}$$

отличен от нуля. В противном случае связка называется *несобственной*.

Для собственной связки определены три плоскости

$$\begin{aligned} A_0x + B_0y + C_0z + D_0 &= 0, \\ A_1x + B_1y + C_1z + D_1 &= 0, \\ A_2x + B_2y + C_2z + D_2 &= 0, \end{aligned}$$

пересекающиеся в единственной точке. Эта точка называется *центром* собственной связки.

Предложение 1. *Любая собственная связка состоит из всех плоскостей, проходящих через ее центр.*

Доказательство. Ясно, что каждая плоскость собственной связки проходит через центр $M_0(x_0, y_0, z_0)$ связки (его координаты удовлетворяют уравнению этой плоскости). Обратно, пусть

$$Ax + By + Cz + D = 0 \tag{4}$$

— произвольная плоскость, проходящая через точку $M_0(x_0, y_0, z_0)$. Выберем на этой плоскости две точки $M_1(x_1, y_1, z_1)$ и $M_2(x_2, y_2, z_2)$, не лежащие на одной прямой с точкой M_0, и попробуем найти в связке плоскость, проходящую через эти точки. Параметры μ, ν, ξ, отвечающие этой плоскости, должны удовлетворять соотношениям

$$\begin{aligned} &(\mu A_0 + \nu A_1 + \xi A_2)\, x_1 + (\mu B_0 + \nu B_1 + \xi B_2)\, y_1 + \\ &\qquad + (\mu C_0 + \nu C_1 + \xi C_2)\, z_1 + (\mu D_0 + \nu D_1 + \xi D_2) = 0, \\ &(\mu A_0 + \nu A_1 + \xi A_2)\, x_2 + (\mu B_0 + \nu B_1 + \xi B_2)\, y_2 + \\ &\qquad + (\mu C_0 + \nu C_1 + \xi C_2)\, z_2 + (\mu D_0 + \nu D_1 + \xi D_0) = 0, \end{aligned}$$

т. е. соотношениям

$$\begin{aligned} &(A_0x_1 + B_0y_1 + C_0z_1 + D_0)\,\mu + (A_1x_1 + B_1y_1 + C_1z_1 + D_1)\,\nu + \\ &\qquad + (A_2x_1 + B_2y_1 + C_2z_1 + D_2)\,\xi = 0, \\ &(A_0x_2 + B_0y_2 + C_0z_2 + D_0)\,\mu + (A_1x_2 + B_1y_2 + C_1z_2 + D_1)\,\nu + \\ &\qquad + (A_2x_2 + B_2y_2 + C_2z_2 + D_2)\,\xi = 0. \end{aligned}$$

Эти равенства представляют собой два однородных уравнения относительно трех неизвестных μ, ν и ξ. Из алгебры известно, что у таких уравнений всегда имеется нетривиальное реше-

ние [1]) $\mu_0 : \nu_0 : \xi_0$. Соответствующая этим значениям параметров плоскость связки и является плоскостью, проходящей через точки M_1 и M_2. Являясь плоскостью связки, эта плоскость проходит, кроме того, и через точку M_0. Но через три не лежащие на одной прямой точки проходит только одна плоскость. Поэтому найденная плоскость совпадает с данной плоскостью (4).

Замечание 1. Тот факт, что плоскость (4) принадлежит связке, можно доказать и по-иному. Действительно, рассмотрим систему уравнений

$$\begin{aligned} A_0\mu + A_1\nu + A_2\xi &= A, \\ B_0\mu + B_1\nu + B_2\xi &= B, \\ C_0\mu + C_1\nu + C_2\xi &= C \end{aligned}$$

относительно неизвестных μ, ν, ξ. По условию, определитель этой системы отличен от нуля и потому она имеет единственное решение μ_0, ν_0, ξ_0. Ясно, что хотя бы одно из чисел μ_0, ν_0, ξ_0 отлично от нуля (ибо в противном случае все три числа A, B, C были бы равны нулю). Плоскость связки, соответствующая этим значениям параметра, параллельна плоскости (4) и имеет с ней общую точку M_0. Следовательно, она совпадает с этой плоскостью.

Задание. Докажите, что значения μ_0, ν_0, ξ_0 параметров, дающие плоскость (4), определены этой плоскостью однозначно с точностью до пропорциональности.

Подчеркнем, что

любая точка пространства является центром некоторой собственной связки.

Для доказательства этого утверждения достаточно выбрать три плоскости, пересекающиеся только в этой точке, и рассмотреть собственную связку, ими определенную.

Предложение 2. *Любая несобственная связка состоит из всех плоскостей, параллельных некоторой прямой.*

Доказательство. Для несобственной связки ранг матрицы

$$\begin{pmatrix} A_0 & B_0 & C_0 \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{pmatrix} \tag{5}$$

меньше трех. Этот ранг не может быть равен единице, поскольку тогда ранг матрицы (2) будет, вопреки предположению, меньше трех. Следовательно, ранг матрицы (5) в точности равен

[1]) В пространстве с координатами μ, ν, ξ эти уравнения определяют две плоскости, проходящие через начало координат. Поэтому им удовлетворяют координаты любой точки прямой, являющейся пересечением этих плоскостей.

двум. Это означает, что матрица (5) содержит две непропорциональные строки, линейной комбинацией которых является третья строка. Пусть, для определенности, не пропорциональны первые две строки. Тогда соответствующие плоскости

$$\begin{aligned} A_0x + B_0y + C_0z + D_0 &= 0, \\ A_1x + B_1y + C_1z + D_1 &= 0 \end{aligned} \tag{6}$$

не параллельны и потому пересекаются по некоторой прямой (уравнениями которой и являются уравнения (6)).

Рассмотрим теперь произвольную плоскость

$$Ax + By + Cz + D = 0, \tag{7}$$

принадлежащую нашей связке. В матрице

$$\begin{pmatrix} A_0 & B_0 & C_0 \\ A_1 & B_1 & C_1 \\ A & B & C \end{pmatrix} \tag{8}$$

первые две строки не пропорциональны, а третья строка, линейно выражаясь по условию через строки матрицы (5), линейно выражается через первые две строки. Следовательно, ранг матрицы (8) равен двум. Поэтому (см. п. 2 § 3 гл. 3) плоскость (7) параллельна прямой (6).

Обратно, пусть

$$Ax + By + Cz + D = 0 \tag{9}$$

— произвольная плоскость, параллельная прямой (6). Тогда ранг матрицы (8) равен двум, и потому ее третья строка линейно выражается через первые две строки (по условию, линейно независимые). Следовательно, ранг матрицы

$$\begin{pmatrix} A_0 & B_0 & C_0 \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A & B & C \end{pmatrix} \tag{10}$$

также равен двум (дополнительная строка A_2, B_2, C_2 линейно выражается, как мы знаем, через первые две строки).

Рассмотрим теперь матрицу

$$\begin{pmatrix} A_0 & B_0 & C_0 & D_0 \\ A_1 & B_1 & C_1 & D_1 \\ A_2 & B_2 & C_2 & D_2 \\ A & B & C & D \end{pmatrix}. \tag{11}$$

Она получается из матрицы (10) прибавлением одного столбца, и потому ее ранг может увеличиться только на единицу. Следовательно, ранг матрицы (11) не превосходит трех. Но, по условию, первые три строки матрицы (11) линейно независимы (они составляют матрицу (2), ранг которой равен трем). Следовательно, ранг матрицы (11) в точности равен трем и ее последняя строка линейно выражается через предыдущие. Это означает, что

$$A = \mu A_0 + \nu A_1 + \xi A_2,$$
$$B = \mu B_0 + \nu B_1 + \xi B_2,$$
$$C = \mu C_0 + \nu C_1 + \xi C_2,$$
$$D = \mu D_0 + \nu D_1 + \xi D_2,$$

т. е. плоскость (9) принадлежит рассматриваемой связке.

3. Расширенное пространство

Аналогично расширенной плоскости мы можем построить *расширенное пространство*, точками которого являются все точки аффинного пространства плюс некоторое множество «несобственных точек». При этом каждой прямой λ аффинного пространства сопоставлена некоторая несобственная точка — *несобственная точка* этой прямой, причем так, что любая несобственная точка является несобственной точкой некоторой прямой и двум различным прямым тогда и только тогда сопоставлена одна и та же точка, когда эти прямые параллельны. Предполагая такое соответствие заданным, мы назовем *прямой расширенного пространства* либо произвольную прямую λ аффинного пространства, к которой присоединена ее несобственная точка, либо множество несобственных точек всевозможных прямых некоторой плоскости (*несобственную прямую этой плоскости*). Аналогично, *плоскостями расширенного пространства* мы будем называть либо плоскости аффинного пространства, расширенные присоединением их несобственных прямых, либо множество всех несобственных точек (*несобственную плоскость пространства*).

Аналогично аффинной геометрии в геометрии расширенного пространства

через любые две различные точки проходит единственная прямая, и через любые три точки, не лежащие на одной прямой, проходит единственная плоскость.

С другой стороны, в противоположность аффинной геометрии, в расширенном пространстве

любые две различные плоскости пересекаются по единственной прямой, и любые три плоскости, из которых ни одна не содержит прямую пересечения двух других, пересекаются в единственной точке.

Пучком плоскостей в расширенном пространстве называется множество всех плоскостей, содержащих данную прямую — *центральную прямую пучка.* Пучок называется *собственным*, когда эта прямая — собственная, и *несобственным*, когда эта прямая — несобственная. Аналогично, *связкой плоскостей* в расширенном пространстве называется множество всех плоскостей, содержащих данную точку — *центр связки.* Если эта точка — собственная, связка называется *собственной*, а если она — несобственная, то связка также называется *несобственной.*

Ясно, что каждый собственный пучок и каждая собственная связка состоит из тех же плоскостей (лишь пополненных несобственными прямыми), что и соответствующие пучок или связка в аффинном пространстве. Что же касается несобственных пучков и связок, то они получаются из соответствующих пучков и связок в аффинном пространстве присоединением дополнительно несобственной плоскости (и конечно, добавлением к каждой плоскости ее несобственной прямой). При этом центральной прямой несобственного пучка будет несобственная прямая плоскости, которой в аффинном пространстве параллельны плоскости пучка, а центром несобственной связки будет несобственная точка прямой, которой параллельны плоскости связки.

Автоматическая проверка показывает, что в расширенном пространстве

любые две различные плоскости содержатся в единственном пучке, и любые три плоскости, не принадлежащие одному пучку, содержатся в единственной связке.

Кроме того,

любые две различные связки содержат единственный общий пучок, и любые три связки, не содержащие общего пучка, содержат единственную общую плоскость.

Мы не будем здесь давать формально-аксиоматические определения расширенного пространства, поскольку это делается в точной аналогии со случаем плоскости (см. п. 2 § 1), и лишь отметим, что при расширении аффинного пространства получается *аффинно-проективное пространство*, а при расширении евклидова пространства — *евклидово-проективное пространство.* Кроме того, конечно, все это можно делать над любым полем $\mathbb{K}$, причем в случае $\mathbb{K}=\mathbb{C}$ — в двух известных нам вариантах («чисто комплексном» и «вещественно-комплексном»).

Задание. Дайте формальные определения аффинно-проективного и евклидово-проективного пространств над произвольным полем $\mathbb{K}$ и, в частности над $\mathbb{R}$ и $\mathbb{C}$.

Чтобы построить «аффинную» модель расширенного пространства, мы рассмотрим в аффинном пространстве множество всевозможных связок. Примем эти связки за «точки» нашей мо-

дели. Множество таких точек мы назовем «прямой», если оно состоит из всех связок, содержащих некоторый фиксированный пучок. Тем самым «прямые» нашей модели и пучки плоскостей аффинного пространства будут находиться в естественном биективном соответствии. Наконец, множество точек мы назовем «плоскостью», если оно либо состоит из всех несобственных связок, либо является множеством всех связок, содержащих некоторую фиксированную плоскость аффинного пространства.

Без труда проверяется (ср. п. 3 § 1), что тем самым мы действительно получим модель расширенного пространства. Собственными точками этой модели являются собственные связки и только они. Аналогично, собственным прямым соответствуют собственные пучки плоскостей, а несобственным прямым — несобственные пучки.

Замечание 1. Никакого аналога «пространственной» или «векторной» интерпретации расширенной плоскости мы для расширенного пространства получить не можем (поскольку в нашем распоряжении нет «четырехмерного пространства»).

Однако отсутствие «векторной» модели не мешает введению в расширенном пространстве однородных координат (и, следовательно, построению «числовой модели»).

Пусть в пространстве (нерасширенном) задана произвольная аффинная координатная система $Oxyz$.

Для любой точки M расширенного (аффинно-проективного) пространства мы будем считать ее *однородными аффинными координатами* числа X, Y, Z, T, определенные с точностью до пропорциональности соотношениями

$$X : Y : Z : T = x : y : z : 1,$$

если точка M — собственная и имеет аффинные координаты x, y, z, и соотношениями

$$X : Y : Z : T = l : m : n : 0,$$

если точка M — несобственная и является несобственной точкой прямой

$$\frac{x - x_0}{l} = \frac{y - y_0}{m} = \frac{z - z_0}{n}. \qquad (1)$$

Ясно, что тем самым мы получим взаимно однозначное соответствие между точками расширенного пространства и элементами множества $\mathbb{R}P^3$.

Если координаты x, y, z являются прямоугольными (евклидовыми) координатами, координаты $X : Y : Z : T$ называются *однородными евклидовыми координатами* (конечно, здесь уже речь идет о расширенном евклидовом, т. е. евклидово-проективном, пространстве).

По определению, собственные точки характеризуются условием $T \neq 0$, а несобственные — условием $T = 0$. Следовательно, *уравнение*

$$T = 0$$

является в координатах $X : Y : Z : T$ *уравнением несобственной плоскости.*

Аффинные координаты собственных точек выражаются через координаты $X : Y : Z : T$ по формулам

$$x = \frac{X}{T}, \quad y = \frac{Y}{T}, \quad z = \frac{Z}{T}.$$

Подставив эти выражения в уравнение

$$Ax + By + Cz + D = 0 \tag{2}$$

произвольной плоскости, мы получим (после умножения на T) уравнение

$$AX + BY + CZ + DT = 0. \tag{3}$$

Координаты $X : Y : Z : T$ собственной точки тогда и только тогда удовлетворяют этому уравнению, когда точка принадлежит плоскости (2). Что же касается несобственных точек, то для любой такой точки ее координаты $X : Y : Z : T = l : m : n : 0$ тогда и только тогда удовлетворяют уравнению (3), когда

$$Al + Bm + Cn = 0,$$

т. е. когда прямая (1) параллельна плоскости (2), и, значит, рассматриваемая несобственная точка (являющаяся несобственной точкой прямой (1)) принадлежит плоскости (2).

Этим доказано, что

уравнение (3) *выражает в расширенном пространстве плоскость* (2), *пополненную всеми ее несобственными точками.*

Тем самым, нами доказано следующее

Предложение 1. *В расширенном пространстве любое уравнение вида*

$$AX + BY + CZ + DT = 0$$

(*хотя бы один коэффициент которого отличен от нуля*) *выражает некоторую плоскость: собственную, если отличен от нуля хотя бы один из коэффициентов* A, B, C, *и несобственную, если* $A = B = C = 0$.

Поэтому прямые в расширенном пространстве задаются двумя (не пропорциональными) уравнениями вида

$$A_0X + B_0Y + C_0Z + D_0T = 0,$$
$$A_1X + B_1Y + C_1Z + D_1T = 0.$$

Теперь уже ясно (в полной аналогии с п. 4 § 1), что «числовой моделью» расширенного пространства будет стандартное

(арифметическое) аффинно-проективное пространство $\mathbb{R}P^3$, точками которого являются классы $(X : Y : Z : T)$ пропорциональных четверок $(X, Y, Z, T) \neq (0, 0, 0, 0)$. Несобственные точки в этом пространстве характеризуются условием $T = 0$.

Можно, конечно, дать и независимое определение аффинно-проективного (и евклидово-проективного) пространства с помощью (очевидным образом определяемых) групп $\mathrm{Aff}(\mathbb{R}P^3)$ и $\mathrm{Ort}(\mathbb{R}P^3)$ (ср. п. 4 § 1).

Задание. Определите группы $\mathrm{Aff}(\mathbb{R}P^3)$ и $\mathrm{Ort}(\mathbb{R}P^3)$ и сформулируйте аксиомы аффинно-проективного и евклидово-проективного пространств.

4. Проективное пространство

Проективным пространством называется аффинно-проективное пространство, в котором игнорируется различие между собственными и несобственными точками.

Задание. Введите группу $\mathrm{Proj}(\mathbb{R}P^3)$ и дайте формальное определение проективного пространства (ср. п. 5 § 1).

Обратим внимание на то, что каждая плоскость проективного пространства автоматически является проективной плоскостью в смысле п. 5 § 1 и, аналогично, каждая собственная плоскость аффинно-проективного пространства является аффинно-проективной плоскостью. Напротив, несобственная плоскость аффинно-проективного пространства является лишь проективной плоскостью.

Идея считать прямые и плоскости самостоятельными объектами (а не множеством точек) приводит, аналогично случаю плоскости (см. п. 7 § 1), к следующему определению:

Определение 1. *Конфигурационным пространством* называется множество $\mathfrak{P}$, состоящее из объектов трех сортов: *точек*, *прямых*, *плоскостей*, связанных отношением *инцидентности*, удовлетворяющим следующим аксиомам:

Инц. 1. *Любые две различные точки инцидентны одной и только одной прямой.*

Инц. 2. *Любые две различные прямые, инцидентные одной плоскости, инцидентны одной и только одной точке.*

Инц. 3. *Любая прямая инцидентна по крайней мере трем различным точкам.*

Инц. 4. *Любая точка инцидентна по меньшей мере трем различным прямым.*

Инц. 5. *Любые две различные плоскости инцидентны единственной прямой.*

Инц. 6. *Любые две различные прямые, инцидентные одной точке, инцидентны единственной плоскости.*

Инц. 7. *Любая прямая инцидентна по крайней мере трем различным плоскостям.*

Инц. 8. *Любая плоскость инцидентна по крайней мере трем различным прямым.*

Инц. 9. *Любая плоскость и любая не инцидентная ей прямая инцидентны единственной точке.*

Инц. 10. *Любая точка и любая не инцидентная ей прямая инцидентны единственной плоскости.*

Обратим внимание на то, что аксиомы Инц. 1—4 совпадают (с точностью до оговорки в аксиоме Инц. 2) с аксиомами Инц. 1—4 из определения 1 п. 7 § 1. Следовательно,

множество всех точек и прямых, инцидентных некоторой плоскости конфигурационного пространства, является конфигурационной плоскостью.

Определение 2. Конфигурационное пространство называется *паппповым пространством*, если любая его плоскость является паппповой плоскостью (см. определение 2 п. 7 § 1).

Ясно, что

проективное пространство над любым полем $\mathbb{K}$ *является паппповым пространством.*

Обратное утверждение мы (как и для случая плоскости) приведем без доказательства:

Теорема 1. *Любое паппово пространство изоморфно проективному пространству над некоторым полем* $\mathbb{K}$.

Поэтому аксиомы паппова пространства *полны* в том смысле, что каждая конфигурационная теорема, верная в проективной геометрии над любым полем $\mathbb{K}$, может быть выведена из этих аксиом.

Мы видим, что ситуация в «пространственном» случае, казалось бы, вполне аналогична ситуации, имеющей место в «плоском» случае. Однако более глубокое исследование вскрывает замечательное различие между этими двумя случаями.

Пусть $\mathbb{K}$ — произвольное *тело*[1]) (некоммутативное поле). Назовем две четверки (X_1, Y_1, Z_1, T_1) и (X_2, Y_2, Z_2, T_2) элементов тела $\mathbb{K}$, отличные от нулевой четверки (0, 0, 0, 0), *пропорциональными*, если существует такой элемент ρ тела $\mathbb{K}$, что

$$X_2 = X_1\rho, \quad Y_2 = Y_1\rho, \quad Z_2 = Z_1\rho, \quad T_2 = T_1\rho.$$

(Обратите внимание на то, что порядок множителей здесь существен: четверки (X, Y, Z, T) и $(\rho X, \rho Y, \rho Z, \rho T)$, вообще говоря, не пропорциональны). Без труда проверяется (так же, как для случая, когда $\mathbb{K}$ коммутативно), что отношение пропорциональности является отношением эквивалентности. Соответствующие классы эквивалентности обозначаются символом $(X : Y : Z : T)$, а их множество — символом $\mathbb{K}P^3$.

Аналогично вводится множество $\mathbb{K}P^2$ классов $(X : Y : Z)$ пропорциональных троек и, вообще, множество $\mathbb{K}P^n$ при любом $n \geqslant 1$.

[1]) Примером тела является рассматриваемое в п. 4 § 2 гл. 7 тело кватернионов.

Элементы $(X:Y:Z:T)$ множества $\mathbb{K}P^3$ мы будем называть *точками*. Множества точек, удовлетворяющих уравнениям вида

$$AX+BY+CZ+DT=0,$$

мы назовем *плоскостями* [1]), а множества, являющиеся пересечением двух (различных) плоскостей,— *прямыми*. За *отношение инцидентности* между точками, прямыми и плоскостями мы примем (как и в коммутативном случае) отношение теоретико-множественной принадлежности (например, точка инцидентна плоскости, если она содержится в ней).

Простая выкладка показывает, что

множество так определенных точек, прямых и плоскостей является относительно указанного отношения инцидентности конфигурационным пространством.

Задание. Докажите это утверждение (не забывайте следить за порядком множителей!).

Мы будем это пространство (допуская определенную вольность) обозначать тем же символом $\mathbb{K}P^3$ и будем называть его (а также любое конфигурационное пространство, ему изоморфное) *проективным пространством над телом* $\mathbb{K}$.

Аналогично определяется *проективная плоскость над телом* $\mathbb{K}$ [2]).

Оказывается, что справедлива следующая замечательная теорема, которую мы приведем без доказательства:

Теорема 2. *Любое конфигурационное пространство является проективным пространством над некоторым телом* $\mathbb{K}$.

Таким образом, папповость нам нужна только для того, чтобы тело $\mathbb{K}$ было коммутативным.

Из теоремы 2 вытекает, что конфигурационная теорема тогда и только тогда может быть выведена из аксиом Инц. 1—10, когда она выполняется в проективном пространстве над любым телом $\mathbb{K}$.

Примером такой теоремы является, скажем, теорема Дезарга (см. п. 7 § 1).

Упражнение. Докажите теорему Дезарга двумя способами: из аксиом Инц. 1—10 (указание: рассмотрите сначала случай, когда плоскости **123** и **456** различны) и алгебраической выкладкой (не забывайте о порядке множителей!).

---

[1]) Обратим внимание на то, что уравнение вида

$$XA+YB+ZC+TD=0,$$

вообще говоря, не является уравнением плоскости.

[2]) Конечно, мы могли бы ввести проективное пространство (и проективную плоскость) над телом $\mathbb{K}$ не как конфигурационное пространство (конфигурационную плоскость), а как множество, снабженное координатными системами (ср. п. 5 § 1). Мы от этого уклонились, во-первых, потому что это нам сейчас не нужно, а, во-вторых, потому, что при этом возникают некоторые осложнения в описании элементов групп $\mathrm{Proj}(\mathbb{K}P^n)$ (ввиду трудностей в построении теории определителей над некоммутативным телом).

Глубокое различие между случаями $n = 3$ и $n = 2$ (о котором мы упоминали выше) состоит в том, что для конфигурационных плоскостей аналог теоремы 2 не верен: не любая конфигурационная плоскость является проективной плоскостью над некоторым телом. Действительно, поскольку проективная плоскость над произвольным телом $\mathbb{K}$ является плоскостью проективного пространства над телом $\mathbb{K}$, в ней, согласно сказанному выше, выполнена теорема Дезарга. Называя конфигурационные плоскости, в которых выполнена теорема Дезарга, *дезарговыми плоскостями,* мы, следовательно, можем сказать, что

проективная плоскость над произвольным телом $\mathbb{K}$ является дезарговой плоскостью.

Оказывается, что это необходимое условие является также и достаточным, т. е. справедлива следующая теорема (которую мы также приведем без доказательства):

Теорема 3. *Конфигурационная плоскость тогда и только тогда является проективной плоскостью над некоторым телом $\mathbb{K}$, когда она дезаргова.*

Замечание 1. В алгебре доказывается, что любое конечное тело является полем. Поэтому

любая дезаргова плоскость, содержащая конечное число точек, является папповой плоскостью.

Удивительно, что чисто геометрического («синтетического») доказательства этого утверждения до сих пор неизвестно.

Определение 3. Конфигурационная теорема $\mathscr{A}'$, называется (*пространственно*) *двойственной* теореме $\mathscr{A}$, если она получается из теоремы $\mathscr{A}$ заменами слова «точка» словом «плоскость» и наоборот.

Обратившись к аксиомам Инц. 1—10, мы немедленно обнаружим, что аксиомы Инц. 1, Инц. 2, Инц. 3, Инц. 4, Инц. 9, двойственны, соответственно, аксиомам Инц. 5, Инц. 6, Инц. 7, Инц. 8, Инц. 10, т. е. что аксиомы Инц. 1—10 разбиваются на пары двойственных друг другу аксиом.

Отсюда очевидным образом (ср. п. 7 § 1) вытекает следующий общий принцип:

Принцип пространственной двойственности. *Теорема, пространственно двойственная верной конфигурационной теореме, также является верной теоремой.*

Здесь под «верной теоремой» понимается, естественно, теорема, вытекающая из аксиом Инц. 1—10. Все такие теоремы верны в проективном пространстве над любым телом $\mathbb{K}$.

Однако если тело $\mathbb{K}$ фиксировано, то под «верной теоремой» можно понимать конфигурационную теорему, справедливую в проективном пространстве именно над этим телом $\mathbb{K}$, и встает вопрос, справедлив ли принцип пространственной двойственности при таком понимании термина «верная теорема», т. е. справедлив ли этот принцип над данным телом $\mathbb{K}$?

Оказывается, что ответ на этот вопрос, вообще говоря, отрицательный.

Выясним, в чем тут дело.

Определение 4. *Антиавтоморфизмом* конфигурационного пространства $\mathfrak{P}$ называется его преобразование, переводящее точки в плоскости, прямые в прямые, плоскости в точки и сохраняющее отношение инцидентности (ср. определение 3 п. 7 § 1).

Ясно, что

если конфигурационное пространство обладает хотя бы одним антиавтоморфизмом, то в нем выполнен принцип двойственности.

Имея это в виду, постараемся построить для конфигурационного пространства $\mathbb{K}P^3$ антиизоморфизм тем же способом, как это мы делали в п. 7 § 1 в случае плоскости, т. е. сопоставив точке $(A:B:C:D)$ плоскость с теми же координатами, т. е. с уравнением

$$AX+BY+CZ+DT=0.$$

Однако легко видеть, что полученное (очевидно, биективное) соответствие не будет, вообще говоря, сохранять отношение инцидентности, поскольку факт инцидентности точки $(X:Y:Z:T)$ и плоскости $(A:B:C:D)$ выражается уравнением

$$AX+BY+CZ+DT=0, \tag{1}$$

а факт инцидентности точки $(A:B:C:D)$ и плоскости $(X:Y:Z:T)$ — уравнением

$$XA+YB+ZC+TD=0,$$

вообще говоря, отличным от уравнения (1).

Тем не менее, если тело $\mathbb{K}$ является полем, то уравнения (1) и (2) равносильны. Следовательно,

принцип двойственности выполнен в проективном пространстве над любым полем $\mathbb{K}$.

Замечание 2. Это утверждение можно обобщить. Напомним с этой целью, что биективное отображение $a\longmapsto\bar{a}$ тела $\mathbb{K}$ на себя называется *антиавтоморфизмом*, если для любых элементов $a,\ b\in\mathbb{K}$ имеют место равенства

$$\overline{a+b}=\bar{a}+\bar{b},$$
$$\overline{ab}=\bar{b}\bar{a},$$
$$\bar{\bar{a}}=a.$$

Предполагая, что тело $\mathbb{K}$ обладает таким антиавтоморфизмом [1]), сопоставим произвольной точке $(X:Y:Z:T)$ плоскость с координатами $(\bar{X}:\bar{Y}:\bar{Z}:\bar{T})$ (а плоскости $(A:B:C:D)$ — точку

[1]) Например, является телом кватернионов (см. п. 5 § 2 гл. 7).

$(\bar{A} : \bar{B} : \bar{C} : \bar{D})$). Поскольку факт инцидентности точки $(\bar{A} : \bar{B} : \bar{C} : \bar{D})$ и плоскости $(\bar{X} : \bar{Y} : \bar{Z} : \bar{T})$ выражается уравнением

$$\bar{X}\bar{A} + \bar{Y}\bar{B} + \bar{Z}\bar{C} + \bar{T}\bar{D} = 0,$$

переходящим при рассматриваемом антиавтоморфизме в уравнение (1), мы видим, что построенное (очевидно, биективное) соответствие сохраняет отношение инцидентности. Следовательно,

если тело $\mathbb{K}$ *обладает антиавтоморфизмом, то антиизоморфизмом обладает проективное пространство над* $\mathbb{K}$, *и потому в этом пространстве выполнен принцип двойственности.*

Ясно, что аналогичное утверждение справедливо и для плоскостей.

Можно доказать, что если $\mathbb{K}P^3$ (или $\mathbb{K}P^2$) обладает антиавтоморфизмом, то антиавтоморфизмом обладает и тело $\mathbb{K}$.

Дополнение. О геометрии прямых в пространстве

Геометрия прямых в пространстве значительно интереснее и сложнее геометрии плоскостей в пространстве и геометрии прямых на плоскости. Мы не имеем здесь возможности заняться этой геометрией сколько-нибудь подробно и потому ограничимся лишь тем, что объясним, почему эта геометрия более сложна.

Пусть

$$\begin{aligned} A_0x + B_0y + C_0z + D_0 &= 0, \\ A_1x + B_1y + C_1z + D_1 &= 0 \end{aligned} \tag{1}$$

— произвольная прямая в пространстве. Что можно принять за ее координаты? Считать ими коэффициенты уравнений (1) явно не очень удобно, поскольку ту же самую прямую можно задать любой парой плоскостей

$$\begin{aligned} A'_0x + B'_0y + C'_0z + D'_0 &= 0, \\ A'_1x + B'_1y + C'_1z + D'_1 &= 0 \end{aligned} \tag{2}$$

из соответствующего пучка и потому совсем другими коэффициентами. Хотелось бы связать с прямой (1) инвариантные выражения, т. е. функции коэффициентов, не зависящие от выбора уравнений (1).

Рассмотрим с этой целью матрицу коэффициентов

$$\begin{pmatrix} A_0 & B_0 & C_0 & D_0 \\ A_1 & B_1 & C_1 & D_1 \end{pmatrix} \tag{3}$$

уравнений (1). Она обладает шестью минорами второго порядка

$$\begin{aligned} p_{12} &= \begin{vmatrix} A_0 & B_0 \\ A_1 & B_1 \end{vmatrix}, & p_{23} &= \begin{vmatrix} B_0 & C_0 \\ B_1 & C_1 \end{vmatrix}, \\ p_{13} &= \begin{vmatrix} A_0 & C_0 \\ A_1 & C_1 \end{vmatrix}, & p_{24} &= \begin{vmatrix} B_0 & D_0 \\ B_1 & D_1 \end{vmatrix}, \\ p_{14} &= \begin{vmatrix} A_0 & D_0 \\ A_1 & D_1 \end{vmatrix}, & p_{34} &= \begin{vmatrix} C_0 & D_0 \\ C_1 & D_1 \end{vmatrix}. \end{aligned} \tag{4}$$

Поскольку ранг матрицы (3) равен двум (в противном случае плоскости (1) совпадают и не определяют никакой прямой), хотя бы один из этих миноров отличен от нуля. Посмотрим, как эти миноры меняются при замене плоскостей (1) плоскостями (2).

По условию, плоскости (2) принадлежат пучку плоскостей, определенному плоскостями (1). Следовательно, существуют такие числа μ_0, ν_0, μ_1, ν_1, что

$$\begin{aligned} A'_0 &= \mu_0 A_0 + \nu_0 A_1, & A'_1 &= \mu_1 A_0 + \nu_1 A_1, \\ B'_0 &= \mu_0 B_0 + \nu_0 B_1, & B'_1 &= \mu_1 B_0 + \nu_1 B_1, \\ C'_0 &= \mu_0 C_0 + \nu_0 C_1, & C'_1 &= \mu_1 C_0 + \nu_1 C_1, \\ D'_0 &= \mu_0 D_0 + \nu_0 D_1, & D'_1 &= \mu_1 D_0 + \nu_1 D_1. \end{aligned} \tag{5}$$

Пусть

$$\begin{aligned} p'_{11} &= \begin{vmatrix} A'_0 & B'_0 \\ A'_1 & B'_1 \end{vmatrix}, & p'_{23} &= \begin{vmatrix} B'_0 & C'_0 \\ B'_1 & C'_1 \end{vmatrix}, \\ p'_{13} &= \begin{vmatrix} A'_0 & C'_0 \\ A'_1 & C'_1 \end{vmatrix}, & p'_{24} &= \begin{vmatrix} B'_0 & D'_0 \\ B'_1 & D'_1 \end{vmatrix}, \\ p'_{14} &= \begin{vmatrix} A'_0 & D'_0 \\ A'_1 & D'_1 \end{vmatrix}, & p'_{34} &= \begin{vmatrix} C'_0 & D'_0 \\ C'_1 & D'_1 \end{vmatrix}. \end{aligned} \tag{6}$$

Подставив выражения (5) в формулы (6), мы немедленно получим (воспользовавшись, например, теоремой об умножении определителей), что

$$\begin{aligned} p'_{12} &= \Delta p_{12}, & p'_{23} &= \Delta p_{23}, \\ p'_{13} &= \Delta p_{13}, & p'_{24} &= \Delta p_{24}, \\ p'_{14} &= \Delta p_{14}, & p'_{34} &= \Delta p_{34}, \end{aligned}$$

где

$$\Delta = \begin{vmatrix} \mu_0 & \nu_0 \\ \mu_1 & \nu_1 \end{vmatrix}.$$

Заметим, что $\Delta \neq 0$, поскольку в противном случае плоскости (2) совпадают.

Тем самым мы доказали, что

числа p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34} однозначно с точностью до пропорциональности определены прямой (1).

Эти числа называются *плюккеровыми координатами* прямой (1). Подчеркнем, что это — однородные координаты.

Теперь уже ясна одна из причин сложности геометрии прямых: прямые описываются шестью однородными координатами, т. е. могут рассматриваться как точки пятимерного пространства.

Правда, пока не ясно, нет ли среди плюккеровых координат лишних и, вообще, не связаны ли эти координаты каким-нибудь соотношением. Чтобы

ответить на этот вопрос, рассмотрим определитель

$$\begin{vmatrix} A_0 & B_0 & C_0 & D_0 \\ A_1 & B_1 & C_1 & D_1 \\ A_0 & B_0 & C_0 & D_0 \\ A_1 & B_1 & C_1 & D_1 \end{vmatrix}.$$

Этот определитель равен нулю, поскольку он имеет одинаковые строки. С другой стороны, разложив его по первым двум строкам (мы здесь пользуемся известной из алгебры теоремой Лапласа о разложении определителей), мы немедленно получим, что он равен $2(p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23})$. Следовательно,

плюккеровы координаты любой прямой удовлетворяют соотношению

$$p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23} = 0. \tag{7}$$

Это соотношение известно как соотношение Плюккера[1]).

Соотношение Плюккера является соотношением второй степени. Это означает, что прямые представляют собой точки пятимерного пространства, принадлежащие некоторой «четырехмерной гиперповерхности второго порядка». Это вторая — и главная — причина сложности геометрии прямых.

Но является ли соотношение Плюккера единственным соотношением, связывающим плюккеровы координаты, т. е., иными словами, можно ли для любых чисел p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}, удовлетворяющих соотношению Плюккера, найти прямую, плюккеровыми координатами которой они являются, т. е. можно ли решить уравнения (4) относительно неизвестных A_0, B_0, C_0, D_0, A_1, B_1, C_1, D_1? Простая выкладка (которую мы предоставили читателю) показывает, что это действительно возможно. Однако при этом могут получиться уравнения параллельных плоскостей. Чтобы этого не было, следует, очевидно, дополнительно потребовать, чтобы хотя бы одно из чисел

$$p_{12}, \quad p_{13}, \quad p_{23} \tag{8}$$

было отлично от нуля.

Таким образом, резюмируя, мы получаем, что

прямые в аффинном пространстве описываются шестью однородными плюккеровыми координатами p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}, *подчиненными соотношению Плюккера* (7) *и обладающими тем свойством, что хотя бы одна из трех координат* (8) *отлична от нуля.*

Замечание 1. Полученные в п. 1 § 3 гл. 3 формулы показывают, что плюккеровы координаты p_{23}, $-p_{13}$, p_{12} являются координатами l, m, n направляющего вектора прямой. Следовательно,

две прямые тогда и только тогда параллельны, когда их координаты (8) *пропорциональны.*

В случае, когда координаты (8) равны нулю, получаются пары параллельных плоскостей, т. е. несобственные прямые. Поэтому в проективном (или аффинно-проективном) пространстве прямые, как и следовало ожидать, описываются несколько проще: требование, чтобы координаты (8) были отличны от нуля, уже не нужно.

---

[1]) Читатель, не знакомый с теоремой Лапласа, может проверить соотношение Плюккера прямым вычислением.

Заметим в заключение, что плюккеровы координаты p_{ij} можно считать определенными для всех значений $i, j = 1, 2, 3, 4$, предполагая при этом, что $p_{ji} = -p_{ij}$ (и, в частности, что $p_{ii} = 0$). Приняв это соглашение, мы получим возможность записывать плюккеровы координаты любой прямой в виде кососимметрической матрицы четвертого порядка:

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix} = \begin{pmatrix} 0 & p_{12} & p_{13} & p_{14} \\ -p_{12} & 0 & p_{23} & p_{24} \\ -p_{13} & -p_{23} & 0 & p_{34} \\ -p_{14} & -p_{24} & -p_{34} & 0 \end{pmatrix}.$$

§ 3. ГЕОМЕТРИЯ ОКРУЖНОСТЕЙ НА ПЛОСКОСТИ

1. Степень точки относительно окружности

Мы знаем, что в прямоугольных координатах Oxy уравнение окружности радиуса R с центром в точке (a, b) имеет вид

$$(x-a)^2 + (y-b)^2 = R^2.$$

Раскрыв скобки в этом уравнении и положив

$$c = \frac{R^2 - a^2 - b^2}{2}, \tag{1}$$

мы приведем уравнение окружности к виду

$$x^2 + y^2 - 2ax - 2by - 2c = 0. \tag{2}$$

Определение 1. Для любой точки $M_0(x_0, y_0)$ число, получающееся в результате подстановки координат этой точки в левую часть уравнения (2), называется *степенью* точки M_0 относительно данной окружности и обозначается символом $\mathrm{ст}\, M_0$. Таким образом, по определению,

$$\mathrm{ст}\, M_0 = x_0^2 + y_0^2 - 2ax_0 - 2by_0 - 2c = (x_0 - a)^2 + (y_0 - b)^2 - R^2. \tag{3}$$

Ясно, что

$\mathrm{ст}\, M_0 < 0 \Leftrightarrow$ *точка* M_0 *лежит внутри окружности;*
$\mathrm{ст}\, M_0 = 0 \Leftrightarrow$ *точка* M_0 *принадлежит окружности;*
$\mathrm{ст}\, M_0 > 0 \Leftrightarrow$ *точка* M_0 *лежит вне окружности.*

В частности,

степень центра окружности отрицательна и равна $-R^2$, *где* R *— радиус окружности.*

Степень начала координат (0, 0) равна, очевидно, $-2c$.

Степень допускает красивое геометрическое истолкование. Чтобы получить его, мы рассмотрим произвольную прямую

$$\begin{aligned} x &= x_0 + lt, \\ y &= y_0 + mt, \end{aligned} \tag{4}$$

проходящую через точку $M_0(x_0, y_0)$. Будем, для упрощения вычислений, считать, что направляющий вектор $\boldsymbol{a}(l, m)$ этой прямой является ортом (т. е. что $l^2 + m^2 = 1$). Точки пересечения прямой (4) с окружностью (2) определяются из уравнения

$$(x_0 + lt)^2 + (y_0 + mt)^2 - 2a(x_0 + lt) - 2b(y_0 + mt) - 2c = 0,$$

т. е. из уравнения

$$t^2 + 2[l(x_0 - a) + m(y_0 - b)]t + \text{ст}\, M_0 = 0. \qquad (5)$$

Пусть прямая (4) на самом деле пересекает окружность, т. е. пусть уравнение (5) имеет вещественные (быть может, совпадающие) корни t_1 и t_2. Через эти корни координаты точек пересечения $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$ выражаются формулами

$$x_1 = x_0 + lt_1, \quad x_2 = x_0 + lt_2,$$
$$y_1 = y_0 + mt_1, \quad y_2 = y_0 + mt_2.$$

Рассмотрим теперь векторы $\overrightarrow{M_0M_1}$ и $\overrightarrow{M_0M_2}$. Ясно, что

$$\overrightarrow{M_0M_1} = t_1\boldsymbol{a}, \quad \overrightarrow{M_0M_2} = t_2\boldsymbol{a}.$$

Следовательно, в ориентации прямой (4), определенной вектором $\boldsymbol{a}$, величина (см. п. 2 § 2 гл. 1) вектора $\overrightarrow{M_0M_1}$ равна t_1, а величина вектора $\overrightarrow{M_0M_2}$ равна t_2 (напомним, что вектор $\boldsymbol{a}$ мы предполагаем единичным). Поэтому согласно формуле Вьета для произведения корней квадратного уравнения произведение t_1t_2 величин этих векторов равно свободному члену уравнения (5), т. е. равно степени $\text{ст}M_0$ точки M_0. Тем самым мы доказали следующее

Предложение 1. *Степень точки M_0 относительно окружности равна произведению величин векторов $\overrightarrow{M_0M_1}$ и $\overrightarrow{M_0M_2}$, где M_1 и M_2 — точки пересечения окружности с произвольной прямой, проходящей через точку M_0.*

Заметим, что хотя величины векторов на прямой зависят от ориентации этой прямой (от выбора направляющего вектора), их произведение от этой ориентации уже не зависит.

В частности, мы видим, что

для любой прямой, проходящей через точку M_0 и пересекающей данную окружность, произведение длин отрезков $\overline{M_0M_1}$ и $\overline{M_0M_2}$ не зависит от прямой.

Этот факт должен быть известен читателю из школьного курса геометрии («теорема о произведении длин секущих»).

При $t_1 = t_2$, т. е. когда прямая (4) пересекает окружность только в одной точке $M_1 = M_2$ (и потому *касается* ее в этой точке), мы получаем

Следствие. *Степень точки M_0 относительно окружности равна квадрату длины касательной, проведенной из точки M_0 к окружности.*

Здесь предполагается, что точка M_0 лежит вне окружности (поскольку только из такой точки можно провести к окружности касательную).

Тот факт, что квадрат длины касательной равен произведению длин секущих, также должен быть известен читателю из школьного курса геометрии.

Пусть нам даны две неконцентрические окружности

$$\Sigma_1\colon\ x^2+y^2-2a_1x-2b_1y-2c_1=0$$

и

$$\Sigma_2\colon\ x^2+y^2-2a_2x-2b_2y-2c_2=0.$$

Степень точки $M(x, y)$ относительно первой окружности мы будем обозначать через $\mathrm{ст}_1M$, а относительно второй — через $\mathrm{ст}_2M$. Рассмотрим все точки $M(x, y)$, имеющие относительно этих окружностей равные степени

$$\mathrm{ст}_1 M=\mathrm{ст}_2 M.$$

Поскольку

$$\mathrm{ст}_1 M=x^2+y^2-2a_1x-2b_1y-2c_1,$$
$$\mathrm{ст}_2 M=x^2+y^2-2a_2x-2b_2y-2c_2,$$

координаты этих точек удовлетворяют соотношению

$$(a_2-a_1)\,x+(b_2-b_1)\,y+(c_2-c_1)=0. \tag{6}$$

Так как по условию $(a_1, b_1)\neq(a_2, b_2)$, равенство (6) является уравнением некоторой прямой.

Определение 2. Прямая (6) называется *радикальной осью* окружностей Σ_1 и Σ_2.

Прямая, соединяющая центры окружностей Σ_1 и Σ_2, имеет уравнение

$$\frac{x-a_1}{a_2-a_1}=\frac{y-b_1}{b_2-b_1}$$

и, очевидно, перпендикулярна прямой (6). Таким образом,

радикальная ось двух (неконцентрических) окружностей перпендикулярна прямой, соединяющей их центры.

Замечание 1. Если от евклидовой плоскости (в которой мы до сих пор работали) перейти к евклидово-проективной плоскости (см. п. 2 § 1), то уравнение (6) (переписанное, конечно, в евклидово-проективных координатах) будет иметь смысл и для концентрических (не совпадающих) окружностей и будет в этом случае уравнением несобственной прямой. Таким образом, мы можем считать, что

радикальной осью концентрических (несовпадающих) окружностей является несобственная прямая.

Для совпадающих окружностей уравнение (6) вырождается в тождество, которому удовлетворяет каждая точка плоскости.

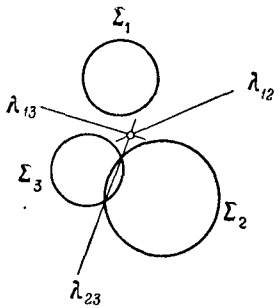
Предложение 2. *Радикальной осью двух пересекающихся окружностей является прямая, проходящая через их точки пересечения, а радикальной осью двух касающихся окружностей является их общая касательная.*

Доказательство. Если окружности пересекаются, то степени их общих точек относительно обеих окружностей равны нулю и потому одинаковы. Поэтому радикальная ось проходит через эти точки. Если же окружности касаются, то их общая касательная проходит через точку касания и перпендикулярна прямой, соединяющей центры окружностей. Поэтому она является радикальной осью.

Обратно, если радикальная ось двух окружностей пересекает одну из этих окружностей, то каждая точка пересечения принадлежит второй окружности (ибо степень этой точки относительно второй окружности равна ее степени относительно первой окружности и потому равна нулю). Следовательно,

радикальная ось двух непересекающихся окружностей лежит вне этих окружностей (их не пересекает).

Рассмотрим три попарно неконцентрические окружности Σ_1, Σ_2 и Σ_3, центры O_1, O_2 и O_3 которых не коллинеарны. Пусть



λ_{12} — радикальная ось окружностей Σ_1 и Σ_2, а λ_{23} — радикальная ось окружностей Σ_2 и Σ_3. Поскольку прямые O_1O_2 и O_2O_3, по условию, различны (и не параллельны), а прямые λ_{12} и λ_{23} им перпендикулярны, то прямые λ_{12} и λ_{23} пересекаются в единственной точке M_0. Для этой точки выполнены равенства

$$\text{ст}_1 M_0 = \text{ст}_2 M_0, \quad \text{ст}_2 M_0 = \text{ст}_3 M_0.$$

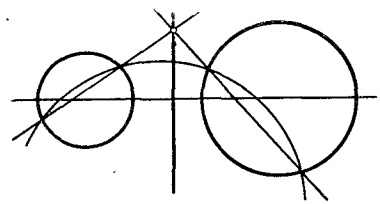
Поэтому

$$\text{ст}_1 M_0 = \text{ст}_3 M_0,$$

т. е. точка M_0 принадлежит и радикальной оси λ_{13} окружностей Σ_1 и Σ_3.

Таким образом, мы доказали, что

для любых трех окружностей Σ_1, Σ_2 и Σ_3 с неколлинеарными центрами три радикальные оси каждой пары окружностей проходят через одну точку.



Определение 3. Точка пересечения радикальных осей λ_{12}, λ_{23}, λ_{13} окружностей Σ_1, Σ_2, Σ_3 называется *радикальным центром* этих окружностей.

Из доказанного утверждения вытекает простой способ построения радикальной оси λ_{12} двух непересекающихся (и неконцентричных) окружностей Σ_1 и Σ_2: достаточно провести произвольную окружность Σ_3, пересекающую окружности Σ_1 и Σ_2,

построить радикальные оси λ_{13} и λ_{23} (см. предложение 2) и провести через точку их пересечения прямую, перпендикулярную прямой, соединяющей центры окружностей Σ_1 и Σ_2. Эта прямая и будет радикальной осью λ_{12}.

Если центры окружностей Σ_1, Σ_2 и Σ_3 коллинеарны, то радикальные оси λ_{12}, λ_{23} и λ_{13} параллельны, так что для таких окружностей радикального центра не существует.

На евклидово-проективной плоскости мы можем считать (в случае, когда все оси λ_{12}, λ_{23} и λ_{13} различны), что радикальным центром окружностей с коллинеарными центрами является несобственная точка прямой, перпендикулярной прямой, соединяющей их центры.

Если две из трех радикальных осей совпадают, то третья ось также с ними совпадает. В этом случае говорят, что окружности Σ_1, Σ_2, Σ_3 *имеют общую радикальную ось*. Никакого определенного радикального центра (собственного или несобственного) в этом случае не имеется.

На евклидово-проективной плоскости три концентрические окружности также имеют общую радикальную ось (несобственную прямую).

2. Связки окружностей

Три числа a, b, c полностью определяют окружность

$$\Sigma\colon\ x^2 + y^2 - 2ax - 2by - 2c = 0.$$

Определение 1. Числа a, b, c называются *координатами* окружности Σ (в данной системе прямоугольных координат Oxy). Окружность с координатами a, b, c мы будем обозначать символом $\Sigma(a, b, c)$.

Выбрав в пространстве некоторую систему прямоугольных координат $Oxyz$, мы каждой окружности $\Sigma(a, b, c)$ можем сопоставить точку пространства, имеющую координаты (a, b, c). Будем называть эту точку точкой, *изображающей* данную окружность.

Поскольку радиус

$$R = \sqrt{a^2 + b^2 + 2c}$$

окружности $\Sigma(a, b, c)$ должен быть вещественным числом,

числа a, b, c тогда и только тогда являются координатами некоторой окружности, когда

$$a^2 + b^2 + 2c > 0. \tag{1}$$

Другими словами,

отображение

«окружность» $\longmapsto$ *«изображающая точка»*

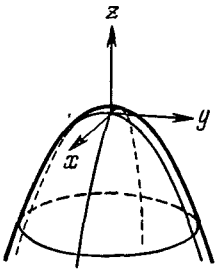
определяет биективное соответствие между множеством всех окружностей на плоскости и множеством всех точек пространства, координаты которых удовлетворяют неравенству

$$x^2 + y^2 + 2z > 0 \qquad (2)$$

Замечание 1. Рассмотрим в пространстве поверхность Q с уравнением

$$x^2 + y^2 + 2z = 0.$$

Эта поверхность называется *параболоидом вращения*. В следующей главе мы покажем, что он получается вращением параболы $y^2 = -2z$ (расположенной в плоскости Oyz) вокруг оси Oz. Этот параболоид разбивает все пространство на две области, одна из которых называется *внешней областью* и характеризуется неравенством $x^2 + y^2 + 2z > 0$, а вторая называется *внутренней областью* параболоида и характеризуется неравенством $x^2 + y^2 + 2z < 0$. Таким образом, мы можем сказать, что



Параболоид Q

точки, изображающие окружности, заполняют внешнюю область параболоида Q.

Определение 1. *Связкой окружностей* называется множество всех окружностей, изображающие точки которых принадлежат некоторой плоскости

$$Ax + By + Cz + D = 0. \qquad (3)$$

О плоскости (3) мы будем говорить, что она *изображает* связку. Связка называется *собственной*, если $C \neq 0$, т. е. если плоскость (3) не параллельна оси Oz. В противном случае связка называется *несобственной*.

Замечание 1. Не нужно думать, что любая точка плоскости (3) изображает некоторую окружность связки. Этим свойством обладают только точки, удовлетворяющие условию (2).

Другими словами, окружности связки изображаются точками плоскости (3), принадлежащими внешней области параболоида Q.

Для несобственных связок уравнение (3) означает, что центры (a, b) окружностей связки принадлежат прямой

$$Ax + By + D = 0.$$

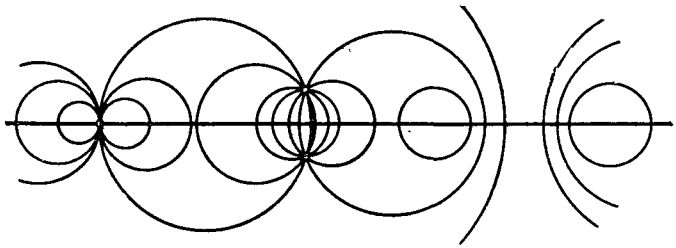
При этом на третью координату $z = c$ окружностей связки это уравнение никаких ограничений не накладывает. Этим доказано следующее

Предложение 1. *Каждая несобственная связка является множеством всех окружностей, центры которых принадлежат некоторой фиксированной прямой.*

Ясно, что обратное утверждение также справедливо, т. е.

множество всех окружностей, центры которых принадлежат данной прямой, является несобственной связкой.

Определение 2. Прямая, содержащая центры окружностей несобственной связки, называется *фундаментальной прямой* этой связки.



Несобственная связка

Чтобы получить пример собственной связки, мы вспомним, что степень p произвольной точки $M_0(a_0, b_0)$ плоскости относительно некоторой окружности $\Sigma(a, b, c)$ выражается, по определению, формулой

$$p = a_0^2 + b_0^2 - 2a_0a - 2b_0b - 2c.$$

Это показывает, что

множество всех окружностей, относительно которых данная точка $M_0(a_0, b_0)$ имеет данную степень p, является собственной связкой с уравнением

$$a_0x + b_0y + z + \frac{p - a_0^2 - b_0^2}{2} = 0. \qquad (4)$$

Оказывается, что таким образом можно задать любую собственную связку, т. е. имеет место следующее

Предложение 2. *Любая собственная связка* (3) *является множеством всех окружностей, относительно которых фиксированная точка $M_0(a_0, b_0)$ имеет фиксированную степень p.*

Доказательство. Чтобы привести уравнение (3) к виду (4), достаточно положить

$$a_0 = \frac{A}{C}, \quad b_0 = \frac{B}{C}, \quad p = \frac{A^2 + B^2 + 2CD}{C^2}. \qquad (5)$$

Заметим, что числа a_0, b_0 и p однозначно определены связкой (не меняются при умножении ее уравнения на любое число).

Определение 3. Точка $M_0(a_0, b_0)$ называется *центром* связки (5), а число p — ее *степенью.*

Заметим, что

центр и степень однозначно определяют связку.

На евклидово-проективной плоскости центр можно приписать и несобственным связкам, а именно, центром несобственной связки, изображающейся плоскостью

$$Ax + By + D = 0,$$

следует считать точку с однородными координатами $(A:B:0)$, т. е. несобственную точку прямых, перпендикулярных прямой фундаментальной связки.

Степень несобственной связки считается, по определению, бесконечной.

Из определений немедленно вытекает, что

если для трех окружностей связки радикальный центр существует, то он совпадает с центром связки, а если радикальный центр не существует (окружности имеют общую радикальную ось), то их радикальная ось проходит через центр связки.

На евклидово-проективной плоскости это утверждение верно и для несобственных связок.

Задание. Докажите, что

если из трех окружностей две принадлежат связке с центром в точке M_0 и если точка M_0 является радикальным центром этих окружностей (или лежит на их общей радикальной оси), то третья окружность также принадлежит связке.

В случае, когда точка $M_0(a_0, b_0)$ совпадает с началом координат $O(0,0)$, уравнение (4) сводится к равенству $z = -p/2$. Следовательно,

каждая окружность собственной связки степени p, имеющей центр в точке $O(0, 0)$, задается уравнением вида

$$x^2 + y^2 - 2ax - 2by + p = 0,$$

где a и b — параметры, определяющие окружность.

Эти параметры представляют собой не что иное, как координаты центра окружности. Они должны выбираться так, чтобы было выполнено условие (1), т. е. чтобы

$$a^2 + b^2 > p. \tag{6}$$

Таким образом,

каждая точка плоскости, координаты (a, b) которой удовлетворяют неравенству (6), является центром единственной окружности, принадлежащей собственной связке степени p с центром в точке $O(0, 0)$.

Этим собственные связки отличаются от несобственных (центры окружностей которых принадлежат фундаментальной прямой и не заполняют никакой части плоскости).

Определение 4. Собственная связка называется

эллиптической, *параболической*, *гиперболической*, если ее степень p { отрицательна: $p < 0$, равна нулю: $p = 0$, положительна: $p > 0$.

Из неравенства (6) немедленно вытекает следующее

Предложение 3. *Если собственная связка*

а) *эллиптична,*

б) *параболична,*

в) *гиперболична,*

то центры ее окружностей заполняют

а) *всю плоскость,*

б) *всю плоскость, за исключением центра связки,*

в) *всю плоскость, за исключением точек, не лежащих вне окружности радиуса* $\sqrt{p}$ *с центром в центре связки.*

Определение 5. При $p > 0$ окружность радиуса $\sqrt{p}$ с центром в центре связки называется *фундаментальной окружностью* гиперболической связки. Аналогично, при $p < 0$ окружность радиуса $\sqrt{-p}$ с центром в центре связки называется *главной окружностью* эллиптической связки.

Ясно, что

фундаментальная (главная) окружность гиперболической (эллиптической) связки однозначно определяет эту связку.

При этом

любая окружность является фундаментальной (главной) окружностью некоторой гиперболической (эллиптической) связки.

Заметим, что

главная окружность эллиптической связки принадлежит этой связке, а фундаментальная окружность гиперболической связки этой связке не принадлежит.

Так как каждая точка является центром некоторой окружности эллиптической связки, то

каждая точка плоскости, изображающей эллиптическую связку, изображает некоторую окружность связки.

Таким образом, плоскость, изображающая эллиптическую связку, целиком лежит вне параболоида Q, т. е. его *не пересекает*.

Аналогично,

каждая точка плоскости, изображающей параболическую связку, изображает, за исключением одной точки, некоторую окружность связки.

Эта исключительная точка характеризуется тем, что при проектировании параллельно оси Oz на плоскость Oxy она переходит в центр $M_0(a_0, b_0)$ связки, т. е. тем, что ее координаты имеют вид (a_0, b_0, c_0), где c_0 — некоторое число. Подставив координаты (a_0, b_0, c_0) в уравнение связки (4) (при $p = 0$), мы немедленно получим, что

$$a_0^2 + b_0^2 + 2c_0 = 0.$$

Таким образом, исключительная точка (a_0, b_0, c_0) принадлежит параболоиду Q (и является единственной точкой рассматриваемой плоскости, лежащей на параболоиде Q). Это, по определению, означает, что плоскость, изображающая параболическую связку, *касается* параболоида Q (в точке (a_0, b_0, c_0)).

Рассмотрим теперь гиперболические связки (задаваемые уравнением (4) при $p < 0$).

Так же, как и выше, мы немедленно получаем, что

среди точек плоскости, изображающей гиперболическую связку, только точки, проектирующиеся в точки, лежащие вне фундаментальной окружности связки, изображают окружности связки.

Координаты (x, y, z) точек, проектирующихся в точки фундаментальной окружности, должны удовлетворять уравнению связки (4), а первые две координаты (x, y) — уравнению фундаментальной окружности

$$(x - a_0)^2 + (y - b_0)^2 = p,$$

т. е. уравнению

$$x^2 + y^2 - 2a_0x - 2b_0y + a_0^2 + b_0^2 - p = 0.$$

Прибавив к этому уравнению уравнение (4), умноженное на 2, мы получим уравнение

$$x^2 + y^2 + 2z = 0. \tag{7}$$

Аналогично показывается, что точки, проектирующиеся внутрь фундаментальной окружности, удовлетворяют соотношению

$$x^2 + y^2 + 2z < 0. \tag{8}$$

Таким образом, первые точки принадлежат параболоиду Q, а вторые лежат в его внутренней области. Другими словами, плоскость, изображающая гиперболическую связку, *пересекает* гиперболоид Q.

Условие (7) выделяет на плоскости (4) некоторую замкнутую линию [1]), а условие (8) выделяет точки, лежащие внутри этой линии.

Собирая вместе все доказанное (и используя принцип обращения), мы получаем следующую теорему.

Теорема 1. *Собственная связка окружностей тогда и только тогда является*

а) *эллиптической связкой,*

б) *параболической связкой,*

в) *гиперболической связкой,*

когда изображающая ее плоскость

а) *не пересекает параболоид Q,*

б) *касается этого параболоида (и в этом случае точка касания проектируется в центр связки на плоскости Oxy),*

в) *пересекает параболоид Q по замкнутой линии, проектирующейся в фундаментальную окружность связки.*

---

[1]) С помощью результатов п. 2 § 1 гл. 5 легко показать, что эта линия является эллипсом. Впрочем, это непосредственно вытекает (см. п. 8 § 2 гл. 5) также из того, что она является образом некоторой окружности (фундаментальной окружности связки) при параллельном проектировании одной плоскости (плоскости Oxy) на другую плоскость (плоскость, изображающую связку).

Замечание 2. В частности, мы видим, что

для плоскости в пространстве, не параллельной оси Oz, возможны следующие три взаимно исключающие друг друга случая:

1) *плоскость не пересекает параболоид Q (и тогда целиком лежит в его внешней области);*

2) *плоскость касается этого параболоида (и тогда вся, за исключением точки касания, лежит во внешней области);*

3) *плоскость пересекает параболоид по некоторой линии (и тогда имеет точки как во внешней, так и во внутренней областях параболоида Q).*

Упражнение. Дайте прямое доказательство этого утверждения. (Указание: перейдите к координатам, для которых данная плоскость является координатной плоскостью.)

Задание. Докажите, что плоскости, параллельные оси *Oz*, т. е. изображающие несобственные связки, также пересекают параболоид *Q* по некоторой линии, но уже незамкнутой (параболе).

Чтобы получить наглядное описание собственных связок, нам понадобится следующее

Определение 6. Две окружности $\Sigma_0 = \Sigma(a_0, b_0, c_0)$ и $\Sigma = \Sigma(a, b, c)$ называются *ортогональными*, если они пересекаются под прямым углом, т. е. если они имеют две общие точки и их радиусы в этих точках перпендикулярны. Говорят, что окружность Σ *диаметрально пересекает* окружность Σ_0, если либо $\Sigma = \Sigma_0$, либо окружности Σ и Σ_0 пересекаются и их точки пересечения являются диаметрально противоположными точками окружности Σ_0.

Пусть R_0 и R — радиусы окружностей Σ_0 и Σ, а d — расстояние между их центрами. Очевидные элементарно-геометрические соображения показывают, что окружности тогда и только тогда ортогональны, когда

$$d^2 = R_0^2 + R^2.$$

Но

$$d^2 = (a - a_0)^2 + (b - b_0)^2 = a^2 + b^2 + a_0^2 + b_0^2 - 2(a_0 a + b_0 b)$$

и

$$R_0^2 = a_0^2 + b_0^2 + 2c, \quad R^2 = a^2 + b^2 + 2c.$$

Следовательно,

окружности $\Sigma(a_0, b_0, c_0)$ и $\Sigma(a, b, c)$ тогда и только тогда ортогональны, когда

$$a_0 a + b_0 b + c + c_0 = 0. \tag{9}$$

Отсюда непосредственно вытекает следующее

Предложение 4. *Множество всех окружностей, ортогональных данной окружности $\Sigma(a_0, b_0, c_0)$, является собственной связкой с уравнением*

$$a_0 x + b_0 y + z + c_0 = 0. \tag{10}$$

Далее, ясно, что окружность $\Sigma(a, b, c)$ тогда и только тогда диаметрально пересекает окружность $\Sigma(a_0, b_0, c_0) \neq \Sigma(a, b, c)$, когда радикальная ось

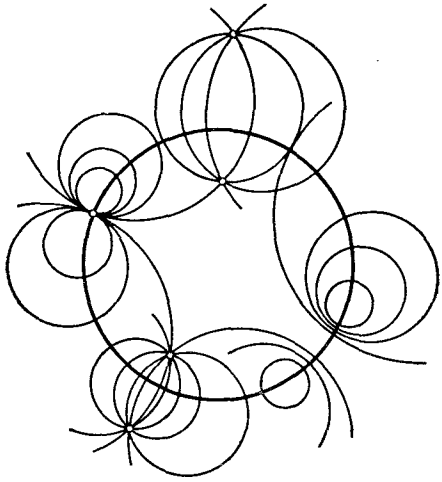
$$(a - a_0)x + (b - b_0)y + (c - c_0) = 0$$

этих окружностей проходит через центр (a_0, b_0) окружности $\Sigma(a_0, b_0, c_0)$ (ибо радикальной осью двух пересекающихся окружностей является прямая, проходящая через их точки пересечения), т. е. когда имеет место соотношение

$$(a - a_0)a_0 + (b - b_0)b_0 + (c - c_0) = 0,$$

равносильное соотношению

$$a_0a + b_0b + c - (a_0^2 + b_0^2 + c_0) = 0. \quad (11)$$

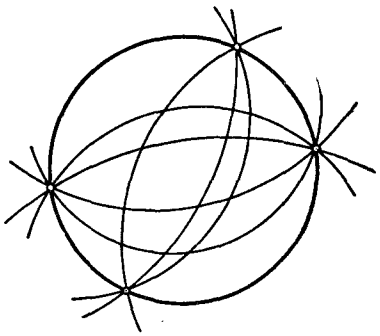


Гиперболическая связка

Поскольку при $\Sigma(a_0, b_0, c_0) = \Sigma(a, b, c)$ это соотношение также справедливо, тем самым доказано следующее

Предложение 5. *Множество всех окружностей, диаметрально пересекающих данную окружность* $\Sigma(a_0, b_0, c_0)$, *является собственной связкой с уравнением*

$$a_0x + b_0y + z - (a_0^2 + b_0^2 + c_0) = 0. \quad (12)$$



Эллиптическая связка

Еще один пример собственной связки мы получим, рассмотрев для произвольной точки $M_0(a_0, b_0)$ множество всех окружностей, проходящих через эту точку. Условие, что окружность

$$x^2 + y^2 - 2ax - 2by - 2c = 0$$

проходит через точку $M_0(a_0, b_0)$, выражается равенством

$$a_0^2 + b_0^2 - 2aa_0 - 2bb_0 - 2c = 0.$$

Следовательно, имеет место

Предложение 6. *Множество всех окружностей, проходящих через фиксированную точку* $M_0(a_0, b_0)$, *является собственной связкой с уравнением*

$$a_0x + b_0y + z - \frac{a_0^2 + b_0^2}{2} = 0. \quad (13)$$

Вычисляя по формулам (5) центр и степень связок (10), (12) и (13), мы немедленно получим следующее

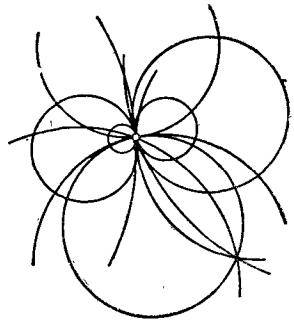
Предложение 7. *Центром связки* (10) *является центр* $M_0(a_0, b_0)$ *окружности* $\Sigma(a_0, b_0, c_0)$, *а ее степень выражается формулой*

$$p = a_0^2 + b_0^2 + 2c_0 = R_0^2$$

и, следовательно, положительна (*связка гиперболична*).

Центром связки (12) *является центр* $M_0(a_0, b_0)$ *окружности* $\Sigma(a_0, b_0, c_0)$, *а ее степень* p *выражается формулой*

$$p = a_0^2 + b_0^2 - 2(a_0^2 + b_0^2 + c_0) = \\ = 2c_0 - a_0^2 - b_0^2 = -R_0^2$$



Параболическая связка

и, следовательно, отрицательна (*связка эллиптична*).

Центром связки (13) *является точка* $M_0(a_0, b_0)$, *а ее степень равна нулю*:

$$p = (a_0^2 + b_0^2) - (a_0^2 + b_0^2) = 0$$

(*связка параболична*).

Кроме того, мы видим, что

окружность $\Sigma(a_0, b_0, c_0)$ *является фундаментальной окружностью связки* (10) *и главной окружностью связки* (12).

При этом ясно, что в виде (10) можно представить любую гиперболическую связку (достаточно принять за окружность $\Sigma(a_0, b_0, c_0)$ фундаментальную окружность связки), в виде (12) — любую эллиптическую связку (достаточно за окружность $\Sigma(a_0, b_0, c_0)$ принять главную окружность связки), в виде (13) — любую параболическую связку (достаточно за точку $M_0(a_0, b_0)$ принять центр связки).

Резюмируя, мы получаем следующую теорему:

Теорема 2. *Связка окружностей тогда и только тогда является*

а) *эллиптической связкой,*

б) *параболической связкой,*

в) *гиперболической связкой,*

г) *несобственной связкой,*

когда она состоит из окружностей,

а) *диаметрально пересекающих некоторую окружность* (*являющуюся главной окружностью связки*),

б) *проходящих через некоторую точку* (*являющуюся центром связки*),

в) *ортогонально пересекающих некоторую окружность* (*являющуюся фундаментальной окружностью связки*),

г) *имеющих центры на данной прямой.*

3. Пучки окружностей

Определение 1. *Пучком окружностей* называется множество всех окружностей, изображающие точки которых принадлежат некоторой прямой

$$\frac{x-a_0}{A}=\frac{y-b_0}{B}=\frac{z-c_0}{C} \tag{1}$$

в пространстве. Об этой прямой мы будем говорить, что она *изображает* пучок. Если $A=B=0$, т. е. если прямая (1) параллельна оси Oz, пучок окружностей называется *несобственным*, а в противном случае — *собственным*.

Поскольку прямая однозначно определяется любыми ее двумя (различными) точками, *пучок окружностей однозначно определен, когда известны две его окружности.*

Координаты центров окружностей пучка удовлетворяют первому из уравнений (1), т. е. уравнению

$$\frac{x-a_0}{A}=\frac{y-b_0}{B}. \tag{2}$$

Для собственного пучка это уравнение выражает собой прямую.

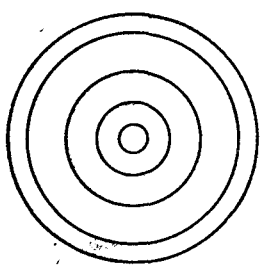
Определение 2. Прямая (2) называется *линией центров* собственного пучка.

Соглашение. Мы раз и навсегда условимся, что координаты x, y, z в пространстве *согласованы* с координатами x, y на рассматриваемой нами плоскости в том смысле, что эта плоскость является координатной плоскостью Oxy (и ее точка (x, y) имеет в пространстве координаты $(x, y, 0)$).

В силу этого соглашения,

линия центров собственного пучка является проекцией на плоскость Oxy прямой (1), *изображающей в пространстве этот пучок.*

Для несобственного пучка линия центров вырождается в одну точку $M_0(a_0, b_0)$. Следовательно, все окружности несобственного пучка концентричны. Более того, ясно, что для несобственного пучка уравнение (1) не накладывает никаких ограничений на третью координату $z=c$. Это означает, что справедливо следующее



Несобственный пучок

Предложение 1. *Каждый несобственный пучок состоит из всех окружностей, имеющих данный центр $M_0(a_0, b_0)$ (или, по-другому: из всех окружностей, концентрических данной).*

Определение 3. Общий центр M_0 всех окружностей несобственного пучка называется его *центром.*

Ясно, что

пучок окружностей тогда и только тогда несобствен, когда он является пересечением двух несобственных связок с непа-

раллельными фундаментальными прямыми (точка пересечения этих прямых и будет центром пучка).

Чтобы получить пример собственного пучка, мы рассмотрим на плоскости произвольную прямую

$$Ax + By + C = 0 \tag{3}$$

и произвольную окружность $\Sigma_0 = \Sigma(a_0, b_0, c_0)$. Как мы знаем (см. п. 1), для любой окружности $\Sigma = \Sigma(a, b, c)$ (не концентричной с окружностью Σ_0) радикальной осью окружностей Σ и Σ_0 является прямая

$$(a - a_0)x + (b - b_0)y + (c - c_0) = 0. \tag{4}$$

Поэтому радикальная ось окружностей Σ и Σ_0 тогда и только тогда совпадает с данной прямой (4), когда

$$\frac{a - a_0}{A} = \frac{b - b_0}{B} = \frac{c - c_0}{C}.$$

Это показывает, что

множество, состоящее из окружности $\Sigma(a_0, b_0, c_0)$ и всех окружностей, имеющих с этой окружностью данную радикальную ось (3), *является пучком с уравнением*

$$\frac{x - a_0}{A} = \frac{y - b_0}{B} = \frac{z - c_0}{C}.$$

Сравнивая это уравнение с уравнением (1), мы немедленно получаем следующее

Предложение 2. *Каждый собственный пучок является множеством всех окружностей, состоящим из некоторой окружности $\Sigma(a_0, b_0, c_0)$ и всех окружностей, имеющих с этой окружностью данную радикальную ось* (3).

Отсюда вытекает

Следствие. *Три (неконцентрические) окружности Σ_1, Σ_2 и Σ_3 тогда и только тогда принадлежат одному пучку (изображающие их точки коллинеарны), когда они имеют общую радикальную ось.*

Доказательство. Если окружности Σ_1, Σ_2 и Σ_3 принадлежат пучку (1), то радикальной осью каждой из них и окружности Σ_0 является прямая (3). Поэтому радикальная ось любых двух из этих окружностей совпадает с прямой (3). Обратно, если окружности Σ_1, Σ_2 и Σ_3 имеют общую радикальную ось, то, приняв за прямую (3) эту ось, а за окружность Σ_0 — одну из данных окружностей (например, окружность Σ_1), мы, очевидно, получим пучок, содержащий все три окружности.

Определение 4. Прямая (3) называется *осью* собственного пучка (1). Она перпендикулярна его линии центров (2). Точка пересечения оси и линии центров называется *центром* собственного пучка. Поскольку центр принадлежит оси пучка, его

степень относительно любой окружности пучка — одна и та же для всех окружностей. Эта степень p называется *степенью* пучка.

На евклидово-проективной плоскости осью каждого несобственного пучка можно считать несобственную прямую. Степень несобственного пучка не определяется (при желании ее можно считать бесконечной).

Ясно, что

собственный пучок однозначно определен, если известны его ось, линия центров и степень.

Все другие способы задания пучка легко сводятся к этому. Например, при задании пучка двумя (неконцентрическими) окружностями линия центров определяется как прямая, проходящая через центры данных окружностей, ось — как радикальная ось этих окружностей, а степень — как степень точки пересечения линии центров и оси относительно любой из данных окружностей.

Решив совместно уравнения (2) и (3) линии центров и оси пучка, мы получим, что

центр пучка (1) *имеет координаты*

$$x_0 = -\frac{AC + B(Ab_0 - Ba_0)}{A^2 + B^2}, \quad y_0 = -\frac{BC - A(Ab_0 - Ba_0)}{A^2 + B^2}. \tag{5}$$

Подставив эти выражения в уравнение произвольной окружности пучка, мы, после тривиальных алгебраических преобразований, получим, что

степень пучка (1) *равна*

$$p = \left(\frac{Aa_0 + Bb_0 + C}{\sqrt{A^2 + B^2}}\right)^2 - (a_0^2 + b_0^2 + 2c_0). \tag{6}$$

Задание. Проведите подробно все вычисления (помните, что не любая точка прямой (1) изображает окружность; в частности, точка (a_0, b_0, c_0) может окружности не изображать).

Замечание 1. В предположении, что точка (a_0, b_0, c_0) изображает некоторую окружность, формулу (6) можно доказать и более геометрическим способом. Действительно, как мы знаем (п. 5 § 1 гл. 3), число

$$d_0^2 = \left(\frac{Aa_0 + Bb_0 + C}{\sqrt{A^2 + B^2}}\right)^2$$

является квадратом расстояния точки (a_0, b_0) от прямой $Ax + By + C = 0$ (оси пучка). Но поскольку точка (a_0, b_0) принадлежит линии центров, перпендикулярной оси пучка и пересекающей эту ось в центре пучка, число d_0^2 равно квадрату расстояния центра пучка от точки (a_0, b_0), т. е. до центра окружности $\Sigma(a_0, b_0, c_0)$. Остается заметить, что степень p центра пучка относительно окружности $\Sigma(a_0, b_0, c_0)$, по определению, равна $d_0^2 - R_0^2$.

Определение 5. Систему прямоугольных координат x, y мы будем называть *канонической* относительно рассматриваемого

(собственного) пучка окружностей, если ее осью абсцисс является линия центров пучка, а осью ординат — ось пучка (и значит, началом координат — центр пучка).

Каноническая система координат характеризуется, таким образом, тем, что уравнение (1), прямой, изображающей этот пучок, имеет (в согласованной системе координат $Oxyz$) вид

$$\frac{x-a_0}{1}=\frac{y-0}{0}=\frac{z-c_0}{0}.$$

Формула (6) в канонической системе координат приобретает вид

$$p=-2c_0.$$

Обозначая a_0 через t, мы получаем, таким образом, следующее предложение, более или менее удовлетворительно описывающее все окружности собственного пучка:

Предложение 3. *В системе координат, канонической относительно собственного пучка степени p, уравнение произвольной окружности этого пучка имеет вид*

$$x^2+y^2-2xt+p=0, \tag{7}$$

где t — параметр, определяющий данную окружность. Центром этой окружности является точка $(t, 0)$.

Мы видим, в частности, что с точностью до выбора системы координат существует только один пучок степени p.

Наглядно это означает, что

любые два пучка одной и той же степени можно перевести друг в друга некоторым движением.

Определение 6. Собственный пучок окружностей называется

эллиптическим, *параболическим*, *гиперболическим*, если его степень p отрицательна: $p=-e^2$, равна нулю: $p=0$, положительна: $p=e^2$.

Поскольку радиус окружности (7) равен, очевидно,

$$\sqrt{t^2-p}=\begin{cases}\sqrt{t^2+e^2}, & \text{если пучок эллиптичен,}\\ |t|, & \text{если пучок параболичен,}\\ \sqrt{t^2-e^2}, & \text{если пучок гиперболичен,}\end{cases}$$

справедливо следующее

Предложение 4. *Если собственный пучок*

а) *эллиптичен,*

б) *параболичен,*

в) *гиперболичен,*

то центры его окружностей заполняют

а) *всю линию центров,*

б) *всю линию центров, за исключением центра пучка,*

в) *всю линию центров, за исключением точек отрезка, концами которого являются точки, удаленные от центра пучка на расстояние* $e = \sqrt{p}$.

Определение 7. При $p > 0$ точки, имеющие канонические координаты $(-e, 0)$ и $(e, 0)$, называются *фундаментальными точками* гиперболического пучка. Аналогично, при $p < 0$ точки, имеющие канонические координаты $(0, e)$ и $(0, -e)$, называются *главными точками* эллиптического пучка.

Ясно, что

фундаментальные (главные) точки гиперболического (эллиптического) пучка однозначно определяют этот пучок.

При этом

любые две точки являются фундаментальными (главными) точками некоторого (однозначно определенного) гиперболического (эллиптического) пучка.

Заметим еще, что

любая окружность эллиптического пучка проходит через его главные точки и никакая окружность гиперболического пучка не проходит через его фундаментальные точки.

Действительно, точки $(0, e)$ и $(0, -e)$ удовлетворяют при любом t уравнению

$$x^2 + y^2 - 2xt - e^2 = 0,$$

а точки $(-e, 0)$ и $(e, 0)$ удовлетворяют уравнению

$$x^2 + y^2 - 2xt + e^2 = 0$$

только при $t = \pm e$.

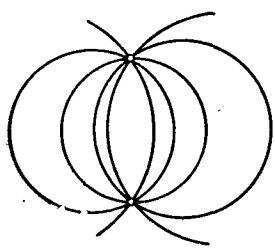
В частности, мы видим, что справедливо следующее

Предложение 5. *Любые две окружности эллиптического пучка пересекаются (в его главных точках).*

Более того, ясно, что

любая окружность Σ, *проходящая через главные точки эллиптического пучка, принадлежит этому пучку.*

Действительно, пусть (a, b) — центр окружности Σ. Так как главные точки симметричны относительно линии центров, то точка (a, b) принадлежит этой линии и, следовательно, является центром некоторой окружности Σ' пучка. Окружности Σ и Σ' концентричны и имеют общие точки (обе главные точки пучка). Следовательно, $\Sigma' = \Sigma$, так что окружность Σ принадлежит пучку.



Эллиптический пучок

Мы сформулируем доказанное утверждение в виде следующей теоремы:

Теорема 1. *Каждый эллиптический пучок состоит из всех окружностей, проходящих через две фиксированные (различные) точки (главные точки пучка).*

Что касается гиперболических пучков, то для них справедливо следующее

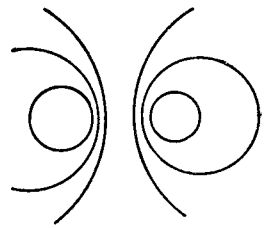
Предложение 6. *Никакие две (различные) окружности гиперболического пучка не имеют общих точек (не пересекаются).*

Доказательство. Чтобы найти общие точки двух окружностей

$$\begin{aligned} x^2 + y^2 - 2xt_1 + e^2 = 0, \\ x^2 + y^2 - 2xt_2 + e^2 = 0 \end{aligned} \qquad (8)$$

гиперболического пучка, надо совместно решить эти два уравнения. Но, вычитая одно уравнение из другого, мы при $t_1 \neq t_2$ немедленно получим, что $x = 0$, и следовательно, что

$$y^2 + e^2 = 0.$$



Гиперболический пучок

Следовательно, уравнения (8) решений (вещественных) не имеют, и потому соответствующие окружности не пересекаются.

Наконец, для параболических пучков имеет место

Предложение 7. *Каждая окружность параболического пучка проходит через его центр и касается в этой точке оси пучка, так что любые две окружности параболического пучка касаются (в его центре).*

Доказательство. В канонических координатах произвольная окружность параболического пучка имеет уравнение вида

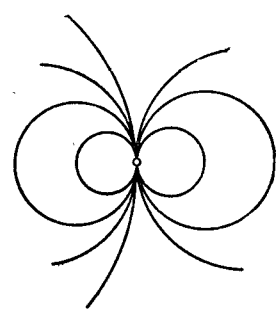
$$x^2 + y^2 - 2xt = 0,$$

т. е. уравнение

$$(x - t)^2 + y^2 = t^2.$$

Следовательно, она проходит через точку (0, 0) (центр пучка), а так как ее центр $(t, 0)$ лежит на оси абсцисс, то ось ординат (ось пучка) является касательной в точке (0, 0) к этой окружности.

Задание. Докажите, что
любая окружность, касающаяся в центре параболического пучка его оси, принадлежит пучку.



Параболический пучок

Таким образом, имеет место следующая теорема:

Теорема 2. *Каждый параболический пучок состоит из всех окружностей, проходящих через фиксированную точку (центр пучка) и касающихся в этой точке фиксированной прямой (оси пучка).*

Сопоставим теперь предложения 5, 6, 7 и воспользуемся принципом обращения. В результате мы получим следующую теорему:

Теорема 3. *Собственный пучок окружностей тогда и только тогда является*

а) *эллиптическим пучком,*

б) *параболическим пучком,*

в) *гиперболическим пучком,*

когда его любые две окружности

а) *пересекаются,*

б) *касаются,*

в) *не имеют ни одной общей точки.*

Так как каждая точка линии центров эллиптического пучка является, согласно предложению 4, центром некоторой окружности пучка, то

каждая точка прямой, изображающей эллиптический пучок, является изображением некоторой окружности пучка.

Таким образом, прямая, изображающая эллиптический пучок, лежит во внешней области параболоида Q и, следовательно, не пересекает этот параболоид.

Аналогично,

каждая точка прямой, изображающей параболический пучок, за исключением одной точки (проектирующейся в центр пучка), является изображением некоторой окружности пучка.

Эта исключительная точка имеет координаты (a_0, b_0, c_0), где (a_0, b_0) — координаты центра пучка. Так как $p = 0$ и $Aa_0 + Bb_0 + C = 0$, то из формулы (6) следует, что

$$a_0^2 + b_0^2 + 2c_0 = 0.$$

Таким образом, точка (a_0, b_0, c_0) принадлежит параболоиду Q. Поскольку все остальные точки прямой, изображающей параболический пучок, лежат во внешней области параболоида Q, это, по определению, означает, что прямая, изображающая параболический пучок, *касается* параболоида Q (в точке (a_0, b_0, c_0)).

Наконец,

точка прямой, изображающей гиперболический пучок, тогда и только тогда является изображением некоторой окружности пучка, когда она проектируется в точку линии центров, не лежащую на отрезке с концами в фундаментальных точках пучка.

Другими словами,

если $M_1(a_1, b_1, c_1)$ *и* $M_2(a_2, b_2, c_2)$ *— точки прямой*

$$\frac{x - a_0}{A} = \frac{y - b_0}{B} = \frac{z - c_0}{C},$$

изображающей гиперболический пучок, проектирующиеся в его фундаментальные точки, то точка этой прямой тогда и только тогда изображает окружность пучка, когда она не принадлежит отрезку $\overline{M_1M_2}$.

Чтобы найти точки M_1 и M_2, мы воспользуемся тем, что расстояние их проекций (фундаментальных точек пучка) до оси

$$Ax + By + C = 0$$

пучка равно $\sqrt{p}$, так что

$$\left| \frac{Aa_1 + Bb_1 + C}{\sqrt{A^2 + B^2 + C^2}} \right| = \left| \frac{Aa_2 + Bb_2 + C}{\sqrt{A^2 + B^2 + C^2}} \right| = p.$$

Сравнивая эти формулы с формулой (6) (и учитывая, что за точку (a_0, b_0, c_0) в формуле (6) можно принять любую точку изображающей прямой,— в частности, точку M_1 или точку M_2), мы немедленно получаем, что

$$a_1^2 + b_1^2 + 2c_1 = 0 \quad \text{и} \quad a_2^2 + b_2^2 + 2c_2 = 0.$$

Аналогично показывается, что для точек (a, b, c), являющихся внутренними точками отрезка $\overline{M_1M_2}$, имеет место неравенство

$$a^2 + b^2 + 2c < 0.$$

Таким образом, точки M_1 и M_2 принадлежат параболоиду Q, а внутренние точки отрезка $\overline{M_1M_2}$ принадлежат внутренней области этого параболоида. Это означает, что прямая, изображающая гиперболический пучок, пересекает параболоид Q в двух точках.

Собирая вместе все доказанное (и используя принцип обращения), мы получаем следующую теорему:

Теорема 4. *Собственный пучок окружностей тогда и только тогда является*

а) *эллиптическим пучком,*

б) *параболическим пучком,*

в) *гиперболическим пучком,*

когда изображающая его прямая

а) *не пересекает параболоид Q,*

б) *касается этого параболоида,*

в) *пересекает его в двух точках.*

Что касается несобственных пучков, то, по определению, они изображаются прямыми, параллельными оси Oz.

Задание. Докажите, что

любая прямая, параллельная оси Oz, пересекает параболоид Q в одной точке, причем одна из соответствующих полупрямых лежит во внешней области параболоида Q, а другая — во внутренней.

Замечание 2. Это вскрывает неожиданное сходство между гиперболическими и несобственными пучками: для обоих пучков изображающие их прямые имеют бесконечно много точек во внутренней области параболоида. На этом основании иногда бывает удобно причислять несобственные пучки к гиперболическим.

Замечание 3. При доказательстве теоремы 4 мы по ходу дела доказали также, что

для любой прямой в пространстве, не параллельной оси Oz, возможны следующие три взаимоисключающих друг друга случая:

1) *прямая не пересекает параболоид Q (и в этом случае целиком лежит в его внешней области);*

2) *прямая касается параболоида Q (и в этом случае вся, за исключением точки касания, лежит в его внешней области);*

3) *прямая пересекает параболоид Q в двух точках M_1 и M_2 (и в этом случае отрезок M_1M_2 прямой лежит во внутренней области параболоида, а все остальные ее точки — во внешней области).*

Задание. Дайте прямое доказательство этого утверждения. (Указание: рассмотрите параметрические уравнения прямой и составьте уравнение для значений параметра, отвечающих точкам пересечения; ср. п. 2 § 1 гл. 3.)

4. Пучки как пересечения связок

Любая прямая в пространстве является пересечением двух плоскостей и любые две (непараллельные) плоскости пересекаются по прямой. Следовательно,

любой пучок окружностей является пересечением (общей частью) двух связок окружностей и любые две связки (имеющие хотя бы одну общую окружность) пересекаются по пучку.

Отсюда и из теоремы 2 п. 2 немедленно вытекает

Теорема 1. *Если существуют окружности,*

1) *диаметрально пересекающие две данные (различные) окружности;*

2) *диаметрально пересекающие данную окружность и проходящие через данную точку;*

3) *диаметрально пересекающие данную окружность и ортогональные другой данной окружности (отличной от первой);*

4) *диаметрально пересекающие данную окружность и имеющие центры на данной прямой;*

5) *проходящие через две данные (различные) точки;*

6) *проходящие через данную точку и ортогональные данной окружности;*

7) *проходящие через данную точку и имеющие центры на данной прямой;*

8) *ортогональные двум данным (различным) окружностям;*

9) *ортогональные данной окружности и имеющие центры на данной прямой;*

10) *центры которых принадлежат одновременно двум данным (различным) прямым,*

то эти окружности составляют пучок.

При этом любой пучок может быть получен таким образом (вообще говоря, разными способами).

Задание. Покажите, что

окружности, удовлетворяющие условию 4), 5), 7) *или* 9), *существуют всегда*; *удовлетворяющие условию* 1), 3) *или* 8), — *тогда и только тогда, когда данные окружности не концентричны*; *удовлетворяющие условию* 2) *или* 6), — *тогда и только тогда, когда данная точка не является центром данной окружности*; *удовлетворяющие условию* 10), — *тогда и только тогда, когда данные прямые не параллельны.*

Из теоремы 5 п. 3 и теоремы 1 п. 2 немедленно вытекает следующее

Предложение 1. *Эллиптическая связка содержит только эллиптические пучки. Параболическая связка содержит только эллиптические и параболические пучки. Гиперболическая связка содержит пучки всех трех типов (но только собственные). Несобственная связка содержит как несобственные пучки, так и собственные пучки всех трех типов.*

Задание. Дайте прямое доказательство предложения 1, использующее только теорему 4 п. 3 и теорему 1 п. 2.

Следствие. *Пересечение двух связок, из которых одна — эллиптическая, может быть только эллиптическим пучком.*

Пересечение двух связок, из которых одна — параболическая, может быть только эллиптическим или параболическим пучком.

Пересечение двух связок, из которых одна — собственная, может быть только собственным пучком.

Отсюда вытекает, что

описанные в теореме 1 *способы* 1), 2), 3), 4) *и* 5) *построения пучков приводят только к эллиптическим пучкам*;

способы 6) *и* 7) *могут приводить как к эллиптическим, так и к параболическим пучкам*;

способы 8) *и* 9) *могут приводить к любым собственным пучкам*;

способ 10) *приводит только к несобственным пучкам.*

Упражнение. Докажите, что

способы 6) *и* 7) *тогда и только тогда приводят к эллиптическим пучкам, когда данная точка не лежит на данной окружности (данной прямой)*;

способы 8) *и* 9) *тогда и только тогда приводят к*

а) *эллиптическим пучкам,*

б) *параболическим пучкам,*

в) *гиперболическим пучкам,*

когда данные окружности (данные окружность и прямая)

а) *не пересекаются,*

б) *касаются,*

в) *пересекаются в двух точках.*

В частности, мы видим, что

несобственный пучок может быть представлен лишь как пересечение несобственных связок (фундаментальные прямые которых пересекаются в центре пучка) и, обратно, пересечение любых двух несобственных связок (с непараллельными фундаментальными прямыми) является несобственным пучком.

Обратим внимание на то, что

пересечение двух (различных) параболических связок обязательно является эллиптическим пучком.

При этом

любой эллиптический пучок единственным образом представляется в виде пересечения двух параболических связок.

Действительно, это является лишь другой формулировкой теоремы 1 п. 3.

Аналогично, теорема 2 п. 3 означает, что

любой параболический пучок однозначно представляется в виде пересечения некоторой параболической связки и множества всех окружностей, касающихся некоторой прямой.

Подчеркнем, что последнее множество связкой не является.

Чтобы представить параболический пучок в виде пересечения связок, достаточно заметить, что окружность Σ тогда и только тогда касается в точке M_0 некоторой прямой λ, когда она ортогонально пересекает произвольную окружность, проходящую через точку M_0 и имеющую центр на прямой λ. Поэтому теорему 2 п. 3 мы можем переформулировать в следующем виде:

Теорема 2'. *Любой параболический пучок состоит из всех окружностей, проходящих через фиксированную точку и ортогонально пересекающих в этой точке фиксированную окружность. Осью этого пучка является прямая, проходящая через данную точку и центр данной окружности.*

Эта теорема означает, что

любой параболический пучок является пересечением (однозначно определенной) параболической связки и некоторой гиперболической связки.

Заметим, однако, что эта гиперболическая связка однозначно не определена. С другой стороны, произвольной она тоже быть не может: необходимо (и достаточно), чтобы ее фундаментальная окружность проходила через центр пучка, а центр этой окружности лежал на оси пучка.

Чтобы достичь однозначного представления, мы переформулируем теорему 2 п. 3 еще раз:

Теорема 2''. *Любой параболический пучок состоит из всех окружностей, проходящих через фиксированную точку, центры которых лежат на фиксированной прямой (проходящей через эту точку). Осью пучка является прямая, проходящая через данную точку перпендикулярно данной прямой.*

Это означает, что

любой параболический пучок однозначно представляется в виде пересечения параболической и несобственной связок.

Теоремы 2' и 2'' отвечают способам 6) и 7) построения пучков из теоремы 1. Формулировки теорем, отвечающих способам 8) и 9), мы оставляем читателю (эти теоремы мало интересны, поскольку они, подобно теореме 2', не дают однозначного представления).

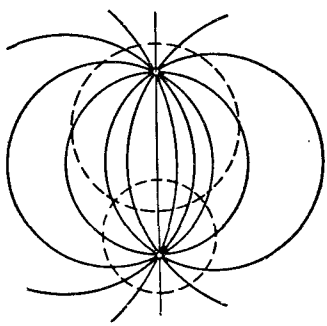
Обратим внимание на то, что однозначность несобственной связки в представлении параболического пучка по способу 7) (теорема 2″) является общим фактом, имеющим место для любых собственных пучков, т. е.

каждый собственный пучок содержится в единственной несобственной связке.

Действительно, ось пучка должна быть фундаментальной прямой этой связки. (Другое доказательство: каждая прямая, не параллельная оси Oz, содержится в единственной плоскости, параллельной этой оси.)

Для эллиптических пучков помимо теоремы 1 п. 3, отвечающей способу 5) построения пучков, возможны еще восемь (!) аналогичных теорем, отвечающих другим способам из теоремы 1. Доказательство их не совсем тривиально, поскольку необходимо показать, что каждым из этих способов можно получить все эллиптические пучки. Мы здесь ограничимся способом 1), т. е. докажем следующую теорему:

Теорема 3. *Каждый эллиптический пучок состоит из всех окружностей, диаметрально пересекающих две фиксированные неконцентрические окружности Σ_1 и Σ_2. Осью пучка является прямая, соединяющая центры этих окружностей.*



Доказательство. Легко видеть, что эта теорема непосредственно вытекает из следующего предложения:

Предложение 2. *Каждая точка оси эллиптического пучка, лежащая между его главными точками, является центром единственной окружности, которую диаметрально пересекает каждая окружность пучка. Никаких других окружностей, диаметрально пересекаемых всеми окружностями пучка, не существует.*

Действительно, если Σ_1 и Σ_2 — любые две из этих окружностей, то пучок, состоящий из всех окружностей, диаметрально пересекающих окружности Σ_1 и Σ_2, содержит данный эллиптический пучок и потому с ним совпадает (ибо две прямые в пространстве, одна из которых содержится в другой, совпадают).

Таким образом, нам осталось лишь доказать предложение 2.

Запишем с этой целью наш пучок в канонических координатах x, y. Тогда любая окружность этого пучка будет иметь уравнение вида

$$x^2 + y^2 - 2xt - e^2 = 0, \qquad (1)$$

а его главные точки будут иметь координаты $(0, -e)$ и $(0, e)$.

Условие того, что окружность (1) диаметрально пересекает некоторую окружность $\Sigma(a_0, b_0, c_0)$, имеет (см. формулу (11) п. 2) вид

$$a_0 t + \frac{e^2}{2} - (a_0^2 + b_0^2 + c_0) = 0.$$

Чтобы это равенство было выполнено для любых t, необходимо и достаточно, чтобы $a_0 = 0$ и $a_0^2 + b_0^2 + c_0 = \frac{e^2}{2}$. Таким образом, мы доказали, что

все окружности (1) *тогда и только тогда диаметрально пересекают окружность* $\Sigma(a_0, b_0, c_0)$, *когда*

$$a_0 = 0, \quad b_0^2 + c_0 = \frac{e^2}{2}.$$

Для завершения доказательства предложения 2 остается заметить, что этим условиям и условию

$$a_0^2 + b_0^2 + 2c_0 > 0$$

(необходимому для существования окружности $\Sigma(a_0, b_0, c_0)$) можно, очевидно, удовлетворить тогда и только тогда, когда

$$-e < b_0 < e,$$

причем каждому такому значению b_0 будет соответствовать одно вполне определенное значение c_0, т. е. одна вполне определенная окружность $\Sigma(a_0, b_0, c_0)$.

Гиперболические пучки могут получаться только способами 8) и 9). При этом легко видеть, что

если пучок можно получить по способу 8), *то его можно получить и по способу* 9).

Действительно, если прямая в пространстве является пересечением двух плоскостей, не параллельных оси Oz, то она же является пересечением одной из этих плоскостей (безразлично, какой) и (однозначно определенной) плоскости, параллельной оси Oz.

Мы не будем ограничиваться гиперболическими пучками и исследуем представимость по способу 8) любых (собственных) пучков.

Для этого нам понадобится следующее

Предложение 3. *Окружность, ортогональная двум окружностям* $\Sigma(a_0, b_0, c_0)$ *и* $\Sigma(a_1, b_1, c_1)$ *некоторого пучка, ортогональна и любой окружности этого пучка.*

Доказательство. Вспомнив параметрические уравнения прямой в пространстве, проходящей через две точки (п. 1 § 3 гл. 3), мы немедленно получим, что координаты произвольной

окружности $\Sigma(a, b, c)$ рассматриваемого пучка выражаются формулами

$$a = (1-t)\,a_0 + ta_1,$$
$$b = (1-t)\,b_0 + tb_1,$$
$$c = (1-t)\,c_0 + tc_1,$$

где t — некоторое число.

С другой стороны, окружности, ортогональные окружностям $\Sigma(a_0, b_0, c_0)$ и $\Sigma(a_1, b_1, c_1)$, образуют (если они существуют) пучок с изображающей прямой

$$\begin{aligned} a_0x + b_0y + z + c_0 &= 0, \\ a_1x + b_1y + z + c_1 &= 0. \end{aligned} \qquad (2)$$

Для завершения доказательства остается заметить, что если точка (x, y, z) удовлетворяет уравнениям (2), то для любого t она удовлетворяет и уравнению

$$((1-t)\,a_0 + ta_1)\,x + ((1-t)\,b_0 + tb_1)\,y + z + ((1-t)\,c_0 + tc_1) = 0.$$

Уравнения (2) определяют пучок (а не пустое множество) тогда и только тогда, когда окружности $\Sigma(a_0, b_0, c_0)$ и $\Sigma(a_1, b_1, c_1)$ не концентричны, т. е. когда данный пучок — собственный. Тем самым доказано, что

окружности, ортогональные всем окружностям произвольного собственного пучка, составляют пучок (также, очевидно, собственный).

Определение 1. Пучок окружностей, ортогональных всем окружностям данного собственного пучка, называется пучком, *сопряженным* этому пучку.

Поскольку сопряженный пучок — собственный, определен пучок, сопряженный сопряженному пучку. Ввиду симметричности отношения ортогональности окружностей (если окружность Σ_1 ортогональна окружности Σ_2, то окружность Σ_2 ортогональна окружности Σ_1), ясно, что этот пучок совпадает с исходным пучком. Другими словами,

отношение сопряженности пучков симметрично.

Замечание 1. Поскольку пучки окружностей изображаются прямыми в пространстве, отношение сопряженности пучков индуцирует некоторое отношение сопряженности между прямыми в пространстве. Это очень интересное отношение мы в дополнении к п. 3 § 2 гл. 6 рассмотрим с совершенно иных позиций.

Предложение 4.

а) *Осью сопряженного пучка является линия центров данного пучка.*

б) *Линией центров сопряженного пучка является ось данного пучка.*

в) *Центр сопряженного пучка совпадает с центром данного пучка.*

г) *Степень p' сопряженного пучка равна степени p данного пучка, взятой с обратным знаком:*

$$p' = -p.$$

Доказательство. Введем на плоскости координаты x, y, канонические относительно данного пучка, и согласованные с ними координаты x, y, z в пространстве. Тогда для любых двух окружностей пучка их вторые координаты y будут равны нулю, а их третьи координаты будут одинаковы (и равны $-\frac{p}{2}$, где p — степень пучка), т. е. эти окружности будут иметь координаты $\left(a_0,\ 0,\ -\frac{p}{2}\right)$ и $\left(a_1,\ 0,\ -\frac{p}{2}\right)$, где $a_0 \neq a_1$. Поэтому уравнения (2) сопряженного пучка будут иметь вид

$$\begin{aligned} a_0 x + z - \frac{p}{2} &= 0, \\ a_1 x + z - \frac{p}{2} &= 0. \end{aligned} \tag{3}$$

Этим уравнениям удовлетворяет, в частности, точка

$$\left(0,\ 0,\ \frac{p}{2}\right).$$

Поскольку векторное произведение векторов $n_0(a_0,\ 0,\ 1)$ и $n_1(a_1,\ 0,\ 1)$ равно $(0,\ a_1 - a_2,\ 0)$, за направляющий вектор прямой (3) мы можем принять вектор $(0,\ 1,\ 0)$. Следовательно, каноническими уравнениями прямой (3) будут уравнения

$$\frac{x-0}{0} = \frac{y-0}{1} = \frac{z-\frac{p}{2}}{0}. \tag{4}$$

Поэтому осью сопряженного пучка будет прямая

$$0x + 1y + 0 = 0,$$

т. е. ось абсцисс $y = 0$. Поскольку эта ось является, по определению, линией центров данного пучка, утверждение а) тем самым полностью доказано.

Аналогично, линией центров сопряженного пучка будет прямая

$$\frac{x-0}{0} = \frac{y-0}{1},$$

т. е. ось ординат $x = 0$ (ось данного пучка). Тем самым доказано и утверждение б).

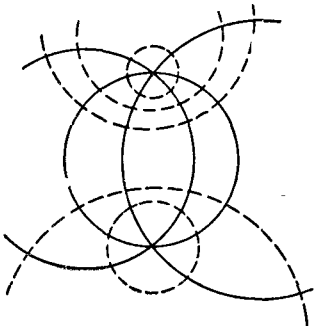
Утверждение в) непосредственно вытекает из утверждений а) и б), а для доказательства утверждения г) достаточно применить к пучку (4) формулу (6) п. 3, согласно которой

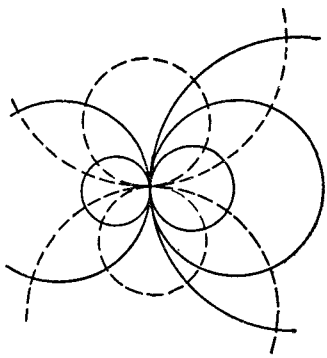
$$p' = \left(\frac{0\cdot 0 + 1\cdot 0 + 0}{\sqrt{0^2 + 1^2}}\right)^2 - \left(0^2 + 0^2 + 2\cdot\frac{p}{2}\right) = -p.$$

Тем самым предложение 4 полностью доказано.

Следствие. *Пучок, сопряженный*

эллиптическому
параболическому
гиперболическому
} *пучку, является* {
гиперболическим
параболическим
эллиптическим
} *пучком.*





Теперь мы уже немедленно получаем теорему, утверждающую универсальность способа 8) построения пучков:

Теорема 4. *Любой собственный пучок состоит из всех окружностей, ортогональных двум фиксированным неконцентрическим окружностям* Σ_1 *и* Σ_2. *Этот пучок*

а) *эллиптический,*

б) *параболический,*

в) *гиперболический,*

тогда и только тогда, когда окружности Σ_1 *и* Σ_2

а) *не пересекаются,*

б) *касаются,*

в) *пересекаются.*

Доказательство. В свете сказанного выше, первое утверждение очевидно: достаточно за окружности Σ_1 и Σ_2 принять любые две окружности сопряженного пучка. Второе утверждение теперь непосредственно вытекает из теоремы 3 п. 3.

Наиболее интересно, конечно, утверждение теоремы 4, относящееся к гиперболическим пучкам (поскольку для других пучков мы обладаем более простыми способами их построения).

5. Прямые как окружности

Многочисленные мало приятные оговорки в предыдущем пункте были вызваны существованием связок с пустым пересечением (т. е. существованием в пространстве параллелльных плоскостей). Естественно, что эти оговорки будут ненужными и вся теория пучков и связок окружностей приобретет большую стройность и законченность, если мы перейдем от евклидова

пространства к евклидово-проективному (см. п. 3 § 2), в котором любые две плоскости пересекаются по прямой (возможно, несобственной). Для этого мы в первую очередь должны, конечно, ответить на вопрос о том, что следует изображать посредством несобственных точек.

Для ответа на этот вопрос мы заметим, что самым общим образом уравнение произвольной окружности может быть записано в следующем виде:

$$M(x^2+y^2)+Ax+By+C=0, \tag{1}$$

где $M\neq 0$. Радиус R этой окружности выражается формулой

$$R^2=\frac{A^2+B^2-2CM}{4M^2}, \tag{2}$$

так что, кроме условия $M\neq 0$, должно быть выполнено также условие

$$A^2+B^2-2CM>0. \tag{3}$$

Поскольку координаты (a, b, c) окружности (1) выражаются формулами

$$a=-\frac{A}{2M},\quad b=-\frac{B}{2M},\quad c=-\frac{C}{2M},$$

мы видим, что точка пространства с однородными координатами $X:Y:Z:T$ (связанными с координатами x, y, z формулами $x=\frac{X}{T}$, $y=\frac{Y}{T}$, $z=\frac{Z}{T}$) изображает окружность (1) с $A=X$, $B=Y$, $C=Z$ и $M=-\frac{1}{2}T$.

Это показывает, что несобственным точкам пространства (для которых $T=0$) естественно сопоставить линии, выражающиеся уравнением (1) с $M=0$, т. е. прямые.

Таким образом,

окружность

$$x^2+y^2-2ax-2by-2c=0$$

мы изображаем собственной точкой $(a:b:c:1)$ *евклидово-проективного пространства, а прямую*

$$Ax+By+C=0$$

— его несобственной точкой $(A:B:C:0)$.

Соглашение 1. Для упрощения формулировок (и только для этого!) целесообразно теперь несколько изменить нашу терминологию и называть *окружностями* произвольные линии, выражающиеся (в прямоугольных координатах x, y) уравнением вида (1), коэффициенты которого подчинены только условию (3) (заметим, что это условие обеспечивает отличие от нуля хотя бы одного из коэффициентов A, B и C, а при $M=0$ — от-

личие от нуля хотя бы одного из коэффициентов A и B). При $M \neq 0$ мы получаем, следовательно, настоящие окружности, а при $M = 0$ — прямые.

Замечание 1. Линия, имеющая уравнение вида (1) в одной системе прямоугольных координат, имеет, очевидно, уравнение такого же вида (и даже с тем же M) в любой другой системе прямоугольных координат. Это показывает, что наше определение окружностей корректно (в рамках координатно-аксиоматического построения евклидовой геометрии; с содержательной точки зрения корректность этого определения очевидна).

При сделанном соглашении

любая окружность (1) *будет изображаться в евклидово-проективном пространстве точкой*

$$(A : B : C : -2M).$$

Тем самым устанавливается биективное соответствие между всеми окружностями на (евклидовой) плоскости и точками $(X : Y : Z : T)$ евклидово-проективного пространства, для которых

$$X^2 + Y^2 + 2ZT \neq 0. \tag{4}$$

Замечание 2. Уравнение

$$X^2 + Y^2 + 2ZT = 0 \tag{5}$$

получается из уравнения

$$x^2 + y^2 + 2z = 0$$

заменами $x = \frac{X}{T}$, $y = \frac{Y}{T}$, $z = \frac{Z}{T}$ (и умножением на T^2). При этом уравнению (5) удовлетворяет, кроме собственных точек, только одна несобственная точка $(0 : 0 : 1 : 0)$, т. е. несобственная точка Z_∞ оси Oz. На этом основании естественно считать (по определению!), что в евклидово-проективном пространстве параболоид Q содержит, кроме собственных точек, еще только одну несобственную точку Z_∞. В соответствии с этим соглашением несобственную плоскость как плоскость, имеющую с параболоидом Q только одну общую точку, естественно считать плоскостью, касающейся параболоида Q в точке Z_∞. Но мы знаем, что если собственная плоскость касается параболоида, то все ее точки, отличные от точки касания, лежат во внешней области параболоида. Поэтому все несобственные точки, отличные от точки Z_∞, также естественно считать лежащими в этой области.

Приняв все эти соглашения, мы можем, таким образом, сказать, что

окружности на евклидовой плоскости находятся в биективном соответствии с точками евклидово-проективного пространства, принадлежащими внешней области параболоида Q.

Замечание 3. Формула (2) показывает, что при $M \to 0$ (и A, B, C постоянных) радиус R стремится к ∞. На этом основании иногда говорят, что «прямые являются окружностями бесконечного радиуса».

При фиксированных A, B, C и любом $M \neq 0$ центр

$$\left(-\frac{A}{2M}, -\frac{B}{2M}\right)$$

окружности (1) принадлежит прямой

$$Bx - Ay = 0, \tag{6}$$

перпендикулярной прямой

$$Ax + By + C = 0.$$

В евклидово-проективной плоскости эти центры имеют однородные координаты $(A : B : -2M)$, а несобственная точка прямой (6) — координаты $(A : B : 0)$. Это показывает, что

в евклидово-проективной плоскости можно считать, что каждая (собственная) прямая обладает «центром», являющимся общей несобственной точкой всех прямых, ей перпендикулярных.

Замечание 4. Поскольку несобственная прямая изображается точкой Z_∞, принадлежащей параболоиду Q, считать ее «окружностью» нецелесообразно. Например, никакого «центра» ей приписать разумным образом нельзя.

Замечание 5. Введение в рассмотрение евклидово-проективной плоскости позволяет также пролить новый свет на соглашение 1.

Действительно, в однородных евклидовых координатах $X : Y : Z$ (связанных с координатами x, y соотношениями $x = \frac{X}{Z}$, $y = \frac{Y}{Z}$) уравнение (1) имеет вид

$$M(X^2 + Y^2) + AXZ + BYZ + CZ^2 = 0,$$

и при $M = 0$ распадается на уравнение собственной прямой

$$AX + BY + CZ = 0$$

и уравнение несобственной прямой

$$Z = 0.$$

Следовательно, на самом деле окружностью, получающейся при $M = 0$, мы должны считать не прямую, а пару, состоящую из этой прямой и несобственной прямой.

Обобщив понятие окружности, мы должны теперь соответствующим образом пополнить связки и пучки окружностей, присоединив к ним окружности (т. е. прямые), изображающиеся несобственными точками.

Так, например, к собственному пучку

$$\frac{x - a_0}{A} = \frac{y - b_0}{B} = \frac{z - c_0}{C}$$

мы должны теперь присоединить прямую, изображающуюся соответствующей несобственной точкой. Эта несобственная точка

имеет по определению (см. п. 3 § 2) однородные координаты $(A:B:C:0)$ и потому изображает прямую

$$Ax+By+C=0,$$

т. е. ось пучка. Таким образом,

ось собственного пучка окружностей является единственной принадлежащей ему прямой.

Несобственные пучки (состоящие из окружностей, имеющих данный центр) никаких прямых не содержат.

Даже на евклидово-проективной плоскости к несобственному пучку ничего не добавляется: единственного кандидата — несобственную прямую — мы окружностью условились не считать.

Аналогично, к связке, изображающейся плоскостью

$$Ax+By+Cz+D=0,$$

мы должны присоединить прямые, изображающиеся несобственными точками этой плоскости, т. е. точками $(X:Y:Z:0)$, для которых

$$AX+BY+CZ=0. \tag{7}$$

Пусть сначала $C=0$, т. е. расматриваемая связка — несобственная (состоит из всех окружностей, центры которых принадлежат прямой $Ax+By+D=0$). В этом случае условию (7) удовлетворяют точки вида $(-B:A:Z:0)$, где Z произвольно (и только такие точки). Эти точки изображают прямые

$$-Bx+Ay+Z=0,$$

перпендикулярные прямой

$$Ax+By+D=0$$

(и только такие прямые). Следовательно,

прямыми, принадлежащими несобственной связке, являются прямые, перпендикулярные ее фундаментальной прямой (и только такие прямые).

Пусть теперь $C\neq 0$ (связка собственная). Так как центр связки имеет координаты

$$a_0=\frac{A}{C},\quad b_0=\frac{B}{C},$$

то любая прямая, проходящая через этот центр, имеет уравнение вида

$$X\left(x-\frac{A}{C}\right)+Y\left(y-\frac{B}{C}\right)=0,$$

т. е. вида

$$Xx+Yy-\frac{AX+BY}{C}=0,$$

где X и Y — произвольные параметры (одновременно не равные нулю). Поэтому эта прямая изображается точкой

$$\left(X : Y : -\frac{AX+BY}{C} : 0\right), \tag{8}$$

удовлетворяющей уравнению (7).

Обратно, любая точка $(X : Y : Z : 0)$, удовлетворяющая уравнению (7), имеет вид (8) и потому изображает прямую, проходящую через центр связки.

Следовательно,

прямыми, принадлежащими собственной связке, являются прямые, проходящие через ее центр, и только эти прямые.

В евклидово-проективной плоскости это утверждение верно и для несобственных связок, поскольку, как мы знаем, центром несобственной связки является несобственная точка прямых, перпендикулярных прямой, на которой расположены центры окружностей связки.

Поскольку в евклидово-проективном пространстве совершенно неестественно ограничиваться лишь его собственными прямыми и плоскостями, мы теперь не можем ограничиваться пучками и связками, ими изображаемыми. Чтобы быть последовательными, мы должны ввести в рассмотрение пучки окружностей (являющихся на самом деле прямыми), изображающиеся несобственными прямыми, и связку, изображающуюся несобственной плоскостью.

Что касается последней, то ясно, что

связка окружностей, изображающаяся несобственной плоскостью, является не чем иным, как множеством всех (собственных) прямых плоскости.

На евклидово-проективной плоскости эту связку можно описать как связку окружностей, центры которых принадлежат несобственной прямой. На этом основании связку всех прямых следует считать несобственной связкой, фундаментальной прямой которой является несобственная прямая.

С другой стороны, поскольку эта связка изображается несобственной плоскостью, касающейся параболоида Q, ее следует считать также и параболической связкой (ср. теорему 1 п. 2).

Рассмотрим теперь пучки, изображающиеся несобственными прямыми.

Каждая несобственная прямая в пространстве является несобственной прямой некоторой плоскости

$$Ax + By + Cz + D = 0.$$

Поэтому пучок, изображающийся этой прямой, состоит из всех прямых, содержащихся в связке, изображающейся этой плоскостью. Следовательно, если эта связка — собственная, то рассматриваемый пучок состоит из всех прямых, проходящих через центр связки, а если эта связка — несобственная, то пучок со-

стоит из всех прямых, перпендикулярных прямой, на которой расположены центры окружностей связки.

Таким образом, мы видим, что

пучки окружностей, изображающиеся несобственными прямыми, являются не чем иным, как пучками прямых в смысле п. 1 § 1.

При этом собственные пучки прямых принадлежат собственным связкам окружностей, а несобственные пучки — несобственным связкам.

Задание. Докажите, что

собственные пучки прямых изображаются несобственными прямыми, не проходящими через несобственную точку Z_∞ оси Oz, а несобственные пучки прямых — несобственными прямыми, проходящими через эту точку.

Посмотрим теперь, что происходит с полученными в предыдущих пунктах результатами при нашем новом понимании пучков и связок окружностей.

В первую очередь нам надо распространить термины «ортогональные окружности» и «диаметрально пересекающиеся окружности» на случай, когда эти окружности являются прямыми.

Определение 1. Говорят, что прямая и окружность (настоящая) *ортогональны*, если прямая проходит через центр окружности. В этом случае мы будем также говорить, что прямая *диаметрально пересекает* окружность. Две прямые мы будем называть *ортогональными*, если они перпендикулярны.

На евклидово-проективной плоскости возможна единая формулировка: прямая (собственная) и произвольная окружность (конечного или «бесконечного» радиуса) ортогональны, если прямая проходит через центр окружности.

Заметим, что отношение «окружность (или прямая) диаметрально пересекает прямую» мы не определяем. В связи с этим мы принимаем следующее

Соглашение 2. В высказывании «окружность Σ_1 диаметрально пересекает окружность Σ_2» всегда будет неявно предполагаться, что окружность Σ_2 является настоящей окружностью (тогда как окружность Σ_1 может быть и прямой).

Предложение 1. *Множество всех окружностей Σ, ортогональных данной окружности Σ_0, является связкой окружностей.*

Это предложение аналогично предложению 4 п. 2. Однако даже в случае, когда Σ_0 является настоящей окружностью, оно утверждает нечто иное, поскольку теперь окружность Σ может быть и прямой. Поэтому оно и в этом случае нуждается в доказательстве.

Доказательство. Пусть Σ_0 — настоящая окружность. В этом случае следует только проверить, что к собственной связке настоящих окружностей, получающейся в силу предложения

4 п. 2, присоединяется точно то, что нужно, т. е. пучок прямых с центром в центре связки. Но это ясно, поскольку согласно предложению 6 п. 2 центром этой связки является центр окружности Σ_0, а согласно определению 1 прямая тогда и только тогда ортогональна окружности Σ_0, когда она проходит через ее центр. Следовательно, предложение 1 в этом случае справедливо.

Пусть окружность Σ_0 является прямой λ. Тогда множество всех окружностей, ортогональных окружности Σ_0, является (предложение 1 п. 2) несобственной связкой (с фундаментальной прямой λ), пополненной (определение 1) всеми прямыми, перпендикулярными прямой λ. Следовательно, предложение 1 справедливо и в этом случае.

Таким образом, мы видим, что предложение 1 объединяет в единой формулировке два ранее никак не связанных предложения: предложение 1 п. 2 и предложение 4 п. 2.

Кроме того, поскольку в зависимости от вида окружности Σ_0 получается либо гиперболическая, либо несобственная связка, мы видим, что *несобственные связки целесообразно причислять к гиперболическим связкам*. В силу этого соглашения каждая гиперболическая связка будет обладать фундаментальной окружностью (которая может оказаться и прямой).

Предложение 2. *Множество всех окружностей, диаметрально пересекающих данную окружность Σ_0, является связкой окружностей.*

Доказательство. В соответствии с соглашением 2 окружность Σ_0 может быть только настоящей окружностью. Но тогда множество всех окружностей, ее диаметрально пересекающих, является (предложение 5 п. 2) собственной связкой настоящих окружностей, пополненной (определение 1) пучком прямых с центром в центре окружности Σ_0, являющемся (предложение 6 п. 2) центром связки.

Предложение 3. *Совокупность всех окружностей, проходящих через данную точку, является связкой окружностей.*

Доказательство. Достаточно заметить, что эта совокупность, кроме пучка прямых с центром в данной точке, содержит еще только настоящие окружности, составляющие (предложения 5 и 6 п. 2) связку с центром в этой точке.

Обратим внимание на то, что предложения 1, 2 и 3 описывают все возможные связки, за исключением несобственной связки всех прямых (которую мы условились считать параболической связкой).

Рассмотрим теперь пучки окружностей.

Следующая теорема (являющаяся аналогом теоремы 1 п. 4) теперь очевидна.

Теорема 1. *Множество всех окружностей,*

а) *диаметрально пересекающих две данные (различные) окружности;*

б) *диаметрально пересекающих данную окружность и проходящих через данную точку*;

в) *диаметрально пересекающих данную окружность и ортогональных другой данной окружности (возможно, совпадающей с первой)*;

г) *проходящих через две данные (различные) точки;*

д) *проходящих через данную точку и ортогональных данной окружности*;

е) *ортогональных двум данным (различным) окружностям, является пучком окружностей.*

Любой пучок окружностей может быть получен хотя бы одним из этих способов.

Способ а) дает эллиптический пучок окружностей (вместе с его осью).

Способ б) дает эллиптический пучок окружностей, когда данная точка не является центром данной окружности (в силу соглашения 2 являющейся настоящей окружностью), и собственный пучок прямых в противном случае. На этом основании собственные пучки прямых естественно причислять к эллиптическим пучкам (что согласуется также с теоремой 4 п. 3).

Способ в) (объединяющий способы 3) и 4) теоремы 1 п. 4) дает эллиптический пучок окружностей, если данные окружности различны (причем вторая окружность может при этом быть прямой), и собственный пучок прямых, если эти окружности совпадают.

Способ г) дает эллиптический пучок окружностей.

Способ д) (объединяющий способы 6) и 7) теоремы 1 п. 4) дает параболический пучок окружностей (конечно, вместе с осью), когда данная окружность (могущая быть и прямой) содержит данную точку, эллиптический пучок, когда данная точка не лежит на данной окружности и не является ее центром, и собственный пучок прямых, когда данная окружность является настоящей окружностью, а данная точка совпадает с ее центром.

Способ е) (объединяющий способы 8), 9) и 10) теоремы 1 п. 4) в случае, когда не более одной из данных окружностей является прямой, дает эллиптический пучок окружностей (или собственный пучок прямых), если данные окружности не пересекаются, параболический пучок, если они касаются, и гиперболический пучок, если данные окружности пересекаются. Если же обе данные окружности являются прямыми, то в случае, когда они пересекаются, получается несобственный пучок окружностей (который мы условились причислять к гиперболическим пучкам; см. замечание 2 п. 4), а в случае, когда они параллельны (но не совпадают), — несобственный пучок прямых.

При этом (см. п. 4)

каждым из способов а) — е) *можно получить любой эллиптический пучок окружностей (в частности, любой собственный пучок прямых)*,

каждым из способов д) *или* е) *можно получить любой параболический пучок окружностей.*

способом е) *можно получить любой гиперболический пучок окружностей (в частности, любой несобственный пучок окружностей; см. замечание* 2 п. 4), *а также любой несобственный пучок прямых.*

Интересено отметить, что предложение 3 п. 4 сохраняется и при теперешнем обобщенном понимании окружностей, т. е. имеет место следующее

Предложение 4. *Любая окружность* Σ, *ортогональная двум окружностям* Σ_1 *и* Σ_2 *некоторого пучка, ортогональна и любой окружности этого пучка.*

Задание. Докажите предложение 4 (не забудьте рассмотреть все четыре случая: когда данный пучок является собственным пучком окружностей, когда он является несобственным пучком окружностей, когда он является собственным пучком прямых и, наконец, когда он является несобственным пучком прямых; кроме того, в первом случае следует отдельно рассмотреть возможности, не покрываемые предложением 1 п. 4, т. е. ситуации, когда либо окружность Σ является прямой, либо прямой (и, следовательно, осью данного пучка) является одна из окружностей Σ_1 и Σ_2.

Согласно предложению 4 определение сопряженного пучка (определение 1 п. 4) имеет теперь смысл для любых пучков. При этом

пучком, сопряженным несобственному пучку окружностей, является собственный пучок прямых с тем же центром,

пучком, сопряженным собственному пучку прямых, является несобственный пучок окружностей (имеющий тот же центр).

пучком, сопряженным несобственному пучку прямых, является несобственный пучок прямых, состоящий из прямых, перпендикулярных прямым пучка.

Мы видим, в частности, что всегда

центры сопряженных пучков совпадают.

Чтобы эта формулировка охватывала и несобственные пучки прямых, мы, естественно, должны перейти в евклидово-проективную плоскость.

Выше мы условились собственный пучок прямых считать эллиптическим пучком, а несобственный пучок окружностей — гиперболическим пучком. Мы видим, следовательно, что для этих пучков остается справедливым следствие из предложения 4 п. 4.

Чтобы это следствие осталось справедливым и для несобственных пучков прямых, нужно, очевидно, считать эти пучки параболическими (что, кстати, согласуется и с теоремой 4 п. 3).

6. Окружности на вещественно-комплексной плоскости

Полную законченность и изящество геометрия окружностей приобретает только на *евклидовой вещественно-комплексной плоскости* (точнее, чтобы специально не оговаривать несоб-

ственные пучки и связки — на евклидово-проективной вещественно-комплексной плоскости).

В свете сказанного в п. 5 мы принимаем

Определение 1. *Окружностью* (*вещественной*) на евклидовой вещественно-комплексной плоскости называется множество точек, координаты x, y которых в некоторой евклидовой координатной системе удовлетворяют уравнению вида

$$M(x^2+y^2)+Ax+By+C=0, \tag{1}$$

где M, A, B, C — произвольные вещественные числа, причем хотя бы одно из чисел M, A, B не равно нулю (ясно, что это определение корректно, т. е. при переходе к другим евклидовым координатам x', y' мы снова получаем уравнение вида (1), хотя, конечно, с другими коэффициентами).

Окружность (1) мы будем изображать точкой $(A:B:C:-2M)$ расширенного (евклидово-проективного) пространства (таким образом, пространство у нас по-прежнему остается вещественным). Тем самым мы, очевидно, получим биективное соответствие между окружностями на вещественно-комплексной плоскости и всеми точками расширенного пространства, отличными от точки $Z_\infty(0:0:1:0)$.

При $M=0$ окружность (1) является прямой

$$Ax+By+C=0$$

и изображается несобственной точкой $(A:B:C:0)$. При $M\neq 0$ окружность (1) называется *настоящей окружностью* и имеет уравнение

$$x^2+y^2-2ax-2by-2c=0, \tag{2}$$

где

$$a=-\frac{A}{2M},\quad b=-\frac{B}{2M},\quad c=-\frac{C}{2M}.$$

Такая окружность изображается собственной точкой с неоднородными (обычными) координатами (a, b, c). Проекция этой точки на плоскость Oxy, т. е. точка (a, b), называется *центром* окружности (2).

Уравнение (2) может быть переписано в виде

$$(x-a)^2+(y-b)^2=R^2, \tag{3}$$

где

$$R^2=a^2+b^2+2c.$$

При $R^2>0$ мы будем называть окружность (3) *действительной*[1]) *окружностью* или *окружностью вещественного радиуса* R. В пересечении с вещественной плоскостью она дает обычную окружность с центром (a, b) и радиусом R.

Действительные окружности изображаются теми же точками, что и соответствующие обычные окружности, т. е. точками внешней области параболоида Q.

[1]) Обратим внимание на различие между понятиями «действительная окружность» и «вещественная окружность».

При $R = 0$ мы будем окружность (3) называть *окружностью нулевого радиуса*. Она проходит через ее центр (a, b) (который является ее единственной вещественной точкой) и состоит из двух комплексно-сопряженных прямых

$$x + iy = a + ib, \qquad x - iy = a - ib.$$

Окружности нулевого радиуса изображаются, очевидно, точками, принадлежащими параболоиду Q.

Определение 2. Прямые (собственные), имеющие уравнения вида

$$x \pm iy + C = 0,$$

называются *изотропными*.

Таким образом,

окружность нулевого радиуса является парой комплексно-сопряженных изотропных прямых, проходящих через ее центр.

При $R^2 < 0$ окружность (3) называется *окружностью мнимого радиуса* iR (или просто *мнимой окружностью*).

Каждая такая окружность изображается точками внутренней области параболоида Q.

Никаких вещественных точек окружность мнимого радиуса вообще не имеет (является нулевой линией).

На евклидово-проективной вещественно-комплексной плоскости мы несобственную прямую теперь будем считать окружностью (поскольку причины, по которым мы ее раньше не считали окружностью, теперь отпали). Изображается эта окружность точкой Z_∞.

Замечание 1. Однако, несмотря на то, что точка Z_∞ принадлежит параболоиду Q, мы не будем причислять несобственную прямую к окружностям нулевого радиуса. Считать ее, подобно собственным прямым, «окружностью бесконечного радиуса» также нет оснований. Поэтому никакого «радиуса» (так же как и никакого «центра») мы этой окружности не приписываем.

Степень $\operatorname{ст} M_0$ произвольной точки M_0 относительно (настоящей) окружности (3) определяется так же, как в вещественном случае, т. е. формулой (3) п. 1. По-прежнему,

точка M_0 тогда и только тогда принадлежит окружности (3), *когда ее степень* $\operatorname{ст} M_0$ *равна нулю.*

Степень любой вещественной точки относительно окружности мнимого радиуса всегда положительна, так что каждая точка вещественной плоскости «лежит вне» произвольной такой окружности.

Для окружности нулевого радиуса степень $\operatorname{ст} M_0$ равна квадрату расстояния от точки M_0 до центра этой окружности.

Предложение 1 п. 1, дающее геометрическую интерпретацию степени, конечно, теперь смысла не имеет, хотя его аналог для модуля $|\operatorname{ст} M_0|$ степени остается справедливым.

Упражнение. Сформулируйте и докажите аналог предложения 1 п. 1 для модуля степени (не забудьте, что длина вектора в комплексной плоскости может быть равна нулю, даже если сам вектор не равен нулю).

Замечание 2. Обратим внимание на то, что степень точки относительно окружности «бесконечного радиуса» (т. е. относительно прямой) мы не определяем.

Радикальная ось двух (неконцентрических) настоящих окружностей определяется точно так же, как в вещественном случае, и она по-прежнему является прямой с уравнением (6) п. 1.

Задание. Покажите, что

радикальной осью двух окружностей нулевого радиуса является медиатриса[1]) *их центров.*

Чтобы идти дальше, мы должны предварительно обсудить вопрос о точках пересечения двух различных (настоящих) окружностей

$$\begin{aligned} x^2 + y^2 - 2a_1x - 2b_1y - 2c_1 = 0, \\ x^2 + y^2 - 2a_2x - 2b_2y - 2c_2 = 0. \end{aligned} \tag{4}$$

По определению, чтобы найти эти точки, надо решить систему (4) двух уравнений с двумя неизвестными. Вычтя первое уравнение из второго (и разделив получающееся уравнение на -2), мы получим, что эта система равносильна следующей системе:

$$\begin{aligned} x^2 + y^2 - 2a_1x - 2b_1y - 2c_1 = 0, \\ (a_2 - a_1)\,x + (b_2 - b_1)\,y + (c_2 - c_1) = 0. \end{aligned} \tag{5}$$

Мы видим, что при $(a_1, b_1) = (a_2, b_2)$ система (5) решений не имеет (ибо, по условию, $(a_1, b_1, c_1) \neq (a_2, b_2, c_2)$). Следовательно,

концентрические окружности не пересекаются и в евклидовой вещественно-комплексной плоскости.

В однородных координатах $X : Y : Z$ уравнения (5) имеют вид

$$\begin{aligned} X^2 + Y^2 - 2a_1XZ - 2b_1YZ - 2c_1Z^2 = 0, \\ (a_2 - a_1)\,X + (b_2 - b_1)\,Y + (c_2 - c_1)\,Z = 0 \end{aligned}$$

и при $(a_1, b_1) = (a_2, b_2)$ сводятся к уравнениям

$$\begin{aligned} X^2 + Y^2 = 0, \\ Z = 0, \end{aligned}$$

имеющим два решения $(i : 1 : 0)$ и $(-i : 1 : 0)$. Следовательно,

в евклидово-проективной вещественно-комплексной плоскости любые две концентрические окружности пересекаются в двух несобственных комплексно-сопряженных точках $(i : 1 : 0)$ *и* $(-i : 1 : 0)$.

Определение 3. Точки $(i : 1 : 0)$ и $(-i : 1 : 0)$ называются *циклическими точками.*

---

[1]) *Медиатрисой* двух точек M_1 и M_2 называется прямая, проходящая через середину отрезка $\overline{M_1M_2}$ перпендикулярно прямой M_1M_2. Она является множеством всех точек, расстояния которых от точек M_1 и M_2 одинаковы.

Заметим, что если точки имеют координаты $(\pm i : 1 : 0)$ в одной однородной евклидовой координатной системе, то они имеют те же координаты и в любой другой однородной евклидовой координатной системе. Поэтому циклические точки определены корректно.

Задание. Покажите, что

прямая тогда и только тогда изотропна, когда она проходит через циклическую точку.

Для неконцентрических окружностей мы можем из второго уравнения (5) выразить одно неизвестное через другое и, подставив это выражение в первое уравнение, получить для второго неизвестного квадратное уравнение с вещественными коэффициентами.

Случай 1. Это квадратное уравнение имеет два различных вещественных корня.

В этом случае данные окружности имеют две различные вещественные точки пересечения. Ясно, что это возможно только тогда, когда сами окружности действительны, и эти точки будут точками пересечения обыкновенных окружностей, высекаемых данными окружностями на вещественной плоскости.

Случай 2. Квадратное уравнение имеет два различных комплексно-сопряженных корня.

В этом случае данные окружности имеют две различные комплексно-сопряженные точки пересечения. Сами же окружности могут быть при этом любыми (с вещественным, нулевым или мнимым радиусом). Если они действительны, то высекаемые ими на вещественной плоскости обыкновенные окружности не пересекаются.

Случай 3. Квадратное уравнение имеет один двойной (вещественный) корень.

В этом случае данные окружности имеют одну точку пересечения, являющуюся вещественной. Принято говорить, что окружности *касаются* друг друга в этой точке, а их точку пересечения принято называть *точкой касания*. Ни одна из касающихся окружностей не может быть мнимой. Касающиеся действительные окружности высекают на вещественной плоскости обыкновенные окружности, касающиеся друг друга в элементарно-геометрическом («школьном») смысле. Действительная окружность касается окружности нулевого радиуса с центром в точке M_0 тогда и только тогда, когда она проходит через точку M_0 (являющуюся в этом случае точкой касания).

Таким образом,

две настоящие неконцентрические окружности на евклидовой вещественно-комплексной плоскости имеют либо две различные точки пересечения (вещественные или комплексно-сопряженные), либо одну общую точку (точку касания).

Ясно, что на евклидово-проективной плоскости

любая (настоящая) окружность проходит через циклические точки $(i : 1 : 0)$ *и* $(-i : 1 : 0)$.

Поэтому на этой плоскости к указанным выше точкам пересечения двух окружностей добавляются еще две циклические точки.

Замечание 3. Таким образом, на евклидово-проективной вещественно-комплексной плоскости две (неконцентрические) окружности имеют, вообще говоря, четыре точки пересечения: две собственные и две несобственные (циклические). Исключением является случай касающихся окружностей, когда имеется только одна общая собственная точка (точка касания). Чтобы этот случай (по крайней мере, «лингвистически») не представлял исключения, удобно считать точку касания кратной (а именно, *двойной*) точкой пересечения и считать ее дважды. Тогда и в случае касающихся окружностей будет четыре точки пересечения.

Другое исключение представляет собой случай концентрических окружностей, когда имеются только две (циклические) точки пересечения. Чтобы сохранить и в этом случае общий закон, приходится, по определению, считать эти точки *двойными точками пересечения* (т. е. считать, что *концентрические окружности касаются друг друга в циклических точках* [1])).

Приняв эти соглашения, мы, следовательно, можем сказать, что

на евклидово-проективной вещественно-комплексной плоскости любые две различные (настоящие) окружности имеют ровно четыре точки пересечения.

Это утверждение является частным случаем теоремы Безу, которую мы докажем в п. 4 § 2 гл. 6.

Возвращаясь к радикальным осям, заметим, что второе из уравнений (5) является как раз уравнением радикальной оси данных окружностей. Следовательно,

радикальная ось двух окружностей проходит через их точки пересечения.

В случае, когда точки пересечения различны, это полностью характеризует радикальную ось (ср. предложение 2 п. 1).

Пусть точки пересечения совпадают (окружности касаются). Тогда радикальная ось будет иметь с каждой из окружностей только одну общую точку. Условившись называть (ср. конец п. 2 § 1 гл. 1) прямую, имеющую с (настоящей) окружностью только одну общую точку, *касательной* к этой окружности в этой точке (*точке касания*), мы можем, следовательно, сказать, что

радикальной осью двух касающихся окружностей является их общая касательная.

Таким образом, предложение 2 п. 1 сохраняется и в вещественно-комплексной плоскости (только без оговорки, что окружности пересекаются).

Радикальный центр трех (настоящих) окружностей на вещественно-комплексной плоскости определяется дословно так же, как в вещественном случае (определение 3 п. 1). Все его свойства также полностью сохраняются.

---

[1]) Таким образом, концентрические (и только концентрические) окружности касаются друг друга в двух точках!

Определение (*вещественных*, т. е. определяемых уравнениями с вещественными коэффициентами) *пучков* и *связок* окружностей также остается прежним. Все их свойства сохраняются. Однако теперь

каждая параболическая связка содержит, кроме действительных окружностей (и прямых), одну окружность нулевого радиуса (с центром в центре связки).

На евклидово-проективной плоскости несобственная связка всех прямых (которую мы условились считать параболической) пополняется теперь несобственной прямой.

Аналогично,

каждая гиперболическая (собственная) связка содержит, кроме действительных окружностей (и прямых), бесконечно много окружностей нулевого радиуса (с центрами в точках фундаментальной окружности) и бесконечно много окружностей мнимого радиуса (с центрами, лежащими внутри фундаментальной окружности).

Это верно и для несобственной связки окружностей (которую мы условились считать гиперболической) с тем лишь отличием, что центры всех ее окружностей принадлежат фундаментальной прямой.

Что же касается эллиптических связок, то

на евклидовой (или евклидово-проективной) вещественно-комплексной плоскости каждая эллиптическая связка содержит только действительные окружности.

Задание. Докажите эти утверждения.

Таким образом,

для любой собственной связки окружностей на евклидовой вещественно-комплексной плоскости каждая вещественная точка плоскости является центром (единственной) окружности связки.

Определение 4. Две (настоящие) окружности $\Sigma(a_1, b_1, c_1)$ и $\Sigma(a_2, b_2, c_2)$ на вещественно-комплексной плоскости называются *сопряженными*, если

а) они концентричны:

$$(a_1, b_1) = (a_2, b_2),$$

б) выполнено равенство

$$a_1a_2 + b_1b_2 + c_1 + c_2 = 0.$$

Окружность, сопряженная главной окружности эллиптической связки, называется *фундаментальной окружностью* этой связки, а окружность, сопряженная фундаментальной окружности гиперболической связки, называется *главной окружностью* связки. Для параболической связки, по определению, главная окружность совпадает с фундаментальной и является окружностью нулевого радиуса с центром в центре связки.

Задание. Покажите, что
окружность, сопряженная с окружностью
а) *вещественного радиуса,*
б) *нулевого радиуса,*
в) *мнимого радиуса,*
является, соответственно, окружностью
а) *мнимого радиуса,*
б) *нулевого радиуса,*
в) *вещественного радиуса.*

Согласно определению 3
каждая собственная связка окружностей на евклидовой вещественно-комплексной плоскости обладает и фундаментальной и главной окружностью.

Для непараболических связок одна из этих окружностей имеет вещественный радиус, а другая — мнимый.

Упражнение. Покажите, что
окружность Σ *тогда и только тогда ортогональна окружности* Σ_1, *когда она диаметрально пересекает сопряженную окружность* Σ_2.
Покажите также, что
окружность тогда и только тогда ортогональна некоторой окружности нулевого радиуса, когда она проходит через центр этой окружности.

Сопоставляя эти утверждения с теоремой 2 п. 2, мы немедленно получаем следующую теорему:

Теорема 1. *Произвольная собственная связка окружностей на евклидовой вещественно-комплексной плоскости состоит из всех окружностей, ортогональных фиксированной окружности (фундаментальной окружности связки), а также из всех окружностей, диаметрально пересекающих фиксированную окружность (главную окружность связки).*

Аналогично унифицируется теперь и теория пучков. Мы не будем этого делать во всех деталях и ограничимся следующей теоремой:

Теорема 2. *На евклидовой вещественно-комплексной плоскости любая точка линии центров каждого собственного пучка окружностей является центром единственной настоящей окружности пучка.*

Каждый собственный непараболический пучок окружностей состоит из всех окружностей, проходящих через две различные точки (вещественные, если пучок эллиптический, и комплексно-сопряженные, если пучок гиперболический).

Задание. Докажите теорему 2.

Таким образом, если мы назовем *главными точками* пучка точки, через которые проходят все окружности пучка (ср. определение 7 п. 3), то мы получим, что
собственный пучок окружностей на евклидовой вещественно-комплексной плоскости тогда и только тогда является
а) *эллиптическим пучком,*
б) *параболическим пучком,*
в) *гиперболическим пучком,*

когда он имеет

а) *две главные вещественные точки,*

б) *одну главную вещественную точку (центр пучка),*

в) *две главные комплексно-сопряженные точки.*

Далее, легко видеть, что

для любой окружности Σ собственного пучка в этом пучке имеется единственная окружность Σ', ортогональная окружности Σ.

Действительно, в канонических координатах окружность Σ (если эта окружность — настоящая) имеет уравнение вида

$$x^2 + y^2 - 2xt + p = 0,$$

где t — некоторый параметр (а p — степень пучка). Условие, что некоторая другая окружность Σ' с уравнением

$$x^2 + y^2 - 2xt' + p = 0$$

ортогональна окружности Σ, имеет вид (см. формулу (9) п. 2)

$$tt' = p. \tag{6}$$

Следовательно, по любому $t \neq 0$ мы однозначно определяем соответствующее t'.

Если $t = 0$, т. е. окружность Σ является главной окружностью пучка, то ей ортогональна, очевидно, ось пучка. Обратно, если окружность Σ является прямой (и, следовательно, осью пучка), ей ортогональна главная окружность пучка.

Поскольку каждая точка линии центров пучка является центром единственной окружности пучка, мы можем, следовательно, определить некоторое отображение T этой линии в себя, сопоставив произвольной точке M точку $M' = T(M)$, являющуюся центром окружности Σ', принадлежащей пучку и ортогональной окружности Σ, имеющей центр в точке M (и также принадлежащей пучку).

Собственно говоря, это верно только в евклидово-проективной плоскости, поскольку отображение T сопоставляет центру пучка несобственную точку линии центров.

Согласно формуле (6) отображение T переводит точку линии центров, имеющую абсциссу t, в точку с абсциссой

$$t' = \frac{p}{t}. \tag{7}$$

Мы будем считать, что эта формула определяет отображение T и для комплексных t (не являющихся абсциссами центров пучка).

Формула (7) показывает, что при $p \neq 0$ (т. е. для непараболического пучка) отображение T биективно (на расширенной вещественно-комплексной прямой) и совпадает с обратным ото-

бражением T^{-1}. При $p = 0$ (для параболического пучка) отображение T переводит всю линию центров в центр пучка.

Определение 5. Биективное отображение, совпадающее со своим обратным, называется *инволюцией.* Отображение, переводящее всю прямую в некоторую ее точку M_0, называется *параболической инволюцией с центром* M_0.

Таким образом, мы можем сказать, что любой пучок окружностей определяет на его линии центров некоторую инволюцию T, параболическую, если пучок параболический.

Точка M называется *неподвижной точкой* инволюции T, если

$$T(M) = M.$$

Точки линии центров собственного пучка, являющиеся неподвижными точками соответствующей инволюции T, называются *фундаментальными точками* этого пучка. Их абсциссы t определяются из уравнения

$$t^2 = p. \tag{8}$$

Для гиперболического пучка ($p = e^2$) это уравнение имеет решения $\pm e$. Следовательно, наше теперешнее определение фундаментальных точек совпадает для гиперболических пучков с определением 7 п. 3.

Для параболических пучков ($p = 0$) уравнение (8) имеет единственное решение $t = 0$, а для эллиптических пучков ($p = -e^2$) — два комплексно-сопряженных чисто мнимых решения $\pm ie$.

Таким образом,

на евклидовой вещественно-комплексной плоскости собственный пучок окружностей тогда и только тогда является

а) *эллиптическим пучком,*

б) *параболическим пучком,*

в) *гиперболическим пучком,*

когда он имеет

а) *две фундаментальные комплексно-сопряженные точки,*

б) *одну фундаментальную вещественную точку (центр пучка),*

в) *две фундаментальные вещественные точки.*

Замечание 3. На этом основании любая (непараболическая) инволюция, имеющая две вещественные неподвижные точки, называется *гиперболической инволюцией,* а инволюция, имеющая две комплексно-сопряженные неподвижные точки (и следовательно, не имеющая вещественных неподвижных точек), называется *эллиптической инволюцией* [1]).

---

[1]) Правду сказать, исторически дело обстояло как раз наоборот. По причинам, которые мы объясним ниже (см. замечание 2 п. 6 § 1 гл. 6), термины «эллиптическая», «параболическая» и «гиперболическая» были первоначально применены к инволюциям и лишь затем перенесены на соответствующие пучки.

Дополнение. **Геометрии параболической и гиперболической связок**

Зафиксировав на вещественной евклидово-проективной плоскости некоторую параболическую связку с центром O, будем называть «точками» всевозможные точки плоскости, отличные от точки O, и «прямыми» — всевозможные окружности связки, из которых удалена точка O. Очевидная проверка показывает, что справедливо следующее

Предложение 1. *Эти «точки» и «прямые» удовлетворяют (по отношению к теоретико-множественному отношению принадлежности) всем планиметрическим аксиомам* I_1—I_4 *первой группы аксиом Гильберта, а также аксиоме параллельности* V.

Заметим, что это верно только тогда, когда мы принимаем соглашение 1 из п. 5, т. е. считаем прямые, проходящие через точку O, окружностями связки.

Пусть A, B и C — три «коллинеарные» (попарно различные) «точки» (т. е. «точки», принадлежащие одной окружности Σ нашей связки). Мы скажем, что «точка» B расположена *между* «точками» A и C, если на окружности Σ точки B и O *разделяют* точки A и C, т. е. из двух дуг окружности Σ с концами в точках A и C одна содержит точку B, а другая — точку O. В случае, когда окружность Σ является прямой, а точки A и C — собственными точками плоскости, это означает, что точка B принадлежит отрезку $\overline{AC}$, если этот отрезок не содержит точки O, и лежит вне этого отрезка — в противном случае[1]). Легко показывается, что справедливо следующее

Предложение 2. *Для отношения «между» выполнены все аксиомы* II_1—II_4 *второй группы аксиом Гильберта и аксиома* IV *Дедекинда.*

Задание. Докажите предложение 2.

Из предложения 2, в частности, следует, что в геометрии наших «точек» и «прямых» могут быть определены понятия «отрезка» и «угла».

Естественно ожидать, что предложения 1 и 2 можно дополнить следующим предложением:

Предложение 3. *Для «отрезков» и «углов» можно ввести отношение «конгруэнтность» так, чтобы были выполнены все аксиомы* III_1—III_6 *третьей группы аксиом Гильберта.*

Это предложение действительно верно, но описание соответствующего отношения конгруэнтности требует понятий, которые мы введем только в гл 7. Поэтому мы его доказывать здесь не будем.

Упражнение. Прочтите определение инверсии относительно окружности (определение 1 п. 1 § 3 гл. 7) и докажите, что предложение 3 будет выполнено, если мы будем считать «отрезки» (или «углы») «конгруэнтными», когда их можно перевести друг в друга некоторой последовательностью инверсий относительно окружностей связки.

Предложения 1, 2, 3 означают, что справедлива следующая

Теорема 1. *Построенная «геометрия параболической связки» является интерпретацией евклидовой геометрии.*

---

[1]) Таким образом, на прямых наше отношение «между» не совпадает с обычным отношением «между».

Замечание 1. Эту теорему можно было бы предугадать (но, конечно, не доказать), заметив, что аналогичное построение для связки всех прямых (на нерасширенной плоскости) приводит к обычной интерпретации евклидовой геометрии (в которой «точки» — это точки, «прямые» — это прямые и т. д.), и приняв во внимание, что связка всех прямых аналогична параболическим связкам окружностей.

Конечно, возможность интерпретировать евклидову геометрию как геометрию параболической связки, возможность сама по себе интересная и полезная (поскольку она позволяет из любой евклидовой теоремы о точках и прямых «бесплатно» получить некоторую теорему о точках и окружностях), большого принципиального значения все же не имеет. Однако она наводит на мысль о возможности построения аналогичных «геометрий гиперболических и эллиптических связок».

В геометрии гиперболической связки «точками» считаются внутренние точки фундаментальной окружности связки, а «прямыми» — дуги окружностей связки, лежащие внутри фундаментальной окружности. Отношение «между» определяется очевидным образом, а отношение «конгруэнтность» — так же, как для геометрии параболической связки (посредством инверсий относительно окружностей связки). Ясно, что в этой геометрии вместо евклидовой аксиомы о параллельных выполнена аксиома Лобачевского. Поэтому естественно ожидать, что справедлива следующая

Теорема 2. *Геометрия гиперболической связки является интерпретацией неевклидовой геометрии Лобачевского.*

Упражнение. Докажите теорему 2.

Замечание 2. В § 3 гл. 7 мы вернемся к теореме 2 с несколько иных позиций и, в частности, полностью ее докажем.

Теорема 2 означает, что геометрия Лобачевского допускает интерпретацию в рамках евклидовой геометрии и, следовательно, непротиворечива (поскольку непротиворечива евклидова геометрия).

Обратим внимание на то, что поскольку неконгруэнтных гиперболических связок существует бесконечно много (они различаются их степенями), существует на самом деле бесконечно много геометрий Лобачевского, различающихся значением некоторого параметра e (называемого *кривизной* геометрии Лобачевского). При $e \to 0$ геометрия Лобачевского переходит в евклидову геометрию.

Поскольку геометрия Лобачевского интерпретируется в гиперболической связке окружностей, она часто называется *гиперболической геометрией.* По аналогичным соображениям, евклидова геометрия называется *параболической геометрией.*

Возможна, конечно, и *эллиптическая геометрия*, интерпретируемая в эллиптической связке окружностей. Эта геометрия была впервые построена Риманом и потому называется *неевклидовой геометрией Римана.* В этой геометрии параллельных прямых нет (любые две прямые пересекаются). Она может быть проинтерпретирована на расширенной (проективной) плоскости.

Глава 5

ЭЛЕМЕНТАРНАЯ ТЕОРИЯ ЛИНИЙ И ПОВЕРХНОСТЕЙ ВТОРОГО ПОРЯДКА

§ 1. ЛИНИИ ВТОРОГО ПОРЯДКА

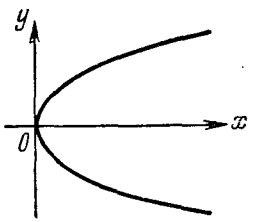
В этом параграфе все координаты x, y предполагаются прямоугольными.

1. Параболы

Определение 1. *Параболой* называется линия, имеющая в некоторой системе прямоугольных координат x, y уравнение

$$y^2 = 2px, \tag{1}$$

где $p > 0$. Уравнение (1) называется *каноническим уравнением* параболы, а координаты, в которых парабола имеет такое уравнение, — *каноническими координатами.*



Замечание 1. Из школьного курса читателю должно быть известно определение параболы как *графика квадратного трехчлена*:

$$y = ax^2 + bx + c. \tag{2}$$

Легко видеть, что переносом начала координат, переименованием координат и, если нужно, изменением знака одной из координат любое уравнение (2) можно привести к виду (1). Следовательно, наше определение совпадает со школьным.

Задание. Докажите последнее утверждение.

Поскольку уравнение (1) не меняется при замене y на $-y$, то вместе с некоторой точкой (x, y) парабола (1) содержит и точку $(x, -y)$, симметричную точке (x, y) относительно оси абсцисс. Это означает, что

ось абсцисс канонической системы координат является осью симметрии параболы.

Поскольку парабола (1) симметрична относительно оси абсцисс, нам достаточно исследовать только ее «верхнюю половину»

$$y = \sqrt{2px} \tag{2}$$

(имеется в виду «арифметический квадратный корень»). Из теории элементарных функций известно, что функция (2) является монотонно возрастающей функцией переменной x и стремится к ∞ при $x \to \infty$. Геометрически это означает, что чем дальше мы будем удаляться вправо в положительном направлении оси Ox, тем выше будет подниматься парабола и при достаточном удалении поднимется сколь угодно высоко.

Однако парабола поднимается вверх довольно «медленно», во всяком случае «медленнее» любой прямой (т. е. функция (2) возрастает медленнее любой линейной функции $y = ax + b$). Действительно, прямая, проходящая через точку $(0, 0)$ и произвольную точку (x, y) параболы (1) (точнее, ее верхней половины (2)), имеет угловой коэффициент

$$k = \frac{y}{x} = \frac{\sqrt{2p}}{\sqrt{x}},$$

монотонно стремящийся к нулю при $x \to \infty$. Это показывает (ввиду симметричности параболы), что

какой бы угол, охватывающий положительную полуось оси абсцисс мы ни взяли, все достаточно далекие (т. е. имеющие достаточно большую абсциссу) точки параболы будут лежать в этом угле.

Таким образом, если смотреть вдоль положительной полуоси оси абсцисс, то парабола будет казаться сходящейся!

Теперь легко видеть, что

ось Ox является единственной осью симметрии параболы (1).

Действительно, оси симметрии, параллельной оси Ox (но не совпадающей с этой осью), существовать не может, потому что иначе на оси Oy, кроме точки $(0, 0)$, была бы и другая (симметричная ей) точка, что невозможно (ибо согласно уравнению (1) при $x = 0$ обязательно $y = 0$). Пусть существует ось симметрии λ, не параллельная оси Ox. Рассмотрим достаточно малый угол, охватывающий положительную полуось оси Ox. При симметрии относительно прямой λ этот угол перейдет в некоторый другой угол, охватывающий полуось, симметричную полуоси оси Ox. Поскольку все достаточно далекие точки параболы содержатся в первом угле, они должны содержаться и во втором угле. Но при достаточной малости угла это, очевидно, невозможно.

Определение 2. Ось симметрии параболы (по доказанному — единственная) называется ее *осью*.

Предложение 1. *С точностью до изменения ориентации оси ординат каноническая система координат однозначно определена параболой.*

Доказательство. Осью абсцисс канонической системы является ось параболы, началом координат канонической системы — *вершина* параболы (точка пересечения ее с осью

симметрии[1])), а ориентацией оси абсцисс — ориентация, по отношению к которой парабола расположена в положительной полуплоскости $x \geqslant 0$. Таким образом, единственным элементом произвола, остающимся в канонической системе, является ориентация ее оси ординат.

Из предложения 1 вытекает, что все понятия и конструкции, использующие каноническую систему координат и не зависящие от ориентации ее оси ординат, инвариантно связаны с параболой (однозначно ею определены). В частности, это верно для числа p, точки $(p/2, 0)$ и прямой $x = -p/2$.

Определение 3. Число p называется *фокальным параметром* параболы, а число $p/2$ — ее *фокусным расстоянием.* Точка $(p/2, 0)$ называется *фокусом* параболы, а прямая $x = -p/2$ — ее *директрисой.*

Ясно, что

параметр p параболы равен расстоянию ее фокуса до ее директрисы, а фокусное расстояние $p/2$ равно расстоянию от ее фокуса до вершины параболы.

На прямой, проходящей через фокус параболы перпендикулярно оси (т. е. параллельно директрисе), парабола высекает некоторый отрезок, называемый *фокальной хордой* параболы. Ординаты концов фокальной хорды определяются, очевидно, из уравнения

$$y^2 = 2p \cdot \frac{p}{2},$$

т. е. равны $\pm p$. Следовательно,

длина фокальной хорды параболы равна $2p$.

Простое геометрическое описание параболы дается следующим предложением:

Предложение 2. *Парабола является геометрическим местом точек, равноудаленных от директрисы и фокуса.*

Задание. Докажите предложение 2.

Указание: расстояние δ_M произвольной точки плоскости $M(x, y)$ до директрисы $x = -p/2$ выражается формулой

$$\delta_M = \left| x + \frac{p}{2} \right|,$$

а ее расстояние r_M до фокуса — формулой

$$r_M = \sqrt{\left(x - \frac{p}{2}\right)^2 + y^2}.$$

В связи с предложением 2 необходимо заметить, что

любая прямая и любая не принадлежащая ей точка являются директрисой и фокусом некоторой (однозначно определенной) параболы.

[1]) Вообще, *вершиной* произвольной линии, имеющей ось симметрии, называется точка ее пересечения с этой осью.

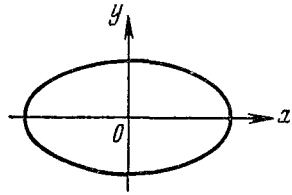
Действительно, пусть p — расстояние между данной точкой и данной прямой. Рассмотрим координатную систему, осью абсцисс которой является прямая, перпендикулярная данной прямой и проходящая через данную точку, а началом координат — середина перпендикуляра, опущенного из данной точки на данную прямую. Пусть ориентация оси абсцисс выбрана так, что данная точка имеет положительную абсциссу. Ориентацию оси ординат мы выберем произвольно. В этой координатной системе данная прямая имеет уравнение $x = -p/2$, а данная точка — координаты $(p/2, 0)$. Поэтому эти прямая и точка являются, соответственно, директрисой и фокусом параболы, задаваемой в построенной координатной системе уравнением

$$y^2 = 2px.$$

2. Эллипсы

Определение 1. *Эллипсом* называется линия, выражающаяся в некоторой системе прямоугольных координат уравнением вида

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad (1)$$



где $a \geqslant b > 0$.

Система координат, в которой эллипс имеет уравнение (1), называется *канонической* (для этого эллипса), а уравнение (1) называется *каноническим уравнением* эллипса.

При $a = b$ уравнение (1) является уравнением окружности радиуса a с центром в точке $(0, 0)$. Таким образом,

окружность является частным случаем эллипса.

Ясно, что любая точка с координатами вида

$$\begin{aligned} x &= a\cos\theta, \\ y &= b\sin\theta, \end{aligned} \qquad -\pi < \theta \leqslant \pi, \qquad (2)$$

удовлетворяет уравнению (1), и обратно, координаты любой точки, удовлетворяющей уравнению (1), могут быть (единственным образом) представлены в виде (2). Это означает, что *уравнения* (2) *являются параметрическими уравнениями эллипса* (1).

Пользуясь уравнениями (2), можно без труда с любой степенью точности начертить эллипс (для данных a и b) или хотя бы составить представление о его виде. Для этого нужно выбрать достаточно много значений θ, вычислить для этих значений по формулам (2) соответствующие координаты x, y и затем соединить найденные точки плавной кривой.

Иным способом представление о форме эллипса (а также все его свойства) можно получить, если ввести в рассмотрение преобразование плоскости,

переводящее произвольную точку $M(x, y)$ в точку $M'(x', y')$ с координатами

$$\begin{aligned} x' &= x, \\ y' &= ky, \end{aligned} \tag{3}$$

где k — некоторое положительное число. Это преобразование называется *сжатием* плоскости к оси Ox с коэффициентом k. (Конечно, настоящим «сжатием» оно является только при $k < 1$.)

Если точка $M(x, y)$ принадлежит окружности

$$x^2 + y^2 = R^2,$$

то, поскольку

$$\begin{aligned} x &= x', \\ y &= \frac{y'}{k}, \end{aligned}$$

точка $M'(x', y')$ будет удовлетворять соотношению

$$(x')^2 + \left(\frac{y'}{k}\right)^2 = R^2,$$

т. е. будет принадлежать эллипсу

$$\frac{x^2}{R^2} + \frac{y^2}{(kR)^2} = 1.$$

Это показывает, что

эллипс (1) *является результатом сжатия к оси Ox окружности*

$$x^2 + y^2 = a^2$$

с коэффициентом $k = \dfrac{b}{a} \leqslant 1$.

Замечание 1. Уравнения (2) не являются, конечно, единственно возможными параметрическими уравнениями эллипса. Например, переходя от параметра θ к параметру

$$t = \operatorname{tg}\frac{\theta}{2},$$

мы немедленно получим, что

эллипс (1) *может быть задан параметрическими уравнениями вида*

$$\begin{aligned} x &= a\,\frac{1 - t^2}{1 + t^2}, \\ y &= b\,\frac{2t}{1 + t^2}, \end{aligned} \qquad -\infty < t < +\infty.$$

Обратим внимание на то, что, таким образом, *любой эллипс может быть параметризован рационально*, т. е. для него существуют параметрические уравнения, правые части которых являются рациональными функциями параметра.

Заметим, что парабола $y^2 = 2px$ также допускает очевидную рациональную параметризацию:

$$\begin{aligned} x &= 2pt^2, \\ y &= 2pt. \end{aligned}$$

Уравнения (2) показывают, что координаты любой точки (x, y) эллипса (1) удовлетворяют неравенствам

$$|x| \leqslant a, \quad |y| \leqslant b. \tag{4}$$

Эти неравенства определяют на плоскости прямоугольник со сторонами длины $2a$ и $2b$, параллельными осям координат, и с центром в начале координат $(0, 0)$.

Определение 2. Прямоугольник (4) называется *основным прямоугольником* эллипса (1).

Как уже было сказано, эллипс (1) целиком содержится в прямоугольнике (4). При этом каждая сторона этого прямоугольника содержит одну (и только одну) точку эллипса. Именно,

сторона $x = a$, $-b \leqslant y \leqslant b$ содержит точку $(a, 0)$,
сторона $x = -a$, $-b \leqslant y \leqslant b$ содержит точку $(-a, 0)$.
сторона $-a \leqslant x \leqslant a$, $y = b$ содержит точку $(0, b)$,
сторона $-a \leqslant x \leqslant a$, $y = -b$ содержит точку $(0, -b)$.

Это означает, что

эллипс (1) *вписан в свой основной прямоугольник* (4).

Так как уравнение эллипса не меняется при замене x на $-x$ или y на $-y$, то

координатные оси канонической системы координат являются осями симметрии эллипса, а следовательно, начало координат — (единственным) центром симметрии эллипса.

Задание. Докажите, что ограниченная фигура (т. е. фигура, помещающаяся в некотором прямоугольнике) не может иметь более одного центра симметрии.

При $a = b$, т. е. для эллипса, являющегося окружностью, любая прямая, проходящая через его центр симметрии $(0, 0)$, является осью симметрии. Напротив, легко видеть, что

при $a > b$, т. е. для эллипса, не являющегося окружностью, координатные оси канонической системы координат являются его единственными осями симметрии.

Действительно, ясно, что любая ось симметрии фигуры, обладающей единственным центром симметрии, проходит через этот центр.

Следовательно, если эллипс (1) обладает осью симметрии, отличной от координатных осей, то уравнение этой оси имеет вид

$$y = kx, \qquad k \neq 0. \tag{5}$$

Но легко видеть, что симметрия относительно прямой (5) любую точку $M(x, y)$ переводит в точку $M'(x', y')$ с координатами

$$x' = \frac{(1 - k^2)\,x + 2ky}{1 + k^2}, \quad y' = \frac{2kx - (1 - k^2)\,y}{1 + k^2}.$$

Задание. Выведите эти формулы.

В частности, точка $A(a, 0)$ при симметрии относительно прямой (5) переходит в точку $A'(a', b')$ с координатами

$$a' = \frac{1-k^2}{1+k^2}a, \quad b' = \frac{2k}{1+k^2}a.$$

Если прямая (5) является осью симметрии эллипса (1), то, поскольку точка A принадлежит этому эллипсу, точка A' также ему принадлежит, т. е.

$$\frac{(a')^2}{a^2} + \frac{(b')^2}{b^2} = 1,$$

и следовательно,

$$\left(\frac{1-k^2}{1+k^2}\right)^2 + \left(\frac{2k}{1+k^2}\right)^2 \frac{a^2}{b^2} = 1.$$

Поскольку

$$1 - \left(\frac{1-k^2}{1+k^2}\right)^2 = \left(\frac{2k}{1+k^2}\right)^2 \neq 0$$

и поскольку, по условию, $a \geqslant b > 0$, тем самым доказано, что, вопреки предположению, $a = b$.

Задание. Докажите аналогичным образом отсутствие у параболы $y^2 = 2px$ осей симметрии, отличных от оси Ox (см. п. 1).

Доказанное утверждение означает, что координатные оси канонических координат однозначно характеризуются (при $a > b$) как оси симметрии эллипса. При этом ось Ox характеризуется как ось симметрии, на которой эллипс высекает отрезок, больший, чем отрезок, высекаемый эллипсом на другой оси (первый отрезок имеет длину $2a$, а другой — длину $2b$).

Следовательно, справедливо

Предложение 1. *С точностью до изменения ориентации координатных осей каноническая система координат однозначно определена эллипсом (не являющимся окружностью).*

Для эллипса, являющегося окружностью, канонической системой может служить произвольная система прямоугольных координат, начало которой совпадает с центром этой окружности.

Ввиду единственности канонической системы координат, все понятия и конструкции, использующие каноническую систему и не зависящие от ориентации ее осей, инвариантно связаны с эллипсом. В частности, это верно для основного прямоугольника (4) и чисел a, b, $c = \sqrt{a^2 - b^2}$, $e = \frac{c}{a}$ и $p = \frac{b^2}{a}$.

Определение 3. Числа a и b называются соответственно *большой полуосью* и *малой полуосью* эллипса [1]), число $c =$

[1]) Таким образом, эти «полуоси» являются числами!

$= \sqrt{a^2 - b^2}$ называется *линейным эксцентриситетом* эллипса, число $2c$ называется *фокусным расстоянием* эллипса, число

$$e = \frac{c}{a} = \sqrt{1 - \frac{b^2}{a^2}}$$

называется (*числовым*) *эксцентриситетом* эллипса, число

$$p = \frac{b^2}{a}$$

называется *фокальным параметром* эллипса.

Для окружности

$$a = b, \quad c = 0, \quad e = 0, \quad p = a.$$

С эллипсом инвариантно связаны также его центр симметрии $(0, 0)$ (называемый обычно просто *центром* эллипса), а при $a > b$ — его оси симметрии Ox и Oy (называемые обычно просто *осями* эллипса: ось Ox — *большой*, а ось Oy — *малой*) и четыре точки пересечения $(\pm a, 0)$, $(0, \pm b)$ эллипса с его осями (*вершины* эллипса). Точки

$$(c, 0) \quad \text{и} \quad (-c, 0)$$

большой оси тоже инвариантно определены эллипсом и называются его *фокусами* (фокус $(c, 0)$ — *правым*, а фокус $(-c, 0)$ — *левым*; заметим, что различение левого и правого фокусов уже не инвариантно: оно зависит от выбора ориентации оси Ox).

Термин «фокусное расстояние» для числа $2c$ объясняется тем, что это число, очевидно, равно расстоянию между фокусами.

Большая ось эллипса содержит его фокусы и потому иногда называется его *фокальной осью*.

Так как $0 < b \leqslant a$, то

$$0 \leqslant c < a \quad \text{и} \quad 0 \leqslant e < 1,$$

причем $c = 0$ (т. е. $e = 0$) тогда и только тогда, когда $a = b$, т. е. когда эллипс является окружностью. Таким образом, во-первых,

для эллипса, не являющегося окружностью, фокусы не совпадают друг с другом и отличны от его центра, тогда как для окружности фокусы совпадают с ее центром.

Во-вторых,

эксцентриситет e эллипса является неотрицательным числом, меньшим единицы:

$$0 \leqslant e < 1,$$

равным нулю тогда и только тогда, когда эллипс является окружностью.

Отрезки, высекаемые эллипсом на прямых, проходящих через фокусы эллипса перпендикулярно оси абсцисс канонической системы координат, называются *фокальными хордами* эллипса. Поскольку при симметрии относительно оси ординат фокальные хорды переходят друг в друга, они имеют одну и ту же длину.

Задание. Докажите, что длина фокальной хорды эллипса равна $2p$.

Для любой точки $M(x, y)$ эллипса длины отрезков, соединяющих ее с фокусами эллипса, называются *фокальными радиусами* этой точки (соответственно — *правым* и *левым*). Иногда «фокальными радиусами» называются сами эти отрезки.

Для точки окружности имеется только один фокальный радиус, совпадающий с ее радиусом.

По определению, левый фокальный радиус r_1 точки $M(x, y)$ эллипса выражается формулой

$$r_1 = \sqrt{(x+c)^2 + y^2}.$$

Следовательно,

$$r_1 = \sqrt{x^2 + 2xc + c^2 + b^2\left(1 - \frac{x^2}{a^2}\right)} =$$

$$= \sqrt{\left(1 - \frac{b^2}{a^2}\right)x^2 + 2cx + c^2 + b^2} =$$

$$= \sqrt{e^2x^2 + 2aex + a^2} = |a + ex|,$$

и потому

$$r_1 = a + ex, \tag{6}$$

ибо $|ex| = e|x| \leqslant ea < a$.

Аналогично, для правого фокального радиуса r_2 получается, что

$$r_2 = a - ex. \tag{7}$$

Из этих формул непосредственно вытекает, что для любой точки эллипса

$$r_1 + r_2 = 2a.$$

Обратно, пусть $M(x, y)$ — такая точка плоскости, что сумма ее расстояний от фокусов эллипса равна $2a$, т. е. такая, что

$$\sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2} = 2a.$$

«Уединяя радикалы» и дважды возводя в квадрат, мы, как легко видеть, получим уравнение

$$(a^2 - c^2)x^2 + a^2y^2 = a^2(a^2 - c^2),$$

равносильное, очевидно, уравнению эллипса (1) (ибо, по определению, $a^2 - c^2 = b^2$). Тем самым доказано, что

точка $M(x, y)$ тогда и только тогда принадлежит эллипсу (1), *когда сумма ее расстояний от фокусов эллипса равна* $2a$.

Это означает, что справедливо следующее

Предложение 2. *Любой эллипс является геометрическим местом точек, сумма расстояний которых от двух данных точек (фокусов эллипса) постоянна (равна длине большой оси эллипса).*

В частности, мы видим, что

эллипс однозначно определен, когда известны его фокусы и его большая полуось a.

При этом ясно, что

для любых двух точек и любого числа a, большего половины расстояния между этими точками, существует (единственный) эллипс, фокусами которого являются данные точки, а большая полуось равна a.

Действительно, если данные точки совпадают, то искомым эллипсом является окружность радиуса a с центром в этих совпадающих точках. Пусть данные точки различны. Рассмотрим систему прямоугольных координат x, y, осью абсцисс которой является прямая, проходящая через данные точки, а осью ординат — их медиатриса. В этой координатной системе данные точки имеют координаты $(-c, 0)$ и $(c, 0)$, где c — половина расстояния между ними, и потому являются фокусами эллипса

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

где $b^2 = a^2 - c^2$. Для завершения доказательства остается заметить, что большой полуосью этого эллипса является, очевидно, число a.

Замечание 2. Обратим внимание на то, что предложение, обратное предложению 2, требует очевидных оговорок: геометрическое место точек, сумма расстояний которых от двух данных точек (находящихся друг от друга на расстоянии $2c$) равна данному числу $2a$, является эллипсом только при $a > c$ (если $a = c$, то геометрическое место представляет собой отрезок, а если $a < c$, то — пустое множество).

Определение 4. Для любого эллипса, не являющегося окружностью, прямые

$$x = -\frac{a}{e} \quad \text{и} \quad x = \frac{a}{e},$$

называются его *директрисами* (соответственно *левой* и *правой*).

Директрисы инвариантно связаны с эллипсом, но различение правой и левой директрис зависит от ориентации оси абсцисс канонической системы координат.

Замечание 3. На евклидово-проективной плоскости директрисы имеют (в однородных координатах $X : Y : Z = x : y : 1$) уравнения

$$eX - aZ = 0 \quad \text{и} \quad eX + aZ = 0.$$

При $e = 0$ эти уравнения превращаются в уравнение $Z = 0$ несобственной прямой. На этом основании *целесообразно считать, что окружность имеет одну «двойную директрису» — несобственную прямую.*

Наглядно (на евклидовой плоскости) это отражается в том, что если эллипс изменяется так, что его большая полуось a остается неизменной, а эксцентриситет e стремится к нулю (т. е., другими словами, если эллипс приближается к окружности $x^2 + y^2 = a^2$), то его директрисы неограниченно удаляются от оси ординат.

Фокус и директриса называются *одноименными*, если они оба — правые или оба — левые. Ясно, что это отношение между фокусом и директрисой геометрически инвариантно.

Расстояние d фокуса до одноименной директрисы выражается, очевидно, формулой

$$d = \frac{a}{e} - c.$$

Поскольку

$$\frac{a}{e} - c = a\left(\frac{1}{e} - e\right) = a\,\frac{1 - e^2}{e} = \frac{1}{e}\,\frac{b^2}{a} = \frac{p}{e},$$

тем самым доказано, что

$$d = \frac{p}{e}. \tag{8}$$

Поскольку для любой точки $M(x, y)$ эллипса $|x| \leqslant a$, а $e < 1$, то весь эллипс расположен в полосе между директрисами $x = \pm \frac{a}{e}$. Поэтому расстояние δ_1 точки $M(x, y)$ эллипса до левой директрисы $x = -\frac{a}{e}$ выражается формулой

$$\delta_1 = \frac{a}{e} + x,$$

а расстояние δ_2 до правой директрисы $x = \frac{a}{e}$ — формулой

$$\delta_2 = \frac{a}{e} - x.$$

Сравнив эти формулы с формулами (6) и (7) для фокальных радиусов r_1 и r_2 точки $M(x, y)$, мы немедленно получим, что для любой точки эллипса

$$\frac{r_1}{\delta_1} = \frac{r_2}{\delta_2} = e.$$

Обратно, пусть $M(x, y)$ — такая точка плоскости, что, скажем, отношение ее расстояния от левого фокуса эллипса к ее расстоянию от левой директрисы эллипса равно e, т. е. такая, что

$$\sqrt{(x + c)^2 + y^2} = e\left|\frac{a}{e} + x\right|.$$

Тогда $(x+c)^2+y^2=(a+ex)^2$, т. е.

$$x^2+2cx+c^2+y^2=a^2+2aex+e^2x^2$$

или

$$(1-e^2)x^2+y^2=a^2-e^2.$$

Поскольку последнее уравнение равносильно, очевидно, уравнению эллипса (1), тем самым доказано, что

точка $M(x, y)$ тогда и только тогда принадлежит эллипсу (1) (*не являющемуся окружностью*), *когда отношение ее расстояния от левого фокуса эллипса к ее расстоянию от левой директрисы эллипса равно эксцентриситету e эллипса.*

Так как эллипс симметричен относительно оси Oy, то аналогичное утверждение верно и по отношению к правым фокусу и директрисе.

Таким образом, нами доказано следующее

Предложение 3. *Любой эллипс* (*не являющийся окружностью*) *представляет собой геометрическое место точек, отношения расстояний которых от данной точки* (*фокуса эллипса*) *и от данной прямой* (*одноименной с фокусом директрисы*) *равны данному положительному числу, меньшему единицы* (*эксцентриситету эллипса*).

В частности, мы видим, что

эллипс однозначно определен, если известны его фокус, одноименная с этим фокусом директриса и его эксцентриситет.

При этом

для любой точки, любой не содержащей эту точку прямой и любого положительного числа e, меньшего единицы, существует (*единственный*) *эллипс с эксцентриситетом e, для которого данная точка является фокусом, а данная прямая — одноименной с этим фокусом директрисой.*

Задание. Докажите это утверждение.

Замечание 4. Предложение 3 аналогично предложению 2 п. 1 и переходит в него при $e=1$. Поэтому целесообразно считать, что

эксцентриситет e параболы равен единице.

Это согласуется также с формулой (8), поскольку для параболы расстояние d равно фокальному параметру p.

Замеченное сходство между эллипсом и параболой можно выявить и по-другому, рассмотрев уравнение эллипса в системе координат x', y', получающейся из канонической системы переносом начала в левую вершину $(-a, 0)$ эллипса. Канонические координаты x, y выражаются через эти координаты по формулам

$$x=x'-a,$$
$$y=y'.$$

Поэтому уравнение эллипса в координатах x', y' имеет вид

$$\frac{(x'-a)^2}{a^2}+\frac{(y')^2}{b^2}=1.$$

Отбросив штрихи и раскрыв скобки, мы немедленно получим отсюда уравнение

$$y^2=2px+qx^2, \tag{9}$$

где

$$q=-\frac{b^2}{a^2}=e^2-1.$$

Тем самым мы доказали, что

в указанной системе прямоугольных координат эллипс выражается уравнением (9), *где* $-1\leqslant q<0$.

Уравнение (9) называется уравнением эллипса «при вершине».

При $q=0$, т. е. при $e=1$, уравнение (9) переходит в каноническое уравнение параболы. Это еще раз оправдывает представление о параболе как о линии, эксцентриситет которой равен единице.

Замечание 5. Пусть эллипс (9) меняется так, что его фокальный параметр p остается неизменным, а величина q стремится к нулю (т. е. эксцентриситет e стремится к единице). Тогда эллипс будет неограниченно вытягиваться в положительном направлении оси Ox и в пределе даст параболу $y^2=2px$. В этом смысле парабола является предельным положением эллипса при $e\to 1$ (и p постоянном).

Замечание 6. Легко видеть, что

при $q<-1$ *уравнение* (9) *также является уравнением некоторого эллипса с вершиной в начале координат.*

Действительно, в координатах x', y', связанных с координатами x, y соотношениями

$$x=y'-\frac{p}{q},$$
$$y=x',$$

линия (9) имеет уравнение

$$x'^2=2p\left(y'-\frac{p}{q}\right)+q\left(y'-\frac{p}{q}\right)^2,$$

т. е. уравнение

$$x'^2=qy'^2-\frac{p^2}{q}.$$

Отбрасывая штрихи и полагая

$$a^2=-\frac{p^2}{q},\quad b^2=\frac{p^2}{q^2}$$

(напомним, что, по условию, $q < -1$), мы получим каноническое уравнение эллипса

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

(это, действительно, каноническое уравнение, поскольку $a > b$). Фокальный параметр этого эллипса равен $\frac{p}{\sqrt{-q}}$, а его эксцентриситет равен $\sqrt{1 + \frac{1}{q}}$.

Отличие случая $q < -1$ от случая $-1 \leqslant q < 0$ состоит в том, что во втором случае эллипс «как и полагается» вытянут в горизонтальном направлении, тогда как в первом случае он вытянут в вертикальном направлении.

Заметим еще, что при $q = -1$ мы получаем окружность

$$y^2 = 2px - x^2$$

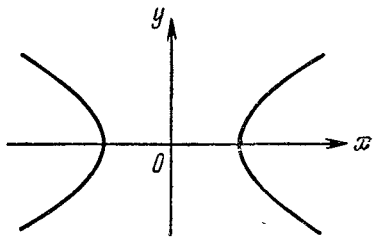
радиуса p с центром в точке $(p, 0)$.

3. Гиперболы

Определение 1. *Гиперболой* называется линия, выражающаяся в некоторой системе прямоугольных координат уравнением вида

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \qquad (1)$$

где $a > 0$, $b > 0$.



Система координат, в которой гипербола имеет уравнение (1), называется *канонической* (для этой гиперболы), а уравнение (1) называется *каноническим уравнением* гиперболы.

Частным случаем гиперболы является линия с уравнением

$$\frac{x^2}{a^2} - \frac{y^2}{a^2} = 1,$$

т. е. с уравнением

$$x^2 - y^2 = a^2. \qquad (2)$$

Эта линия называется *равнобочной гиперболой.*

В координатах x', y', связанных с каноническими координатами x, y соотношениями

$$x' = \frac{\sqrt{2}}{2}(x - y),$$

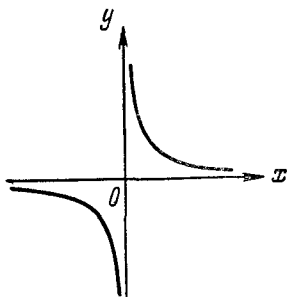
$$y' = \frac{\sqrt{2}}{2}(x + y)$$

(для этих координат оси Ox' и Oy' получаются поворотом по часовой стрелке осей Ox и Oy на $\pi/4$, т. е. являются биссектрисами координатных углов), равнобочная гипербола выражается уравнением

$$x'y' = \frac{a^2}{2},$$

т. е. представляет собой известный из школьного курса график обратной пропорциональности.

Напомним основные факты об этом графике. Для упрощения формул мы уберем у обозначений координат штрихи и положим $A = \frac{a^2}{2}$. Таким образом, речь у нас будет идти о графике функции



$$y = \frac{A}{x}, \qquad A > 0. \tag{3}$$

Так как $y > 0$ при $x > 0$ и $y < 0$ при $x < 0$, то

график функции (3) *расположен в первом и третьем квадрантах.*

При этом на оси ординат Oy нет ни одной точки графика, так что эта ось отделяет часть графика, расположенную в первом квадранте, от его части, расположенной в третьем квадранте. Эти две отдельные части графика называются его *ветвями.*

Поскольку уравнение $xy = A$ не меняется при одновременном изменении знаков у x и y, то

график функции (3) *симметричен относительно начала координат,*

а поскольку это уравнение не меняется и при перестановке координат, то

график функции (3) *симметричен относительно биссектрис координатных углов.*

При этом, при симметрии относительно биссектрисы второго и четвертого координатных углов ветви графика переходят друг в друга, а при симметрии относительно биссектрисы первого и третьего координатных углов каждая ветвь графика переходит в себя. Таким образом,

каждая ветвь графика функции (3) *симметрична относительно биссектрисы первого и третьего координатных углов.*

Определение 2. Пусть Γ — произвольная линия и α — некоторая прямая. Предположим, что линия Γ имеет часть, «уходящую в бесконечность» (т. е. такую, что, двигаясь по этой части линии, точка $M(x, y)$ неограниченно удаляется от начала координат [1])), и что расстояние точки $M(x, y)$ от прямой α

[1]) Строгое описание такого поведения кривой требует рассмотрения ее параметрических уравнений. Мы это предоставим сделать читателю.

стремится при этом к нулю. В этой ситуации прямая α называется *асимптотой* линии Г. Пусть M_0 — основание перпендикуляра, опущенного из точки M на прямую α. Когда точка M удаляется по рассматриваемой части линии Г в бесконечность, точка M_0 также удаляется в бесконечность, двигаясь по прямой α в определенном направлении (хотя, вообще говоря, и не монотонно). Говорят, что прямая α является асимптотой линии Г *в этом направлении.*

Способы разыскания асимптот (когда они существуют) излагаются в курсе анализа. Мы ограничимся лишь случаем, когда асимптотой является ось абсцисс Ox и когда рассматриваемая линия является графиком некоторой функции $y = f(x)$. Так как расстояние от произвольной точки $(x, f(x))$ этого графика до оси Ox равно, очевидно, $|f(x)|$, то утверждение, что ось Ox является в положительном направлении асимптотой графика $y = f(x)$, равносильно утверждению, что $f(x) \to 0$ при $x \to +\infty$.

Для интересующей нас функции $y = \frac{A}{x}$ дело обстоит именно так: $y \to 0$ при $x \to +\infty$. Следовательно,

ось абсцисс является асимптотой графика функции (3) *в положительном направлении.*

В силу указанных выше свойств симметрии графика отсюда следует, что

ось абсцисс является асимптотой графика функции (3) *и в отрицательном направлении,*

а также, что

ось ординат тоже является асимптотой графика функции (3) (*как в положительном, так и в отрицательном направлениях*).

Более точно,

оси абсцисс и ординат являются в положительных направлениях асимптотами ветви графика, расположенной в первом квадранте, а в отрицательных направлениях — асимптотами ветви, расположенной в третьем квадранте.

Поскольку график функции (3) «уходит в бесконечность» только при $x \to \pm\infty$ и $x \to 0$,

никаких других асимптот этот график не имеет.

Возвращаясь к равнобочной гиперболе

$$x^2 - y^2 = a^2 \tag{2}$$

и учитывая, что при повороте на $\pi/2$ координатные оси переходят в биссектрисы координатных углов и наоборот, мы получаем, во-первых, что

оси координат являются осями симметрии равнобочной гиперболы (2), *а начало координат — ее центром симметрии,*

и, во-вторых, что

равнобочная гипербола (2) *имеет две* (*и только две*) *асимптоты, являющиеся биссектрисами координатных углов.*

Кроме того,

гипербола (2) *целиком помещается в двух вертикальных углах, образованных этими биссектрисами и содержащими ось абсцисс.*

Часть гиперболы, содержащаяся в правом вертикальном угле, называется *правой ветвью*, а содержащаяся в левом — *левой ветвью.* (Конечно, различение правой и левой ветви зависит от ориентации оси абсцисс.)

Поскольку уравнение (2) может быть переписано в виде

$$x^2 = a^2 + y^2,$$

абсцисса любой точки гиперболы удовлетворяет неравенству

$$|x| \geqslant a,$$

причем равенство возможно только при $y = 0$.

Таким образом,

равнобочная гипербола пересекает ось абсцисс в точках $(\pm a, 0)$, *причем ее правая ветвь расположена справа от прямой* $x = a$, *а левая — слева от прямой* $x = -a$.

В частности,

в полосе $-a < x < a$ *нет ни одной точки гиперболы.*

Все сказанное доставляет нам уже довольно определенное представление о форме равнобочной гиперболы.

С другой стороны, легко видеть, что

произвольная гипербола

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \tag{1}$$

получается из равнобочной гиперболы

$$\frac{x^2}{a^2} - \frac{y^2}{a^2} = 1 \tag{2}$$

сжатием к оси Ox с коэффициентом $k = \frac{b}{a}$.

Действительно, при сжатии к оси Ox в отношении k каждая точка (x, y) плоскости получается из точки $(x, y/k)$. Поэтому в результате сжатия линия с уравнением (2) переходит в линию с уравнением

$$\frac{x^2}{a^2} - \frac{y^2}{(ka)^2} = 1,$$

т. е. $\left(\text{при } k = \frac{b}{a}\right)$ — в гиперболу (1).

Тем самым мы получаем вполне удовлетворительное представление о форме и произвольной гиперболы.

В частности, мы видим, что

гипербола (1) *имеет две* (*и только две*) *асимптоты, выражающиеся уравнениями*

$$y = \frac{b}{a}x, \quad y = -\frac{b}{a}x.$$

Действительно, мы знаем, что асимптотами равнобочной гиперболы (2) являются биссектрисы координатных углов, т. е. диагонали произвольного квадрата с центром в начале координат и сторонами, параллельными осям координат. Рассмотрим, в частности, квадрат, длины сторон которого равны $2a$. При сжатии к оси Ox с коэффициентом $\frac{b}{a}$ этот квадрат переходит в прямоугольник с вертикальными сторонами длины $2b$ и горизонтальными сторонами прежней длины $2a$, а его диагонали переходят в диагонали этого прямоугольника, т. е. в прямые $y = \pm \frac{b}{a}x$.

Определение 3. Прямоугольник $|x| \leqslant a$, $|y| \leqslant b$ называется *основным прямоугольником* гиперболы (1).

Замечание 1. Ясно, что

асимптоты гиперболы тогда и только тогда перпендикулярны, когда $b = a$, *т. е. когда гипербола равнобочна.*

Далее, мы видим, что

гипербола (1) *целиком расположена в содержащих ось абсцисс вертикальных углах, образованных асимптотами, но вне полосы* $-a < x < a$.

Кроме того,

гипербола (1) *касается сторон* $x = a$ *и* $x = -a$ *основного прямоугольника в точках* $(a, 0)$ *и* $(-a, 0)$ *и состоит из двух ветвей — правой и левой.*

Наконец,

для гиперболы (1) *начало координат является центром симметрии, а оси координат — осями симметрии, причем ось абсцисс пересекает гиперболу, а ось ординат — нет.*

При этом

никаких других осей симметрии гипербола не имеет.

Действительно, поскольку при любой симметрии асимптоты должны переходить в асимптоты, а точки касания гиперболы сторон основного прямоугольника — в точки касания, то основной прямоугольник гиперболы должен переходить сам в себя, причем его вертикальные стороны должны после симметрии остаться вертикальными. Но это, очевидно, возможно только при симметриях относительно средних линий этого прямоугольника, т. е. при симметриях относительно осей координат.

Таким образом, ось абсцисс канонической системы координат мы можем охарактеризовать как ось симметрии гиперболы, пересекающую эту гиперболу, а ось ординат — как ось симметрии гиперболы, не пересекающую гиперболу.

Поэтому справедливо следующее

Предложение 1. *С точностью до ориентации осей каноническая система координат однозначно определена гиперболой.*

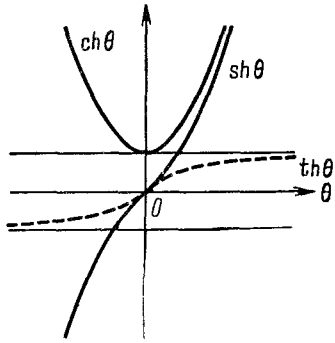
Рассмотрим функции

$$\operatorname{sh}\theta = \frac{e^{\theta} - e^{-\theta}}{2}, \quad \operatorname{ch}\theta = \frac{e^{\theta} + e^{-\theta}}{2},$$

называемые соответственно *гиперболическим синусом* и *гиперболическим косинусом.*

Элементарное исследование (например, методами, известными из курса анализа) показывает, что

обе функции $\operatorname{sh}\theta$ *и* $\operatorname{ch}\theta$ *определены для всех* θ, *причем функция* $\operatorname{sh}\theta$ *нечетна и при изменении* θ *от* $-\infty$ *до* $+\infty$ *монотонно возрастает от* $-\infty$ *до* $+\infty$, *обращаясь в нуль при* $\theta = 0$. *а функция* $\operatorname{ch}\theta$ *четна и при изменении* θ *от* $-\infty$ *до* $+\infty$ *сначала монотонно убывает от* $+\infty$ *до* 1 (*получающейся при* $\theta = 0$), *а затем монотонно возрастает от* 1 *до* $+\infty$.



Далее, так как

$$\left(\frac{e^{\theta} \pm e^{-\theta}}{2}\right)^2 = \frac{e^{2\theta} + e^{-2\theta} \pm 2}{4},$$

то

$$\operatorname{ch}^2\theta - \operatorname{sh}^2\theta = 1. \tag{4}$$

Это соотношение (аналогичное известному тождеству для тригонометрических функций $\cos\theta$ и $\sin\theta$) позволяет построить для гиперболических функций «тригонометрию», весьма похожую на тригонометрию функций $\cos\theta$ и $\sin\theta$. В частности, можно ввести *гиперболический тангенс*

$$\operatorname{th}\theta = \frac{\operatorname{sh}\theta}{\operatorname{ch}\theta} = \frac{e^{\theta} - e^{-\theta}}{e^{\theta} + e^{-\theta}},$$

доказать формулы сложения, формулы двойных и половинных аргументов и т. п. Например, по аналогии с известными формулами тригонометрии, выражающими синус и косинус через тангенс половинного угла, для гиперболических функций справедливы формулы

$$\operatorname{sh}\theta = \frac{2\operatorname{th}\frac{\theta}{2}}{1 - \operatorname{th}^2\frac{\theta}{2}}, \quad \operatorname{ch}\theta = \frac{1 + \operatorname{th}^2\frac{\theta}{2}}{1 - \operatorname{th}^2\frac{\theta}{2}} \tag{5}$$

(проще всего эти формулы проверяются прямым вычислением).

Ввиду соотношения (4) любая точка вида $(a\operatorname{ch}\theta,\ b\operatorname{sh}\theta)$ принадлежит гиперболе

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

или, более точно, ее правой ветви (поскольку всегда $\operatorname{ch}\theta \geqslant 1$). Обратно, любая точка этой правой ветви однозначно представляется в виде $(a\operatorname{ch}\theta,\ b\operatorname{sh}\theta)$. Действительно, поскольку функ-

ция $\operatorname{sh}\theta$ монотонно изменяется от $-\infty$ до $+\infty$ (и непрерывна), для любого y существует единственное θ, удовлетворяющее соотношению $y = b \operatorname{sh}\theta$. Тогда

$$x^2 = a^2\left(1 + \frac{y^2}{b^2}\right) = a^2(1 + \operatorname{sh}^2\theta) = a^2 \operatorname{ch}\theta$$

и, следовательно, $x = a \operatorname{ch}\theta$ (ибо, по условию, $x > 0$).

Тем самым мы доказали, что

уравнения

$$\begin{aligned} x &= a \operatorname{ch}\theta, \\ y &= b \operatorname{sh}\theta, \end{aligned} \qquad -\infty < \theta < +\infty \tag{6}$$

являются параметрическими уравнениями правой ветви гиперболы (1).

Переходя к параметру $t = \operatorname{th}\frac{\theta}{2}$, мы получаем отсюда (см. формулы (4)), что

правая ветвь гиперболы (1) *может быть задана и параметрическими уравнениями вида*

$$\begin{aligned} x &= a\frac{1+t^2}{1-t^2}, \\ y &= b\frac{2t}{1-t^2}, \end{aligned} \qquad -1 < t < 1.$$

Таким образом, аналогично эллипсу и параболе,

гипербола (*точнее, ее правая ветвь*) *допускает рациональную параметризацию.*

Поскольку с точностью до ориентации осей каноническая координатная система однозначно определена гиперболой, все конструкции и понятия, использующие эту координатную систему и не зависящие от ориентации ее осей, инвариантно связаны с гиперболой. В частности, это верно для чисел a, b, $c = \sqrt{a^2 + b^2}$, $e = \frac{c}{a}$ и $p = \frac{b^2}{a}$.

Определение 4. Числа a и b называются соответственно *действительной* и *мнимой полуосями* гиперболы [1]), число

$$c = \sqrt{a^2 + b^2}$$

называется *линейным эксцентриситетом* гиперболы, число $2c$ — *фокусным расстоянием* гиперболы, число

$$e = \frac{c}{a} = \sqrt{1 + \frac{b^2}{a^2}}$$

[1]) Происхождение довольно неудачных, но сохраняющихся по традиции эпитетов «действительная» и «мнимая» в применении к полуосям гиперболы объясняется тем, что уравнение (определяющее точки пересечения гиперболы с осью Ox, имеет вещественные (действительные) решения, а аналогичное уравнение для оси Oy — мнимые решения.

— (*числовым*) *эксцентриситетом* гиперболы, а число

$$p = \frac{b^2}{a}$$

— *фокальным параметром* гиперболы.

Для равнобочной гиперболы

$$a = b, \quad c = \sqrt{2}a, \quad e = \sqrt{2}, \quad p = a.$$

С гиперболой инвариантно связаны также ее (единственный) центр симметрии (0, 0) (называемый обычно просто *центром* гиперболы), ее оси симметрии Ox и Oy (называемые обычно просто *осями*: ось Ox — *действительной осью*, а ось Oy — *мнимой осью*) и две точки пересечения $(\pm a, 0)$ гиперболы с действительной осью (*вершины* гиперболы). Точки

$$(-c, 0) \quad \text{и} \quad (c, 0)$$

тоже инвариантно определены гиперболой и называются ее *фокусами* (соответственно *левым* и *правым*; заметим, что различение левого и правого фокуса зависит от ориентации оси Ox).

Термин «фокусное расстояние» для числа $2c$ объясняется тем, что это число, очевидно, равно расстоянию между фокусами.

Действительная ось гиперболы содержит ее фокусы и потому называется также *фокальной осью*.

Так как $b \neq 0$, то

$$c > a \quad \text{и} \quad e > 1,$$

причем $c = \sqrt{2}a$ (т. е. $e = \sqrt{2}$) тогда и только тогда, когда $a = b$, т. е. когда гипербола равнобочна.

Поскольку $c > a$, прямые $x = \pm c$ пересекают гиперболу. Отрезки, высекаемые гиперболой на этих прямых, называются ее *фокальными хордами*. Поскольку при симметрии относительно оси ординат фокальные хорды переходят друг в друга, они имеют одну и ту же длину.

Задание. Докажите, что длина фокальной хорды гиперболы равна $2p$.

Для любой точки $M(x, y)$ гиперболы длины отрезков, соединяющих ее с фокусами гиперболы, называются *фокальными радиусами* этой точки (соответственно *правым* и *левым*). Иногда «фокальными радиусами» называются сами эти отрезки.

По определению, левый фокальный радиус r_1 точки $M(x, y)$ гиперболы выражается формулой

$$r_1 = \sqrt{(x + c)^2 + y^2}.$$

Следовательно,

$$r_1 = \sqrt{x^2 + 2cx + c^2 + b^2\left(\frac{x^2}{a^2} - 1\right)} =$$

$$= \sqrt{\left(1 + \frac{b^2}{a^2}\right)x^2 + 2cx + c^2 - b^2} =$$

$$= \sqrt{e^2x^2 + 2eax + a^2} = |ex + a|.$$

Поскольку для точек правой ветви $x > 0$, а для точек левой ветви $x < 0$, причем $|ex| > |x| \geqslant a$, тем самым доказано, что

$$r_1 = \begin{cases} a + ex & \text{при} \quad x > 0, \\ -a - ex & \text{при} \quad x < 0. \end{cases} \tag{7}$$

Аналогично, для правого фокального радиуса r_2 получается, что

$$r_2 = |ex - a|,$$

т. е. что

$$r_2 = \begin{cases} -a + ex & \text{при} \quad x > 0, \\ a - ex & \text{при} \quad x < 0. \end{cases} \tag{8}$$

Из этих формул непосредственно вытекает, что

$$r_1 - r_2 = \begin{cases} 2a & \text{при} \quad x > 0, \\ -2a & \text{при} \quad x < 0, \end{cases}$$

т. е. что

$$|r_1 - r_2| = 2a.$$

Задание. Докажите обратное утверждение, т. е. что точка плоскости, для которой разность ее расстояний от фокусов гиперболы равна по абсолютной величине числу $2a$, принадлежит гиперболе.

Таким образом,

точка $M(x, y)$ тогда и только тогда принадлежит гиперболе (1), *когда абсолютная величина разности ее расстояний от фокусов гиперболы равна* $2a$.

Это означает, что справедливо следующее

Предложение 2. *Любая гипербола является геометрическим местом точек, абсолютная величина разности расстояний которых от двух данных точек (фокусов гиперболы) постоянна (равна* $2a$).

Заметим, что

геометрическое место точек, разность расстояний которых от двух данных точек равна данному положительному числу $2a$, *является одной из ветвей гиперболы (а именно, ветвью, одноименной с фокусом, фокальный радиус которого является вычитаемым).*

Задание. Докажите это утверждение.

Замечание 2. Так же как в аналогичной ситуации для эллипса, геометрическое место точек, абсолютная величина разности расстояний которых от двух данных точек (находящихся друг от друга на расстоянии $2c$) равна данному неотрицательному числу $2a$, является гиперболой только при $0 < 2a < 2c$. При $2a = 2c$ это геометрическое место представляет собой внешность (два луча) отрезка с концами в данных точках, а при $2a > 2c$ — пусто. При $a = 0$ оно является медиатрисой данных точек.

Согласно предложению 2

гипербола однозначно определена, когда известны ее фокусы и действительная полуось a.

При этом

для любых двух различных точек и любого положительного числа a, *меньшего половины расстояния между этими точками, существует (единственная) гипербола, фокусами которой являются данные точки, а действительная полуось равна* a.

Задание. Докажите это утверждение (ср. в п. 2 доказательство аналогичного утверждения для эллипса).

Определение 5. Прямые

$$x = -\frac{a}{e}, \quad x = \frac{a}{e}$$

называются *директрисами* (соответственно *левой* и *правой*) гиперболы (1). Директрисы инвариантно связаны с гиперболой, но различение правой и левой директрис зависит от ориентации оси абсцисс.

Заметим, что поскольку $e > 1$,

директрисы отделяют вершины гиперболы от ее центра.

Фокус и директриса называются *одноименными*, если они оба — правые или оба — левые. Ясно, что это отношение между фокусом и директрисой геометрически инвариантно.

Расстояние d фокуса до одноименной директрисы выражается, очевидно, формулой

$$d = c - \frac{a}{e}.$$

Поскольку

$$c - \frac{a}{e} = a\left(e - \frac{1}{e}\right) = a\,\frac{e^2 - 1}{e} = \frac{1}{e}\,\frac{b^2}{a} = \frac{p}{e},$$

тем самым доказано (ср. формулу (8) п. 2), что

$$d = \frac{p}{e}.$$

Так же, как для эллипса, показывается, что

точка $M(x, y)$ *тогда и только тогда принадлежит гиперболе* (1), *когда отношение ее расстояния до некоторого фокуса ги-*

перболы к ее расстоянию до одноименной директрисы гиперболы равно эксцентриситету e гиперболы.

Задание. Докажите это утверждение (не забудьте рассмотреть обе ветви гиперболы).

Таким образом, справедливо следующее

Предложение 3. *Любая гипербола представляет собой геометрическое место точек, отношения расстояний которых от данной точки (фокуса гиперболы) и от данной прямой (одноименной с фокусом директрисы) равно данному положительному числу e, большему единицы (эксцентриситету гиперболы).*

В частности,

гипербола однозначно определена, если известны ее фокус, одноименная с этим фокусом директриса и ее эксцентриситет e.

При этом

для любой точки, любой не содержащей эту точку прямой и любого положительного числа e, большего единицы, существует (единственная) гипербола с эксцентриситетом e, для которой данная точка является фокусом, а данная прямая — одноименной с этим фокусом директрисой.

Задание. Докажите это утверждение.

Замечание 3. Таким образом, гипербола может быть определена той же геометрической конструкцией, что и эллипс и парабола (см. предложение 2 п. 1 и предложение 3 п. 2), только при значении параметра e, большем единицы. Следовательно,

и эллипс (не являющийся окружностью), и парабола, и гипербола являются геометрическими местами точек, отношения расстояний которых от данной точки (фокуса) до данной прямой (одноименной директрисы) постоянны (равны эксцентриситету e).

При этом

значениям e от нуля до единицы соответствуют эллипсы, значению $e = 1$ соответствуют параболы, а значениям $e > 1$ — гиперболы (в частности, значению $e = \sqrt{2}$ — равнобочные гиперболы).

Единство эллипсов, парабол и гипербол можно прояснить и с иной точки зрения. Рассмотрим с этой целью для произвольной гиперболы (1) прямоугольные координаты x', y', получающиеся из канонических координат x, y при переносе начала координат в правую вершину $(a, 0)$ гиперболы. Канонические координаты x, y выражаются через эти координаты по формулам

$$x = x' + a,$$
$$y = y'.$$

Поэтому уравнение гиперболы в координатах x', y' имеет вид

$$\frac{(x'+a)^2}{a^2} - \frac{y'^2}{b^2} = 1.$$

Отбросив штрихи и раскрыв скобки, мы немедленно получим отсюда уравнение

$$y^2 = 2\frac{b^2}{a}x + \frac{b^2}{a^2}x^2,$$

т. е. уравнение

$$y^2 = 2px + qx^2, \tag{9}$$

где

$$q = \frac{b^2}{a^2} = e^2 - 1.$$

Тем самым мы доказали, что

в указанной системе прямоугольных координат гипербола выражается уравнением (9) *с* $q > 0$.

Замечание 4. Пусть гипербола меняется так, что ее фокальный параметр p остается неизменным, а величина q стремится к нулю (т. е. эксцентриситет e стремится к единице). Тогда левая ветвь гиперболы будет неограниченно удаляться влево и в пределе исчезнет, а правая ветвь будет «сходиться» и в пределе перейдет в параболу $y^2 = 2px$. В этом смысле (ср. замечание 4 п. 2) парабола является предельным положением гиперболы, получающимся при $e \to 1$ (и p постоянном).

Сопоставляя полученные результаты с результатами п. 2, мы, в частности, получаем следующее

Предложение 4. *Эллипс, парабола и гипербола могут быть заданы уравнением*

$$y^2 = 2px + qx^2, \tag{10}$$

где $q < 0$ *для эллипса,* $q = 0$ *для параболы и* $q > 0$ *для гиперболы.*

При этом все эллипсы получаются уже при $-1 \leqslant q < 0$ (*окружности — при* $q = -1$).

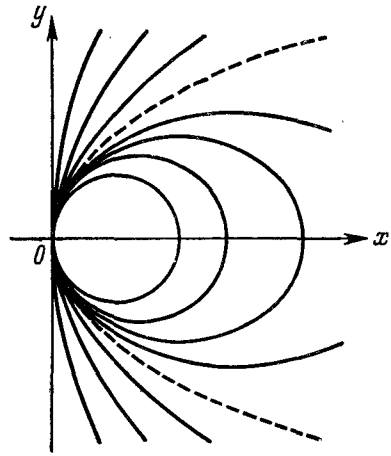
При $-1 \leqslant q$ *эксцентриситет* e *связан с* q *формулой*

$$e = \sqrt{1+q}.$$

При $q < -1$ *получается эллипс, имеющий эксцентриситет* $\sqrt{1+\frac{1}{q}}$ *и фокальный параметр* $\frac{p}{\sqrt{-q}}$.

Замечание 5. Уравнение (10) показывает, что при непрерывном изменении q от $-\infty$ до $+\infty$ (и постоянном p) эллипс при q, близких к $-\infty$, сильно «прижатый» к оси ординат, постепенно «округляется», превращаясь при $q = -1$ в окружность, затем «вытягивается» в положительном направлении оси

Ox, переходя при $q = 0$ в параболу, превращающуюся при $q > 0$ в гиперболу с правой ветвью, «близкой» к параболе, и левой ветвью, далеко сдвинутой влево. При дальнейшем возрастании q ветви гиперболы постепенно «раскрываются» все шире (и приближаются друг к другу), так что при q, близких к $+\infty$, получается гипербола, сильно «прижатая» к оси ординат.



При $q \geqslant -1$ все эти кривые имеют один и тот же фокальный параметр p, а их эксцентриситет e меняется от 0 до $+\infty$. При уменьшении q от -1 до $-\infty$ эксцентриситет растет от нуля до единицы, а фокальный параметр стремится к нулю. Поэтому, хотя в пределе при $q \to -\infty$ и получается значение $e = 1$, все же предельное положение эллипса параболой не является, поскольку для него $p = 0$.

Предельное положение эллипса при $q \to -\infty$ можно при желании рассматривать как дважды пробегаемую ось ординат. Предельное положение гиперболы при $q \to +\infty$ также является дважды пробегаемой осью ординат, но если гипербола «в пределе» является осью ординат, дважды пробегаемой в одном и том же направлении, то эллипс «в пределе» является осью ординат, пробегаемой сначала в одном направлении, а потом в другом.

4. Уравнения эллипса, параболы и гиперболы в полярных координатах

Единство эллипсов, парабол и гипербол проявляется также и в их уравнениях в полярных координатах.

Рассмотрим сначала произвольную параболу

$$y^2 = 2px \qquad (1)$$

и найдем ее уравнение в системе полярных координат r, φ, полюсом которой является фокус параболы, а полярной осью — ось параболы (или точнее, поскольку необходимо указать ориентацию полярной оси, — ось абсцисс канонической системы координат). С этой полярной системой координат согласована система прямоугольных координат x', y', выражающихся через выражение через r и φ, мы немедленно получим, что

$$x' = x - \frac{p}{2},$$
$$y' = y.$$

Поскольку

$$x' = r\cos\varphi,$$

$$y' = r\sin\varphi,$$

то, следовательно,

$$x = \frac{p}{2} + r\cos\varphi,$$

$$y = r\sin\varphi.$$

Подставив эти выражения в уравнение (1) параболы, мы получим, что в координатах r, φ парабола выражается уравнением

$$r^2\sin^2\varphi - 2pr\cos\varphi - p^2 = 0.$$

Решая это уравнение относительно r и учитывая, что величина r должна быть положительна, мы немедленно получим, что

$$r = \frac{p\cos\varphi + \sqrt{p^2\cos^2\varphi + p^2\sin^2\varphi}}{\sin^2\varphi} = \frac{p(1+\cos\varphi)}{\sin^2\varphi} = \frac{p}{1-\cos\varphi}.$$

Таким образом,

в указанной полярной системе координат парабола выражается уравнением

$$r = \frac{p}{1-\cos\varphi}.$$

Рассмотрим теперь произвольный эллипс

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad a \geqslant b > 0.$$

Так же, как для параболы, мы направим полярную ось по оси абсцисс канонической системы координат, а за полюс примем фокус, имеющий отрицательную абсциссу, т. е. левый фокус. Тогда канонические координаты x, y будут выражаться через полярные координаты r, φ по формулам

$$x = -c + r\cos\varphi,$$

$$y = r\sin\varphi.$$

Чтобы найти уравнение эллипса в координатах r, φ, можно подставить эти выражения в каноническое уравнение эллипса и решить получающееся квадратное уравнение относительно r. Однако проще вспомнить формулу (6) п. 2 для левого фокального радиуса произвольной точки эллипса и учесть, что в силу нашего выбора полюса этот фокальный радиус совпадает с полярным радиусом r. Подставив в эту формулу вместо x его выражение через r и φ, мы немедленно получим, что

$$r = a + e(-c + r\cos\varphi),$$

и следовательно, что

$$r = \frac{a - ec}{1 - e\cos\varphi}.$$

Поскольку

$$a - ec = a - \frac{c^2}{a} = \frac{a^2 - c^2}{a} = \frac{b^2}{a} = p,$$

тем самым доказано, что

в указанной полярной системе координат эллипс выражается уравнением

$$r = \frac{p}{1 - e\cos\varphi}.$$

Наконец, для гиперболы

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \qquad a > 0, \quad b > 0,$$

мы рассмотрим систему полярных координат r, φ, полярная ось которой совпадает с осью Ox, а полюс находится в правом фокусе гиперболы (заметим, что для эллипса мы брали левый фокус). Тогда

$$x = c + r\cos\varphi,$$
$$y = r\sin\varphi,$$

причем полярный радиус r произвольной точки $M(x, y)$ гиперболы будет не чем иным, как ее правым фокальным радиусом и потому будет выражаться формулой (8) п. 3. В частности, для точек правой ветви (с $x > 0$) мы получим, что

$$r = -a + ex$$

и потому

$$r = -a + e(c + r\cos\varphi),$$

т. е.

$$r = \frac{ec - a}{1 - e\cos\varphi}.$$

Поскольку

$$ec - a = \frac{c^2}{a^2} - a = \frac{c^2 - a^2}{a} = \frac{b^2}{a} = p,$$

тем самым доказано, что

в указанной системе полярных координат правая ветвь гиперболы выражается уравнением

$$r = \frac{p}{1 - e\cos\varphi}.$$

Сопоставляя полученные результаты, мы немедленно получаем следующее окончательное

Предложение 1. *Уравнение*

$$r = \frac{p}{1 - e\cos\varphi}$$

является

а) *при* $0 \leqslant e < 1$ — *уравнением эллипса,*

б) *при* $e = 1$ — *уравнением параболы,*

в) *при* $e > 1$ — *уравнением одной ветви гиперболы.*

Во всех трех случаях число e является эксцентриситетом линии, число p — фокальным параметром, а полюс $r = 0$ — фокусом линии (в случае $e > 1$ — одноименным с рассматриваемой ветвью).

§ 2. НЕКОТОРЫЕ ДОПОЛНИТЕЛЬНЫЕ СВОЙСТВА ЛИНИЙ ВТОРОГО ПОРЯДКА

1. Эллипс, парабола и гипербола как конические сечения

Определение 1. *Прямым круговым конусом* называется поверхность, имеющая в некоторых прямоугольных координатах $Oxyz$ уравнение вида

$$x^2 + y^2 = R^2 z^2. \tag{1}$$

Плоскость $z = 1$ пересекает поверхность (1) по линии с уравнением

$$x^2 + y^2 = R^2, \tag{2}$$

т. е. по окружности радиуса R (как мы знаем, x, y являются прямоугольными координатами и на плоскости $z = 1$). С другой стороны, если точка $(x, y, 1)$ принадлежит окружности (2), то любая точка вида (xt, yt, t) принадлежит поверхности (1), и обратно. Следовательно, поверхность (1) образована всевозможными прямыми, проходящими через точку $O(0, 0, 0)$ и произвольную точку окружности (2). Это показывает, что поверхность (1) состоит из двух бесконечных конусов, имеющих общую вершину O.

В пересечении с произвольной плоскостью

$$x\cos\alpha + y\cos\beta + z\cos\gamma - p = 0, \qquad p \geqslant 0, \tag{3}$$

конус (1) дает некоторую линию.

Определение 2. В случае, когда плоскость (3) не проходит через начало координат O, т. е. в случае, когда $p > 0$, линия, высекаемая на плоскости (3) конусом (1), называется *коническим сечением.*

Упражнение. Какие линии пересечения получаются при $p = 0$? Что изменится, если мы вместо евклидова пространства будем рассматривать евклидово вещественно-комплексное пространство?

Чтобы найти уравнения конических сечений, мы заметим, что в силу круговой симметрии конуса можно, без ограничения общности, считать, что плоскость (3) параллельна оси Oy, т. е. что $\cos\beta = 0$. Тогда $\cos\gamma = \sin\alpha$, и потому уравнение плоскости (3) может быть записано в виде

$$x\cos\alpha + z\sin\alpha - p = 0. \tag{4}$$

Более того, по аналогичным соображениям мы можем считать, что

$$0 \leqslant \alpha \leqslant \frac{\pi}{2}. \tag{5}$$

Чтобы найти пересечение плоскости (4) с конусом (1), мы рассмотрим векторы

$$e_1(-\sin\alpha, 0, \cos\alpha), \quad e_2(0, 1, 0), \quad e_3(\cos\alpha, 0, \sin\alpha).$$

Первые два из этих векторов параллельны плоскости (4), а третий ей перпендикулярен. Кроме того, эти векторы составляют ортонормированный базис пространства, разноименный с исходным координатным базисом.

Пусть x', y', z' — прямоугольные координаты в пространстве, определенные репером $Oe_1e_2e_3$. Согласно общим формулам п. 2 § 1 гл. 2 координаты x, y, z выражаются через координаты x', y', z' по формулам

$$\begin{aligned} x &= -x' \sin\alpha + z' \cos\alpha, \\ y &= y', \\ z &= x' \cos\alpha + z' \sin\alpha. \end{aligned}$$

Поэтому в координатах x', y', z' конус (3) имеет уравнение

$$(-x' \sin\alpha + z' \cos\alpha)^2 + y'^2 - R^2 (x' \cos\alpha + z' \sin\alpha)^2 = 0,$$

т. е. уравнение

$$y'^2 = 2x'z' (1 + R^2) \cos\alpha \sin\alpha + x'^2 (R^2 \cos^2\alpha - \sin^2\alpha) + z'^2 (R^2 \sin^2\alpha - \cos^2\alpha). \quad (6)$$

С другой стороны, плоскость (4) имеет в координатах x', y', z' уравнение

$$z' - p = 0,$$

и потому координаты x', y' являются прямоугольными координатами на этой плоскости, причем, чтобы получить в этих координатах уравнение линии, являющейся пересечением конуса (1) с плоскостью (4), достаточно в уравнении (6) положить $z' = p$. Таким образом, эта линия будет иметь в координатах x', y' уравнение

$$y'^2 = 2x'p (1 + R^2) \cos\alpha \sin\alpha + x'^2 (R^2 \cos^2\alpha - \sin^2\alpha) + p^2 (R^2 \sin^2\alpha - \cos^2\alpha). \quad (7)$$

Перенесем теперь начало координат в точку

$$\left(-p \frac{R \sin\alpha - \cos\alpha}{R \cos\alpha + \sin\alpha}, 0\right),$$

т. е. перейдем от координат x', y' к координатам x'', y'', выражающимся формулами

$$\begin{aligned} x'' &= x' + p \frac{R \sin\alpha - \cos\alpha}{R \cos\alpha + \sin\alpha}, \\ y'' &= y'. \end{aligned}$$

(Заметим, что в силу условия (5) величина $R \cos\alpha + \sin\alpha$ отлична от нуля.)

В координатах x'', y'' линия (7) имеет уравнение

$$y''^2 = 2p(1+R^2)\left(x'' - p\frac{R\sin\alpha - \cos\alpha}{R\cos\alpha + \sin\alpha}\right)\cos\alpha\sin\alpha +$$

$$+\left(x'' - p\frac{R\sin\alpha - \cos\alpha}{R\cos\alpha + \sin\alpha}\right)^2 (R^2\cos^2\alpha - \sin^2\alpha) + p^2(R^2\sin^2\alpha - \cos^2\alpha).$$

Раскрыв скобки, мы после очевидных преобразований получим, как легко видеть, следующее уравнение:

$$y''^2 = 2Rpx'' + x''^2(R^2\cos^2\alpha - \sin^2\alpha). \tag{8}$$

Сравнив это уравнение с уравнением (10) п. 3 § 1, мы немедленно получаем, что

каждое коническое сечение является либо эллипсом, либо параболой, либо гиперболой.

При этом фокальным параметром этой линии является число Rp, а коэффициент q из уравнения (19) п. 3 выражается формулой

$$q = R^2\cos^2\alpha - \sin^2\alpha. \tag{9}$$

Уравнение (8) является уравнением параболы тогда и только тогда, когда

$$R^2\cos^2\alpha - \sin^2\alpha = 0,$$

т. е. когда

$$R = \operatorname{tg}\alpha. \tag{10}$$

Поскольку фокальный параметр этой параболы равен Rp, мы получаем, следовательно, что

каждая парабола является коническим сечением.

Более того, ясно, что

каждая парабола является сечением произвольного фиксированного конуса (1), (*т. е. конуса, отвечающего данному фиксированному* R; *например*, $R = 1$).

Плоскости, на которых данный конус (1) высекает параболы, с точностью до параллельности определяются из условия (10). Чтобы получить данную параболу (с данным фокальным параметром), нужно соответствующим образом подобрать расстояние этой плоскости до вершин конуса (например, при $R=1$ это расстояние должно быть равно фокальному параметру).

Выражая из соотношения (9) угол α через величины q и R, мы немедленно получим, что

$$\sin^2\alpha = \frac{R^2 - q}{1 + R^2}. \tag{11}$$

Отсюда следует, что для любого q, удовлетворяющего неравенствам

$$-1 \leqslant q < 0,$$

существует угол α, удовлетворяющий уравнению (9). Следовательно,

каждый эллипс является коническим сечением.

Более того,

каждый эллипс является сечением произвольного фиксированного конуса (1).

Для соответствующей плоскости угол α определяется из соотношения (11), а свободный член (расстояние до вершины конуса) должен быть равен $\frac{p}{R}$, где p — фокальный параметр эллипса.

При $q > 0$ уравнение (11) имеет решение только при

$$0 < q \leqslant R^2. \tag{12}$$

Поэтому в отличие от парабол и эллипсов

не каждая гипербола является сечением данного конуса (1).

Тем не менее,

каждая гипербола является коническим сечением.

Действительно, при любом $q > 0$ условие (12) удовлетворяется, если R достаточно велико.

Замечание 1. Последнее утверждение может быть доказано проще, если заметить, что, скажем, плоскость

$$x = b$$

пересекает конус (1) по кривой с уравнением

$$b^2 + y^2 = R^2 z^2,$$

т. е. с уравнением

$$\frac{z}{\left(\frac{b}{R}\right)^2} - \frac{y^2}{b^2} = 1,$$

являющимся уравнением гиперболы с полуосями $a = \frac{b}{R}$ и b.

Резюмируя, мы получаем следующее предложение:

Предложение 1. *Линия на плоскости тогда и только тогда является коническим сечением, когда она представляет собой либо эллипс, либо параболу, либо гиперболу.*

Замечание 2. Описанное в замечании 5 п. 3 § 1 постепенное вытягивание окружности во все более и более «длинный» эллипс, переход этого эллипса в параболу, а затем в гиперболу мы можем теперь получить, вращая вокруг оси Oy плоскость (4).

Действительно, при $\alpha = \frac{\pi}{2}$ мы получаем в пересечении с конусом (1) окружность радиуса R. При уменьшении угла α мы будем получать все более и более «длинные» эллипсы, пока, наконец, при $\alpha = \operatorname{arctg} R$ мы не получим параболу. При дальнейшем уменьшении угла α вплоть до $\alpha = 0$ мы будем

получать гиперболы. Однако всех гипербол мы при этом не получим. Чтобы получить остальные гиперболы, можно, например, оставляя плоскость (с $\alpha = 0$) неподвижной, постепенно «раскрывать» конус (1), неограниченно увеличивая R. При этом будет оставаться неизменным уже не фокальный параметр, а мнимая полуось гиперболы. (Чтобы сохранить неизменным фокальный параметр, следует одновременно с раскрытием конуса двигать плоскость параллельно самой себе, приближая ее к оси Oy.)

2. Взаимное расположение конических сечений и прямых

Эллипсы, гиперболы и параболы задаются уравнениями второй степени, т. е. являются *линиями второго порядка*. Чтобы не повторяться, нам будет удобно рассмотреть вопрос о взаимном расположении этих линий и прямых сначала в общем виде, т. е. для любых линий второго порядка (впрочем, как мы увидим в следующей главе, эллипсами, гиперболами и параболами исчерпываются, по существу, все линии второго порядка).

Чтобы найти точки пересечения произвольной прямой

$$\begin{aligned} x &= x_0 + lt, \\ y &= y_0 + mt \end{aligned} \tag{1}$$

с некоторой линии второго порядка

$$F(x, y) \equiv a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0, \tag{2}$$

нужно (см. п. 2 § 1 гл. 3) найти корни уравнения

$$\begin{aligned} F(x_0 + lt, y_0 + mt) \equiv{}& (a_{11}l^2 + 2a_{12}lm + a_{22}m^2)t^2 + \\ &+ 2(a_{11}lx_0 + a_{12}(ly_0 + mx_0) + a_{22}my_0 + a_{13}l + a_{23}m)t + \\ &+ a_{11}x_0^2 + 2a_{12}x_0y_0 + a_{22}y_0^2 + 2a_{13}x_0 + 2a_{23}y_0 + a_{33} = 0. \end{aligned} \tag{3}$$

Определение 1. Прямая (1) называется прямой *неасимптотического направления* (по отношению к линии (2)), если

$$a_{11}l^2 + 2a_{12}lm + a_{22}m^2 \neq 0.$$

Для такой прямой уравнение (3) является уравнением второй степени и потому имеет либо два вещественных корня, либо два комплексных (невещественных) корня, либо один двукратный вещественный корень. В первом случае прямая (1) пересекает линию (2) в двух точках, во втором случае прямая (1) и линия (2) не пересекаются [1]), а в третьем случае прямая (1) и линия (2) имеют одну общую точку.

[1]) В вещественной плоскости. В вещественно-комплексной плоскости этому случаю отвечают две комплексно-сопряженные точки пересечения.

Определение 2. Если уравнение (3) имеет один двукратный вещественный корень, т. е. прямая (1) неасимптотического направления имеет с линией (2) одну общую точку, то говорят, что в этой точке прямая (1) *касается* линии (2).

Замечание 1. В анализе касательная к произвольной линии в точке M_0 определяется как предельное положение секущей M_0M_1 (где M_1 — вторая точка пересечения секущей с данной линией) при $M_1 \to M_0$. Для линии второго порядка (2) точка M_0 определяется некоторым корнем t_0 уравнения (3), а точка M_1 — другим его корнем t_1. Стремление точки M_1 к точке M_0 означает такое изменение коэффициентов уравнений (1), при котором корень t_0 уравнения (3) не меняется, а корень t_1 стремится к корню t_0. Следовательно, в пределе мы получим такую прямую, которой будет соответствовать двукратный корень уравнения (3). Таким образом, данное нами алгебраическое определение касательной (пригодное только для линий второго порядка) является частным случаем общего аналитического определения касательной (пригодного для любых линий).

Для прямой *асимптотического направления*, т. е. для прямой (1), направляющий вектор $\boldsymbol{a}(l, m)$ которой удовлетворяет условию

$$a_{11}l^2 + 2a_{12}lm + a_{22}m^2 = 0,$$

уравнение (3) вырождается в линейное уравнение (конечно, если коэффициент при t также не обращается в нуль), и потому каждая такая прямая пересекает кривую (2) в одной точке [1]). Тем не менее, эта прямая касательной не считается.

Определение 3. Прямая (1) называется *асимптотой* кривой (2), если в уравнении (3) равны нулю коэффициенты при t^2 и при t, а свободный член $F(x_0, y_0)$ отличен от нуля.

Такая прямая не имеет с линией второго порядка ни одной общей точки (даже в комплексной плоскости).

Замечание 2. Это определение асимптот (для линий второго порядка) по форме отличается от определения 2 п. 3 § 1 (пригодного для любых линий). Мы дальше на примерах увидим, что оно и по существу отлично от определения 2 п. 3 § 1 (прямая, являющаяся асимптотой линии второго порядка в смысле определения 2 п. 3, будет асимптотой и в нашем теперешнем смысле, но не наоборот).

Если для прямой (1) обращаются в нуль все коэффициенты уравнения (3), то такая прямая целиком принадлежит линии (2). Подчеркнем, что, таким образом, любая прямая, целиком принадлежащая линии второго порядка, является, по определению, прямой асимптотического направления.

Замечание 3. Термин «прямая асимптотического направления» согласуется с использованием термина «направление», объясненном в подстрочном примечании на стр. 17. Согласно этому примечанию направлением на плоскости (или в пространстве) называется класс параллельных прямых. Все прямые одного направления имеют одни и те же направляющие

[1]) Даже в комплексной плоскости!

векторы и потому каждое направление (на плоскости) однозначно характеризуется отношением $l:m$ координат этих векторов (здесь «отношение» $l:m$ следует, конечно, понимать не как число, — иначе будет исключен случай $m=0$, — а как элемент множества $\mathbb{R}P^1$; см. п. 6 § 1 гл. 2). На этом основании мы будем иногда использовать сокращенный термин «направление $l:m$», понимая под этим направление, характеризующееся отношением $l:m$.

Замечание 4 (очень важное!). Во всем сказанном выше мы нигде не использовали предположения об евклидовости координат x, y. Поэтому все это сохраняет силу для любых аффинных координат. Это показывает, что *введенные понятия (касательные, асимптоты и т. д.) принадлежат аффинной геометрии.*

Для параболы

$$y^2=2px, \qquad p>0, \tag{4}$$

уравнение (3) имеет вид

$$m^2t^2+2(my_0-2pl)t+y_0^2-2px_0=0.$$

Следовательно, асимптотические направления характеризуются равенством

$$m=0.$$

Другими словами,

прямая тогда и только тогда имеет относительно параболы асимптотическое направление, когда она параллельна оси этой параболы.

Если $m=0$ и $my_0-2pl=0$, то $l=0$, что невозможно (вектор $\boldsymbol{a}(l, m)$, по условию, отличен от нуля). Следовательно,

асимптот парабола не имеет.

Кроме того,

парабола не содержит ни одной прямой.

Каждая прямая неасимптотического направления (т. е. не параллельная оси Ox) либо вообще не пересекает параболу, либо касается ее в некоторой точке, либо пересекает параболу в двух различных точках. Чтобы выяснить, какой из этих двух случаев имеет место, удобно перейти к уравнению прямой, разрешенному относительно x, т. е. к уравнению вида

$$x=ky+\tau \tag{5}$$

(это всегда возможно, ибо, по условию, рассматриваемая прямая не параллельна оси абсцисс). Тогда ординаты точек пересечения прямой с параболой будут определяться из уравнения

$$y^2=2p(ky+\tau),$$

т. е. будут выражаться формулой

$$y=pk\pm\sqrt{p^2k^2+2p\tau}.$$

Таким образом,

если $\tau < -\frac{pk^2}{2}$, *то прямая* (5) *с параболой не пересекается*;

если $\tau = -\frac{pk^2}{2}$, *то прямая* (5) *касается параболы* (*в точке с ординатой* pk);

если $\tau > -\frac{pk^2}{2}$, *то прямая* (5) *пересекает параболу в двух различных точках.*

Согласно принципу обращения обратные утверждения также справедливы.

В частности, мы видим, что для любого неасимптотического направления прямые этого направления, пересекающие параболу, заполняют целую полуплоскость. Следовательно,

у параболы имеются хорды любого наперед заданного неасимптотического направления.

Для эллипса

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad a \geqslant b > 0, \tag{6}$$

уравнение (3) имеет вид

$$\left(\frac{l^2}{a^2} + \frac{m^2}{b^2}\right)t^2 + 2\left(\frac{lx_0}{a^2} + \frac{my_0}{b^2}\right)t + \left(\frac{x_0^2}{a^2} + \frac{y_0^2}{b^2} - 1\right) = 0.$$

Поскольку при $(l, m) \neq (0, 0)$

$$\frac{l^2}{a^2} + \frac{m^2}{b^2} \neq 0,$$

эллипс не имеет прямых асимптотического направления.

В частности,

эллипс не имеет асимптот и не содержит ни одной прямой.

По аналогии со случаем параболы мы предпочтем сейчас исследовать пересечения эллипса с прямыми, задавая их не параметрическими уравнениями (1), а уравнениями, разрешенными относительно y:

$$y = kx + \tau \tag{7}$$

(заметим, что если в случае параболы нам было удобно рассматривать прямые, заданные уравнениями (5) — разрешенными относительно x, — то здесь уже нет в этом необходимости).

Конечно, ограничиваясь прямыми вида (7), мы исключаем из рассмотрения прямые $x = x_0$, параллельные оси ординат. Пересечения таких прямых с эллипсом будут поэтому требовать дополнительного исследования.

Чтобы найти точки пересечения эллипса (6) с прямой (7), мы должны решить (относительно x) уравнение

$$\frac{x^2}{a^2}+\frac{(kx+\tau)^2}{b^2}=1,$$

т. е. уравнение

$$\left(\frac{1}{a^2}+\frac{k^2}{b^2}\right)x^2+\frac{2k\tau}{b^2}x+\frac{\tau^2}{b^2}-1=0.$$

Сделав это, мы получим для абсцисс x точек пересечения формулу:

$$x=\frac{-a^2k\tau\pm ab\sqrt{a^2k^2+b^2-\tau^2}}{a^2k^2+b^2}.$$

Следовательно,

если $\tau^2>a^2k^2+b^2$, *то прямая* (7) *не пересекается с эллипсом*;

если $\tau^2=a^2k^2+b^2$, *то прямая* (7) *касается эллипса*;

если $\tau^2<a^2k^2+b^2$, *то прямая* (7) *пересекает эллипс в двух точках.*

Для прямых вида $x=x_0$ ординаты точки их пересечения с эллипсом определяются из уравнения

$$\frac{x_0^2}{a^2}+\frac{y^2}{b^2}=1,$$

т. е. выражаются формулой

$$y=\pm\frac{b}{a}\sqrt{a^2-x_0^2}.$$

Следовательно,

если $|x_0|>a$, *то прямая* $x=x_0$ *не пересекается с эллипсом*;

если $|x_0|=a$, *то прямая* $x=x_0$ *касается эллипса*;

если $|x_0|<a$, *то прямая* $x=x_0$ *пересекает эллипс в двух точках.*

Мы видим, таким образом, что в любом пучке параллельных прямых прямые, пересекающие эллипс в двух точках, заполняют целую полосу. Следовательно, как и у параболы,

у эллипса имеются хорды любого наперед заданного направления (*заведомо, неасимптотического*).

Для гиперболы

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1,\qquad a>0,\quad b>0, \tag{8}$$

уравнение (3) имеет вид

$$\left(\frac{l^2}{a^2}-\frac{m^2}{b^2}\right)t^2+2\left(\frac{lx_0}{a^2}-\frac{my_0}{b^2}\right)t+\left(\frac{x_0^2}{a^2}-\frac{y_0^2}{b^2}-1\right)=0.$$

Уравнение асимптотических направлений

$$\frac{l^2}{a^2} - \frac{m^2}{b^2} = 0$$

имеет теперь два решения:

$$l : m = \pm a : b.$$

Таким образом,

гипербола имеет два асимптотических направления, характеризующихся угловыми коэффициентами $\pm \frac{b}{a}$.

Дальнейшее исследование пересечений гиперболы (8) с прямой мы предпочтем, как и в случае эллипса, производить в предположении, что прямая задана уравнением

$$y = kx + \tau \tag{9}$$

(или уравнением $x = x_0$).

Чтобы найти точки пересечения гиперболы (8) с прямой (9), мы должны решить (относительно x) уравнение

$$\frac{x^2}{a^2} - \frac{(kx+\tau)^2}{b^2} = 1,$$

т. е. уравнение

$$\left(\frac{1}{a^2} - \frac{k^2}{b^2}\right) x^2 - 2\frac{k\tau}{b^2} x - \left(\frac{\tau^2}{b^2} + 1\right) = 0. \tag{10}$$

При $k = \pm \frac{b}{a}$ (прямая (9) имеет асимптотическое направление) это уравнение вырождается в линейное и имеет при $\tau \neq 0$ только один корень

$$x = \frac{\tau^2 + b^2}{2\tau k},$$

а при $\tau = 0$ вообще не может быть удовлетворено. Это означает, что

из всех прямых асимптотических направлений только прямые

$$y = \frac{b}{a} x, \quad y = -\frac{b}{a} x$$

не пересекают гиперболы, т. е. являются ее асимптотами (в смысле определения 3).

Но, как мы знаем, эти прямые являются асимптотами гиперболы и в смысле определения 2 п. 3 § 1. Таким образом,

для гиперболы асимптоты в смысле определения 2 п. 3 § 1 *совпадают с асимптотами в смысле определения* 3.

Для прямой (9) неасимптотического направления уравнение (10) имеет два корня

$$x = \frac{-a^2k\tau \pm ab\sqrt{\tau^2 - a^2k^2 + b^2}}{a^2k^2 - b^2}. \tag{11}$$

Следовательно,

если $\tau^2 < a^2k^2 - b^2$, *то прямая* (9) *не пересекает гиперболу*;

если $\tau^2 = a^2k^2 - b^2$, *то прямая* (9) *касается гиперболы*;

если $\tau^2 > a^2k^2 - b^2$, *то прямая* (9) *пересекает гиперболу в двух точках.*

Замечание 5. Если $|k| < \frac{b}{a}$, т. е. прямая (9) при $\tau = 0$ заключена в той паре вертикальных углов, образованных асимптотами, которая содержит гиперболу, то величина $a^2k^2 - b^2$ отрицательна и потому при любом τ возможен лишь третий случай, т. е. прямая $y = kx + \tau$ пересекает гиперболу. Если же $|k| > \frac{b}{a}$, т. е. если прямая (9) при $\tau = 0$ заключена в дополнительной паре вертикальных углов, то величина $a^2k^2 - b^2$ положительна и возможны все три случая.

Для прямых вида $x = x_0$ ординаты y их точек пересечения с гиперболой определяются из уравнения

$$\frac{x_0^2}{a^2} - \frac{y^2}{b^2} = 1,$$

т. е. выражаются формулой

$$y = \pm \frac{b}{a} \sqrt{x_0^2 - a^2}.$$

Следовательно,

каждая прямая вида $x = x_0$ *является прямой неасимптотического направления, причем*

при $|x_0| < a$ *она не пересекает гиперболу,*

при $|x_0| = a$ *она касается гиперболы,*

при $|x_0| > a$ *она пересекает гиперболу в двух точках.*

В частности, мы видим, что в любом пучке параллельных прямых, имеющих неасимптотическое направление, характеризующееся угловым коэффициентом k (случай $k = \infty$, т. е. случай прямых $x = x_0$ не исключается), прямые, пересекающие гиперболу, исчерпывают при $|k| < \frac{b}{a}$ весь пучок, а при $|k| > \frac{b}{a}$ заполняют две полуплоскости, отделенные полосой, в которой вообще нет точек гиперболы. В обоих случаях

у гиперболы имеются хорды любого наперед заданного неасимптотического направления.

3. Прямые, касающиеся конических сечений

Согласно результатам п. 2 прямая

$$x = ky + \tau \tag{1}$$

тогда и только тогда касается параболы

$$y^2 = 2px, \qquad p > 0,$$

когда

$$\tau = -\frac{pk^2}{2}.$$

При этом координаты точки касания $M_0(x_0, y_0)$ выражаются, очевидно, формулами

$$x_0 = \frac{pk^2}{2}, \quad y_0 = pk.$$

Поэтому уравнение (1) мы можем переписать в следующем виде:

$$x = \frac{y_0}{p} y - x_0.$$

Тривиальным образом преобразовав это уравнение, мы получим, что

касательная к параболе $y^2 = 2px$ *в точке* $M_0(x_0, y_0)$ *имеет уравнение*

$$y_0 y = p(x + x_0). \tag{2}$$

Замечание 1. Строго говоря, мы доказали лишь, что если в точке $M_0(x_0, y_0)$ существует касательная к параболе, то она выражается уравнением (2). Но чтобы доказать существование касательной, достаточно показать, что для любой точки $M_0(x_0, y_0)$ параболы существует такое k, что $x_0 = \frac{pk^2}{2}$ и $y_0 = pk$ (ибо тогда прямая с угловым коэффициентом k, проходящая через точку M_0, и будет, по доказанному выше, касательной), а это очевидно (достаточно, например, за k принять число $\frac{y_0}{p}$).

Рассмотрим теперь эллипс

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad a \geqslant b > 0.$$

Как мы знаем (см. п. 2), прямая

$$y = kx + \tau \tag{3}$$

тогда и только тогда касается эллипса, когда

$$|\tau| = \sqrt{a^2k^2 + b^2}.$$

При этом, согласно формуле (8) п. 2, абсцисса x_0 точки касания выражается формулой

$$x_0 = \frac{-a^2k\tau}{a^2k^2 + b^2} = \frac{-a^2k}{\pm\sqrt{a^2k^2 + b^2}}.$$

Подставив это выражение в уравнение (3), мы найдем и ординату y_0 точки касания:

$$y_0 = kx_0 + \tau = \frac{b^2}{\pm\sqrt{a^2k^2 + b^2}}.$$

Следовательно, $\frac{y_0}{x_0} = -\frac{b^2}{a^2 k}$, т. е.

$$k = -\frac{b^2}{a^2} \cdot \frac{x_0}{y_0}. \tag{4}$$

Таким образом, касательная к эллипсу в точке $M_0(x_0, y_0)$ представляет собой прямую, проходящую через точку M_0 и имеющую угловой коэффициент (4). Поэтому эта касательная выражается уравнением

$$y - y_0 = -\frac{b^2}{a^2} \cdot \frac{x_0}{y_0}(x - x_0), \tag{5}$$

т. е. уравнением

$$\frac{x_0 x}{a^2} + \frac{y_0 y}{b^2} = \frac{x_0^2}{a^2} + \frac{y_0^2}{b^2}.$$

Поскольку $\frac{x_0^2}{a^2} + \frac{y_0^2}{b^2} = 1$ (ибо точка $M_0(x_0, y_0)$ принадлежит эллипсу), тем самым доказано, что

касательная к эллипсу в точке $M_0(x_0, y_0)$ *имеет уравнение*

$$\frac{x_0 x}{a^2} + \frac{y_0 y}{b^2} = 1. \tag{6}$$

Замечание 2. Строго говоря (ср. выше аналогичное замечание 1 для параболы), мы лишь доказали, что если в точке M_0 эллипса существует касательная вида $y = kx + b$, то она выражается уравнением (6) (или, точнее, уравнением (5)). Чтобы доказать существование такой касательной, достаточно, очевидно, доказать, что координаты x_0, y_0 точки M_0 могут быть представлены в виде

$$x_0 = \frac{-a^2 k}{\pm\sqrt{a^2 k^2 + b^2}}, \quad y_0 = \frac{b^2}{\pm\sqrt{a^2 k^2 + b^2}} \tag{7}$$

с некоторым k.

Мы покажем, что

любая точка $M_0(x_0, y_0)$ *эллипса, отличная от вершин* $(\pm a, 0)$, *допускает представление вида* (7).

Этим и будет доказано, что в точке M_0 существует касательная.

Пусть сначала $y_0 > 0$. Найдя k из соотношения

$$y_0 = \frac{b^2}{\sqrt{a^2 k^2 + b^2}},$$

мы получим, что

$$k = \pm \frac{b}{a} \cdot \frac{\sqrt{b^2 - y_0^2}}{y_0}.$$

Знак k мы выберем противоположным знаку x_0. Тогда величина

$$\frac{-a^2 k}{\sqrt{a^2 k^2 + b^2}}$$

будет иметь тот же знак, что и x_0. Поскольку ее квадрат равен

$$\frac{a^4k^2}{a^2k^2+b^2}=\frac{a^4}{b^4}\cdot\frac{b^4}{a^2k^2+b^2}\cdot k^2=\frac{a^4}{b^4}\cdot y_0^2\cdot\frac{b^2}{a^2}\cdot\frac{b^2-y_0^2}{y_0^2}=$$

$$=\frac{a^2}{b^2}(b^2-y_0^2)=x_0^2,$$

этим доказано, что

$$x_0=\frac{-a^2k}{\sqrt{a^2k^2+b^2}}.$$

Таким образом, при $y_0>0$ все доказано.

При $y_0<0$ мы рассмотрим точку $(-x_0, -y_0)$ (также принадлежащую эллипсу) и представим ее в виде (7) (что, по уже доказанному, возможно). Тогда для точки (x_0, y_0) мы, очевидно, получим представление вида (7), но уже с отрицательными знаменателями.

Тем самым существование касательной вида $y=kx+\tau$ в любой точке $M_0(x_0, y_0)$ эллипса, для которой $y_0\neq 0$, т. е. в любой точке, отличной от вершин $(\pm a, 0)$, полностью доказано.

Что же касается этих вершин, то, согласно сказанному в предыдущем пункте, в них касательные также существуют, но являются уже прямыми вида $x=x_0$, параллельными оси ординат, а именно, прямыми

$$x=\pm a. \tag{8}$$

С другой стороны, подставив в уравнение (6) значения $x_0=\pm a$, $y_0=0$, мы, очевидно, получим как раз уравнения (8). Следовательно, уравнение (6) является уравнением касательной и в каждой из вершин $(\pm a, 0)$.

Таким образом, мы доказали, что

касательная к эллипсу существует в каждой его точке $M_0(x_0, y_0)$ *и выражается уравнением* (6).

Аналогичное утверждение имеет место и для гиперболы

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1, \qquad a>0, \quad b>0,$$

т. е.

касательная к гиперболе существует в каждой ее точке $M_0(x_0, y_0)$ *и выражается формулой*

$$\frac{xx_0}{a^2}-\frac{yy_0}{b^2}=1. \tag{9}$$

В частности, угловой коэффициент этой касательной равен (при $y_0\neq 0$)

$$k=\frac{b^2}{a^2}\cdot\frac{x_0}{y_0}. \tag{10}$$

Задание. Докажите формулу (9).

Определение 1. Линия называется *выпуклой*, если она расположена по одну сторону от каждой ее касательной (за исключением, конечно, точки касания).

Предложение 1. *Парабола, эллипс и каждая ветвь гиперболы являются выпуклыми линиями.*

Доказательство. Подставляя координаты x, y произвольной точки $M(x, y)$ параболы в уравнение касательной (2) и учитывая соотношения $y_0^2 = 2px_0$ и $y^2 = 2px$, мы немедленно получим, что

$$y_0 y - p(x + x_0) = -\frac{1}{2}(y - y_0)^2.$$

Таким образом, при подстановке координат любой точки $M(x, y) \neq M_0(x_0, y_0)$ параболы в уравнение касательной в точке $M_0(x_0, y_0)$ мы получаем числа одного знака (отрицательные). Поэтому (см. п. 4 § 1 гл. 3) все эти точки расположены по одну и ту же сторону от касательной.

Аналогичное утверждение для эллипса и гиперболы проще всего доказать, используя их параметрические уравнения (уравнения (2) п. 2 § 1 и уравнения (6) п. 3 § 1). Пусть θ_0 — значение параметра, отвечающее точке касания M_0, а θ — значение параметра, отвечающее произвольной точке M эллипса. Тогда

$$x_0 = a\cos\theta_0, \quad y_0 = b\sin\theta_0,$$
$$x = a\cos\theta, \quad y = b\sin\theta,$$

и потому

$$\frac{x_0 x}{a^2} + \frac{y_0 y}{b^2} - 1 = \cos\theta_0\cos\theta + \sin\theta_0\sin\theta - 1 = \cos(\theta_0 - \theta) - 1 \leqslant 0.$$

Таким образом, при подстановке координат произвольной точки $M \neq M_0$ эллипса в уравнение касательной в точке M_0 действительно получаются числа одного и того же знака (отрицательные).

Случай гиперболы совершенно аналогичен:

$$\frac{x_0 x}{a^2} - \frac{y_0 y}{b^2} - 1 = \operatorname{ch}\theta_0 \operatorname{ch}\theta - \operatorname{sh}\theta_0 \operatorname{sh}\theta - 1 = \operatorname{ch}(\theta - \theta_0) - 1 \geqslant 0,$$

причем равенство нулю имеет место только при $\theta = \theta_0$.

Очень интересное и важное (для оптики) свойство касательных к коническим сечениям указывается в следующем предложении:

Предложение 2. *Касательная к эллипсу или гиперболе образует равные углы с фокальными радиусами точки касания.*

Доказательство. Под углом между двумя прямыми мы понимаем здесь острый (точнее, не тупой) угол из двух смежных углов, образованных этими прямыми. Поэтому для доказательства равенства интересующих нас углов достаточно до-

казать, что они имеют равные синусы, т. е. что отношение расстояния от фокуса до касательной к длине соответствующего фокального радиуса — одно и то же для обоих фокусов.

Мы знаем (п. 5 § 1 гл. 3), что расстояние от некоторой точки до прямой равно абсолютной величине числа, получающегося при подстановке координат этой точки в левую часть нормального уравнения прямой. Поэтому прежде всего мы должны выяснить, что получится при подстановке координат фокусов в нормальное уравнение касательной.

Согласно формулам (6) и (9) нормальное уравнение касательной к эллипсу или к гиперболе имеет вид

$$\left(\frac{x_0x}{a^2} \pm \frac{y_0y}{b^2} - 1\right)N = 0,$$

где $N > 0$ — некоторый нормирующий множитель (точное значение которого нам не понадобится). Подставив в левую часть этого уравнения координаты $(-c, 0)$ левого фокуса, мы получим число

$$\left(-\frac{x_0c}{a^2} - 1\right)N = -\frac{a + ex_0}{a}N,$$

абсолютная величина которого равна

$$\frac{N}{a}|a + ex_0|.$$

Но как для эллипса, так и для гиперболы длина r_1 левого фокального радиуса точки $M_0(x_0, y_0)$ равна $|a + ex_0|$; см. формулы (6) п. 2 § 1 и (7) п. 3 § 1 (заметим, что в пп. 2 и 3 § 1 мы число r_1 называли просто левым фокальным радиусом). Следовательно, синус угла, образованного левым фокальным радиусом с касательной, равен $\frac{N}{a}$.

Аналогично, при подстановке координат $(c, 0)$ правого фокуса в левую часть нормального уравнения касательной получается число

$$\left(\frac{x_0c}{a^2} - 1\right)N = -\frac{a - ex_0}{a}N,$$

абсолютная величина которого равна

$$\frac{N}{a}|a - ex_0| = \frac{N}{a}r_2$$

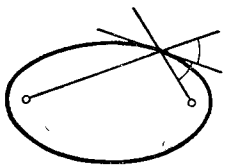
(см. формулы (7) п. 2 § 1 и (8) п. 2 § 1. Поэтому синус угла, образованного правым фокальным радиусом с касательной, равен тому же числу $\frac{N}{a}$.

Тем самым предложение 2 полностью доказано.

В случае эллипса более внимательный анализ показывает, что при подстановке координат любого из фокусов в нормальное уравнение касательной получается отрицательное число. Следовательно (см. п. 5 § 1 гл. 3),

оба фокуса эллипса расположены по одну сторону от касательной (а именно, там, где расположен центр эллипса).

Отсюда следует, что касательная не пересекает треугольник с вершинами в фокусах и точке касания, т. е. содержится в вертикальных углах, внешних к этому треугольнику. Образуя одинаковые углы со сторонами треугольника, она поэтому является биссектрисой этих внешних углов. Таким образом,



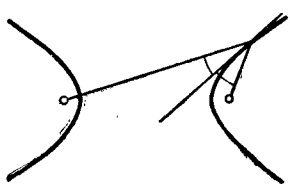
касательная к эллипсу является биссектрисой внешних углов, образованных фокальными радиусами точки касания.

Из этого свойства касательной вытекает, что лучи, исходящие из фокуса вогнутого эллиптического зеркала, собираются в другом его фокусе.

В случае гиперболы касательная (9) пересекает ось Ox в точке с абсциссой $\frac{a^2}{x_0}$. Но, как мы знаем, $|x_0| \geqslant a$, и потому

$$\left|\frac{a^2}{x_0}\right| \leqslant a.$$

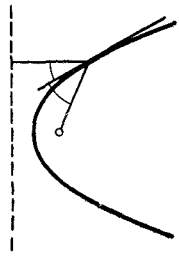
Это означает, что точка пересечения касательной с осью абсцисс лежит между вершинами гиперболы, а потому и между ее фокусами. Другими словами, касательная проходит внутри треугольника, имеющего вершины в фокусах гиперболы и в точке касания. Образуя равные углы со сторонами этого треугольника, она является, следовательно, биссектрисой его внутреннего угла. Таким образом,



касательная к гиперболе является биссектрисой внутреннего угла, образованного фокальными радиусами точки касания.

Чтобы получить аналог предложения 2 для случая параболы, мы рассмотрим угол, образованный фокальным радиусом точки касания M_0 и перпендикуляром, опущенным из точки касания на директрису параболы (точнее, тот из двух смежных углов, образованных этими прямыми, который является углом треугольника, имеющего вершины в фокусе, в основании перпендикуляра и в точке касания). Острый угол, образованный перпендикуляром и касательной, равен, очевидно, углу, образованному касательной и осью абсцисс, т. е. равен углу при вершине A в треугольнике AFM_0, где A — точка пересечения касательной с осью абсцисс, а F — фокус параболы. Но точка A имеет, очевидно, абсциссу $-x_0$. Поэтому ее расстояние до фо-

куса F равно $\frac{p}{2}+x_0$ (напомним, что $x \geqslant 0$), т. е. равно расстоянию δ_{M_0} от точки касания M_0 до директрисы, а значит (см. предложение 1 п. 1), равно расстоянию r_{M_0} от точки M_0 до фокуса F. Следовательно, треугольник AFM_0 — равнобедренный, и потому углы при его вершинах A и M равны. Тем самым доказано, что угол, образованный касательной с перпендикуляром, опущенным из точки касания на директрису, равен углу, образованному касательной с фокальным радиусом точки касания. Другими словами, мы доказали следующее



Предложение 3. *Касательная к параболе является биссектрисой угла, образованного фокальным радиусом точки касания и перпендикуляром, опущенным из точки касания на директрису.*

Замечание 3. В изложенном доказательстве молчаливо предполагалось, что точка M_0 отлична от вершины (0, 0) параболы. Однако доказанное утверждение остается справедливым и в этом случае, поскольку касательная к параболе в ее вершине перпендикулярна оси параболы.

Из доказанного свойства касательной к параболе следует, что лучи, исходящие из фокуса вогнутого параболического зеркала, отражаются параллельным пучком, а именно, параллельно оси зеркала. На этом свойстве параболических зеркал основано устройство прожекторов, зеркальных телескопов, фар автомобилей и т. п.

4. Семейства софокусных эллипсов и гипербол

Пусть на плоскости даны две (различные) точки F_1 и F_2. В этом пункте мы изучим семейство всех эллипсов и гипербол, имеющих точки F_1 и F_2 своими фокусами.

Выберем прямоугольные координаты x, y так, чтобы точка F_1 имела координаты $(-c, 0)$, а точка F_2 — координаты $(c, 0)$, где c — половина расстояния между точками F_1 и F_2.

Ясно, что для любого эллипса (и любой гиперболы) с фокусами F_1 и F_2 прямая, соединяющая фокусы (т. е. ось абсцисс), и их медиатриса (т. е. ось ординат) являются осями симметрии эллипса (гиперболы). Это показывает, что для любой линии нашего «семейства софокусных эллипсов и гипербол» координаты x, y являются каноническими координатами. Следовательно, уравнение этой линии мы можем записать в виде

$$\frac{x^2}{A}+\frac{y^2}{B}=1,$$

где A и B — некоторые числа, обладающие тем свойством, что

$$A-B=c^2.$$

Выбрав в качестве «начального» произвольный эллипс

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1,$$

где

$$a^2-b^2=c^2,$$

$$a>b>0,$$

и положив

$$\lambda=A-a^2=B-b^2,$$

мы можем уравнение произвольной линии рассматриваемого семейства записать в виде

$$\frac{x^2}{a^2+\lambda}+\frac{y^2}{b^2+\lambda}=1, \qquad (1)$$

где λ — некоторый параметр (отличный от $-a^2$ и $-b^2$).

Это уравнение обычно называется *каноническим уравнением семейства софокусных эллипсов и гипербол.*

Возможны следующие случаи:

1. $-\infty<\lambda<a^2$. Уравнение (1) не удовлетворяется ни одной точкой плоскости. Этот случай нам не интересен и потому в дальнейшем мы будем исключать его из рассмотрения.

2. $-a^2<\lambda<-b^2$. Уравнение (1) является уравнением гиперболы с действительной полуосью $\sqrt{a^2+\lambda}$ и мнимой полуосью $\sqrt{-b^2-\lambda}$.

3. $-b^2<\lambda<+\infty$. Уравнение (1) является уравнением эллипса с большой полуосью $\sqrt{a^2+\lambda}$ и малой полуосью $\sqrt{b^2+\lambda}$.

При $\lambda\to-a^2+0$ действительная полуось гиперболы (1) неограниченно убывает, и эта гипербола все теснее прижимается с двух сторон к оси ординат Oy. Уравнение же (1)[1]) при $\lambda\to-a^2+0$ переходит в уравнение $x^2=0$ дважды взятой оси Oy. Таким образом, дважды взятую ось ординат можно рассматривать как предельный случай гипербол (1) при $\lambda\to-a^2+0$.

При $\lambda\to-b^2-0$ действительная полуось гиперболы (1) приближается к $c=\sqrt{a^2-b^2}$, а мнимая полуось неограниченно убывает. Сама же гипербола все теснее прижимается с двух сторон к оси абсцисс, из которой вырезан отрезок между фокусами.

При $\lambda\to-b^2+0$ большая полуось эллипса (1) приближается к $c=\sqrt{a^2-b^2}$, а малая неограниченно убывает. Сам же эллипс все теснее охватывает отрезок оси абсцисс между фокусами.

---

[1]) Точнее, равносильное ему (при $\lambda\neq-a^2, -b^2$) уравнение

$$(b^2+\lambda)x^2+(a^2+\lambda)y^2=(a^2+\lambda)(b^2+\lambda).$$

Поскольку уравнение (1) при $\lambda \to -b^2$ дает дважды взятую ось абсцисс $y^2 = 0$, мы можем, таким образом, сказать, что пределом кривых (1) при $\lambda \to -b^2$ является дважды взятая ось Ox.

При $\lambda \to +\infty$ эллипс (1) неограниченно увеличивается, а его эксцентриситет e стремится к единице (эллипс «округляется»).

Предложение 1. *Через любую точку плоскости (не принадлежащую координатным осям) проходят точно две кривые рассматриваемого семейства: одна — являющаяся эллипсом, и другая — являющаяся гиперболой.*

Доказательство. Условие, что линия (1) проходит через точку (x, y), выражается уравнением

$$x^2(b^2+\lambda) + y^2(a^2+\lambda) = (a^2+\lambda)(b^2+\lambda),$$

т. е. уравнением

$$\lambda^2 + (a^2 + b^2 - x^2 - y^2)\lambda + a^2b^2 - b^2x^2 - a^2y^2 = 0.$$

Поскольку

$$(a^2+b^2-x^2-y^2)^2 - 4(a^2b^2 - b^2x^2 - a^2y^2) = \\ = (b^2 - a^2 + x^2 - y^2)^2 + 4x^2y^2,$$

это уравнение имеет два вещественных корня

$$\lambda_{1,2} = \frac{x^2+y^2-a^2-b^2 \pm \sqrt{(b^2-a^2+x^2-y^2)^2+4x^2y^2}}{2}, \qquad (2)$$

при $x \neq 0$ и $y \neq 0$ отличных от $-a^2$ и $-b^2$.

Так как при $x \neq 0$ и $y \neq 0$ функция

$$f(\lambda) = \lambda^2 + (a^2 + b^2 - x^2 - y^2)\lambda + a^2b^2 - b^2x^2 - a^2y^2$$

положительна при $\lambda = -a^2$ (равна $(a^2 - b^2)x^2$), отрицательна при $\lambda = -b^2$ (равна $(b^2 - a^2)y^2$) и положительна при $\lambda \to +\infty$, то один из корней (2) расположен на отрезке $(-a^2, -b^2)$ (и потому ему соответствует гипербола), а другой — на отрезке $(-b^2, +\infty)$ (и потому ему соответствует эллипс).

Определение 1. Числа λ_1 и λ_2, определенные формулой (2), называются *эллиптическими координатами* (на плоскости).

Чтобы найти выражения прямоугольных координат x, y через эллиптические координаты λ_1, λ_2, положим в тождестве

$$f(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)$$

сначала $\lambda = -a^2$, а затем $\lambda = -b^2$. Тогда мы получим, что

$$(a^2 - b^2)x^2 = (a^2 + \lambda_1)(a^2 + \lambda_2),$$
$$-(a^2 - b^2)y^2 = (b^2 + \lambda_1)(b^2 + \lambda_2),$$

откуда следует, что

$$x^2 = \frac{(a^2+\lambda_1)(a^2+\lambda_2)}{a^2-b^2},$$
$$y^2 = -\frac{(b^2+\lambda_1)(b^2+\lambda_2)}{a^2-b^2}. \qquad (3)$$

Обратим внимание на то, что эти формулы определяют лишь квадраты прямоугольных координат, так что для каждой пары значений

$$\lambda_1,\ \lambda_2,\quad -a^2 < \lambda_1 < -b^2,\quad -b^2 < \lambda_2 < +\infty,$$

мы получим по четыре точки. Собственно говоря, этого и следовало ожидать, поскольку софокусные эллипс и гипербола пересекаются в четырех точках. Таким образом, эллиптические координаты являются «настоящими» координатами только в каждом координатном квадранте.

Координатными линиями, т. е. линиями $\lambda_1 = \text{const}$ и $\lambda_2 = \text{const}$, для эллиптических координат являются, по определению, софокусные эллипсы и гиперболы (линии $\lambda_1 = \text{const}$ — гиперболы, а линии $\lambda_2 = \text{const}$ — эллипсы). Как мы знаем (см. п. 3), касательная к эллипсу одинаково наклонена к фокальным радиусам точки касания и является внешней биссектрисой угла между ними, тогда как для гиперболы касательная является внутренней биссектрисой угла между фокальными радиусами точки касания. Отсюда следует, что

софокусные эллипс и гипербола пересекаются под прямым углом.

Координаты на плоскости называются *ортогональными*, если их координатные линии пересекаются под прямым углом. Таким образом, последнее утверждение означает, что

эллиптические координаты являются ортогональными координатами.

Мы не можем здесь вдаваться в исследование многих замечательных свойств эллиптических координат. Отметим лишь следующее свойство «равнодиагональности»:

любой «криволинейный четырехугольник», образованный двумя линиями $\lambda_1 = \text{const}$ *и двумя линиями* $\lambda_2 = \text{const}$, *имеет диагонали равной длины.*

Упражнение. Докажите свойство равнодиагональности.

Замечание 1. Иногда *эллиптическими координатами* называются величины u и v, связанные с координатами λ_1 и λ_2 соотношениями

$$\sin u = \frac{\sqrt{a^2+\lambda_1}}{c},\quad \operatorname{ch} v = \frac{\sqrt{a^2+\lambda_2}}{c},$$
$$0 \leqslant u < 2\pi,\quad 0 \leqslant v < \infty.$$

Задание. Докажите, что прямоугольные координаты x, y выражаются через координаты u и v по формулам

$$x = c \sin u \operatorname{ch} v,$$
$$y = \pm c \cos u \operatorname{sh} v.$$

Координатная линия $u = \text{const}$ является гиперболой

$$\frac{x^2}{c^2 \sin^2 u} - \frac{y^2}{c^2 \cos^2 u} = 1,$$

а координатная линия $v = \text{const}$ — эллипсом

$$\frac{x^2}{c^2 \operatorname{ch}^2 v} + \frac{y^2}{c^2 \operatorname{sh}^2 v} = 1.$$

Упражнение. Рассмотрите аналогичным образом семейства софокусных парабол и постройте соответствующие *параболические координаты.*

5. Диаметры конических сечений

Рассмотрим произвольное семейство параллельных хорд параболы $y^2 = 2px$, т. е. семейство отрезков, получающихся при пересечении параболы прямыми

$$x = ky + \tau$$

с одним и тем же k и различными τ (конечно, удовлетворяющими условию $\tau > -\frac{pk^2}{2}$; см. п. 2). Координаты (x_1, y_1) и (x_2, y_2) концов любой хорды этого семейства связаны соотношениями

$$y_1^2 = 2px_1, \quad y_2^2 = 2px_2, \quad \frac{x_2 - x_1}{y_1 - y_2} = k$$

(первые два соотношения выражают тот факт, что концы хорды принадлежат параболе, а третье — что хорда имеет угловой коэффициент k). Следовательно,

$$y_2^2 - y_1^2 = 2p(x_2 - x_1),$$

т. е.

$$\frac{y_1 + y_2}{2} = p \frac{x_2 - x_1}{y_2 - y_1}$$

и потому

$$\frac{y_1 + y_2}{2} = pk.$$

Но число $\frac{y_1 + y_2}{2}$ является, как мы знаем, ординатой середины отрезка с концами в точках (x_1, y_1) и (x_2, y_2), т. е. середины рассматриваемой хорды. Тем самым мы доказали, что

середины параллельных хорд параболы (имеющих относительно оси ординат угловой коэффициент k) принадлежат прямой

$$y = pk,$$

параллельной оси параболы.

Определение 1. Если для некоторой линии и некоторого семейства параллельных хорд этой линии середины хорд семейства принадлежат одной прямой, то эта прямая называется *диаметром* данной линии, *сопряженным* с направлением данных хорд.

В этой терминологии доказанное свойство параболы означает, во-первых, что

с любым неасимптотическим направлением сопряжен некоторый диаметр параболы,

и, во-вторых, что

все диаметры параболы параллельны ее оси (и, таким образом, имеют асимптотическое направление).

Кроме того, поскольку в виде $y = pk$ можно (при соответствующем выборе k) представить уравнение любой прямой, параллельной оси параболы, мы видим, что

любая прямая, параллельная оси параболы, является ее диаметром,

а поскольку представление этой прямой в виде $y = pk$ единственно, то

любой диаметр параболы сопряжен с однозначно определенным (неасимптотическим) направлением.

О диаметре, сопряженном с асимптотическим направлением, говорить не приходится, поскольку хорд асимптотического направления не существует.

Определение 2. Диаметр линии называется *главным*, если он перпендикулярен хордам, с направлением которых он сопряжен.

Поскольку любой диаметр параболы параллелен оси Ox, каждый главный диаметр параболы (если он существует) должен быть сопряжен с хордами, параллельными оси Oy. Но для таких хорд $k = 0$, и потому диаметр имеет уравнение $y = 0$, т. е. является осью параболы. Таким образом,

парабола обладает единственным главным диаметром, являющимся ее осью.

Не нужно думать, что любая точка диаметра является серединой некоторой хорды. Действительно, тот факт, что точка (x, y) диаметра $y = pk$ является серединой некоторой хорды (с угловым коэффициентом k) означает, что

$$x = \frac{x_1 + x_2}{2}, \quad y = \frac{y_1 + y_2}{2} = pk,$$

где x_1, y_1 и x_2, y_2 — такие числа, что

$$y_1^2 = 2px_1, \quad y_2^2 = 2px_2.$$

В силу тождества

$$\frac{y_1^2 + y_2^2}{2} = \left(\frac{y_1 - y_2}{2}\right)^2 + \left(\frac{y_1 + y_2}{2}\right)^2,$$

мы имеем:

$$x = \frac{x_1 + x_2}{2} = \frac{y_1^2 + y_2^2}{2p} = \frac{1}{p}\left(\frac{y_1 - y_2}{2}\right)^2 + \frac{pk^2}{2}, \qquad (1)$$

откуда непосредственно вытекает, что

точка (x, y) диаметра $y = pk$ тогда и только тогда является серединой некоторой хорды (с угловым коэффициентом k), когда

$$x > \frac{pk^2}{2}. \qquad (2)$$

Действительно, если точка (x, y) является серединой хорды, то $x > \frac{pk^2}{2}$ согласно формуле (1). Обратно, если $x > \frac{pk^2}{2}$, то из соотношения (1) мы получим для $\frac{y_1 - y_2}{2}$ вещественное значение $\sqrt{p\left(x - \frac{pk^2}{2}\right)}$. Решая затем уравнения

$$\frac{y_1 + y_2}{2} = pk,$$

$$\frac{y_1 - y_2}{2} = \sqrt{p\left(x - \frac{pk^2}{2}\right)},$$

мы найдем ординаты концов хорды, серединой которой является данная точка (x, y).

Таким образом, мы видим, что

на диаметре $y = pk$, сопряженном с хордами, имеющими угловой коэффициент k, середины этих хорд заполняют полупрямую, характеризующуюся неравенством (2).

Иногда *диаметром* параболы, сопряженным с хордами данного направления, называют геометрическое место середин этих хорд, т. е. полупрямую (2).

Концевая точка

$$(x_0, y_0) = \left(\frac{pk^2}{2}, pk\right)$$

полупрямой (2) принадлежит параболе (ибо $y_0^2 = 2px_0$), и прямая, проходящая через эту точку параллельно хордам рассматриваемого направления, имеет уравнение

$$x = ky + x_0 - ky_0.$$

Поскольку $y_0 = pk$, мы можем это уравнение переписать в следующем виде:

$$x = \frac{y_0}{p} y + x_0 - \frac{y_0^2}{p},$$

т. е. в виде

$$x = \frac{y_0}{p} y - x_0.$$

Сравнив это уравнение с уравнением касательной к параболе в точке (x_0, y_0) (см. уравнение (2) п. 8), мы немедленно получим, что они выражают одну и ту же прямую. Тем самым доказано, что

касательная к параболе параллельна хордам, с направлением которых сопряжен диаметр, проходящий через точку касания.

Похожие результаты имеют место для эллипса и гиперболы. Например, рассмотрим для гиперболы

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \qquad a > 0, \quad b > 0, \tag{3}$$

семейство параллельных хорд, получающихся при пересечении гиперболы прямыми

$$y = kx + \tau$$

с одним и тем же $k \neq \pm \frac{b}{a}$ и различными τ (конечно, удовлетворяющими условию $\tau^2 > a^2k^2 - b^2$; см. п. 2).

Координаты (x_1, y_1) и (x_2, y_2) концов любой хорды этого семейства связаны соотношениями

$$\frac{x_1^2}{a^2} - \frac{y_1^2}{b^2} = 1, \quad \frac{x_2^2}{a^2} - \frac{y_2^2}{b^2} = 1, \quad \frac{y_2 - y_1}{x_2 - x_1} = k$$

(первые два соотношения выражают тот факт, что концы хорды принадлежат гиперболе, а третье — что хорда имеет данный угловой коэффициент k). Следовательно,

$$\frac{x_2^2 - x_1^2}{a^2} - \frac{y_2^2 - y_1^2}{b^2} = 0,$$

т. е.

$$\frac{y_1 + y_2}{2} = \frac{b^2}{a^2} \cdot \frac{x_2 - x_1}{y_2 - y_1} \cdot \frac{x_1 + x_2}{2},$$

и потому

$$\frac{y_1 + y_2}{2} = \frac{b^2}{a^2} \cdot \frac{1}{k} \cdot \frac{x_1 + x_2}{2}.$$

Это показывает, что

середины параллельных хорд гиперболы (имеющих неасимптотическое направление, характеризуемое угловым коэффи-

циентом k) принадлежат прямой

$$y = \frac{b^2}{a^2 k} x, \tag{4}$$

проходящей через центр гиперболы.

Другими словами,

прямая (4) *является диаметром гиперболы* (3), *сопряженным с хордами, имеющими угловой коэффициент k.*

Середины хорд, параллельных одной из координатных осей, принадлежат, очевидно, другой координатной оси. Следовательно,

координатные оси (*оси симметрии гиперболы*) *также являются ее диаметрами, причем каждый из них сопряжен с хордами, имеющими направление другого.*

Таким образом, единственными прямыми, проходящими через центр гиперболы, которые, возможно, не являются его диаметрами, являются прямые с угловым коэффициентом вида $\frac{b^2}{a^2 k}$, где $k = \pm \frac{b}{a}$ — угловой коэффициент некоторого асимптотического направления. Но ясно, что эти прямые и на самом деле не являются диаметрами, поскольку хорд асимптотического направления не существует и, тем более, не существует их середин.

Поскольку при $k = \pm \frac{b}{a}$ число $\frac{b^2}{a^2 k}$ равно $\pm \frac{b}{a}$, мы видим, таким образом, что этими исключительными прямыми являются асимптоты гиперболы. Тем самым, мы доказали, что

любая прямая, проходящая через центр гиперболы и не являющаяся ее асимптотой, является диаметром гиперболы.

Определение 3. Два направления, характеризующиеся угловыми коэффициентами k и k', называются *сопряженными* (относительно гиперболы (3)), если

$$kk' = \frac{b^2}{a^2}. \tag{5}$$

Кроме того, направления координатных осей также считаются сопряженными.

В этой терминологии доказанное выше утверждение означает, что

диаметром гиперболы, сопряженным с хордами данного неасимптотического направления, является прямая, проходящая через центр и имеющая (*неасимптотическое*!) *направление, сопряженное с данным.*

Диаметры, имеющие сопряженные направления, называются *сопряженными.*

Ясно, что

отношение сопряженности направлений (*и диаметров*) *симметрично, т. е. если одно направление сопряжено с другим, то второе сопряжено с первым.*

Действительно, для направлений координатных осей это очевидно, а для остальных направлений непосредственно вытекает из симметричности формулы (5) относительно k и k'.

Другими словами,

середины хорд, параллельных некоторому диаметру, принадлежат диаметру, параллельному хордам, середины которых лежат на данном диаметре.

Коротко можно сказать, что

из двух сопряженных диаметров каждый сопряжен с хордами, параллельными другому диаметру.

При $k = k'$ равенство (5) возможно только при $k = \pm \frac{b}{a}$.

Следовательно,

асимптотические направления гиперболы (и только они) сопряжены сами себе.

На этом основании асимптоты гиперболы иногда называются ее «самосопряженными диаметрами».

Напомним, что диаметр называется *главным*, если хорды, с которыми он сопряжен, ему перпендикулярны. Таким образом,

диаметр гиперболы тогда и только тогда является главным, когда он перпендикулярен сопряженному диаметру.

В частности,

диаметр, сопряженный с главным диаметром, сам является главным.

Одна пара главных диаметров очевидна: это — пара, состоящая из осей гиперболы. Поскольку (при $k, k' \neq 0, \infty$) условие перпендикулярности

$$kk' = -1$$

не может быть выполнено одновременно с условием сопряженности

$$kk' = \frac{b^2}{a^2},$$

никаких других главных диаметров гипербола не имеет. Таким образом,

оси гиперболы и только они являются главными диаметрами гиперболы.

Пусть $M_0(x_0, y_0)$ — произвольная точка диаметра (4). Прямая, проходящая через точку M_0 и имеющая угловой коэффициент k, выражается уравнением

$$x = kx + t_0,$$

где $t_0 = y_0 - kx_0$. Но мы знаем (см. п. 2), что эта прямая тогда и только тогда пересекает гиперболу в двух точках, когда

$$t_0^2 > a^2k^2 - b^2,$$

т. е. когда

$$(y_0 - kx_0)^2 > a^2k^2 - b^2. \qquad (6)$$

Следовательно, это неравенство является необходимым и достаточным условием того, чтобы точка $M_0(x_0, y_0)$ принадлежала некоторой хорде гиперболы, имеющей угловой коэффициент k (и потому, являясь точкой диаметра (4), была серединой этой хорды). Но, по условию,

$$y_0 = \frac{b^2}{a^2k} x_0,$$

и потому

$$y_0 - kx_0 = -\frac{a^2k^2 - b^2}{a^2k} x_0.$$

Подставляя это выражение в неравенство (6) (и опуская индексы), мы окончательно получаем, что

точка $M(x, y)$, принадлежащая диаметру, сопряженному с хордами, имеющими угловой коэффициент k, тогда и только тогда является серединой одной из этих хорд, когда

$$x^2 > \frac{a^4k^2}{a^2k^2 - b^2}. \qquad (7)$$

При $|k| < \frac{b}{a}$ это неравенство удовлетворяется тождественно, т. е. при всех значениях x, а при $|k| > \frac{b}{a}$ оно определяет на диаметре (4) две полупрямые, получающиеся при удалении из диаметра некоторого отрезка, концами которого являются точки с координатами

$$x_0 = \frac{a^2k}{\pm\sqrt{a^2k^2 - b^2}}, \quad y_0 = \frac{b^2}{\pm\sqrt{a^2k^2 - b^2}}, \qquad (8)$$

удовлетворяющими уравнению гиперболы (3). Тем самым доказано, что

на каждом диаметре гиперболы точки, являющиеся серединами хорд, имеющих направление, сопряженное с направлением диаметра, либо заполняют весь диаметр, либо принадлежат внешности отрезка, высекаемого на диаметре гиперболой.

Заметив, что при $|k| < \frac{b}{a}$ диаметр (4) вообще не пересекает гиперболы и потому в этом случае «внешность отрезка, высекаемого на диаметре гиперболой» представляет собой весь диаметр, мы получаем, что во всех случаях можно считать, что середины указанных хорд заполняют внешность отрезка, высекаемого на диаметре гиперболой.

Без труда проверяется, что это утверждение остается верным и для диаметров, не представимых в виде (4), т. е. для осей гиперболы. При этом на действительной оси середины

перпендикулярных хорд заполняют внешность отрезка с концами в вершинах гиперболы, а на мнимой оси исчерпывают всю эту ось.

Согласно формуле (10) п. 3 угловой коэффициент касательной в каждой из точек с координатами (8) равен

$$\frac{b^2}{a^2} \cdot \frac{x_0}{y_0} = k.$$

Таким образом, так же как для параболы,

касательная к гиперболе параллельна хордам, с направлением которых сопряжен диаметр, проходящий через точку касания.

Очевидно, что это утверждение справедливо и для диаметров, являющихся осями гиперболы.

Совершенно аналогичные результаты имеют место и для эллипса

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad a \geqslant b > 0. \tag{9}$$

Во-первых,

диаметром эллипса (9), *сопряженным с хордами, имеющими угловой коэффициент* k, *является прямая*

$$y = -\frac{b^2}{a^2 k} x, \tag{10}$$

проходящая через центр эллипса, причем любая прямая, проходящая через центр эллипса, является ее диаметром.

В соответствии с формулой (10) направления, характеризуемые угловыми коэффициентами k и k', называются *сопряженными* относительно эллипса (9), если

$$kk' = -\frac{b^2}{a^2} \tag{11}$$

(направления координатных осей также, конечно, считаются сопряженными).

Так же как и для гиперболы,

отношение сопряженности направлений (и диаметров) взаимно, т. е. из двух сопряженных диаметров каждый сопряжен с хордами, параллельными другому диаметру.

Самосопряженных направлений (диаметров) эллипс не имеет.

Для эллипса, являющегося окружностью ($a = b$), условие (11) сопряженности совпадает с условием

$$kk' = -1 \tag{12}$$

перпендикулярности. Поэтому

диаметры окружности тогда и только тогда сопряжены, когда они перпендикулярны.

В частности,

любой диаметр окружности является главным диаметром.

Напротив, поскольку при $a > b$ условия (11) и (12) одновременно выполнены быть не могут,

единственными главными диаметрами эллипса, не являющегося окружностью, являются его оси.

Простое вычисление показывает, что

точка $M(x, y)$ диаметра (10) *тогда и только тогда является серединой одной из сопряженных хорд, когда*

$$|x| < \frac{a^2|k|}{\sqrt{a^2k^2 + b^2}}. \qquad (13)$$

Неравенство (13) определяет на диаметре некоторый отрезок, концы которого принадлежат эллипсу.

Далее, так же как для гиперболы,

касательная к эллипсу параллельна хордам, с направлением которых сопряжен диаметр, проходящий через точку касания.

Задание. Докажите (аналитически) все сформулированные выше свойства диаметров эллипса.

Замечание 1. Все основные свойства диаметров эллипса можно получить чисто геометрически без всяких выкладок, воспользовавшись тем, что любой эллипс получается сжатием окружности к ее диаметру, и заметив, что при любом сжатии параллельные прямые переходят в параллельные прямые, а середины отрезков — в середины отрезков. Действительно, тот факт, что середины параллельных хорд окружности принадлежат одной прямой, очевиден. С другой стороны, любое семейство параллельных хорд эллипса получается при рассматриваемом сжатии из некоторого семейства параллельных хорд окружности. Следовательно, геометрическое место середин параллельных хорд эллипса является образом при этом сжатии геометрического места середин параллельных хорд окружности и потому принадлежит некоторой прямой. С этой точки зрения, сопряженные диаметры эллипса — это диаметры, являющиеся образами при сжатии перпендикулярных диаметров окружности, а главные диаметры — это перпендикулярные диаметры, остающиеся после сжатия перпендикулярными, и т. д.

6. Теоремы Аполлония

Согласно формуле (8) п. 2 диаметр

$$y = kx$$

пересекает эллипс

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad a \geqslant b > 0,$$

в точках с абсциссами

$$x = \pm \frac{ab}{\sqrt{a^2k^2 + b^2}}$$

и, следовательно, с ординатами

$$y = \pm \frac{abk}{\sqrt{a^2k^2 + b^2}}.$$

Рассмотрим теперь сопряженный диаметр

$$y = k'x.$$

Поскольку $k' = -\frac{b^2}{a^2k}$, то

$$a^2k'^2 + b^2 = \frac{b^4}{a^2k^2} + b^2 = b^2\frac{a^2k^2 + b^2}{a^2k^2}, \qquad abk' = -\frac{b^3}{ak}.$$

Следовательно, этот сопряженный диаметр пересекает эллипс в точках с координатами

$$x = \pm \frac{a^2k}{\sqrt{a^2k^2 + b^2}}, \qquad y = \mp \frac{b^2}{\sqrt{a^2k^2 + b^2}}.$$

Определение 1. Отрезок, соединяющий центр эллипса (или гиперболы) с некоторой его (ее) точкой, называется *радиусом* эллипса (гиперболы).

Определение 2. Два радиуса эллипса, имеющие сопряженные направления, называются *сопряженными.*

Из только что доказанных формул непосредственно вытекает, что квадрат a'^2 длины a' радиуса, имеющего угловой коэффициент k, выражается формулой

$$a'^2 = \left(\frac{ab}{\sqrt{a^2k^2 + b^2}}\right)^2 + \left(\frac{abk}{\sqrt{a^2k^2 + b^2}}\right)^2 = \frac{a^2b^2(1 + k^2)}{a^2k^2 + b^2},$$

а квадрат b'^2 длины b' сопряженного радиуса — формулой

$$b'^2 = \left(\frac{a^2k}{\sqrt{a^2k^2 + b^2}}\right)^2 + \left(\frac{b^2}{\sqrt{a^2k^2 + b^2}}\right)^2 = \frac{a^4k^2 + b^4}{a^2k^2 + b^2}.$$

Следовательно,

$$a'^2 + b'^2 = \frac{a^2b^2 + a^2b^2k^2 + a^4k^2 + b^4}{a^2k^2 + b^2} = a^2 + b^2.$$

Тем самым мы доказали, что

сумма площадей квадратов, построенных на произвольной паре сопряженных радиусов эллипса, равна сумме площадей квадратов, построенных на его полуосях:

$$a'^2 + b'^2 = a^2 + b^2.$$

Найдем теперь площадь параллелограмма, построенного на паре сопряженных радиусов. Как мы знаем (п. 4 § 6 гл. 1),

площадь параллелограмма, построенного на векторах (a^1, a^2) и (b^1, b^2), равна абсолютной величине определителя

$$\begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix}.$$

В нашем случае этот определитель имеет вид

$$\begin{vmatrix} \pm\dfrac{ab}{\sqrt{a^2k^2+b^2}} & \dfrac{abk}{\sqrt{a^2k^2+b^2}} \\ \pm\dfrac{a^2k}{\sqrt{a^2k^2+b^2}} & \mp\dfrac{b^2}{\sqrt{a^2k^2+b^2}} \end{vmatrix}$$

и потому равен

$$\frac{ab}{a^2k^2+b^2}\begin{vmatrix} 1 & k \\ a^2k & -b^2 \end{vmatrix} = -ab.$$

Тем самым мы доказали, что

площадь параллелограмма, построенного на произвольной паре сопряженных радиусов эллипса, равна площади ab прямоугольника, построенного на его полуосях.

Чтобы доказать аналогичные теоремы для гиперболы, удобно наряду с данной гиперболой

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \qquad a > 0, \quad b > 0, \tag{1}$$

рассматривать также гиперболу

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = -1, \qquad a > 0, \quad b > 0, \tag{2}$$

для которой каноническими координатами являются переставленные координаты x, y.

Определение 3. Гипербола (2) называется гиперболой, *сопряженной* с гиперболой (1).

Ясно, что

центр, асимптоты и основной прямоугольник сопряженной гиперболы (2) *совпадают с центром, асимптотами и основным прямоугольником исходной гиперболы* (1).

Кроме того,

действительной (*мнимой*) *осью гиперболы* (2) *является мнимая* (*действительная*) *ось гиперболы* (1).

В частности,

вершинами гиперболы (2) *являются точки* $(0, b)$ *и* $(0, -b)$ *оси ординат.*

Наконец, легко видеть, что

гиперболы (1) *и* (2) *имеют одни и те же диаметры, причем два диаметра тогда и только тогда сопряжены относительно гиперболы* (1), *когда они сопряжены относительно гиперболы* (2).

Действительно, тот факт, что диаметры обеих гипербол совпадают, очевиден (ибо диаметром является любая прямая, проходящая через центр гиперболы и отличная от ее асимптот, а центр и асимптоты у наших гипербол совпадают). Далее, условие сопряженности диаметров $y = kx$ и $y = k'x$ относительно гиперболы (1) имеет вид

$$kk' = \frac{b^2}{a^2}. \tag{3}$$

Чтобы написать условие сопряженности этих диаметров относительно гиперболы (2), мы должны их уравнения переписать в виде $x = \frac{1}{k}y$ и $x = \frac{1}{k'}y$ (ибо абсциссой канонической системы координат гиперболы (2) является координата y, а ординатой — координата x) и учесть, что a является мнимой полуосью гиперболы (2), а b — действительной. Поэтому условие сопряженности рассматриваемых диаметров относительно гиперболы (2) имеет вид

$$\frac{1}{k} \cdot \frac{1}{k'} = \frac{a^2}{b^2},$$

что, очевидно, совпадает с (3). Наконец, оси координат, очевидно, сопряжены относительно обеих гипербол.

Диаметр

$$y = kx$$

пересекает гиперболу (1), если $|k| < \frac{b}{a}$, и пересекает сопряженную гиперболу (2), если $|k| > \frac{b}{a}$. Поскольку для сопряженного диаметра

$$y = k'x$$

неравенство $|k'| < \frac{b}{a}$ имеет место тогда и только тогда, когда $|k| > \frac{b}{a}$, мы видим, следовательно, что

для любой пары сопряженных диаметров один из диаметров пересекает исходную гиперболу, а другой — сопряженную.

Согласно формуле (11) п. 2 диаметр

$$y = kx, \quad |k| < \frac{b}{a},$$

пересекает гиперболу (1) в точках с абсциссами

$$x = \pm \frac{ab}{\sqrt{b^2 - a^2k^2}}$$

и, следовательно, с ординатами

$$y = \pm \frac{abk}{\sqrt{b^2 - ak^2}}.$$

Поэтому квадрат a'^2 длины a' радиуса гиперболы (1), имеющего угловой коэффициент k, выражается формулой

$$a'^2 = \left(\frac{ab}{\sqrt{b^2 - a^2k^2}}\right)^2 + \left(\frac{abk}{\sqrt{b^2 - a^2k^2}}\right)^2 = \frac{a^2b^2(1+k^2)}{b^2 - a^2k^2}.$$

Аналогично, сопряженный диаметр $y = k'x$ пересекает сопряженную гиперболу (2) в точках с абсциссами

$$x = \pm \frac{ab}{\sqrt{a^2k'^2 - b^2}}$$

и ординатами

$$y = \pm \frac{abk'}{\sqrt{a^2k'^2 - b^2}}.$$

Но, по определению, $k' = \frac{b^2}{a^2k}$, и потому

$$a^2k'^2 - b^2 = \frac{b^4}{a^2k^2} - b^2 = b^2\frac{b^2 - a^2k^2}{a^2k^2}, \qquad abk' = \frac{b^3}{ak}.$$

Следовательно, координаты точек пересечения сопряженного диаметра с сопряженной гиперболой могут быть записаны в следующем виде:

$$x = \pm \frac{a^2|k|}{\sqrt{b^2 - a^2k^2}}, \quad y = \pm \frac{\varepsilon b^2}{\sqrt{b^2 - a^2k^2}},$$

где $\varepsilon = +1$, если $k > 0$, и $\varepsilon = -1$, если $k < 0$.

Эти формулы получены в предположении, что сопряженный диаметр имеет уравнение вида $y = k'x$, т. е. в предположении, что $k \neq 0$. Но при $k = 0$ сопряженным диаметром является ось ординат, пересекающая сопряженную гиперболу в точках $(0, \pm b)$. Поэтому полученные формулы справедливы и при $k = 0$, если только мы условимся, что в этом случае $\varepsilon = 1$ (или, что приводит к тому же результату, $\varepsilon = -1$).

Определение 4. Радиусы сопряженных гипербол (1) и (2), имеющие сопряженные направления, называются *сопряженными* (ср. определение 2).

Из доказанных выше формул непосредственно вытекает, что квадрат b'^2 длины b' радиуса гиперболы (2), сопряженного с радиусом гиперболы (1), имеющим угловой коэффициент k, выражается формулой

$$b'^2 = \left(\frac{a^2|k|}{\sqrt{b^2 - a^2k^2}}\right)^2 + \left(\frac{\varepsilon b^2}{\sqrt{b^2 - a^2k^2}}\right)^2 = \frac{a^4k^2 + b^4}{b^2 - a^2k^2}.$$

Следовательно,

$$b'^2 - a'^2 = \frac{a^4k^2 + b^4 - a^2b^2(1+k^2)}{b^2 - a^2k^2} = b^2 - a^2.$$

Таким образом,

разность площадей квадратов, построенных на любой паре сопряженных радиусов, равна разности площадей квадратов, построенных на полуосях гиперболы:

$$b'^2 - a'^2 = b^2 - a^2.$$

Наконец, площадь параллелограмма, построенного на паре сопряженных радиусов, равна абсолютной величине определителя

$$\begin{vmatrix} \pm\dfrac{ab}{\sqrt{b^2 - a^2k^2}} & \pm\dfrac{abk}{\sqrt{b^2 - a^2k^2}} \\ \pm\dfrac{a^2|k|}{\sqrt{b^2 - a^2k^2}} & \pm\dfrac{\varepsilon b^2}{\sqrt{b^2 - a^2k^2}} \end{vmatrix} = \varepsilon ab$$

(напомним, что, по определению, $k|k| = \varepsilon k^2$). Таким образом, как и для эллипса,

площадь параллелограмма, построенного на любой паре сопряженных радиусов гиперболы, равна площади ab прямоугольника, построенного на полуосях гиперболы.

Доказанные свойства сопряженных радиусов эллипсов и гипербол известны как теоремы Аполлония.

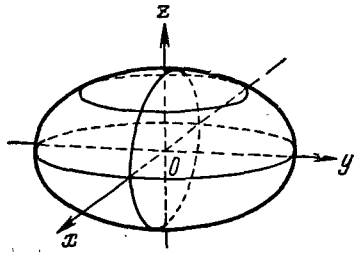
§ 3. ПОВЕРХНОСТИ ВТОРОГО ПОРЯДКА

1. Эллипсоиды

Определение 1. *Эллипсоидом* называется поверхность, выражающаяся в некоторой системе прямоугольных координат x, y, z уравнением вида

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1, \qquad a \geqslant b \geqslant c > 0, \tag{1}$$

где $a \geqslant b \geqslant c > 0$. Эта система координат называется *канонической* (относительно данного эллипсоида), а уравнение (1) называется *каноническим уравнением* эллипсоида.



При $a = b = c$ уравнение (1) является уравнением сферы радиуса a с центром в начале координат.

Таким образом,

сфера является частным случаем эллипсоида.

Пусть $a = b$. Перейдем к цилиндрическим координатам ρ, φ, z (см. п. 5 § 1 гл. 2), согласованным с координатами x, y, z. Эти координаты связаны с координатами x, y, z формулами

$$x = \rho\cos\varphi, \qquad y = \rho\sin\varphi, \qquad z = z.$$

Подставив эти выражения в уравнение (1) (при $a = b$), мы получим уравнение

$$\frac{\rho^2}{a^2} + \frac{z^2}{c^2} = 1.$$

Это уравнение не зависит от φ, и потому вместе с некоторой точкой $M_0(\rho_0, \varphi_0, z_0)$ ему удовлетворяет также любая точка $M(\rho_0, \varphi, z_0)$, где φ произвольно. Геометрически это означает, что вместе с точкой M_0 эллипсоиду принадлежит и любая точка M, получающаяся из точки M_0 поворотом вокруг оси Oz. Поскольку плоскость $y = 0$ пересекает эллипсоид (1) по эллипсу

$$\frac{x^2}{a^2} + \frac{z^2}{c^2} = 1,$$

тем самым доказано, что

эллипсоид

$$\frac{x^2 + y^2}{a^2} + \frac{z^2}{c^2} = 1 \qquad (2)$$

получается из эллипса

$$\frac{x^2}{a^2} + \frac{z^2}{c^2} = 1,$$

расположенного в плоскости Oxz, вращением вокруг оси Oz.

Это дает уже вполне отчетливое представление о форме эллипсоида (2) (называемого обычно *эллипсоидом вращения*).

Аналогично,

при $a = c$ или $b = c$ эллипсоид (1) также является эллипсоидом вращения.

Например, при $b = c$ он получается из эллипса

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

расположенного в плоскости Oxy, вращением вокруг оси Oy.

Замечание 1. Если в плоскости Oxz задана произвольная кривая

$$f(x, z) = 0,$$

то при вращении этой кривой вокруг оси Oz получается поверхность, имеющая в цилиндрических координатах ρ, φ, z уравнение

$$f(\rho, z) = 0, \qquad (3)$$

не зависящее от φ. Таким образом, уравнение (3) является общим уравнением *поверхностей вращения*. В координатах x, y, z оно имеет вид

$$f(\sqrt{x^2 + y^2}, z) = 0.$$

Чтобы представить себе вид произвольного эллипсоида (1), рассмотрим преобразование пространства, переводящее произвольную точку $M(x, y, z)$ в точку $M'(x', y', z')$ с координатами

$$x' = x,$$
$$y' = ky,$$
$$z' = z,$$

где k — некоторое положительное число. Это преобразование называется *сжатием* к плоскости Oxz с коэффициентом k.

Поскольку при таком сжатии точка (x, y, z) получается из точки $\left(x, \frac{y}{k}, z\right)$, эллипсоид вращения (2) переходит в результате этого сжатия в поверхность с уравнением

$$\frac{x^2}{a^2} + \frac{y^2}{(ka)^2} + \frac{z^2}{c^2} = 1,$$

т. е. $\left(\text{при } k = \frac{b}{a}\right)$ в эллипсоид (1). Таким образом,

произвольный эллипсоид (1) *получается из эллипсоида вращения* (2) *сжатием к плоскости* Oxz *с коэффициентом* $\frac{b}{a}$.

Замечание 2. Эллипсоид вращения (2) получается из сферы

$$x^2 + y^2 + z^2 = a^2$$

сжатием пространства к плоскости Oxy с коэффициентом $\frac{c}{a}$. Следовательно,

любой эллипсоид получается из некоторой сферы двумя сжатиями к взаимно перпендикулярным плоскостям.

Другой способ (называемый методом «плоских сечений») представить себе форму эллипсоида (1) заключается в том, чтобы рассмотреть его сечения плоскостями, параллельными координатным плоскостям.

Рассмотрим, например, плоскости $z = h$, параллельные координатной плоскости Oxy. При пересечении этой плоскостью эллипсоида (1) получается линия, имеющая (в координатах x, y, являющихся координатами на плоскости Oxy) уравнение

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 - \frac{h^2}{a^2},$$

т. е. уравнение

$$\frac{x^2}{\left(a\sqrt{1 - \frac{h^2}{c^2}}\right)^2} + \frac{y^2}{\left(b\sqrt{1 - \frac{h^2}{c^2}}\right)^2} = 1.$$

Это показывает, что

при $|h| > c$ *плоскость* $z = h$ *не пересекает эллипсоид* (1);

при $|h| = c$ *плоскость* $z = h$ *имеет с эллипсоидом* (1) *одну общую точку (соответственно* $(0, 0, c)$ *или* $(0, 0, -c)$*)*;

при $|h| < c$ *плоскость* $z = h$ *пересекает эллипсоид* (1) *по эллипсу с полуосями*

$$a\sqrt{1 - \frac{h^2}{c^2}}, \quad b\sqrt{1 - \frac{h^2}{c^2}},$$

наибольшими (и равными a *и* b*) при* $h = 0$ *и монотонно уменьшающимися до нуля, когда* $|h|$ *возрастает от нуля до* c.

Аналогично показывается, что

при $|h| > b$ *плоскость* $y = h$ *не пересекает эллипсоид* (1);

при $|h| = b$ *плоскость* $y = h$ *имеет с эллипсоидом* (1) *одну общую точку* (*соответственно* $(0, b, 0)$ *или* $(0, -b, 0)$);

при $|h| < b$ *плоскость* $y = h$ *пересекает эллипсоид* (1) *по эллипсу с полуосями*

$$a\sqrt{1 - \frac{h^2}{b^2}}, \quad c\sqrt{1 - \frac{h^2}{b^2}},$$

наибольшими (*и равными* a *и* c) *при* $h = 0$ *и монотонно уменьшающимися до нуля, когда* $|h|$ *возрастает от нуля до* b;

и что

при $|h| > a$ *плоскость* $x = a$ *не пересекает эллипсоид* (1);

при $|h| = a$ *плоскость* $x = a$ *имеет с эллипсоидом* (1) *одну общую точку* (*соответственно* $(a, 0, 0)$ *или* $(-a, 0, 0)$);

при $|h| < a$ *плоскость* $x = a$ *пересекает эллипсоид* (1) *по эллипсу с полуосями*

$$b\sqrt{1 - \frac{h^2}{a^2}}, \quad c\sqrt{1 - \frac{h^2}{a^2}},$$

наибольшими (*и равными* b *и* c) *при* $h = 0$ *и монотонно уменьшающимися до нуля, когда* $|h|$ *возрастает от нуля до* a.

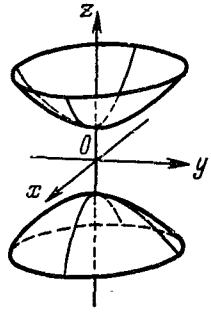
Все это дает уже вполне удовлетворительное представление о форме эллипсоида (1). В частности, мы видим, что

эллипсоид (1) *целиком расположен внутри прямоугольного параллелепипеда с центром в точке* $(0, 0, 0)$, *с гранями, параллельными координатным плоскостям, и со сторонами, имеющими длины* $2a$, $2b$ и $2c$.

Можно еще добавить, что так как уравнение (1) не меняется при изменении знаков координат x, y, z, то

координатные плоскости являются плоскостями симметрии эллипсоида (1), *а начало координат — его центром симметрии.*

Упражнение. Докажите, что если числа a, b и c все различны, то никаких других плоскостей симметрии эллипсоид (1) не имеет.



2. Двуполостные гиперболоиды

Определение 1. *Двуполостным гиперболоидом* называется поверхность, имеющая в некоторой системе прямоугольных координат уравнение вида

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1, \qquad (1)$$

где $a \geqslant b > 0$ и $c > 0$. Эта система координат называется *канонической* (относительно данного гиперболоида), а уравнение (1) называется *каноническим уравнением* двуполостного гиперболоида.

Замечание 1. Весьма часто каноническим уравнением двуполостного гиперболоида называется его уравнение вида

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1.$$

Каноническое в этом смысле уравнение отличается от канонического уравнения в нашем смысле только обозначением координат.

При $a = b$ уравнение двуполостного гиперболоида в цилиндрических координатах ρ, φ, z имеет, очевидно, вид

$$\frac{\rho^2}{a^2} - \frac{z^2}{c^2} = -1.$$

Отсюда вытекает (см. в п. 1 аналогичное рассуждение для эллипсоида), что

гиперболоид

$$\frac{x^2 + y^2}{a^2} - \frac{z^2}{c^2} = -1 \qquad (2)$$

получается из гиперболы

$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = -1,$$

расположенной в плоскости Oxz, вращением вокруг оси Oz (являющейся ее действительной осью).

Это дает уже вполне отчетливое представление о форме гиперболоида (2) (называемого обычно *двуполостным гиперболоидом вращения*).

При сжатии к плоскости Oxz с коэффициентом k гиперболоид (2) переходит, очевидно, в поверхность с уравнением

$$\frac{x^2}{a^2} + \frac{y^2}{(ka)^2} - \frac{z^2}{c^2} = -1,$$

т. е. (при $k = \frac{a}{b}$) — в гиперболоид (1). Таким образом,

произвольный двуполостный гиперболоид (1) *получается из двуполостного гиперболоида вращения* (2) *сжатием к плоскости Oxz с коэффициентом* $\frac{b}{a}$.

Исследуя форму гиперболоида (1) методом плоских сечений, мы немедленно получим, что в пересечении с плоскостью $z = h$ гиперболоид (1) дает линию с уравнением

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{h^2}{c^2} - 1,$$

т. е. с уравнением

$$\frac{x^2}{\left(a\sqrt{\frac{h^2}{c^2} - 1}\right)^2} + \frac{y^2}{\left(b\sqrt{\frac{h^2}{c^2} - 1}\right)^2} = 1.$$

Следовательно,

при $|h|<c$ *плоскость* $z=h$ *не пересекает гиперболоид* (1);

при $|h|=c$ *плоскость* $z=h$ *имеет с гиперболоидом* (1) *одну общую точку (соответственно* $(0,0,c)$ *или* $(0,0,-c)$*)*;

при $|h|>c$ *плоскость* $z=h$ *пересекает гиперболоид* (1) *по эллипсу с полуосями*

$$a\sqrt{\frac{h^2}{c^2}-1},\quad b\sqrt{\frac{h^2}{c^2}-1},$$

монотонно возрастающими (соответственно от a *и* b *до* $+\infty$*), когда* $|h|$ *возрастает от* c *до* $+\infty$.

В пересечении с плоскостью $y=h$ гиперболоид (1) дает линию с уравнением

$$\frac{x^2}{a^2}-\frac{z^2}{c^2}=-1-\frac{h^2}{b^2},$$

т. е. с уравнением

$$\frac{z^2}{\left(c\sqrt{1+\frac{h^2}{b^2}}\right)^2}-\frac{x^2}{\left(a\sqrt{1+\frac{h^2}{b^2}}\right)^2}=1.$$

Следовательно,

любая плоскость $y=h$ *пересекает гиперболоид* (1) *по гиперболе с полуосями*

$$c\sqrt{1+\frac{h^2}{b^2}},\quad a\sqrt{1+\frac{h^2}{b^2}},$$

монотонно возрастающими (соответственно от c *и* a *до* $+\infty$*), когда* $|h|$ *возрастает от нуля до* $+\infty$.

Аналогично показывается, что

любая плоскость $x=h$ *пересекает гиперболоид* (1) *по гиперболе с полуосями*

$$c\sqrt{1+\frac{h^2}{a^2}},\quad b\sqrt{1+\frac{h^2}{a^2}},$$

монотонно возрастающими (соответственно от c *и* b *до* $+\infty$*), когда* $|h|$ *возрастает от нуля до* $+\infty$.

Все это дает уже вполне удовлетворительное представление о форме двуполостного гиперболоида (1).

В частности, мы видим, что, так как ни одна плоскость $z=h$ с $|h|<c$ не пересекает гиперболоид и так как он имеет точки по обе стороны от каждой из этих плоскостей, то

гиперболоид (1) *состоит из двух не связанных друг с другом частей, расположенных соответственно в полупространствах* $z\geqslant c$ *и* $z\leqslant -c$.

Этим и объясняется эпитет «двуполостный» в названии гиперболоида (1).

Заметим в заключение, что по тем же соображениям, что и для эллипсоида,

координатные плоскости являются плоскостями симметрии гиперболоида (1), *а начало координат — его центром симметрии.*

Упражнение. Докажите, что при $a > b$ никаких других плоскостей симметрии гиперболоид (1) не имеет.

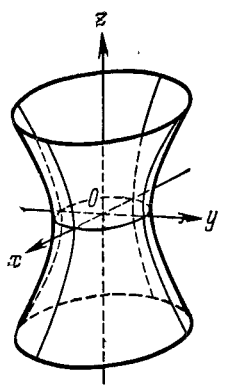
3. Однополостные гиперболоиды

Определение 1. *Однополостным гиперболоидом* называется поверхность, имеющая в некоторой системе прямоугольных координат уравнение вида

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1, \tag{1}$$

где $a \geqslant b > 0$ *и* $c > 0$. Эта система координат называется *канонической* (относительно данного гиперболоида), а уравнение (1) называется *каноническим уравнением* однополостного гиперболоида.

При $a = b$ уравнение однополостного гиперболоида в цилиндрических координатах ρ, φ, z имеет вид



$$\frac{\rho^2}{a^2} - \frac{z^2}{c^2} = 1,$$

Следовательно (ср. пп. 1 и 2), *гиперболоид*

$$\frac{x^2 + y^2}{a^2} - \frac{z^2}{c^2} = 1 \tag{2}$$

получается из гиперболы

$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1,$$

расположенной в плоскости Oxz, вращением вокруг оси Oz (являющейся ее мнимой осью).

Это дает уже вполне отчетливое представление о форме гиперболоида (2) (называемого обычно *однополостным гиперболоидом вращения*).

При сжатии к плоскости Oxz с коэффициентом k гиперболоид (2) переходит, очевидно, в поверхность с уравнением

$$\frac{x^2}{a^2} + \frac{y^2}{(ka)^2} - \frac{z^2}{c^2} = 1,$$

т. е. $\left(\text{при } k = \frac{b}{a}\right)$ — в гиперболоид (1). Таким образом,

произвольный однополостный гиперболоид (1) *получается из однополостного гиперболоида вращения* (2) *сжатием к плоскости Oxz с коэффициентом* $\frac{b}{a}$.

Исследуя форму гиперболоида (1) методом плоских сечений, мы немедленно получим, что в пересечении с плоскостью $z = h$ гиперболоид (1) дает линию с уравнением

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 + \frac{h^2}{c^2},$$

т. е. с уравнением

$$\frac{x^2}{\left(a\sqrt{1+\frac{h^2}{c^2}}\right)^2} + \frac{y^2}{\left(b\sqrt{1+\frac{h^2}{c^2}}\right)^2} = 1.$$

Следовательно,

любая плоскость $z = h$ пересекает гиперболоид (1) по эллипсу с полуосями

$$a\sqrt{1+\frac{h^2}{c^2}}, \quad b\sqrt{1+\frac{h^2}{c^2}},$$

монотонно возрастающими (соответственно от a и b до $+\infty$), когда $|h|$ возрастает от нуля до $+\infty$.

Определение 2. Эллипс

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

получающийся при $h = 0$ (и расположенный в плоскости Oxy), называется *горловым эллипсом* однополостного гиперболоида (1).

В пересечении с плоскостью $y = h$ гиперболоид (1) дает линию с уравнением

$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1 - \frac{h^2}{b^2},$$

т. е. либо с уравнением

$$\frac{x^2}{\left(a\sqrt{1-\frac{h^2}{b^2}}\right)^2} - \frac{z^2}{\left(c\sqrt{1-\frac{h^2}{b^2}}\right)^2} = 1$$

(если $|h| < b$), либо с уравнением

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0 \tag{3}$$

(если $|h| = b$), либо, наконец, с уравнением

$$\frac{z^2}{\left(c\sqrt{\frac{h^2}{b^2}-1}\right)^2} - \frac{x^2}{\left(a\sqrt{\frac{h^2}{b^2}-1}\right)^2} = 1$$

(если $|h| > b$).

Заметив, что точка (x, y) тогда и только тогда удовлетворяет уравнению (3), когда она удовлетворяет либо уравнению

$$\frac{x}{a} - \frac{z}{c} = 0,$$

либо уравнению

$$\frac{x}{a}+\frac{z}{c}=0,$$

мы получаем, следовательно, что

при $|h|<b$ *плоскость* $y=h$ *пересекает гиперболоид* (1) *по гиперболе с полуосями*

$$a\sqrt{1-\frac{h^2}{b^2}},\quad c\sqrt{1-\frac{h^2}{b^2}},$$

монотонно убывающими (соответственно от a *и* c *до нуля), когда* $|h|$ *возрастает от нуля до* b;

при $|h|=b$ *плоскость* $y=h$ *пересекает гиперболоид* (1) *по прямым*

$$\begin{gathered}\frac{x}{a}-\frac{z}{c}=0,\\ y=h\end{gathered}\qquad(4)$$

и

$$\begin{gathered}\frac{x}{a}+\frac{z}{c}=0,\\ y=h;\end{gathered}\qquad(5)$$

при $|h|>b$ *плоскость* $y=h$ *пересекает гиперболоид* (1) *по гиперболе с полуосями*

$$c\sqrt{\frac{h^2}{b^2}-1},\quad a\sqrt{\frac{h^2}{b^2}-1},$$

монотонно возрастающими от нуля до $+\infty$, *когда* $|h|$ *возрастает от* b *до* $+\infty$.

Заметим, что действительные (мнимые) оси гипербол, получающихся при $|h|<b$, параллельны мнимым (действительным) осям гипербол, получающихся при $|h|>b$.

Аналогично показывается, что

при $|h|<a$ *плоскость* $x=h$ *пересекает гиперболоид* (1) *по гиперболе с полуосями*

$$b\sqrt{1-\frac{h^2}{a^2}},\quad c\sqrt{1-\frac{h^2}{a^2}},$$

монотонно убывающими (соответственно от b *и* c *до нуля), когда* $|h|$ *возрастает от нуля до* a;

при $|h|=a$ *плоскость* $x=h$ *пересекает гиперболоид* (1) *по прямым*

$$\frac{y}{b}-\frac{z}{c}=0,\quad x=h \qquad \text{и} \qquad \frac{y}{b}+\frac{z}{c}=0,\quad x=h;$$

при $|h|>a$ *плоскость* $x=h$ *пересекает гиперболоид* (1) *по гиперболе с полуосями*

$$c\sqrt{\frac{h^2}{a^2}-1},\quad b\sqrt{\frac{h^2}{a^2}-1},$$

монотонно возрастающими от нуля до $+\infty$, когда $|h|$ возрастает от a до $+\infty$.

Все это дает вполне удовлетворительное представление о форме однополостного гиперболоида (1).

Можно также добавить, что по тем же соображениям, что и для эллипсоида и двуполостного гиперболоида,

координатные плоскости являются плоскостями симметрии однополостного гиперболоида (1), *а начало координат — его центром симметрии.*

Упражнение. Докажите, что при $a > b$ никаких других плоскостей симметрии гиперболоид (1) не имеет.

4. Прямолинейные образующие однополостного гиперболоида

Обратим внимание на тот замечательный факт, что, несмотря на «искривленность», однополостный гиперболоид (1) п. 3 содержит целые прямые: например, прямую (4) п. 3. Более того, легко видеть, что этот гиперболоид содержит целое семейство таких прямых. Действительно, если гиперболоид является гиперболоидом вращения, то он переходит в себя при любом повороте вокруг оси Oz. Поэтому при этом повороте прямая (4) п. 3 переходит в некоторую прямую, также принадлежащую гиперболоиду. Таким образом, мы можем сказать, что гиперболоид вращения «заметается» (или «зачерчивается») некоторой прямой[1]), вращающейся вокруг оси Oz. При сжатии к плоскости Oxz прямые переходят, очевидно, в прямые, а вращение — в некоторое непрерывное преобразование пространства (описывать которое более подробно нам нет нужды). Следовательно, мы видим, что

произвольный однополостный гиперболоид заметается некоторой непрерывно перемещающейся прямой.

Определение 1. Поверхность в пространстве, являющаяся геометрическим местом точек некоторой непрерывно перемещающейся прямой, называется *линейчатой поверхностью.*

Таким образом, мы доказали, что

каждый однополостный гиперболоид является линейчатой поверхностью.

Определение 1 линейчатой поверхности является не столько строгим математическим определением, сколько наглядным описанием. Однако его можно без особого труда превратить во вполне строгое определение.

С этой целью мы заметим, что любая прямая в пространстве однозначно определена, если известна ее точка $M_0(x_0, y_0, z_0)$ и ее направляющий вектор $\boldsymbol{a}(l, m, n)$. Поэтому, чтобы задать некоторое непрерывное семейство прямых, зависящих, скажем, от параметра τ, пробегающего отрезок $[\tau_1, \tau_2]$, достаточно

---

[1]) То, что при этом получается весь гиперболоид, наглядно очевидно. Строгое доказательство будет дано ниже.

задать непрерывные функции

$$\begin{matrix} x_0(\tau), & y_0(\tau), & z_0(\tau), \\ l(\tau), & m(\tau), & n(\tau), \end{matrix} \tag{1}$$

первые три из которых для любого τ определяют точку на прямой семейства, соответствующей данному значению параметра, а остальные три — направляющий вектор этой прямой (при этом, конечно, требуется, чтобы $l(\tau)^2 + m(\tau)^2 + n(\tau)^2 \neq 0$). Тогда произвольная точка, принадлежащая некоторой прямой этого семейства, будет иметь координаты вида

$$\begin{aligned} x &= x_0(\tau) + l(\tau)\, t, \\ y &= y_0(\tau) + m(\tau)\, t, \\ z &= z_0(\tau) + n(\tau)\, t, \end{aligned} \tag{2}$$

где t — некоторое число (параметр на прямой). Теперь мы видим, что данное выше наглядное определение линейчатой поверхности может быть переформулировано следующим образом:

Поверхность называется линейчатой, если ее можно задать параметрическими уравнениями вида (2), *в которых параметр* t *изменяется от* $-\infty$ *до* $+\infty$, *а параметр* τ — *на некотором отрезке* $[\tau_1, \tau_2]$. *При этом, конечно, предполагается, что все функции* (1) *непрерывны и что* $l(\tau)^2 + m(\tau)^2 + n(\tau)^2 \neq 0$ *ни при одном* $\tau \in [\tau_0, \tau_1]$.

Определение 2. Прямые, из которых состоит данная линейчатая поверхность, называются ее *прямолинейными образующими.* Уравнение каждой прямолинейной образующей получается из уравнений (2) при некотором фиксированном τ.

Таким образом, уравнения (2) мы можем рассматривать не только как уравнения линейчатой поверхности, но и как уравнения произвольной прямолинейной образующей этой поверхности (в зависимости от того, считаем ли мы параметр τ изменяющимся или фиксированным).

С наглядной точки зрения прямолинейные образующие — это мгновенные положения прямой, перемещением которой образована данная линейчатая поверхность.

Замечание 1. Подчеркнем, что понятие «прямолинейной образующей» связано со способом построения поверхности (или, иными словами, с выбором ее уравнения (2)). Может случиться, что та же поверхность будет образована движением совсем другой прямой, т. е. будет геометрическим местом точек совсем другой системы прямолинейных образующих. В этом случае иногда говорят, что рассматриваемая поверхность *дважды линейчата.* (Мыслимы, конечно, и поверхности, линейчатые трижды, четырежды и т. д.) Возможно также, что поверхность (не обязательно линейчатая) содержит некоторую прямую, не включающуюся ни в какое семейство прямолинейных образующих. Такую прямую прямолинейной образующей считать, конечно, нельзя.

Тем не менее во многих учебниках аналитической геометрии прямолинейная образующая определяется просто как прямая, принадлежащая поверх-

иости. Это может быть оправдано тем, что для интересующих аналитическую геометрию поверхностей второго порядка каждая прямая, принадлежащая поверхности, является ее прямолинейной образующей (но уже для поверхностей третьего порядка это не так).

Наше доказательство того, что однополостный гиперболоид является линейчатой поверхностью, было, по существу, чисто синтетическим. Однако его легко можно алгебраизировать, и тем самым не только заново доказать линейчатость однополостного гиперболоида, но и найти его параметрические уравнения вида (2), т. е., иными словами, найти параметрические уравнения его прямолинейных образующих.

Для гиперболоида вращения (2) п. 3 прямая (4) п. 3 (при $h = b$) имеет, очевидно, параметрические уравнения вида

$$x = at,$$
$$y = b,$$
$$z = ct.$$

При повороте вокруг оси Oz на угол τ, сопровождаемом сжатием к плоскости Oxz с коэффициентом $\frac{b}{a}$, эта прямая, как можно показать, переходит в прямую с уравнениями

$$\begin{aligned} x &= -a\sin\tau + a\cos\tau \cdot t, \\ y &= b\cos\tau + b\sin\tau \cdot t, \\ z &= ct. \end{aligned} \qquad (3)$$

Мы не будем этого доказывать, а непосредственно проверим, что при любом τ прямая (3) целиком расположена на гиперболоиде (1), т. е. что при любых t и τ числа x, y, z, даваемые уравнениями (3), удовлетворяют уравнению (1) п. 3 однополостного гиперболоида. Но, действительно,

$$\frac{(-a\sin\tau + a\cos\tau \cdot t)^2}{a^2} + \frac{(b\cos\tau + b\sin\tau \cdot t)^2}{b^2} - \frac{(ct^2)}{c^2} = 1.$$

Чтобы доказать, что уравнения (3) на самом деле определяют семейство прямолинейных образующих гиперболоида, остается только показать, что для любой его точки $M(x, y, z)$ можно подобрать такие (однозначно определенные) значения τ и t, что ее координаты имеют вид (3). Другими словами, нужно показать, что уравнения (3) однозначно разрешимы относительно τ и t (при условии, что x, y и z удовлетворяют уравнению (1) п. 3).

Но последнее уравнение (3) однозначно определяет t:

$$t = \frac{z}{c}.$$

Подставляя это значение в первые два уравнения (3), мы получаем уравнения

$$\frac{x}{a} = -\sin\tau + \cos\tau \cdot \frac{z}{c},$$

$$\frac{y}{b} = \cos\tau + \sin\tau \cdot \frac{z}{c}.$$

Следовательно,

$$\cos\tau = \frac{\frac{xz}{ac} + \frac{y}{b}}{1 + \frac{z^2}{c^2}}, \qquad \sin\tau = \frac{\frac{yz}{bc} - \frac{x}{a}}{1 + \frac{z^2}{c^2}}.$$

Поскольку

$$\left(\frac{\frac{xz}{ac} + \frac{y}{b}}{1 + \frac{z^2}{c^2}}\right)^2 + \left(\frac{\frac{yz}{bc} - \frac{x}{a}}{1 + \frac{z^2}{c^2}}\right)^2 = \frac{\frac{x^2}{a^2}\left(1 + \frac{z^2}{c^2}\right) + \frac{y^2}{b^2}\left(1 + \frac{z^2}{c^2}\right)}{\left(1 + \frac{z^2}{c^2}\right)^2} = 1,$$

эти соотношения однозначно определяют угол τ.

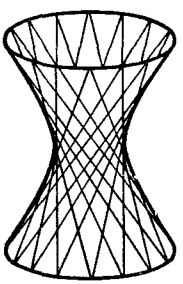
Тем самым мы доказали, что

уравнения (3) *являются уравнениями некоторого семейства прямолинейных образующих однополостного гиперболоида, т. е. при любом* τ *они представляют собой уравнения некоторой прямолинейной образующей гиперболоида* (1), п. 3, *и любая прямолинейная образующая* (*принадлежащая рассматриваемому семейству*) *получается таким образом.*

Иначе можно сказать, что

уравнения (3) *являются параметрическими уравнениями однополостного гиперболоида, имеющими вид* (2).

Заметим теперь, что вместо того, чтобы исходить из прямой (4) п. 3, мы можем с тем же правом отправляться от прямой (5) п. 3 (при $h = b$). В результате мы получим другую систему прямолинейных образующих однополостного гиперболоида с уравнениями



$$\begin{aligned} x &= a\sin\tau + a\cos\tau \cdot t, \\ y &= -b\cos\tau + b\sin\tau \cdot t, \\ z &= ct. \end{aligned} \tag{4}$$

Задание. Докажите это утверждение.

Предложение 1. *Однополостный гиперболоид является дважды линейчатой поверхностью: одно семейство его прямолинейных образующих имеет вид* (3), *а другое — вид* (4).

Доказательство. В свете сказанного выше нам остается только доказать, что не существует третьего семейства прямолинейных образующих, т. е., что

через любую точку однополостного гиперболоида (1) п. 3 *проходят точно две прямые, целиком лежащие на этом гипер-*

болоиде: одна, являющаяся прямолинейной образующей из семейства (3), *а другая, являющаяся прямолинейной образующей из семейства* (4).

Пусть $M(x_0, y_0, z_0)$ — произвольная точка однополостного гиперболоида (1) п. 3 и пусть

$$x = x_0 + lt,$$
$$y = y_0 + mt,$$
$$z = z_0 + nt$$

— произвольная прямая, проходящая через эту точку. Чтобы эта прямая целиком принадлежала гиперболоиду, необходимо и достаточно, чтобы тождественно по t имело место равенство

$$\frac{(x_0 + lt)^2}{a^2} + \frac{(y_0 + mt)^2}{b^2} - \frac{(z_0 + nt)^2}{c^2} = 1,$$

т. е. равенство

$$\left(\frac{l^2}{a^2} + \frac{m^2}{b^2} - \frac{n^2}{c^2}\right)t^2 + 2\left(\frac{lx_0}{a^2} + \frac{my_0}{b^2} - \frac{nz_0}{c^2}\right)t = 0,$$

что возможно тогда и только тогда, когда

$$\frac{l^2}{a^2} + \frac{m^2}{b^2} - \frac{n^2}{c^2} = 0$$

и

$$\frac{lx_0}{a^2} + \frac{my_0}{b^2} - \frac{nz_0}{c^2} = 0. \tag{5}$$

Из первого уравнения мы видим, что $n \neq 0$ (ибо, в противном случае, $l = m = 0$, что невозможно). Поскольку числа l, m, n нам заданы только с точностью до пропорциональности, мы, следовательно, без ограничения общности можем считать, что $n = c$. Но тогда для l и m мы получим уравнение

$$\frac{l^2}{a^2} + \frac{m^2}{b^2} = 1,$$

общее решение которого имеет вид

$$\begin{aligned} l &= a\cos\tau, \\ m &= b\sin\tau, \end{aligned} \qquad 0 \leqslant \tau < 2\pi.$$

Таким образом, параметрические уравнения рассматриваемой прямой мы можем записать в следующем виде:

$$\begin{aligned} x &= x_0 + a\cos\tau \cdot t, \\ y &= y_0 + b\sin\tau \cdot t, \\ z &= z_0 + ct, \end{aligned} \tag{6}$$

а соотношение (5) (которое мы еще не использовали) — в виде

$$\frac{x_0}{a}\cos\tau + \frac{y_0}{b}\sin\tau = \frac{z_0}{c}. \tag{7}$$

Перейдем теперь в уравнениях (6) от параметра t к параметру

$$t' = t + \frac{z_0}{c};$$

Тогда мы получим уравнения

$$\begin{aligned} x &= a\left(\frac{x_0}{a} - \cos\tau \cdot \frac{z_0}{c}\right) + a\cos\tau \cdot t', \\ y &= b\left(\frac{y_0}{b} - \sin\tau \cdot \frac{z_0}{c}\right) + b\sin\tau \cdot t', \\ z &= ct'. \end{aligned} \tag{8}$$

Но согласно соотношениям (7)

$$\cos\tau \cdot \left(\frac{x_0}{a} - \cos\tau \cdot \frac{z_0}{c}\right) = -\frac{y_0}{b}\sin\tau + (1 - \cos^2\tau)\frac{z_0}{c} =$$
$$= -\sin\tau \cdot \left(\frac{y_0}{b} - \sin\tau \cdot \frac{z_0}{c}\right),$$

откуда вытекает, что

$$\frac{x_0}{a} - \cos\tau \cdot \frac{z_0}{c} = -u\sin\tau, \qquad \frac{y_0}{b} - \sin\tau \cdot \frac{z_0}{c} = u\cos\tau, \tag{9}$$

где u — некоторое число. Поэтому уравнения (8) приобретают вид

$$\begin{aligned} x &= -au\sin\tau + a\cos\tau \cdot t', \\ y &= bu\cos\tau + b\sin\tau \cdot t', \\ z &= ct'. \end{aligned}$$

Сравнивая эти уравнения с уравнениями (3) и (4), мы немедленно замечаем, что при $u = \pm 1$ они совпадают с этими уравнениями (с точностью до обозначения параметра t'). Поэтому для завершения доказательства нам остается лишь показать, что $u = \pm 1$, т. е. что $u^2 = 1$.

Но из соотношений (9) непосредственно вытекает, что

$$u^2 = \left(\frac{x_0}{a} - \cos\tau \cdot \frac{z_0}{c}\right)^2 + \left(\frac{y_0}{b} - \sin\tau \cdot \frac{z_0}{c}\right)^2,$$

т. е. что

$$u^2 = \frac{x_0^2}{a^2} + \frac{y_0^2}{b^2} + \frac{z_0^2}{c^2} - 2\frac{z_0}{c}\left(\frac{x_0}{a}\cos\tau + \frac{y_0}{b}\sin\tau\right).$$

Учитывая соотношения (7) и тот факт, что точка $M_0(x_0, y_0, z_0)$ принадлежит гиперболоиду (1) п. 3, мы получаем, следовательно, что

$$u^2 = \frac{x_0^2}{a^2} + \frac{y_0^2}{b^2} - \frac{z_0^2}{c^2} = 1.$$

Тем самым предложение 1 полностью доказано.

Замечание 2. Обратим внимание на то, что мы фактически заново вывели уравнения (3) и (4) прямолинейных образующих однополостного гиперболоида. Этот вывод уравнений (3) и (4) более громоздок, но он доказывает, что никаких других прямолинейных образующих однополостный гиперболоид не имеет.

Замечание 3. Если не гнаться за параметрическими уравнениями прямолинейных образующих (а также за доказательством отсутствия третьего семейства прямолинейных образующих), то можно предложить очень короткий, хотя и искусственный, вывод их непараметрических уравнений (что одновременно еще раз докажет линейчатость однополостного гиперболоида).

Очевидно, что уравнение однополостного гиперболоида можно записать в следующем виде:

$$\left(\frac{x}{a}-\frac{z}{c}\right)\left(\frac{x}{a}+\frac{z}{c}\right)=\left(1-\frac{y}{b}\right)\left(1+\frac{y}{b}\right). \tag{10}$$

Это наводит на мысль рассмотреть для произвольного числа $k \neq 0$ прямую с уравнениями

$$\begin{aligned} \frac{x}{a}-\frac{z}{c} &= k\left(1-\frac{y}{b}\right), \\ \frac{x}{a}+\frac{z}{c} &= \frac{1}{k}\left(1+\frac{y}{b}\right). \end{aligned} \tag{11}$$

Ясно, что уравнение (10) является следствием уравнений (11). Это означает, что любая точка прямой (11) принадлежит гиперболоиду (10). Кроме того, для любой точки $M_0(x_0, y_0, z_0)$ гиперболоида (10) прямая (11) с

$$k=\frac{\frac{x_0}{a}-\frac{z_0}{c}}{1-\frac{y_0}{b}} \tag{12}$$

проходит через эту точку. Этим доказано, что

прямые (11) *составляют семейство прямолинейных образующих однополостного гиперболоида.*

Замечание 4. При $\frac{x_0}{a}-\frac{z_0}{c}=0$ формула (12) дает для k значение 0, а при $1-\frac{y_0}{b}=0$ — «значение» ∞. Ясно, что этим случаям отвечают прямые

$$\frac{x}{a}-\frac{z}{c}=0, \qquad 1-\frac{y}{b}=0,$$

и

$$1+\frac{y}{b}=0, \qquad \frac{x}{a}+\frac{z}{c}=0,$$

которые мы также должны считать принадлежащими нашему семейству.

Чтобы получить этим методом второе семейство прямолинейных образующих гиперболоида, следует рассмотреть

прямые

$$\frac{x}{a}-\frac{z}{c}=k\left(1+\frac{y}{b}\right),$$
$$\frac{x}{a}+\frac{z}{c}=\frac{1}{k}\left(1-\frac{y}{b}\right). \tag{13}$$

Задание. Докажите, что прямые (13) действительно являются прямолинейными образующими однополостного гиперболоида, отличными от образующих (11).

Уравнения (11) и (13) не очень удобны для исследования взаимного расположения прямолинейных образующих. Поэтому мы будем пользоваться исключительно параметрическими уравнениями (3) и (4). Впрочем, целесообразно и эти уравнения несколько преобразовать.

Каждая прямая (3) пересекает плоскость $z=0$, т. е. горловой эллипс

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$

гиперболоида, в точке с координатами

$$x_0=-a\sin\tau, \quad y_0=b\cos\tau.$$

Поэтому

$$a\cos\tau=\frac{a}{b}y_0, \qquad b\sin\tau=-\frac{b}{a}x_0.$$

Следовательно, уравнения (3) мы можем переписать в следующем виде:

$$\begin{aligned} x&=x_0+\frac{a}{b}y_0t,\\ y&=y_0-\frac{b}{a}x_0t,\\ z&=ct. \end{aligned} \tag{14}$$

Тем самым доказано, что

уравнения (14) *представляют собой параметрические уравнения некоторой прямолинейной образующей однополостного гиперболоида, проходящей через точку* (x_0, y_0) *горлового эллипса.*

Аналогично показывается, что

параметрические уравнения другой прямолинейной образующей, проходящей через точку (x_0, y_0) *горлового эллипса, имеют вид*

$$\begin{aligned} x&=x_0-\frac{a}{b}y_0t,\\ y&=y_0+\frac{b}{a}x_0t,\\ z&=ct. \end{aligned} \tag{15}$$

Задание. Выведите уравнения (15).

Теперь у нас все готово для доказательства основных свойств прямолинейных образующих однополостного гиперболоида.

Свойство 1. *Через любую точку гиперболоида проходит одна и только одна прямолинейная образующая каждого семейства.*

Это свойство мы уже доказывали (и притом трижды). Здесь мы его привели лишь для полноты списка свойств.

Свойство 2. *Любые две образующие однополостного гиперболоида, принадлежащие разным семействам, компланарны (лежат в одной плоскости).*

Действительно, пусть образующая первого семейства проходит через точку (x_0, y_0) горлового эллипса, а образующая второго семейства — через точку (x_1, y_1). Тогда согласно уравнениям (14) и (15) направляющий вектор первой образующей имеет вид

$$\left(\frac{a}{b}y_0,\ -\frac{b}{a}x_0,\ c\right),$$

а направляющий вектор второй образующей — вид

$$\left(-\frac{a}{b}y_1,\ \frac{b}{a}x_1,\ c\right).$$

Но мы знаем (п. 2 § 3 гл. 3), что две прямые, имеющие соответственно направляющие векторы (l_0, m_0, n_0) и (l_1, m_1, n_1) и проходящие через точки (x_0, y_0, z_0) и (x_1, y_1, z_1), тогда и только тогда компланарны, когда определитель

$$\begin{vmatrix} x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ l_0 & m_0 & n_0 \\ l_1 & m_1 & n_1 \end{vmatrix}$$

равен нулю. В нашем случае этот определитель имеет вид

$$\begin{vmatrix} x_1 - x_0 & y_1 - y_0 & 0 \\ \frac{a}{b}y_0 & -\frac{b}{a}x_0 & c \\ -\frac{a}{b}y_1 & \frac{b}{a}x_1 & c \end{vmatrix} = abc\left[\left(\frac{x_0^2}{a^2} + \frac{y_0^2}{b^2}\right) - \left(\frac{x_1^2}{a^2} + \frac{y_1^2}{b^2}\right)\right]$$

и, следовательно, на самом деле равен нулю (поскольку $\frac{x_0^2}{a^2} + \frac{y_0^2}{b^2} = 1$ и $\frac{x_1^2}{a^2} + \frac{y_1^2}{b^2} = 1$).

Тем самым свойство 2 полностью доказано.

Можно еще заметить, что рассмотренные образующие тогда и только тогда параллельны, когда

$$\frac{\frac{a}{b}y_0}{-\frac{a}{b}y_1} = \frac{-\frac{b}{a}x_0}{\frac{b}{a}x_1} = \frac{c}{c},$$

т. е. когда

$$x_1 = -x_0, \quad y_1 = -y_0.$$

Это означает, что

две прямолинейные образующие однополостного гиперболоида, принадлежащие разным семействам, тогда и только тогда параллельны, когда они проходят через диаметрально противоположные точки горлового эллипса.

Рассмотрим теперь образующие, принадлежащие одному семейству.

Свойство 3. *Две образующие однополостного гиперболоида, принадлежащие одному семейству, скрещиваются (не лежат в одной плоскости).*

Действительно, пусть (x_0, y_0) и (x_1, y_1) — точки горлового эллипса, через которые проходят данные образующие. Ясно, что эти точки различны, поскольку через любую точку горлового эллипса проходит только одна образующая каждого семейства. Пусть, для определенности, рассматриваемые образующие принадлежат семейству (14). Тогда их направляющие векторы имеют вид

$$\left(\frac{a}{b}y_0, -\frac{b}{a}x_0, c\right), \quad \left(\frac{a}{b}y_1, -\frac{b}{a}x_1, c\right), \tag{16}$$

и потому условие их скрещивания состоит в том, чтобы не был равен нулю определитель

$$\begin{vmatrix} x_1 - x_0 & y_1 - y_0 & 0 \\ \frac{a}{b}y_0 & -\frac{b}{a}x_0 & c \\ \frac{a}{b}y_1 & -\frac{b}{a}x_1 & c \end{vmatrix} = c\left[\frac{b}{a}(x_1 - x_0)^2 + \frac{a}{b}(y_1 - y_0)^2\right].$$

Но поскольку точки (x_0, y_0) и (x_1, y_1) различны, этот определитель на самом деле отличен от нуля.

Для прямолинейных образующих (15) доказательство аналогично.

Замечание 5. Свойство 3 можно доказать без всяких выкладок, если заметить, что прямолинейные образующие одного семейства не могут пересекаться (ибо тогда вопреки свойству 1 через некоторую точку гиперболоида проходили бы две образующие этого семейства). Если же они параллельны, т. е. векторы (16) коллинеарны, то

$$\frac{\frac{a}{b}y_0}{\frac{a}{b}y_1} = \frac{-\frac{b}{a}x_0}{-\frac{b}{a}x_1} = \frac{c}{c},$$

и потому $x_0 = x_1$, $y_0 = y_1$ и, значит, эти образующие совпадают.

Свойство 4. *Никакие три образующие одного семейства прямолинейных образующих однополостного гиперболоида не параллельны одной и той же плоскости.*

Действительно, пусть (x_0, y_0), (x_1, y_1) и (x_2, y_2) — точки горлового эллипса, через которые проходят данные три образующие одного и того же семейства. Согласно свойству 1 эти точки попарно различны, а согласно формулам (14) и (15) направляющие векторы этих образующих имеют вид

$$\left(\pm\frac{a}{b}y_0,\quad \mp\frac{b}{a}x_0,\quad c\right),\quad \left(\pm\frac{a}{b}y_1,\quad \mp\frac{b}{a}x_1,\quad c\right),\quad \left(\pm\frac{a}{b}y_2,\quad \pm\frac{b}{a}x_2,\quad c\right). \tag{17}$$

Поскольку три прямые тогда и только тогда параллельны одной плоскости, когда их направляющие векторы компланарны, для доказательства свойства 4 достаточно доказать, что векторы (17) не компланарны, т. е. определитель, составленный из их координат, не равен нулю. Но этот определитель равен

$$c\begin{vmatrix} 1 & x_0 & y_0 \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{vmatrix}$$

и потому на самом деле отличен от нуля (ибо точки (x_0, y_0), (x_1, y_1) и (x_2, y_2) не коллинеарны).

Замечание 6. Из доказанных свойств прямолинейных образующих вытекает красивое геометрическое описание произвольного однополостного гиперболоида.

Возьмем три произвольных прямолинейных образующих однополостного гиперболоида, принадлежащих одному семейству. Согласно свойству 2 любая образующая другого семейства пересекает каждую из этих образующих в некоторой точке (или ей параллельна, т. е. пересекает ее в несобственной точке). Но ясно, что если нам даны три скрещивающиеся прямые, то существует только одна прямая, проходящая через данную точку одной прямой и пересекающая две другие прямые (возможно, в несобственной точке). Следовательно, прямолинейные образующие некоторого семейства однозначно характеризуются как прямые, пересекающие три произвольно выбранные образующие другого семейства. Это означает, что

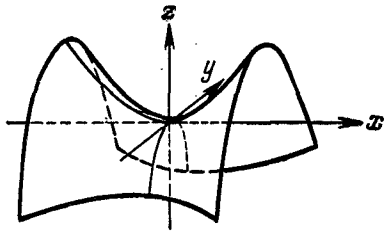
произвольный однополостный гиперболоид представляет собой геометрическое место точек, принадлежащих прямым, пересекающим (возможно, в несобственной точке) каждую из трех фиксированных скрещивающихся прямых, не параллельных никакой плоскости.

Упражнение. Докажите, что

какие бы три скрещивающиеся прямые, не параллельные ни одной плоскости, мы ни взяли, геометрическое место точек, принадлежащих прямым, пересекающим эти прямые, является однополостным гиперболоидом.

5. Гиперболические параболоиды

Определение 1. *Гиперболическим параболоидом* называется поверхность, имеющая в некоторой прямоугольной системе координат уравнение вида



$$\frac{x^2}{p} - \frac{y^2}{q} = 2z, \qquad (1)$$

где $p > 0$, $q > 0$. Эта система координат называется *канонической* (относительно данного параболоида), а уравнение (1) — *каноническим уравнением.*

Представить себе форму гиперболического параболоида довольно трудно. Проще всего это можно сделать методом плоских сечений.

Плоскость $z = h$ пересекает параболоид (1) по линии с уравнением

$$\frac{x^2}{p} - \frac{y^2}{q} = 2h,$$

т. е. либо с уравнением

$$\frac{x^2}{(\sqrt{2hp})^2} - \frac{y^2}{(\sqrt{2hq})^2} = 1$$

(если $h > 0$), либо с уравнением

$$\frac{x^2}{p} - \frac{y^2}{q} = 0$$

(если $h = 0$), либо, наконец, с уравнением

$$\frac{y^2}{(\sqrt{-2qh})^2} - \frac{x^2}{(\sqrt{-2ph})^2} = 1$$

(если $h < 0$).

В первом случае эта линия является гиперболой с полуосями $\sqrt{2hp}$, $\sqrt{2hq}$, во втором случае она состоит из двух прямых

$$\frac{x}{\sqrt{p}} - \frac{y}{\sqrt{q}} = 0 \quad \text{и} \quad \frac{x}{\sqrt{p}} + \frac{y}{\sqrt{q}} = 0,$$

а в третьем случае — снова является гиперболой, но уже с полуосями $\sqrt{-2hq}$, $\sqrt{-2hp}$.

Таким образом,

при $h<0$ плоскость $z=h$ пересекает гиперболический параболоид (1) *по гиперболе с полуосями*

$$\sqrt{-2hq},\quad \sqrt{-2hp},$$

монотонно убывающими от $+\infty$ до нуля, когда h возрастает от $-\infty$ до нуля;

при $h=0$ плоскость $z=h$ пересекает гиперболический параболоид (1) *по паре пересекающихся прямых*

$$\frac{x}{\sqrt{p}}-\frac{y}{\sqrt{q}}=0,\quad z=0 \qquad \text{и} \qquad \frac{x}{\sqrt{p}}+\frac{y}{\sqrt{q}}=0,\quad z=0;$$

при $h>0$ плоскость $z=h$ пересекает гиперболический параболоид (1) *по гиперболе с полуосями*

$$\sqrt{2hp},\quad \sqrt{2hq},$$

монотонно возрастающими от нуля до $+\infty$, когда h возрастает от нуля до $+\infty$.

Обратим внимание на то, что действительные (мнимые) оси гипербол, получающихся при $h<0$, параллельны мнимым (действительным) осям гипербол, получающихся при $h>0$.

Плоскость $y=h$ пересекает параболоид (1) по линии с уравнением

$$\frac{x^2}{p}-\frac{h^2}{q}=2z,$$

т. е. с уравнением

$$x^2=2p\left(z+\frac{h^2}{2q}\right).$$

Это означает, что

плоскость $y=h$ пересекает гиперболический параболоид (1) *по параболе с фокальным параметром p, с вершиной в точке $\left(0,\ h,\ -\frac{h^2}{2q}\right)$ и с «рогами», направленными в положительном направлении оси Oz («вверх»).*

Аналогично показывается, что

плоскость $x=h$ пересекает гиперболический параболоид (1) *по параболе с фокальным параметром q, с вершиной в точке $\left(h,\ 0,\ \frac{h^2}{2p}\right)$ и с «рогами», направленными в отрицательном направлении оси Oz («вниз»).*

Отметим, что вершины парабол, получающихся в пересечении с плоскостями $y=h$ (плоскостями $x=h$), принадлежат параболе, получающейся в пересечении с плоскостью $x=0$ (соответственно с плоскостью $y=0$).

При известном усилии пространственного воображения теперь уже можно представить себе форму гиперболического параболоида. Он «поднимается» в положительном и отрицательном направлениях оси Ox и «опускается» в положительном и отрицательном направлениях оси Oy. Вблизи начала координат он имеет форму седла.

Поскольку уравнение (1) не меняется при изменении знаков у x и y (но не у z!),

координатные плоскости Oxz и Oyz являются плоскостями симметрии гиперболического параболоида.

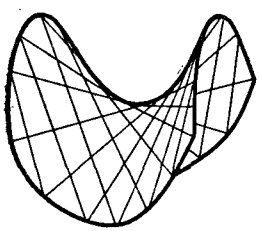
Упражнение. Докажите, что кроме плоскостей Oxz и Oyz других плоскостей симметрии гиперболический параболоид не имеет. Не имеет он и центра симметрии.

Предложение 1. *Гиперболический параболоид является дважды линейчатой поверхностью.*

Доказательство. Пусть $M_0(x_0, y_0, z_0)$ — произвольная точка параболоида (1) и

$$\begin{aligned} x &= x_0 + lt, \\ y &= y_0 + mt, \\ z &= z_0 + nt \end{aligned}$$

— произвольная прямая, проходящая через точку $M_0(x_0, y_0, z_0)$. Чтобы эта прямая целиком принадлежала параболоиду, необходимо и достаточно, чтобы тождественно по t выполнялось равенство



$$\frac{(x_0 + lt)^2}{p} - \frac{(y_0 + mt)^2}{q} = 2(z_0 + nt),$$

т. е. равенство

$$\left(\frac{l^2}{p} - \frac{m^2}{q}\right) t^2 + 2\left(\frac{lx_0}{p} - \frac{my_0}{q} - n\right) t = 0,$$

что возможно только тогда, когда

$$\frac{l^2}{p} - \frac{m^2}{q} = 0 \quad \text{и} \quad \frac{lx_0}{p} - \frac{my_0}{q} - n = 0.$$

Отсюда мы находим:

$$m = \pm\sqrt{\frac{q}{p}}\, l, \quad n = \left(\frac{x_0}{p} \mp \frac{y_0}{\sqrt{pq}}\right) l.$$

Поскольку величины l, m и n нам заданы только с точностью до пропорциональности, мы без ограничения общности можем считать, что $l = \sqrt{p}$. Тем самым доказано, что

прямая, проходящая через точку $M(x_0, y_0, z_0)$ гиперболического параболоида (1), *тогда и только тогда целиком принад-*

лежит этому параболоиду, когда она может быть задана параметрическими уравнениями либо вида

$$\begin{aligned} x &= x_0 + \sqrt{p}\,t, \\ y &= y_0 + \sqrt{q}\,t, \\ z &= z_0 + \left(\frac{x_0}{\sqrt{p}} - \frac{y_0}{\sqrt{q}}\right)t, \end{aligned} \tag{2}$$

либо вида

$$\begin{aligned} x &= x_0 + \sqrt{p}\,t, \\ y &= y_0 - \sqrt{q}\,t, \\ z &= z_0 + \left(\frac{x_0}{\sqrt{p}} + \frac{y_0}{\sqrt{q}}\right)t. \end{aligned} \tag{3}$$

Займемся сначала уравнениями (2). Чтобы преобразовать их к более удобному виду, мы заметим, что точке пересечения прямой (2) с плоскостью $z = 0$ отвечает значение параметра

$$t_0 = -\frac{z_0}{\frac{x_0}{\sqrt{p}} - \frac{y_0}{\sqrt{q}}} = -\frac{1}{2}\left(\frac{x_0}{\sqrt{p}} + \frac{y_0}{\sqrt{q}}\right).$$

Поэтому, если мы перейдем к новому параметру $t' = t - t_0$, то уравнения (2) перейдут в уравнения

$$\begin{aligned} x &= \left(x_0 - \frac{\sqrt{p}}{2}\left(\frac{x_0}{\sqrt{p}} + \frac{y_0}{\sqrt{q}}\right)\right) + \sqrt{p}\,t', \\ y &= \left(y_0 - \frac{\sqrt{q}}{2}\left(\frac{x_0}{\sqrt{p}} + \frac{y_0}{\sqrt{q}}\right)\right) + \sqrt{q}\,t', \\ z &= \left(\frac{x_0}{\sqrt{p}} - \frac{y_0}{\sqrt{q}}\right)t', \end{aligned}$$

т. е. в уравнения (мы опускаем в обозначении параметра штрих)

$$\begin{aligned} x &= \frac{\sqrt{p}}{2}\left(\frac{x_0}{\sqrt{p}} - \frac{y_0}{\sqrt{q}}\right) + \sqrt{p}\,t, \\ y &= -\frac{\sqrt{q}}{2}\left(\frac{x_0}{\sqrt{p}} - \frac{y_0}{\sqrt{q}}\right) + \sqrt{q}\,t, \\ z &= \left(\frac{x_0}{\sqrt{p}} - \frac{y_0}{\sqrt{q}}\right)t. \end{aligned}$$

Следовательно, полагая

$$\tau = \frac{1}{2}\left(\frac{x_0}{\sqrt{p}} - \frac{y_0}{\sqrt{q}}\right), \tag{4}$$

мы окончательно получим из уравнений (2) уравнения

$$\begin{aligned} x &= \sqrt{p}(t+\tau), \\ y &= \sqrt{q}(t-\tau), \\ z &= 2\tau t. \end{aligned} \tag{5}$$

Таким образом, мы показали, что при некоторых значениях параметра τ (а именно, при значениях, представимых формулой (4)) уравнения (5) являются параметрическими уравнениями прямой, целиком принадлежащей параболоиду (1). При этом для любой точки $M_0(x_0, y_0, z_0)$ параболоида существует (единственное) значение параметра τ (а именно, значение, даваемое формулой (4)), при котором прямая (5) проходит через точку $M_0(x_0, y_0, z_0)$. Все это означает, что

уравнения (5) *являются параметрическими уравнениями некоторого семейства прямолинейных образующих гиперболического параболоида* (1).

Выясним теперь возможные значения параметра τ (и заодно его геометрический смысл).

Пусть сначала $\tau \neq 0$. Тогда прямая (5) пересекает плоскость $z = 0$ в точке, отвечающей значению параметра $t = 0$, т. е. в точке с координатами $(\sqrt{p}\,\tau, -\sqrt{q}\,\tau, 0)$. Но ясно, что эта точка принадлежит прямой

$$\left.\begin{aligned} \frac{x}{\sqrt{p}} + \frac{y}{\sqrt{q}} &= 0, \\ z &= 0 \end{aligned}\right\} \tag{6}$$

(лежащей, как мы знаем, на параболоиде (1); см. выше) и число τ является координатой этой точки на прямой (6) в координатной системе с репером Oe, где O — точка $(0, 0, 0)$, а e — направляющий вектор $(\sqrt{p}, -\sqrt{q}, 0)$. При $\tau = 0$ прямая (5) целиком расположена в плоскости $z = 0$ (она является первой из указанных выше прямых, по которым плоскость $z = 0$ пересекает параболоид (1)) и пересекает прямую (6) в точке $(0, 0, 0)$. Поэтому и в этом случае точка пересечения прямых (5) и (6) имеет координату τ в координатной системе на прямой (6) с репером Oe.

Возьмем теперь на прямой (6) произвольную точку M_0 и проведем через нее образующую (5). По только что сказанному параметр τ, соответствующий этой образующей, будет равен координате точки M_0 на прямой (6) в реперe Oe. Так как эта координата может принимать любые значения, то

в уравнениях (5) *параметр* τ *может принимать любые значения от* $-\infty$ *до* $+\infty$.

Одновременно мы доказали, что

гиперболический параболоид (1) *заметается движущейся прямой, все время пересекающей фиксированную прямую* (6).

Чтобы найти второе семейство прямолинейных образующих гиперболического параболоида, можно было бы повторить шаг за шагом все наши рассуждения, но отправляясь не от прямой (2), а от прямой (3). Однако теперь можно поступить проще и написать по аналогии с уравнениями (5) уравнения

$$\begin{aligned} x &= \sqrt{p}\,(\tau + t), \\ y &= \sqrt{q}\,(\tau - t), \\ z &= 2\tau t. \end{aligned} \tag{7}$$

При любых τ и t числа x, y, z, получающиеся из этих уравнений, удовлетворяют уравнению (1) параболоида, ибо

$$(\tau + t)^2 - (\tau - t)^2 = 4\tau t.$$

Поэтому при любом τ, $-\infty < \tau < +\infty$, уравнения (7) являются параметрическими уравнениями некоторой прямой, целиком лежащей на параболоиде. Следовательно, чтобы доказать, что уравнения (7) являются уравнениями прямолинейных образующих, остается только проверить, что для любой точки $M_0(x_0, y_0, z_0)$ параболоида можно подобрать такое τ, что соответствующая прямая (7) проходит через эту точку, т. е. что координаты x_0, y_0, z_0 получаются при некотором значении параметра t. Другими словами, нужно доказать, что для любой точки $M_0(x_0, y_0, z_0)$ параболоида (1) уравнения (7) (в которых положено $x = x_0$, $y = y_0$, $z = z_0$) однозначно разрешимы относительно τ и t. Но это очевидно: решение дается формулами

$$\tau = \frac{1}{2}\left(\frac{x_0}{\sqrt{p}} + \frac{y_0}{\sqrt{q}}\right), \quad t = \frac{1}{2}\left(\frac{x_0}{\sqrt{p}} - \frac{y_0}{\sqrt{q}}\right).$$

Тем самым мы доказали, что

уравнения (7) *также являются параметрическими уравнениями некоторого семейства прямолинейных образующих гиперболического параболоида* (1).

Задание. Проверьте, что каждая прямая (7) пересекает прямую

$$\begin{aligned} \frac{x}{\sqrt{p}} - \frac{y}{\sqrt{q}} &= 0, \\ z &= 0 \end{aligned} \tag{8}$$

и что параметр τ является не чем иным, как координатой точки пересечения прямых (7) и (8) в координатной системе на прямой (8) с репером Oe', где O, как и выше, — точка $(0, 0, 0)$, а e' — направляющий вектор $(\sqrt{p}, \sqrt{q}, 0)$ прямой (8).

Поскольку найденные два семейства прямолинейных образующих явно различны (направляющие векторы $(\sqrt{p}, -\sqrt{q}, 2\tau)$ и $(\sqrt{p}, -\sqrt{q}, 2\tau)$ этих образующих даже не коллинеарны) и поскольку третьего семейства прямолинейных образующих не существует (ибо по доказанному в начале этого пункта

через любую точку параболоида (1) проходят только две прямые, целиком лежащие на этом параболоиде), тем самым предложение 1 полностью доказано.

Замечание 1. Если не доказывать отсутствия третьего семейства прямолинейных образующих, то все сказанное выше можно существенно сократить, начав сразу же с уравнений (5) и доказав, что они дают прямолинейные образующие (так же как мы это выше сделали для уравнений (7)). Правда, при этом остается неясным, как до этих уравнений догадаться. (Впрочем, догадаться на самом деле совсем не сложно, если помнить формулу $(\tau + t)^2 - (\tau - t)^2 = 4\tau t$.)

Очень просто уравнения обоих семейств прямолинейных образующих (правда, в непараметрическом виде и без доказательства отсутствия третьего семейства) можно получить (ср. п. 4), представив уравнение (1) в виде

$$\left(\frac{x}{\sqrt{p}} + \frac{y}{\sqrt{q}}\right)\left(\frac{x}{\sqrt{p}} - \frac{y}{\sqrt{q}}\right) = 2z.$$

Тогда ясно, что любая прямая с уравнениями

$$\begin{aligned} \frac{x}{\sqrt{p}} + \frac{y}{\sqrt{q}} &= 2k, \\ \frac{x}{\sqrt{p}} - \frac{y}{\sqrt{q}} &= \frac{z}{k}, \end{aligned} \tag{9}$$

где $k \neq 0$, а также прямая с уравнениями

$$\begin{aligned} \frac{x}{\sqrt{p}} + \frac{y}{\sqrt{q}} &= 0, \\ z &= 0 \end{aligned} \tag{9a}$$

(соответствующими уравнениям (9) при $k = 0$) целиком расположены на параболоиде (1). Кроме того, при $k = \frac{1}{2}\left(\frac{x_0}{\sqrt{p}} + \frac{y_0}{\sqrt{q}}\right)$ прямая (9) (или (9а), если $k = 0$) проходит через произвольную наперед данную точку $M(x_0, y_0, z_0)$ этого параболоида. Следовательно,

прямые (9) (*вместе с прямой* (9а)) *составляют семейство прямолинейных образующих гиперболического параболоида* (1).

Аналогично показывается, что

прямые

$$\begin{aligned} \frac{x}{\sqrt{p}} - \frac{y}{\sqrt{q}} &= 2k, \\ \frac{x}{\sqrt{p}} + \frac{y}{\sqrt{q}} &= \frac{z}{k}, \end{aligned} \qquad k \neq 0,$$

вместе с прямой

$$\begin{aligned} \frac{x}{\sqrt{p}} - \frac{y}{\sqrt{q}} &= 0, \\ z &= 0 \end{aligned}$$

составляют другое семейство прямолинейных образующих гиперболического параболоида (1).

Свойства прямолинейных образующих гиперболического параболоида во многом аналогичны свойствам прямолинейных образующих однополостного гиперболоида (см. п. 4).

Свойство 1. *Через любую точку гиперболического параболоида проходит одна и только одна образующая каждого семейства.*

Это свойство мы выше уже доказали.

Свойство 2. *Любые две образующие гиперболического параболоида, принадлежащие разным семействам, пересекаются и потому компланарны.*

Действительно, пусть

$$\begin{aligned} x &= \sqrt{p}(t+\tau_1), \\ y &= \sqrt{q}(t-\tau_1), \\ z &= 2\tau_1 t \end{aligned} \qquad \text{и} \qquad \begin{aligned} x &= \sqrt{p}(\tau_2+t), \\ y &= \sqrt{q}(\tau_2-t), \\ z &= 2\tau_2 t \end{aligned}$$

— произвольные образующие первого и второго семейств.

Поскольку

$$\begin{vmatrix} \sqrt{p}(\tau_2-\tau_1) & \sqrt{q}(\tau_2+\tau_1) & 0 \\ \sqrt{p} & \sqrt{q} & 2\tau_1 \\ \sqrt{p} & -\sqrt{q} & 2\tau_2 \end{vmatrix} = 2\sqrt{pq} \begin{vmatrix} \tau_2-\tau_1 & \tau_2+\tau_1 & 0 \\ 1 & 1 & \tau_1 \\ 1 & -1 & \tau_2 \end{vmatrix} = 0,$$

эти образующие не скрещиваются. Так как

$$\begin{vmatrix} \sqrt{p} & \sqrt{q} \\ \sqrt{p} & -\sqrt{q} \end{vmatrix} = 2\sqrt{pq} \neq 0,$$

они и не параллельны. Следовательно, эти образующие пересекаются.

Заметим, что это свойство сильнее соответствующего свойства прямолинейных образующих однополостного гиперболоида (п. 4, свойство 2).

Свойство 3. *Две образующие гиперболического параболоида, принадлежащие одному семейству, скрещиваются.*

Действительно, пусть, например, эти образующие принадлежат семейству (5) и отвечают значениям параметра τ_1 и τ_2. Поскольку

$$\begin{vmatrix} \sqrt{p}(\tau_2-\tau_1) & \sqrt{q}(\tau_1-\tau_2) & 0 \\ \sqrt{p} & \sqrt{q} & 2\tau_1 \\ \sqrt{p} & \sqrt{q} & 2\tau_2 \end{vmatrix} = 4\sqrt{pq}(\tau_2-\tau_1)^2 \neq 0,$$

эти образующие скрещиваются.

Свойство 4. *Все образующие гиперболического параболоида, принадлежащие одному семейству, параллельны некоторой плоскости.*

Действительно, если образующие принадлежат, скажем, семейству (5), они имеют направляющие векторы вида

$$(\sqrt{p},\ -\sqrt{q},\ 2\tau)$$

и потому параллельны плоскости

$$\frac{x}{\sqrt{p}}+\frac{y}{\sqrt{q}}=0.$$

Аналогично, образующие семейства (7) параллельны плоскости

$$\frac{x}{\sqrt{p}}-\frac{y}{\sqrt{q}}=0.$$

Так же как и для однополостного гиперболоида, из этих свойств непосредственно вытекает, что

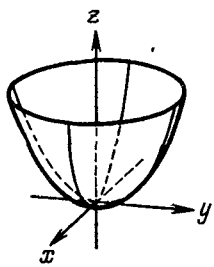
произвольный гиперболический параболоид представляет собой геометрическое место точек, принадлежащих прямым, пересекающим три фиксированные скрещивающиеся прямые, параллельные одной плоскости.

Упражнение. Докажите, что

какие бы три скрещивающиеся прямые, параллельные одной плоскости, мы ни взяли, геометрическое место точек прямых, пересекающих эти прямые, является гиперболическим параболоидом.

6. Эллиптические параболоиды

Определение 1. *Эллиптическим параболоидом* называется поверхность, имеющая в некоторой системе прямоугольных координат уравнение вида



$$\frac{x^2}{p}+\frac{y^2}{q}=2z, \qquad (1)$$

где $p>0$, $q>0$. Эта система координат называется *канонической* (относительно данного параболоида), а уравнение (1) называется *каноническим уравнением* этого эллиптического параболоида.

Форму эллиптического параболоида изучить очень легко.

При $p=q$ параболоид (1) выражается в цилиндрических координатах ρ, φ, z уравнением

$$\frac{\rho^2}{p}=2z,$$

не зависящим от φ. Следовательно,

параболоид

$$\frac{x^2+y^2}{p}=2z \qquad (2)$$

получается из параболы

$$x^2 = 2pz,$$

расположенной в плоскости Oxz, вращением вокруг оси Oz.

На этом основании параболоид (2) называется обычно *параболоидом вращения.*

Заметим, что рассмотренный в § 3 гл. 4 параболоид Q является как раз параболоидом вращения, получающимся при $p = 1$ (правда, в координатной системе с обращенной ориентацией оси Oz).

При сжатии к плоскости Oxz с коэффициентом $\sqrt{\frac{q}{p}}$ параболоид (2) переходит в параболоид (1) (соответствующее рассуждение мы проводили выше для эллипсоида и двуполостного гиперболоида, и потому повторять его здесь третий раз нужды нет). Таким образом,

произвольный эллиптический параболоид получается из параболоида вращения сжатием к некоторой плоскости.

Этого уже вполне достаточно, чтобы отчетливо представить себе форму эллиптического параболоида. Он имеет вид чаши эллиптического сечения (в случае параболоида вращения — кругового), охватывающей положительное направление оси Oz.

Столь же просто форма эллиптического параболоида исследуется методом плоских сечений. Мы приведем лишь окончательный результат:

При $h < 0$ плоскость $z = h$ не пересекает эллиптический параболоид (1);

при $h = 0$ плоскость $z = h$ имеет с параболоидом (1) *одну общую точку;*

при $h > 0$ плоскость $z = h$ пересекает параболоид (1) *по эллипсу с полуосями*

$$\sqrt{2hp}, \quad \sqrt{2hq},$$

монотонно возрастающими от нуля до $+\infty$, когда h возрастает от нуля до $+\infty$.

Плоскость $y = h$ пересекает эллиптический параболоид (1) *по параболе с фокальным параметром p, с вершиной в точке* $\left(0, h, \frac{h^2}{2q}\right)$ *и с «рогами», направленными «вверх»;*

плоскость $x = h$ пересекает эллиптический параболоид (1) *по параболе с фокальным параметром q, с вершиной в точке* $\left(h, 0, \frac{h^2}{2p}\right)$ *и с «рогами», также направленными «вверх».*

Легко, далее, видеть, что в отличие от гиперболического параболоида

эллиптический параболоид линейчатой поверхностью не является.

Более того,

эллиптический параболоид не содержит ни одной прямой.

Действительно, поскольку параболоид (1) расположен в полупространстве $z \geqslant 0$, на нем могут лежать лишь прямые, параллельные плоскости $z = 0$, т. е. расположенные в одной из плоскостей вида $z = h$. Но, как мы видели, каждая такая плоскость пересекает параболоид (при $h > 0$) по эллипсу, и потому никакая прямая, принадлежащая этой плоскости, не может принадлежать параболоиду.

Замечание 1. Аналогично доказывается, что

ни эллипсоид, ни двуполостный гиперболоид не могут содержать ни одной прямой и потому не являются линейчатыми поверхностями.

Действительно, для эллипсоида это вытекает из того, что он целиком расположен в некотором параллелепипеде (см. п. 1). Поскольку плоскость $z = 0$ не пересекает гиперболоид

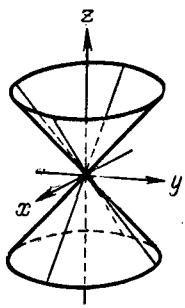
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1,$$

каждая прямая, расположенная на этом гиперболоиде, должна принадлежать плоскости вида $z = h$. Но, как мы знаем (п. 2), каждая такая плоскость либо не пересекает гиперболоид, либо имеет с ним одну общую точку, либо пересекает его по эллипсу.

Однако это не означает, что однополостными гиперболоидами и гиперболическими параболоидами исчерпываются все линейчатые поверхности второго порядка. Линейчатые поверхности второго порядка, не являющиеся ни гиперболоидами, ни параболоидами, мы изучим в следующих пунктах.

7. Конусы второго порядка

Определение 1. *Конусом второго порядка* называется поверхность, имеющая в некоторой системе прямоугольных координат уравнение вида



$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0, \qquad (1)$$

где $a \geqslant b \geqslant 0$, $c > 0$ и $\frac{1}{a^2} + \frac{1}{b^2} + \frac{1}{c^2} = 1$. Эта система координат называется *канонической* (относительно данного конуса), а уравнение (1) называется *каноническим уравнением* этого конуса.

Легко видеть, что

плоскость $z = h$ с $h \neq 0$ пересекает конус (1) по эллипсу с полуосями

$$\frac{a}{c}|h|, \quad \frac{b}{c}|h|,$$

монотонно увеличивающимися при $h \to \infty$ и стремящимися к нулю при $h \to 0$;

плоскость $z = 0$ *пересекает конус* (1) *в одной точке;*

плоскость $y = h$ *с* $h \neq 0$ *пересекает конус* (1) *по гиперболе с полуосями*

$$\frac{c}{b}|h|, \quad \frac{a}{b}|h|,$$

монотонно увеличивающимися при $h \to \infty$ *и стремящимися к нулю при* $h \to 0$;

плоскость $x = h$ *с* $h \neq 0$ *пересекает конус* (1) *по гиперболе с полуосями*

$$\frac{c}{a}|h|, \quad \frac{b}{a}|h|,$$

монотонно увеличивающимися при $h \to \infty$ *и стремящимися к нулю при* $h \to 0$;

плоскости $y = 0$ *и* $x = 0$ *пересекают конус по паре пересекающихся прямых.*

Задание. Докажите все эти утверждения.

Пусть $M_0(x_0, y_0, z_0)$ — произвольная точка конуса (1). Прямая OM_0, проходящая через начало координат $O(0, 0, 0)$ и точку M_0, имеет параметрические уравнения вида

$$x = x_0 t,$$
$$y = y_0 t,$$
$$z = z_0 t.$$

Поскольку

$$\frac{(x_0 t)^2}{a^2} + \frac{(y_0 t)^2}{b^2} - \frac{(z_0 t)^2}{c^2} = \left(\frac{x_0^2}{a^2} + \frac{y_0^2}{b^2} - \frac{z_0^2}{c^2}\right) t^2 = 0,$$

мы видим, что

все точки прямой OM_0 *принадлежат конусу* (1).

Это означает, что справедливо следующее

Предложение 1. *Конус* (1) *является линейчатой поверхностью, обладающей семейством прямолинейных образующих, проходящих через точку* O.

Это оправдывает следующее общее

Определение 2. Линейчатая поверхность, обладающая семейством прямолинейных образующих, проходящих через некоторую точку, называется *конусом* с *вершиной* в этой точке.

Замечание 1. Согласно этому определению, *любая плоскость является конусом* (с вершиной в произвольной ее точке). Более того, конусом будет поверхность, состоящая из произвольного числа плоскостей, проходящих через одну прямую (вершиной такого конуса может служить любая точка этой прямой).

Упражнение. Покажите, что, за исключением этих особых случаев, любой конус имеет только одну вершину и только одно семейство образующих.

Определение 3. Линия, расположенная на конусе, называется его *направляющей*, если любая прямолинейная образующая конуса пересекает ее в одной и только одной точке.

Примером направляющей является линия, получающаяся при пересечении конуса произвольной плоскостью, не проходящей через вершину конуса. Направляющие такого вида называются *плоскими*.

Для конуса второго порядка (1) плоскими направляющими являются, например, все указанные выше эллипсы и гиперболы.

Замечание 2. Интересно, что

любой конус второго порядка (1) *обладает плоской направляющей, являющейся окружностью.*

Упражнение. Докажите это утверждение.

На этом основании конусы (1) иногда называются *круговыми конусами* (*косыми*, если прямая, соединяющая вершину конуса с центром направляющей окружности, не перпендикулярна плоскости этой окружности, и *прямыми* — в противном случае).

Задание. Докажите, что

конус (1) *тогда и только тогда является прямым круговым конусом, когда* $a = b$.

Это возвращает нас к определению 1 п. 5 § 1, поскольку при $a = b$ уравнение (1) может быть переписано в виде

$$x^2 + y^2 = R^2 z^2,$$

где $R = a/c$.

Упражнение. Покажите, что любая не проходящая через начало координат плоскость пересекает конус (1) либо по эллипсу, либо по параболе, либо по гиперболе.

Напомним, что функция $F(x, y, z)$ называется *однородной*, если существует такое число ρ, что для любых x, y, z и t имеет место тождество

$$F(tx, ty, tz) = t^{\rho} F(x, y, z)$$

(число ρ называется при этом *степенью* однородной функции $F(x, y, z)$).

Выше при доказательстве предложения 1 мы по существу использовали лишь тот факт, что функция

$$F(x, y, z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2}$$

является однородной функцией (степени 2). Поэтому то же самое рассуждение показывает, что

любое уравнение

$$F(x, y, z) = 0 \qquad (2)$$

с однородной функцией $F(x, y, z)$ *определяет в пространстве некоторый конус с вершиной в начале координат* O.

Замечание 3. Здесь, конечно, предполагается, что уравнение (2) действительно определяет некоторую поверхность (а не удовлетворяется, скажем, только точкой $(0, 0, 0)$).

Замечание 4. Обратим внимание на то, что уравнение (2) определяет конус не только в прямоугольных, но в любых аффинных координатах x, y, z.

Замечательно, что обратное утверждение также верно, т. е. *любой конус с вершиной O может быть задан уравнением* (2) *с однородной функцией* $F(x, y, z)$.

Действительно, пусть данный конус задается уравнением

$$G(x, y, z) = 0. \tag{3}$$

Рассмотрим его пересечение со сферой

$$x^2 + y^2 + z^2 = 1 \tag{4}$$

радиуса 1 с центром в точке O. Ясно, что это пересечение является направляющей конуса (3).

Определим теперь функцию $F(x, y, z)$ формулой

$$F(x, y, z) = G\left(\frac{x}{\sqrt{x^2+y^2+z^2}}, \frac{y}{\sqrt{x^2+y^2+z^2}}, \frac{z}{\sqrt{x^2+y^2+z^2}}\right),$$

если $(x, y, z) \neq (0, 0, 0)$. При $(x, y, z) = (0, 0, 0)$ мы положим

$$F(0, 0, 0) = 0.$$

Функция $F(x, y, z)$ является, очевидно, однородной функцией (нулевой степени) и потому уравнение

$$F(x, y, z) = 0 \tag{5}$$

определяет некоторый конус с вершиной в точке O.

Функция $F(x, y, z)$, по построению, совпадает с функцией $G(x, y, z)$ на сфере (4). Поэтому конус (5) пересекается со сферой (4) по той же линии, что и конус (3). Таким образом, конусы (3) и (5) имеют одну и ту же вершину и одну и ту же направляющую. Следовательно, они совпадают.

Тем самым мы доказали, что конус (3) может быть задан, как и утверждалось, однородным уравнением.

Замечание 5. Доказанное утверждение, конечно, не означает, что конус (с вершиной в начале координат) нельзя задать неоднородным уравнением.

Рассмотрим, например, функцию вида

$$F(x, y, z) = (x^2 + y^2 + z^2 + 1)\, G(x, y, z),$$

где $G(x, y, z)$ — некоторая однородная функция. Функция $F(x, y, z)$ заведомо не однородна и вместе с тем уравнение

$$F(x, y, z) = 0$$

определяет ту же поверхность (тот же конус), что и уравнение

$$G(x, y, z) = 0.$$

Интересно, что для алгебраических уравнений первой и второй степени это уже не так, т. е.

если многочлен $F(x, y, z)$ первой или второй степени обладает тем свойством, что уравнение

$$F(x, y, z) = 0$$

определяет конус с вершиной в начале координат, то этот многочлен однороден.

Для многочленов первой степени это утверждение очевидно. Докажем его для многочленов второй степени.

Произвольный многочлен второй степени от переменных x, y, z имеет вид

$$F(x, y, z) = F_0 + F_1(x, y, z) + F_2(x, y, z),$$

где F_0 — некоторое число, $F_1(x, y, z)$ — однородный многочлен первой степени относительно x, y, z и $F_2(x, y, z)$ — однородный многочлен второй степени относительно x, y, z (выписывать который в явном виде нам нет нужды).

Мы хотим доказать, что если уравнение

$$F(x, y, z) = 0 \tag{6}$$

является уравнением конуса с вершиной в начале координат и если это уравнение действительно является уравнением второй степени, т. е. многочлен $F_2(x, y, z)$ тождественно не равен нулю, то многочлен $F(x, y, z)$ совпадает с многочленом $F_2(x, y, z)$, т. е. число F_0 и многочлен $F_1(x, y, z)$ равны нулю.

Утверждение относительно числа F_0, по существу, очевидно. Действительно, если уравнение (6) является уравнением конуса с вершиной в начале координат, то ему удовлетворяют, в частности, координаты $(0, 0, 0)$ этой вершины. Но поскольку, в силу однородности, $F_1(0, 0, 0) = 0$ и $F_2(0, 0, 0) = 0$, то это возможно только при $F_0 = 0$.

Утверждение относительно многочлена $F_1(x, y, z)$ доказывается значительно сложнее. Мы проведем доказательсто «от противного». Пусть многочлен $F_1(x, y, z) = Ax + By + Cz$ не равен тождественно нулю, т. е. пусть хотя бы один из его коэффициентов A, B, C, скажем, для определенности, коэффициент C, отличен от нуля. Тогда формулы

$$\begin{aligned} x &= x, \\ y &= y, \\ u &= Ax + By + Cz \end{aligned}$$

будут определять в пространстве некоторые новые аффинные координаты x, y, u (вообще говоря, не прямоугольные, даже если исходные координаты x, y, z были прямоугольны). В координатах x, y, u уравнение (6) конуса принимает, очевидно, вид

$$G_2(x, y, u) + u = 0,$$

где $G_2(x, y, u)$ — некоторый (однородный) многочлен второй степени от x, y, u.

Расположив многочлен $G_2(x, y, u)$ по степеням u, мы можем написать

$$G_2(x, y, u) = au^2 + H_1(x, y)u + H_2(x, y),$$

где $H_1(x, y)$ и $H_2(x, y)$ — однородные многочлены (соответственно первой и

второй степени) от переменных x, y, а a — некоторое число. В этих обозначениях уравнение (6) приобретает вид

$$au^2 + (H_1(x, y) + 1)u + H_2(x, y) = 0. \tag{7}$$

Поскольку это уравнение является, по условию, уравнением конуса с вершиной в начале координат, то вместе с некоторой точкой (x_0, y_0, u_0) ему удовлетворяет и любая точка вида (x_0t, y_0t, u_0t), где t — произвольное число, т. е. тождественно по t имеет место равенство

$$\left[au_0^2 + H_1(x_0, y_0)u_0 + H_2(x_0, y_0)\right]t^2 + tu_0 = 0,$$

что возможно тогда и только тогда, когда $u_0 = 0$ и $H_2(x_0, y_0) = 0$. Таким образом, мы доказали, что любое решение уравнения (7) имеет вид $(x_0, y_0, 0)$, где (x_0, y_0) — некоторое решение уравнения

$$H_2(x, y) = 0. \tag{8}$$

В частности, при $x_0 = 0$, $y_0 = 0$ мы должны обязательно получить $u_0 = 0$. Но полагая в уравнении (7) $x = 0$, $y = 0$, мы получим уравнение

$$au^2 + u = 0.$$

Чтобы это уравнение имело только корень $u_0 = 0$, необходимо, чтобы $a = 0$. Таким образом, уравнение (7) имеет на самом деле вид

$$(H_1(x, y) + 1)u + H_2(x, y) = 0.$$

Чтобы каждое решение этого уравнения имело вид $(x_0, y_0, 0)$, где x_0, y_0 — некоторое решение уравнения (8), необходимо, чтобы для любых чисел (x_1, y_1), для которых $H_2(x_1, y_1) \neq 0$ (т. е. которые не являются решениями уравнения (8)), имело место равенство

$$H_1(x_1, y_1) + 1 = 0.$$

Рассмотрим поэтому более внимательно уравнение (8). Поскольку многочлен $H_2(x, y)$ является однородным многочленом второй степени, это уравнение имеет вид

$$c_0x^2 + 2c_1xy + c_2y^2 = 0. \tag{9}$$

Если многочлен $c_0x^2 + 2c_1xy + c_2y^2$ разлагается на линейные множители, то этому уравнению удовлетворяют точки двух прямых (возможно, совпадающих), проходящих через точку $(0, 0)$. Если же многочлен $c_0x^2 + 2c_1xy + c_2y^2$ отличен от нуля и не разлагается на линейные множители, то уравнению (9) удовлетворяет только точка $(0, 0)$. Наконец, если $c_0 = c_1 = c_2 = 0$, то уравнению (9) удовлетворяет любая точка плоскости.

Таким образом, условию $H_2(x, y) \neq 0$ либо не удовлетворяют координаты ни одной точки плоскости Oxy, либо, напротив, удовлетворяют координаты любой точки этой плоскости, отличной от точки $(0, 0)$ или не принадлежащей двум (возможно, совпадающим) прямым, проходящим через точку $(0, 0)$.

С другой стороны, поскольку многочлен $H_1(x, y)$ является однородным многочленом первой степени, уравнение

$$H_1(x, y) + 1 = 0$$

определяет в плоскости Oxy прямую, не проходящую через точку $(0, 0)$ (либо — если многочлен $H_1(x, y)$ тождественно равен нулю — вообще не может быть удовлетворено).

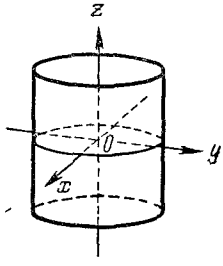
Следовательно, равенство $H_1(x_1, y_1) + 1 = 0$ тогда и только тогда имеет место для любых чисел x_1, y_1, для которых $H_2(x_1, y_1) \neq 0$, когда таких чисел вообще нет, т. е. когда многочлен $H_2(x, y)$ тождественно равен нулю.

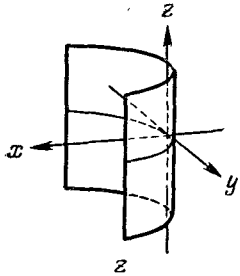
Тем самым мы доказали, что уравнение нашего конуса имеет на самом деле вид

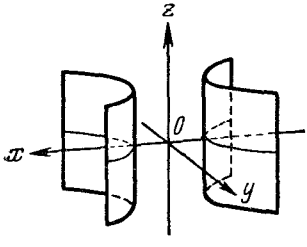
$$(H_1(x, y) + 1)\, u = 0$$

и потому для любых чисел x_0, y_0, для которых $H_1(x_0, y_0) + 1 = 0$, имеет решение вида (x_0, y_0, u) с произвольным u. Поскольку это невозможно (ибо, как мы знаем, для любой точки нашего конуса координата u должна быть равна нулю), отсюда следует, что многочлен $H_1(x, y)$ тождественно равен нулю.

Таким образом, оба многочлена $H_1(x, y)$ и $H_2(x, y)$ тождественно равны нулю. Поскольку $a = 0$, это означает, что тождественно равен нулю многочлен $G_2(x, y, u)$, а значит, вопреки предположению, и многочлен $F_2(x, y, z)$.







8. Цилиндры второго порядка

Определение 1. Поверхность называется

а) *эллиптическим цилиндром*,

б) *параболическим цилиндром*,

в) *гиперболическим цилиндром*,

если в некоторой системе прямоугольных координат x, y, z (называемой *канонической* относительно данной поверхности) она имеет следующее (*каноническое*) уравнение:

$$\text{а)}\ \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad a \geqslant b > 0,$$

$$\text{б)}\ y^2 = 2px, \qquad p > 0, \qquad (1)$$

$$\text{в)}\ \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \qquad a > 0,\ b > 0.$$

Задание. Опишите линии, получающиеся в пересечении этих цилиндров плоскостями $x = h$, $y = h$ и $z = h$.

Поскольку уравнения (1) не зависят от z, вместе с некоторой точкой (x_0, y_0, z_0) им удовлетворяет и любая точка вида (x_0, y_0, z). Это означает, что

вместе с некоторой точкой M_0 каждый из цилиндров (1) *содержит всю прямую, проходящую через эту точку параллельно оси Oz.*

Другими словами, справедливо следующее

Предложение 1. *Каждый из цилиндров* (1) *является линейчатой поверхностью, обладающей семейством прямолинейных образующих, параллельных прямой Oz.*

Это оправдывает следующее общее

Определение 2. Линейчатая поверхность, обладающая семейством прямолинейных образующих, параллельных некоторой фиксированной прямой, называется *цилиндром.* Направление этой прямой называется *осевым направлением* цилиндра.

Замечание 1. Согласно этому определению любая плоскость и, более общо, любая поверхность, состоящая из любого числа плоскостей, проходящих через одну прямую (или параллельных друг другу), является цилиндром. Таким образом, все «особые» конусы, описанные в замечании 1 п. 7, одновременно являются и цилиндрами.

Упражнение. Покажите, что, за исключением цилиндров, являющихся плоскостями или несколькими параллельными плоскостями, любой цилиндр имеет единственное семейство параллельных образующих (и, следовательно, единственное осевое направление).

Замечание 2. Во многих отношениях цилиндры подобны конусам. Это объясняется тем, что в евклидово-проективном пространстве цилиндры являются частным случаем конусов (а именно, представляют собой конусы с вершиной в несобственной точке).

Следующее определение вполне аналогично соответствующему определению для конусов:

Определение 3. Линия, расположенная на цилиндре, называется его *направляющей*, если любая прямолинейная образующая цилиндра пересекает ее в одной и только одной точке. Направляющая называется *плоской*, если она является пересечением цилиндра с некоторой плоскостью.

Можно без особого труда показать, что

любая плоская направляющая

эллиптического		*эллипсом*
параболического	*цилиндра является*	*параболой*
гиперболического		*гиперболой.*

Для эллиптических цилиндров верно и более точное утверждение:

любой эллиптический цилиндр обладает плоской направляющей, являющейся окружностью.

Упражнение. Докажите эти утверждения.

На основании последнего утверждения эллиптические цилиндры иногда называются также *косыми круговыми цилиндрами.* Эпитет «косой» здесь подчеркивает, что плоскость направляющей окружности может быть не перпендикулярна осе-

вому направлению цилиндра. В случае, когда эта плоскость перпендикулярна осевому направлению цилиндра, цилиндр называется *прямым круговым цилиндром* (это в точности — «цилиндры», известные читателю из школы). Ясно, что эллиптический цилиндр тогда и только тогда является прямым круговым цилиндром, когда в его каноническом уравнении $a = b$.

Соображения, использованные выше для доказательства предложения 1, немедленно показывают, что

любое уравнение вида

$$F(x, y) = 0, \tag{2}$$

левая часть которого не зависит от z, определяет в пространстве цилиндр, осевым направлением которого является направление оси Oz.

При этом ясно, что

линия на плоскости Oxy, определяемая уравнением (2), *является направляющей этого цилиндра.*

Замечание 3. Как и в аналогичном утверждении для конусов (см. п. 7), здесь, конечно, предполагается, что уравнение (2) действительно определяет на плоскости Oxy некоторую линию.

Замечание 4. Уравнение вида (2) определяет цилиндр в любых аффинных координатах.

Верно и обратное утверждение, т. е.

любой цилиндр, прямолинейные образующие которого параллельны оси Oz, может быть задан уравнением (2) *с левой частью, не зависящей от z.*

Действительно, пусть (2) — уравнение направляющей данного цилиндра, получающейся при пересечении его координатной плоскостью Oxy. Ясно тогда, что это же уравнение (но рассматриваемое в пространстве, т. е. как уравнение от x, y, z) будет и уравнением данного цилиндра.

Упражнение. Покажите (ср. п. 7), что

если многочлен $F(x, y, z)$ степени не выше второй обладает тем свойством, что уравнение

$$F(x, y, z) = 0$$

определяет цилиндр с прямолинейными образующими, параллельными оси Oz, то этот многочлен не зависит от z.

Как показывает пример уравнения

$$x(x^2 + y^2 + z^2 + 1) = 0,$$

для многочленов степени выше второй соответствующее утверждение неверно.

Глава 6

ОБЩАЯ ТЕОРИЯ ЛИНИЙ ВТОРОГО ПОРЯДКА

§ 1. КЛАССИФИКАЦИЯ ЛИНИЙ ВТОРОГО ПОРЯДКА

1. Линии второго порядка на евклидовой плоскости

Самое общее уравнение линии второго порядка имеет (в прямоугольных или любых аффинных координатах) вид

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0, \qquad (1)$$

где хотя бы одно из чисел a_{11}, a_{12}, a_{22} отлично от нуля.

Предполагая координаты x, y прямоугольными, мы попробуем найти другие (также прямоугольные) координаты x', y', в которых уравнение линии (1) имеет возможно более простой вид.

В первую очередь мы постараемся поворотом координатных осей уничтожить член, содержащий произведение координат (конечно, если этот член имеется, т. е. если $a_{12} \neq 0$).

Повернув оси координат на угол θ, мы получим новые координаты x', y', связанные со старыми координатами соотношениями

$$\begin{aligned} x &= x'\cos\theta - y'\sin\theta, \\ y &= x'\sin\theta + y'\cos\theta \end{aligned} \qquad (2)$$

(см. п. 4 § 1 гл. 2). В этих координатах линия (1) задается уравнением

$$\begin{aligned} a_{11}(x'\cos\theta - y'\sin\theta)^2 + 2a_{12}(x'\cos\theta - y'\sin\theta)(x'\sin\theta + y'\cos\theta) + \\ + a_{22}(x'\sin\theta + y'\cos\theta)^2 + 2a_{13}(x'\cos\theta - y'\sin\theta) + \\ + 2a_{23}(x'\sin\theta + y'\cos\theta) + a_{33} = 0, \end{aligned}$$

т. е. уравнением

$$a'_{11}x'^2 + 2a'_{12}x'y' + a'_{22}y'^2 + 2a'_{13}x' + 2a'_{23}y' + a'_{33} = 0, \qquad (1')$$

где

$$
\begin{aligned}
a'_{11} &= a_{11}\cos^2\theta + 2a_{12}\cos\theta\sin\theta + a_{22}\sin^2\theta,\\
a'_{12} &= (a_{22} - a_{11})\cos\theta\sin\theta + a_{12}(\cos^2\theta - \sin^2\theta),\\
a'_{22} &= a_{11}\sin^2\theta - 2a_{12}\cos\theta\sin\theta + a_{22}\cos^2\theta,\\
a'_{13} &= a_{13}\cos\theta + a_{23}\sin\theta,\\
a'_{23} &= -a_{13}\sin\theta + a_{23}\cos\theta.\\
a'_{33} &= a_{33}.
\end{aligned}
\tag{3}
$$

Условие $a'_{12} = 0$ дает нам для угла θ уравнение

$$(a_{22} - a_{11})\cos\theta\sin\theta + a_{12}(\cos^2\theta - \sin^2\theta) = 0,$$

т. е. уравнение

$$(a_{22} - a_{11})\sin 2\theta + 2a_{12}\cos 2\theta = 0.$$

Следовательно,

$$\operatorname{ctg} 2\theta = \frac{a_{11} - a_{22}}{2a_{12}}. \tag{4}$$

Тем самым доказано, что
при повороте координатных осей на угол

$$\theta = \frac{1}{2}\operatorname{arctg}\frac{a_{11} - a_{22}}{2a_{12}}$$

уравнение (1) *преобразуется в аналогичное уравнение с коэффициентом* a_{12}, *равным нулю.*

Таким образом, без ограничения общности, мы можем считать, что наша линия с самого начала задана уравнением

$$a_{11}x^2 + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0. \tag{5}$$

Дальнейшее упрощение этого уравнения мы произведем, перенося начало координат в некоторую точку (x_0, y_0), т. е. вводя новые координаты x', y', связанные с координатами x, y формулами

$$
\begin{aligned}
x &= x' + x_0,\\
y &= y' + y_0.
\end{aligned}
$$

В координатах x', y' линия (1) задается уравнением

$$
\begin{aligned}
a_{11}(x' + x_0)^2 + 2a_{12}(x' + x_0)(y' + y_0) + a_{22}(y' + y_0)^2 +\\
+ 2a_{13}(x' + x_0) + 2a_{23}(y' + y_0) + a_{33} = 0,
\end{aligned}
$$

т. е. уравнением

$$a'_{11}x'^2 + 2a'_{12}x'y' + a'_{22}y'^2 + 2a'_{13}x' + 2a'_{23}y' + a'_{33} = 0,$$

где

$$\begin{aligned} &a'_{11} = a_{11}, \quad a'_{12} = a_{12}, \quad a'_{22} = a_{22}, \\ &a'_{13} = a_{11}x_0 + a_{12}y_0 + a_{13}, \\ &a'_{23} = a_{12}x_0 + a_{22}y_0 + a_{23}, \\ &a'_{33} = a_{11}x_0^2 + 2a_{12}x_0y_0 + a_{22}y_0^2 + 2a_{13}x_0 + 2a_{23}y_0 + a_{33}. \end{aligned} \tag{6}$$

В частности, для упрощенного уравнения (5) мы получаем, что

$$\begin{aligned} &a'_{11} = a_{11}, \quad a'_{12} = a_{12} = 0, \quad a'_{22} = a_{22}, \\ &a'_{13} = a_{11}x_0 + a_{13}, \\ &a'_{23} = a_{22}y_0 + a_{23}, \\ &a'_{33} = a_{11}x_0^2 + a_{22}y_0^2 + 2a_{13}x_0 + 2a_{23}y_0 + a_{33}. \end{aligned}$$

Постараемся теперь подобрать точку (x_0, y_0) так, чтобы коэффициенты при x' и y' обратились в нуль. Это дает нам уравнения

$$\begin{aligned} a_{11}x_0 + a_{13} &= 0, \\ a_{22}y_0 + a_{23} &= 0. \end{aligned} \tag{7}$$

Если $a_{11} \neq 0$ и $a_{22} \neq 0$, то эти уравнения имеют единственное решение

$$x_0 = -\frac{a_{13}}{a_{11}}, \quad y_0 = -\frac{a_{23}}{a_{22}}. \tag{8}$$

Этим доказано, что

при переносе начала координат в точку с координатами (8) *уравнение* (5) *с* $a_{11} \neq 0$ *и* $a_{22} \neq 0$ *преобразуется в аналогичное уравнение с коэффициентами* a_{13} *и* a_{23}, *равными нулю.*

Пусть теперь один из коэффициентов a_{11} и a_{22} равен нулю (оба они равны нулю быть не могут, поскольку тогда мы имели бы линию не второго, а первого порядка). Переименовывая, если нужно, координаты, мы можем считать, что $a_{22} \neq 0$. Тогда мы можем удовлетворить второму из уравнений (7) (решение дается второй из формул (8)), и в нашем распоряжении еще останется x_0. Постараемся им распорядиться так, чтобы обратить в нуль свободный член. Это даст нам уравнение

$$a_{11}x_0^2 + a_{22}y_0^2 + 2a_{13}x_0 + 2a_{23}y_0 + a_{33} = 0,$$

т. е. $\left(\text{поскольку } a_{11} = 0 \text{ и } y_0 = -\frac{a_{23}}{a_{22}}\right)$ уравнение

$$2a_{13}x_0 = \frac{a_{23}^2}{a_{22}} - a_{33}.$$

При $a_{13} \neq 0$ это уравнение имеет единственное решение; вместе со вторым уравнением (8) это дает нам точку с координатами

$$x_0 = \frac{a_{23}^2 - a_{22}a_{33}}{2a_{13}a_{22}}, \quad y_0 = -\frac{a_{23}}{a_{22}}. \tag{9}$$

Таким образом,

при переносе начала координат в точку с координатами (9) *уравнение* (5) *с* $a_{11} = 0$, $a_{22} \neq 0$ *и* $a_{13} \neq 0$ *переходит в аналогичное уравнение с коэффициентами* a_{23} *и* a_{33}, *равными нулю.*

Если же $a_{13} = 0$, то выбор координаты x_0 никак не влияет на уравнение, и мы получаем, что

при переносе начала координат в точку $\left(x_0, -\frac{a_{23}}{a_{22}}\right)$, *где* x_0 — *произвольно, уравнение* (5) *с* $a_{11} = a_{13} = 0$ *и* $a_{22} \neq 0$ *переходит в аналогичное уравнение с коэффициентом* a_{23}, *равным нулю.*

Резюмируя, мы получаем следующее

Предложение 1. *Для любой линии второго порядка существует система прямоугольных координат, в которой ее уравнение имеет либо вид*

$$a_{11}x^2 + a_{22}y^2 + a_{33} = 0, \tag{10}$$

где $a_{11} \neq 0$, $a_{22} \neq 0$, *либо вид*

$$a_{22}y^2 + 2a_{13}x = 0, \tag{11}$$

где $a_{22} \neq 0$, $a_{13} \neq 0$, *либо вид*

$$a_{22}y^2 + a_{33} = 0, \tag{12}$$

где $a_{22} \neq 0$.

Определение 1. Уравнения (10)—(12) называются *приведенными уравнениями* линии второго порядка. Линии, имеющие приведенные уравнения вида (10), мы будем называть линиями *типа* I, имеющие приведенные уравнения вида (11) — линиями *типа* II, а имеющие приведенные уравнения вида (12) — линиями *типа* III.

Замечание 1. Хотя априори не исключено, что одна и та же линия может принадлежать различным типам (т. е. что при одном способе приведения получится, скажем, уравнение (10), а при другом — уравнение (11)), на самом деле это невозможно. Мы этого доказывать здесь не будем, поскольку ниже мы тип линии охарактеризуем другим способом, из которого невозможность для линии иметь два различных типа будет вытекать автоматически.

Рассмотрим более внимательно уравнения (10) линий типа I.

Случай 1. *Коэффициенты* a_{11} *и* a_{22} *имеют различные знаки.* Умножив (если нужно) уравнение (10) на —1, мы можем, без ограничения общности, считать, что коэффициент a_{33} неположителен:

$$a_{33} \leqslant 0.$$

Кроме того (переименовав, если нужно, координаты), мы можем также считать, что коэффициент a_{11} положителен, а коэффициент a_{22} отрицателен:

$$a_{11} > 0, \quad a_{22} < 0.$$

Случай 1а. *Коэффициент a_{33} отличен от нуля.* Тогда, полагая

$$a^2 = -\frac{a_{33}}{a_{11}}, \quad b^2 = \frac{a_{33}}{a_{22}},$$

мы приведем уравнение (10) к виду

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \qquad a > 0, \quad b > 0, \tag{13}$$

являющемуся каноническим уравнением *гиперболы* (см. п. 3 § 1 гл. 5).

Случай 1б. *Коэффициент a_{33} равен нулю.* Тогда без ограничения общности можно считать, что $a_{11} \leqslant -a_{22}$. Полагая

$$a^2 = \frac{a_{11} - a_{22}}{a_{11}}, \quad b^2 = -\frac{a_{11} - a_{22}}{a_{22}},$$

мы приведем уравнение (10) к виду

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0, \quad \text{где} \quad a \geqslant b > 0 \quad \text{и} \quad \frac{1}{a^2} + \frac{1}{b^2} = 1, \tag{14}$$

т. е. к виду

$$\left(\frac{x}{a} + \frac{y}{b}\right)\left(\frac{x}{a} - \frac{y}{b}\right) = 0.$$

Этому уравнению удовлетворяют точки двух прямых

$$\frac{x}{a} + \frac{y}{b} = 0 \quad \text{и} \quad \frac{x}{a} - \frac{y}{b} = 0,$$

пересекающихся в точке $(0, 0)$. Таким образом, в этом случае наша линия второго порядка состоит из *двух пересекающихся прямых.*

Определение 2. Случай 1 принято называть *гиперболическим* (при $a_{33} \neq 0$ — *невырожденным*, а при $a_{33} = 0$ — *вырожденным*).

Случай 2. *Коэффициенты a_{11} и a_{22} имеют один и тот же знак.* Умножая (если нужно) уравнение (10) на -1, мы, без ограничения общности, можем считать эти коэффициенты положительными:

$$a_{11} > 0, \quad a_{22} > 0.$$

Кроме того, переставляя (если нужно) координаты, мы можем считать, что

$$a_{11} \leqslant a_{22}.$$

Случай 2а. *Коэффициент a_{33} отрицателен.* Тогда, полагая

$$a^2 = -\frac{a_{33}}{a_{11}}, \quad b^2 = -\frac{a_{33}}{a_{22}},$$

мы приведем уравнение (10) к виду

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \ a \geqslant b > 0, \tag{15}$$

являющемуся каноническим уравнением *эллипса* (см. п. 2 § 1 гл. 5).

Случай 2б. *Коэффициент a_{33} равен нулю.* Тогда, полагая

$$a^2 = \frac{a_{11} + a_{22}}{a_{11}}, \quad b^2 = \frac{a_{11} + a_{22}}{a_{22}},$$

мы приведем уравнение (10) к виду

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad \text{где} \quad a \geqslant b > 0 \quad \text{и} \quad \frac{1}{a^2} + \frac{1}{b^2} = 1. \tag{16}$$

Этому уравнению удовлетворяет только *точка* $(0, 0)$.

Случай 2в. *Коэффициент a_{33} положителен.* Тогда, полагая

$$a^2 = \frac{a_{33}}{a_{11}}, \quad b^2 = \frac{a_{33}}{a_{22}},$$

мы приведем уравнение (10) к виду

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1, \ a \geqslant b > 0. \tag{17}$$

Этому уравнению не удовлетворяет ни одна точка плоскости, т. е. оно определяет *пустое множество.*

Определение 3. Случай 2 принято называть *эллиптическим* (при $a_{33} < 0$ — *действительным невырожденным*, при $a_{33} > 0$ — *мнимым невырожденным* и при $a_{33} = 0$ — *вырожденным*).

Рассмотрим теперь уравнения (11) кривых типа II.

Меняя (если нужно) ориентацию оси Ox, мы можем, без ограничения общности, считать, что коэффициенты a_{22} и a_{13} имеют разные знаки (напомним, что по условию $a_{22} \neq 0$ и $a_{13} \neq 0$). Тогда, полагая

$$p = -\frac{a_{13}}{a_{22}},$$

мы приведем уравнение (11) к виду

$$y^2 = 2px, \ p > 0, \tag{18}$$

являющемуся каноническим уравнением *параболы* (см. п. 1 § 1 гл. 5).

Для уравнений (12) кривых типа III возможны три различные случая.

Случай А. *Коэффициент* a_{33} *равен нулю.* Тогда после сокращения на a_{11} уравнение (12) приобретает вид

$$y^2 = 0. \tag{19}$$

Это уравнение определяет ту же прямую (ось Ox), что и уравнение $y = 0$. Однако чтобы оттенить различие между уравнениями $y^2 = 0$ и $y = 0$, мы будем говорить, что первое из них — *дважды взятая прямая* $y = 0$ или что оно определяет *пару совпадающих прямых.*

Случай Б. *Коэффициенты* a_{22} *и* a_{33} *имеют разные знаки.* Тогда, полагая

$$b^2 = -\frac{a_{33}}{a_{22}},$$

мы приведем уравнение (12) к виду

$$y^2 - b^2 = 0, \quad b > 0, \tag{20}$$

т. е. к виду

$$(y + b)(y - b) = 0.$$

Этому уравнению удовлетворяют координаты точек двух параллельных прямых

$$y + b = 0 \quad \text{и} \quad y - b = 0.$$

Таким образом, в рассматриваемом случае наша линия второго порядка состоит из *двух параллельных прямых.*

Случай В. *Коэффициенты* a_{22} *и* a_{33} *имеют одинаковые знаки.* Тогда, полагая

$$b^2 = \frac{a_{33}}{a_{22}},$$

мы приведем уравнение (12) к виду

$$y^2 + b^2 = 0, \quad b > 0. \tag{21}$$

Этому уравнению не удовлетворяет ни одна точка плоскости.

Определение 4. Случаи уравнений (11) и (12) называются *параболическими* (случай уравнения (11) — *невырожденным*, а уравнения (12) — *вырожденным*).

Таким образом, линии типа I могут быть двух видов: гиперболического и эллиптического (с дальнейшим подразделением на вырожденные и невырожденные, а в случае эллиптических линий — еще не действительные и мнимые). Напротив, линии типов II и III объединяются в один вид параболических линий (невырожденных, т. е. парабол, — в случае типа II, и вырожденных, — в случае типа III).

Определение 5. Уравнения (13)—(21) называются (*евклидово*) *каноническими уравнениями*, а описанная процедура — *приведением уравнения* (1) *к каноническому виду.*

Заметим, что эта терминология согласована с употреблением термина «каноническое уравнение» для эллипса, гиперболы и параболы в § 1 гл. 5.

Замечание 2 (очень важное!). Мы видим, что оставаясь в обыкновенной евклидовой плоскости, мы в качестве «линий второго порядка» получаем, в частности, точки (уравнение (16)) и пустые множества (уравнения (17) и (21)). При этом «линии», задаваемые уравнениями (17) и (21), оказываются одинаковыми, хотя сами эти уравнения различны (не сводятся одно к другому заменой координат). Чтобы устранить эти недостатки, целесообразно перейти в евклидову вещественно-комплексную плоскость (см. п. 5 § 3 гл. 2), являющуюся «естественной областью», в которой следует развивать теорию линий второго порядка и в которой эта теория приобретает законченный и изящный вид.

Начиная с этого места, мы будем всегда предполагать (*если только явно не оговорено противное*), *что все линии рассматриваются в евклидовой вещественно-комплексной плоскости.*

Конечно, мы при этом будем по-прежнему интересоваться лишь *вещественными линиями* второго порядка, т. е. линиями, задаваемыми уравнениями (1) с вещественными коэффициентами.

Поскольку уравнение (16) может быть представлено в виде

$$\left(\frac{x}{a}+i\frac{y}{b}\right)\left(\frac{x}{a}-i\frac{y}{b}\right)=0,$$

мы видим, что на вещественно-комплексной плоскости оно определяет *пару мнимых* (*т. е. невещественных*) *комплексно-сопряженных прямых*

$$y=\pm i\frac{b}{a}x,$$

пересекающихся в вещественной точке $(0,0)$.

Аналогично, уравнение (21) определяет *пару параллельных* (*т. е. непересекающихся*) *мнимых комплесно-сопряженных прямых*

$$y=\pm ib.$$

Что же касается уравнения (16), то оно определяет в евклидовой вещественно-комплексной плоскости некоторую новую, у нас еще не встречавшуюся, линию, называемую «мнимым эллипсом».

Дадим точное

Определение 6. *Мнимым эллипсом* называется линия, имеющая в некоторой (*канонической* для этой линии) евклидовой координатной системе уравнение вида

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=-1,$$

где $a \geqslant b > 0$.

При $a = b$ получаются уже известные нам (см. п. 6 § 3 гл. 4) окружности мнимого радиуса.

Эллипсы в прежнем смысле (т. е. в смысле определения 1 п. 2 § 1 гл. 5) мы будем теперь называть *действительными эллипсами.*

Замечание 3. Строго говоря, действительные эллипсы не являются эллипсами в смысле определения 1 п. 2 § 1 гл. 5. Последние являются лишь пересечениями действительных эллипсов с вещественной плоскостью. То же самое, конечно, относится и к линиям, которые мы будем теперь называть «гиперболами» и «параболами».

Различие между эллипсами на евклидовой вещественной плоскости и действительными эллипсами на евклидовой вещественно-комплексной плоскости проявляется, например, в том, что, в то время как эллипс в прежнем смысле не имеет асимптот, *действительный эллипс асимптоты имеет* (ими будут, очевидно, комплексно-сопряженные прямые $y = \pm \frac{bi}{a} x$).

Вместе с тем легко видеть, что развитая в § 1 гл. 5 теория эллипсов на вещественной плоскости сохраняется (с незначительными и само собой понятными изменениями) и для действительных эллипсов на вещественно-комплексной плоскости. Более того, эта теория может быть теперь построена и для мнимых эллипсов. Например, для мнимых эллипсов можно ввести *оси* (являющиеся вещественными прямыми), *центр* (являющийся вещественной точкой), *фокусы* (также являющиеся вещественными точками), *директрисы* (являющиеся мнимыми комплексно-сопряженными прямыми), доказать аналоги предложений 2 и 3 из п. 2 § 1 гл. 5 и т. д.

Упражнение. Докажите для мнимых эллипсов аналоги предложений 2 и 3 из п. 2 § 1 гл. 5.

Задание. Докажите, что

никакой эллипс (действительный или мнимый) на евклидовой вещественно-комплексной плоскости (а также никакая гипербола или парабола) не содержит ни одной прямой (даже мнимой).

Указание: см. доказательство аналогичного утверждения для гипербол на вещественной плоскости в п. 2 § 2 гл. 5.

Таблица линий второго порядка на евклидовой вещественно-комплексной плоскости

Тип линии	Вид линии	Невырожденные линии	Вырожденные линии
I	гиперболический	гипербола $\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$, $a>0,\quad b>0$	пара вещественных пересекающихся прямых $\frac{x^2}{a^2}-\frac{y^2}{b^2}=0$, $a>0,\quad b>0$, $\frac{1}{a^2}+\frac{1}{b^2}=1$
	эллиптический	эллипс а) действительный $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1,\quad a\geqslant b>0$ б) мнимый $\frac{x^2}{a^2}+\frac{y^2}{b^2}=-1$, $a\geqslant b>0$	пара мнимых (комплексно-сопряженных) прямых, пересекающихся в вещественной точке $\frac{x^2}{a^2}+\frac{y^2}{b^2}=0,\quad a\geqslant b>0$, $\frac{1}{a^2}+\frac{1}{b^2}=1$
II, III	параболический	парабола $y^2=2px,\quad p>0$	пара параллельных прямых а) вещественных и различных $y^2-b^2=0,\quad b>0$ б) мнимых (комплексно-сопряженных) $y^2+b^2=0,\quad b>0$ в) совпадающих (вещественных) $y^2=0$

Мы видим, что произвольная линия второго порядка является *либо эллипсом (действительным или мнимым)*,
либо гиперболой,
либо параболой,
либо парой прямых (мнимых или вещественных, пересекающихся, параллельных или совпадающих).

Таким образом, за исключением вырожденных кривых (распадающихся на две прямые) и мнимых кривых (имеющих не более одной вещественной точки), т. е. за исключением геометрически неинтересных случаев, любая линия второго порядка является одним из трех конических сечений, изученных в §§ 1—2 гл. 5. Впрочем, пересекающиеся прямые (мнимые и вещественные) также можно считать коническими сечениями (получающимися в вещественно-комплексном пространстве при пересечении

конуса плоскостью, проходящей через его вершину). Чтобы получить совпадающие прямые (т. е. дважды взятую прямую), следует рассмотреть плоскость, касающуюся конуса по его образующей, а чтобы получить параллельные прямые, нужно вместо конуса взять цилиндр (который можно рассматривать как конус с вершиной в несобственной точке).

2. Инварианты уравнений линий второго порядка

Изложенный в п. 1 способ приведения решает сразу две задачи: он не только дает каноническое (точнее, приведенное) уравнение, но и указывает каноническую координатную систему.

Однако для многих задач столь полная информация оказывается совершенно излишней: часто достаточно знать лишь приведенное уравнение (без знания соответствующей координатной системы) или даже только вид рассматриваемой линии (т. е. является ли она эллипсом, гиперболой и т. п.). В следующем пункте мы изложим существенно более простой метод, позволяющий найти вид кривой и ее приведенное (а значит, и каноническое) уравнение, не производя самого приведения, т. е. не находя соответствующей координатной системы. Этот метод основывается на некоторых алгебраических результатах, доказательству которых и будет посвящен настоящий пункт.

Коэффициенты уравнения

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0 \qquad (1)$$

линии второго порядка удобно представлять себе расположенными в виде матрицы

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

(для единства обозначений мы условливаемся здесь, что символ a_{ij} обозначает то же самое число, что и символ a_{ji}).

Определение 1. Определитель

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

этой матрицы называется *большим определителем* уравнения (1), а его минор

$$\delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}^2$$

— *малым определителем* уравнения (1). Число

$$S = a_{11} + a_{22}$$

(след матрицы определителя δ) называется *следом* уравнения (1).

Числа Δ, δ и S называются также *инвариантами уравнения* (1).

Предположим, что от координат x, y мы перешли к новым координатам x', y', получающимся поворотом прежних осей координат на некоторый угол θ. Тогда вместо уравнения (1) мы получим уравнение (1′) п. 1, коэффициенты которого выражаются формулами (3) п. 1. В частности, для следа

$$S' = a'_{11} + a'_{22}$$

уравнения (1′) мы получаем отсюда выражение

$$S' = a_{11}\cos^2\theta + 2a_{12}\cos\theta\sin\theta + a_{22}\sin^2\theta + \\ + a_{11}\sin^2\theta - 2a_{12}\cos\theta\sin\theta + a_{22}\cos^2\theta = a_{11} + a_{22} = S.$$

Таким образом, преобразованное уравнение (1) имеет тот же след, что и исходное уравнение (1).

Аналогично, для малого определителя

$$\delta' = \begin{vmatrix} a'_{11} & a'_{12} \\ a'_{21} & a'_{22} \end{vmatrix}$$

мы получаем, что

$$\begin{aligned}
\delta' &= a'_{11}a'_{22} - a'^2_{12} = \\
&= (a_{11}\cos^2\theta + 2a_{12}\cos\theta\sin\theta + a_{22}\sin^2\theta) \times \\
&\times (a_{11}\sin^2\theta - 2a_{12}\cos\theta\sin\theta + a_{22}\cos^2\theta) - \\
&- [(a_{22} - a_{11})\cos\theta\sin\theta + a_{12}(\cos^2\theta - \sin^2\theta)]^2 = \\
&= (a_{11}a_{22} - a_{12}^2)\cos^4\theta + \\
&+ 2[-a_{11}a_{12} + a_{12}a_{22} - (a_{22} - a_{11})a_{12}]\cos^3\theta\sin\theta + \\
&+ [a_{11}^2 - 4a_{12}^2 + a_{22}^2 - (a_{22} - a_{11})^2 + 2a_{12}^2]\cos^2\theta\sin^2\theta + \\
&+ 2[a_{12}a_{11} - a_{22}a_{12} + (a_{22} - a_{11})a_{12}]\cos\theta\sin^3\theta + \\
&+ (a_{22}a_{11} - a_{12}^2)\sin^4\theta = \\
&= (a_{11}a_{22} - a_{12}^2)(\cos^4\theta + 2\cos^2\theta\sin^2\theta + \sin^4\theta) = \\
&= (a_{11}a_{22} - a_{12}^2)(\cos^2\theta + \sin^2\theta)^2 = a_{11}a_{22} - a_{12}^2 = \delta.
\end{aligned}$$

Наконец, большой определитель для уравнения (1) имеет вид

$$\Delta' = \begin{vmatrix} a'_{11} & a'_{12} & a'_{13} \\ a'_{21} & a'_{22} & a'_{23} \\ a'_{31} & a'_{32} & a'_{33} \end{vmatrix} = a'_{31}\begin{vmatrix} a'_{12} & a'_{13} \\ a'_{22} & a'_{23} \end{vmatrix} - a'_{32}\begin{vmatrix} a'_{11} & a'_{13} \\ a'_{21} & a'_{23} \end{vmatrix} + a'_{33}\begin{vmatrix} a'_{11} & a'_{12} \\ a'_{21} & a'_{22} \end{vmatrix}.$$

Но простое вычисление (аналогичное приведенному выше для определителя δ') показывает, что

$$\begin{vmatrix} a'_{12} & a'_{13} \\ a'_{22} & a'_{23} \end{vmatrix} = \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \cos\theta - \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} \sin\theta$$

и

$$\begin{vmatrix} a'_{11} & a'_{13} \\ a'_{21} & a'_{23} \end{vmatrix} = \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \sin\theta + \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} \cos\theta.$$

Отсюда, после тривиальных преобразований мы получаем, что

$$a'_{31}\begin{vmatrix} a'_{12} & a'_{13} \\ a'_{22} & a'_{23} \end{vmatrix} - a'_{32}\begin{vmatrix} a'_{11} & a'_{13} \\ a'_{21} & a'_{23} \end{vmatrix} = a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} - a_{32}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}.$$

Кроме того, так как $a'_{33} = a_{33}$ (см. последнюю формулу (3) п. 1) и $\delta' = \delta$ (по уже доказанному), то

$$a'_{33}\begin{vmatrix} a'_{11} & a'_{12} \\ a'_{21} & a'_{22} \end{vmatrix} = a'_{33}\delta' = a_{33}\delta = a_{33}\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

Следовательно,

$$\Delta' = a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} - a_{32}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} + a_{33}\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} =$$

$$= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \Delta.$$

Посмотрим теперь, что происходит с величинами S, δ и Δ при сдвиге начала координат. Как мы знаем, при переносе начала координат в точку (x_0, y_0) коэффициенты уравнения (1) претерпевают изменение, описываемое формулами (6) п. 1. Из этих формул непосредственно явствует, что при этом переносе величины S и δ не изменяются, а определитель Δ переходит в определитель

$$\Delta' = \begin{vmatrix} a_{11} & a_{12} & a'_{13} \\ a_{21} & a_{22} & a'_{23} \\ a'_{31} & a'_{32} & a'_{33} \end{vmatrix},$$

где

$$a'_{13} = a'_{31} = a_{11}x_0 + a_{21}y_0 + a_{31},$$

$$a'_{23} = a'_{32} = a_{12}x_0 + a_{22}y_0 + a_{32},$$

$$a'_{33} = a_{11}x_0^2 + 2a_{12}x_0y_0 + a_{22}y_0^2 + 2a_{13}x_0 + 2a_{23}y_0 + a_{33} =$$

$$= a'_{13}x_0 + a'_{23}y_0 + (a_{31}x_0 + a_{32}y_0 + a_{33}).$$

Вычтя из третьей строки этого определителя первую строку, умноженную на x_0, и вторую строку, умноженную на y_0, мы получим, следовательно, определитель

$$\begin{vmatrix} a_{11} & a_{12} & a'_{13} \\ a_{21} & a_{22} & a'_{23} \\ a_{31} & a_{32} & a^*_{33} \end{vmatrix},$$

где

$$a^*_{33} = a_{31}x_0 + a_{32}y_0 + a_{33}.$$

Поэтому, вычтя из последнего столбца этого определителя первый столбец, умноженный на x_0, и второй столбец, умноженный на y_0, мы получим определитель

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \Delta.$$

Поскольку при этих преобразованиях определитель не меняется, тем самым доказано, что величины S, δ и Δ не меняются и при переносе начала координат.

Так как любое преобразование координат сводится к повороту координатных осей и к переносу начала координат, тем самым мы доказали следующее

Предложение 1. *При любом преобразовании прямоугольных координат x, y числа S, δ и Δ не меняются.*

Это объясняет термин «инвариант» для чисел S, δ и Δ.

Рассмотрим теперь число

$$K = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = a_{33}(a_{11} + a_{22}) - a_{13}^2 - a_{23}^2.$$

При повороте осей координат оно переходит в число

$$K' = \begin{vmatrix} a'_{11} & a'_{13} \\ a'_{31} & a^1_{33} \end{vmatrix} + \begin{vmatrix} a'_{22} & a'_{23} \\ a'_{32} & a'_{33} \end{vmatrix}.$$

Но, как легко проверить,

$$\begin{vmatrix} a'_{11} & a'_{13} \\ a'_{31} & a'_{33} \end{vmatrix} = (a_{33}a_{11} - a_{13}^2)\cos^2\theta + \\ + 2(a_{33}a_{12} - a_{13}a_{23})\cos\theta\sin\theta + (a_{33}a_{22} - a_{23}^2)\sin^2\theta,$$

и, аналогично,

$$\begin{vmatrix} a'_{22} & a'_{23} \\ a'_{32} & a'_{33} \end{vmatrix} = (a_{33}a_{22} - a_{23}^2)\cos^2\theta - \\ - 2(a_{33}a_{12} - a_{13}a_{23})\cos\theta\sin\theta + (a_{33}a_{11} - a_{13}^2)\sin^2\theta.$$

Поэтому

$$K' = (a_{33}a_{11} - a_{13}^2)(\cos^2\theta + \sin^2\theta) + (a_{33}a_{22} - a_{23}^2)(\cos^2\theta + \sin^2\theta) = \\ = (a_{33}a_{11} - a_{13}^2) + (a_{33}a_{22} - a_{23}^2) = K.$$

При переносе начала координат в точку (x_0, y_0) число K переходит (см. формулы (6) п. 1) в число

$$K' = \begin{vmatrix} a_{11} & a'_{13} \\ a'_{31} & a'_{33} \end{vmatrix} + \begin{vmatrix} a_{22} & a'_{23} \\ a'_{32} & a'_{33} \end{vmatrix},$$

которое равно (проверьте это!)

$$K + (a_{11}a_{22} - a_{12}^2)(x_0^2 + y_0^2) + 2(a_{22}a_{13} - a_{12}a_{23})x_0 + 2(a_{11}a_{23} - a_{12}a_{13})y_0 = \\ = K + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ -2x_0 & -2y_0 & x_0^2 + y_0^2 \end{vmatrix}.$$

Таким образом, справедливо следующее

Предложение 2. *При повороте координатных осей число K не меняется, а при переносе начала координат в точку (x_0, y_0) к нему прибавляется определитель*

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ -2x_0 & -2y_0 & x_0^2 + y_0^2 \end{vmatrix}. \qquad (2)$$

Определение 2. Число K называется *семиинвариантом* уравнения (1) (приставка «семи» означает «полу»).

Замечание 1. Обратим внимание на то, что при умножении уравнения на некоторое число $k \neq 0$ инвариант S умножается на k, инвариант δ и семиинвариант K умножаются на k^2, а инвариант Δ умножается на k^3. Поэтому геометрический смысл может иметь только равенство инвариантов S и Δ нулю (или знак их произведения) и знак инварианта δ и семиинварианта K.

3. Определение вида линии второго порядка по инвариантам ее уравнения

Согласно предложению 1 п. 2 инварианты S, δ и Δ имеют одни и те же значения как для исходного уравнения (1) п. 1, так и для соответствующего приведенного уравнения (см. п. 1, уравнения (10), (11) или (12)).

Но для уравнения (10) п. 1 эти инварианты выражаются формулами

$$S = a_{11} + a_{22},$$
$$\delta = a_{11}a_{22} \neq 0, \qquad (1)$$
$$\Delta = \delta a_{33}.$$

для уравнения (11) п.1 — формулами

$$S = a_{22} \neq 0,$$
$$\delta = 0, \tag{2}$$
$$\Delta = -a_{22}a_{13}^2 \neq 0,$$

и для уравнения (12) п. 2 — формулами

$$S = a_{22} \neq 0,$$
$$\delta = 0, \tag{3}$$
$$\Delta = 0.$$

Следовательно (здесь мы пользуемся принципом обращения), *уравнение* (1) п. 1 *тогда и только тогда принадлежит типу* I, *когда* $\delta \neq 0$.

Другими словами,

гиперболические и эллиптические линии характеризуются тем, что $\delta \neq 0$, *а параболические — тем, что* $\delta = 0$.

Кроме того,

невырожденные линии (*гиперболические, эллиптические или параболические*) *характеризуются тем, что* $\Delta \neq 0$, *а вырожденные — тем, что* $\Delta = 0$.

Таким образом, для типов линий второго порядка мы имеем следующую таблицу:

Тип	Инварианты
I	$\delta \neq 0$
II	$\delta = 0$, $\Delta \neq 0$
III	$\delta = 0$, $\Delta = 0$

Замечание 1. Эта таблица, в частности, доказывает корректность распределения линий второго порядка по типам, т. е. факт, что одно и то же уравнение (1) п. 1 не может быть приведено к уравнениям различных типов.

Чтобы научиться отличать по инвариантам гиперболические линии от эллиптических, рассмотрим более внимательно первые два уравнения (1). Эти уравнения показывают, что коэффициенты a_{11} и a_{22} приведенного уравнения (10) п. 1 являются корнями квадратного уравнения

$$\lambda^2 - S\lambda + \delta = 0. \tag{4}$$

Определение 1. Уравнение (4) называется *характеристическим уравнением* (для уравнения (1)).

Заметим, что *это уравнение имеет лишь вещественные корни* (ибо числа a_{11} и a_{22} вещественны).

По определению, эллиптические линии характеризуются тем, что для них числа a_{11} и a_{22} имеют один и тот же знак, т. е. тем, что их произведение положительно. Но по формулам Вьета это произведение равно свободному члену δ характеристического уравнения (4). Поэтому

для эллиптических линий $\delta > 0$.

Аналогично показывается, что

для гиперболических линий $\delta < 0$.

Кроме того, согласно третьему из уравнений (1) коэффициент a_{33} в уравнении (10) п. 1 равен Δ/δ. Поэтому

приведенное уравнение (10) п. 1 *может быть записано в виде*

$$\lambda_1 x^2 + \lambda_2 y^2 + \frac{\Delta}{\delta} = 0, \tag{5}$$

где λ_1, λ_2 — *корни характеристического уравнения* (4).

Таким образом, при $\delta \neq 0$ мы можем найти уравнение, к которому приводится уравнение (1) п. 1, не производя приведения (т. е. не отыскивая соответствующей системы канонических координат). Для этого достаточно вычислить инварианты S, δ и Δ, составить уравнение (4), найти его корни λ_1, λ_2 и написать уравнение (5).

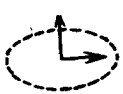
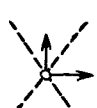
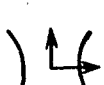
Осталось научиться отличать случай мнимого эллипса от действительного. Но это легко. На самом деле, при $\delta > 0$ знак корней λ_1, λ_2 совпадает, очевидно, со знаком их суммы S. С другой стороны, уравнение (5) показывает, что мы имеем мнимый эллипс, когда знаки корней совпадают со знаком определителя Δ, и действительный, когда эти знаки различны.

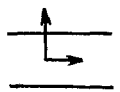
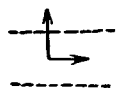
Окончательные результаты для случая $\delta \neq 0$, т. е. случая линий типа I, можно свести в следующую таблицу:

Линия	Характеристика по инвариантам
1. Эллипс	$\delta > 0$, $\Delta \neq 0$
а) действительный	$S\Delta < 0$
б) мнимый	$S\Delta > 0$
2. Гипербола	$\delta < 0$, $\Delta \neq 0$
3. Пара пересекающихся прямых	$\Delta = 0$
а) вещественных	$\delta < 0$
б) мнимых	$\delta > 0$

Рассмотрим теперь параболические линии ($\delta = 0$). Как мы уже знаем, невырожденные линии, т. е. линии, приводящиеся к уравнению (11) п. 1 (тип II), характеризуются условиями

Таблица линий второго порядка

№№ пп	Название	Схематический вид линии	Характеристика	Приведенное уравнение	Каноническое уравнение
[1]	Эллипс (действительный)		$\delta > 0, \quad S\Delta < 0$	$\lambda_1 x^2 + \lambda_2 y^2 + \frac{\Delta}{\delta} = 0$	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad a \geqslant b > 0$
[2]	Мнимый эллипс		$\delta > 0, \quad S\Delta > 0$		$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1, \quad a \geqslant b > 0$
[3]	Точка (пара мнимых пересекающихся прямых)		$\delta > 0, \quad \Delta = 0$		$\frac{x^2}{a^2} + \frac{y^2}{b} = 0, \quad \begin{matrix} a \geqslant b > 0, \\ \frac{1}{a^2} + \frac{1}{b^2} = 1 \end{matrix}$
[4]	Гипербола		$\delta < 0, \quad \Delta \neq 0$		$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \quad a > 0, \ b > 0$

[5]	Пара пересекающихся прямых (вещественных)		$\delta<0, \quad \Delta=0$		$\frac{x^2}{a^2}-\frac{y^2}{b^2}=0, \quad a\geqslant b>0, \quad \frac{1}{a^2}+\frac{1}{b^2}=1$
[6]	Парабола		$\delta=0, \quad \Delta\neq 0$	$Sy^2+2\sqrt{-\frac{\Delta}{S}}\,x=0$	$y^2=2px, \quad p>0$
[7]	Пара параллельных прямых (вещественных)		$\delta=0, \quad \Delta=0, \quad K<0$	$Sy^2+\frac{K}{S}=0$	$y^2-b^2=0, \quad b>0,$
[8]	Пара мнимых параллельных прямых		$\delta=0, \quad \Delta=0, \quad K>0$		$y^2+b^2=0, \quad b>0,$
[9]	Пара совпадающих прямых		$\delta=0, \quad \Delta=0, \quad K=0$		$y^2=0$

$\delta = 0$ и $\Delta \neq 0$. При этом согласно формулам (2) коэффициенты a_{22} и a_{13} приведенного уравнения (11) выражаются формулами

$$a_{22} = S, \quad a_{13} = \sqrt{-\frac{\Delta}{S}}.$$

Таким образом,

приведенное уравнение (11) п. 1 может *быть записано в виде*

$$Sy^2 + 2\sqrt{-\frac{\Delta}{S}}\, x = 0. \tag{6}$$

Следовательно, и для линий типа II мы можем написать приведенное уравнение, не производя фактического приведения.

Осталось рассмотреть лишь линии типа III, приводящиеся к уравнению (12) п. 1. Здесь мы воспользуемся семиинвариантом K.

Вспомним, как мы в п. 1 осуществляли приведение уравнения (1) п. 1 к простейшему виду. Сначала мы поворотом осей координат привели его к виду (5) п. 1. Как мы знаем, семиинвариант K при этом не изменился.

В интересующем нас сейчас случае уравнения (12) п. 1 определитель (2) п. 2 прибавляющийся при переносе начала координат в точку (x_0, y_0) к семиинварианту K, имеет вид

$$\begin{vmatrix} 0 & 0 & 0 \\ 0 & a_{22} & a_{23} \\ -2x_0 & -2y_0 & x_0^2 + y_0^2 \end{vmatrix}$$

и потому равен нулю. Следовательно,

если уравнение (1) п. 1 *приводится к уравнению* (12) п. 1, *то семиинвариант K для обоих уравнений — один и тот же.*

Но для уравнения (12) п. 1 этот семиинвариант выражается формулой

$$K = \begin{vmatrix} a_{22} & 0 \\ 0 & a_{33} \end{vmatrix} = a_{22}a_{33}.$$

Поэтому

уравнение (12) п. 1 *может быть записано в виде*

$$Sy^2 + \frac{K}{S} = 0. \tag{7}$$

Отсюда следует, что

при $K = 0$ мы имеем пару совпадающих прямых,

при $K > 0$ — пару мнимых параллельных прямых,

при $K < 0$ — пару вещественных параллельных прямых.

Окончательные результаты произведенного исследования линий второго порядка подытожены в таблице на стр. 516—517.

Обратим внимание на то, что в этой таблице все линии второго порядка на евклидовой вещественно-комплексной плоскости распределены по д е в я т и классам. Принадлежность линии тому или иному классу определяется знаками инвариантов δ и

$S\Delta$, фактом равенства или неравенства нулю инварианта Δ и (при $\delta = 0$, $\Delta = 0$) знаком семиинварианта K.

Из полученных результатов вытекает, в частности, что *уравнение второй степени*

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0$$

тогда и только тогда является (в прямоугольных координатах x, y)

а) *уравнением окружности (вещественного радиуса), когда*

$$a_{11} = a_{22}, \quad a_{12} = 0, \quad a_{11}a_{33} < a_{13}^2 + a_{23}^2,$$

б) *уравнением равнобочной гиперболы, когда*

$$a_{11} = -a_{22}, \quad \Delta \neq 0,$$

в) *уравнением пары перпендикулярных прямых, когда*

$$a_{11} = -a_{22}, \quad \Delta = 0.$$

Задание. Докажите это утверждение.

4. Линии второго порядка на аффинной плоскости. Теорема единственности

В аффинных координатах x, y произвольная линия второго порядка задается уравнением того же вида

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0, \qquad (1)$$

что и в прямоугольных. Но в аффинных координатах свобода в преобразованиях координат бóльшая, чем в прямоугольных, и потому уравнение (1) можно упрощать сильнее.

Замечание 1. Можно представлять себе, что мы по-прежнему находимся в евклидовой плоскости и лишь позволяем себе пользоваться аффинными координатами. Однако более последовательно считать, что мы перешли теперь в аффинную плоскость (вещественную или вещественно-комплексную), в которой именно аффинные координатные системы наиболее естественны.

Чтобы преобразованием аффинных координат привести уравнение (1) к наиболее простому виду, можно, например, перейти сначала от данных аффинных координат к некоторым прямоугольным (при этом преобразовании уравнение (1) сохранит свой вид, но, конечно, его коэффициенты изменятся), а затем уже известным нам образом перейти от этих прямоугольных координат к прямоугольным каноническим координатам. Если при этом получились канонические уравнения [1]—[5] из таблицы на стр. 516—517, то следует дополнительно сделать (нарушающее прямоугольность координат) преобразование

$$x' = \frac{x}{a},$$

$$y' = \frac{y}{b}.$$

Если получилось уравнение [6], то нужно сделать преобразование

$$x' = px,$$
$$y' = y,$$

а если получились уравнения [7] или [8], то — преобразование

$$x' = x,$$
$$y' = \frac{y}{b}.$$

В случае уравнения [9] можно канонические координаты x, y оставить прежними.

Этим будет доказана следующая теорема об аффинно-канонических уравнениях линий второго порядка:

Теорема 1. *Для любой линии второго порядка существует аффинная координатная система, в которой уравнением линии является одно из следующих девяти уравнений:*

$$[1]\ x^2 + y^2 = 1, \qquad [4]\ x^2 - y^2 = 1, \qquad [7]\ y^2 - 1 = 0,$$
$$[2]\ x^2 + y^2 = -1, \qquad [5]\ x^2 - y^2 = 0, \qquad [8]\ y^2 + 1 = 0,$$
$$[3]\ x^2 + y^2 = 0, \qquad [6]\ y^2 = 2x, \qquad [9]\ y^2 = 0.$$

Определение 1. Уравнения [1]—[9] называются *аффинно каноническими уравнениями*, а аффинные координаты, в которых уравнение данной линии второго порядка является аффинно каноническим уравнением, называются *аффинно каноническими координатами* (относительно этой линии).

Замечание 2. Изложенное доказательство теоремы 1 (хотя и вполне правильное) может быть подвергнуто серьезной критике с методологических позиций. Действительно, теорема 1 имеет чисто аффинный характер (справедлива в аффинной плоскости), а для ее доказательства мы привлекли евклидовы понятия (прямоугольные координаты). Хотелось бы иметь доказательство этой теоремы, не выходящее из рамок аффинной геометрии. Поиск такого доказательства имеет не только методологическое, но и чисто практическое значение, поскольку можно ожидать (и это, как мы увидим ниже, действительно оправдывается), что по пути мы найдем более простой способ приведения, не требующий решения характеристического уравнения (4) из п. 3 (или тригонометрического уравнения (4) из п. 1). Этим вопросом мы займемся в следующих пунктах, а пока обсудим некоторые другие вопросы принципиального характера, связанные с теоремой 1.

Замечание 3. Незначительное формальное усовершенствование изложенного выше доказательства теоремы 1 позволяет сделать его чисто «аффинным». Для этого достаточно заметить, что в п. 1 мы по существу пользовались прямоугольностью координат x, y только для геометрической интерпретации производимых преобразований. Поэтому с формальной точки зрения все приведение из п. 1 можно делать и в аффинных координатах (только теперь угол θ в формулах (2) п. 1 уже не будет иметь никакого явного гео-

метрического смысла). Это позволяет провести доказательство теоремы 1, не употребляя слов «прямоугольные координаты» и, следовательно, целиком в рамках аффинной геометрии.

Именно, применив преобразования из п. 1 к уравнению (1) в аффинных координатах x, y, мы сначала получим описанные в п. 1 «евклидово канонические» уравнения (конечно, это не будут настоящие евклидово канонические уравнения, поскольку соответствующие координаты не будут, вообще говоря, прямоугольными). Затем, применив указанные в доказательстве теоремы 1 дополнительные преобразования (имеющие целью «обратить коэффициенты a, b и p в единицы»), мы и получим аффинно канонические уравнения [1]—[9].

Конечно, этот формальный трюк нисколько не проясняет геометрической основы процесса приведения и не помогает его упростить (правда, некоторого малосущественного упрощения мы все же достигли: предварительный переход от данных аффинных координат к прямоугольным координатам оказывается ненужным).

Тем не менее некоторую пользу из него извлечь можно. Действительно, аналогично евклидовому случаю (см. п. 3) можно поставить задачу об отыскании аффинно канонического уравнения линии второго порядка непосредственно по ее уравнению (1), без того чтобы находить соответствующие аффинно канонические координаты. Чтобы решить эту задачу, мы заметим, что описанные выше «дополнительные» преобразования координат переводят евклидово канонические уравнения, принадлежащие различным классам, в различные аффинно канонические уравнения. С другой стороны, мы знаем, что эти классы могут быть охарактеризованы по инвариантам уравнения данной линии (см. таблицу на стр. 516—517), причем, в силу формально-алгебраического характера этого утверждения, оно верно независимо от того, являются ли наши координаты евклидовыми или любыми аффинными (нужно только, чтобы преобразования координат записывались формулами (2) из п. 1). Поэтому, несмотря на то, что величины S, δ и Δ (не говоря уже о K) при любых преобразованиях аффинных координат, вообще говоря, инвариантными не остаются, мы можем тот факт, что данная линия второго порядка имеет то или иное аффинно каноническое уравнение, охарактеризовать теми же условиями на инварианты:

№№ пп	Аффинно-канонические уравнения	Характеристика по инвариантам
[1]	$x^2+y^2=1$	$\delta>0,\quad S\Delta<0$
[2]	$x^2+y^2=-1$	$\delta>0,\quad S\Delta>0$
[3]	$x^2+y^2=0$	$\delta>0,\quad \Delta=0$
[4]	$x^2-y^2=1$	$\delta<0,\quad \Delta\neq 0$
[5]	$x^2-y^2=0$	$\delta<0,\quad \Delta=0$
[6]	$y^2=2x$	$\delta=0,\quad \Delta\neq 0$
[7]	$y^2-1=0$	$\delta=0,\quad \Delta=0,\quad K<0$
[8]	$y^2+1=0$	$\delta=0,\quad \Delta=0,\quad K>0$
[9]	$y^2=0$	$\delta=0,\quad \Delta=0,\quad K=0$

Происхождение аффинно канонических уравнений [1]—[9] из евклидово канонических уравнений, рассмотренных в предыдущих пунктах, показывает, что (в аффинной вещественно-комплексной плоскости) эти уравнения выражают соответственно следующие линии:

[1] действительные эллипсы,
[2] мнимые эллипсы,
[3] пары мнимых пересекающихся прямых,
[4] гиперболы,
[5] пары вещественных пересекающихся прямых,
[6] параболы,
[7] пары вещественных параллельных (но не совпадающих) прямых,
[8] пары мнимых параллельных прямых,
[9] пары совпадающих вещественных прямых.

Линии, принадлежащие каждому из этих классов, могут быть легко охарактеризованы чисто геометрически.

Действительно, рассмотрим следующие условия:

А) линия имеет более одной вещественной точки;

Б) линия имеет более одной оси симметрии;

В) линия пересекает (в вещественной точке) каждую свою ось симметрии;

Г) линия содержит две различные прямые (возможно, мнимые).

Тогда мы можем составить следующую таблицу:

Класс линий	Выполненные условия	Невыполненные условия
[1]	А, Б, В	Г
[2]	Б	А, В, Г
[3]	Б, В, Г	А
[4]	А, Б	В, Г
[5]	А, Б, В, Г	—
[6]	А, В	Б, Г
[7]	А, Б, Г	В
[8]	Б, Г	А, В
[9]	А, Б, В	Г

Согласно принципу обращения, условия, перечисленные в среднем столбце этой таблицы, однозначно характеризуют соответствующие линии.

Заметим теперь, что тот факт, что некоторая линия второго порядка удовлетворяет тем или иным из условий А—Г, не зависит от того, в какой системе координат она рассматривается. Поэтому

никаким преобразованием аффинных координат ни одно из уравнений [1]—[9] *нельзя перевести в другое.*

Другими словами, «лишних» линий среди линий [1]—[9] нет.

Это означает, что справедлива следующая теорема единственности канонических уравнений:

Теорема 2. *Для любой линии второго порядка ее аффинно-каноническое уравнение однозначно определено.*

Подчеркнем, что различных систем канонических координат может быть, тем не менее, много.

Замечание 4. Поскольку условия Б и В имеют евклидов характер, наше доказательство теоремы 2 — методологически не выдержано (ср. замечание 2).

При более методологически выдержанном подходе следовало бы найти аналогичные характеризующие условия чисто аффинного плана и проверить их выполнение (или невыполнение) непосредственно по уравнениям [1]—[9].

Упражнение. Сделайте это.

Замечание 5. Конечно, еще в пп. 1—3 нам нужно было доказать аналогичную теорему для евклидово канонических уравнений, т. е. доказать, что

никаким преобразованием прямоугольных координат ни одно из уравнений классов [1]—[9] п. 3 (*см. таблицу на стр.* 516—517) *нельзя перевести в другое уравнение.*

Мы этого не сделали в пп. 1—3 только для того, чтобы избежать излишних повторений. Сделаем это теперь.

Поскольку это утверждение для уравнений, принадлежащих различным классам [1]—[9] п. 3, очевидным образом вытекает из уже доказанного утверждения об аффинно канонических уравнениях, нам достаточно теперь лишь показать, что никаким преобразованием прямоугольных координат ни одно из уравнений этих классов нельзя перевести в другое уравнение того же класса.

Ясно, что для этого достаточно показать, что фигурирующие в этих уравнениях коэффициенты внутренним образом связаны с линией, т. е. могут быть геометрически охарактеризованы без использования каких-либо координат. Но это делается без всякого труда.

Например, для линий класса [1] (действительных эллипсов) числа a и b являются расстояниями от центра линии до ее вершин (точек пересечения с осями симметрии), причем если $a \neq b$, то a является бо́льшим из этих расстояний, а b — меньшим.

Для линий [2] (мнимых эллипсов) аналогичные расстояния равны ia и ib, что снова позволяет однозначно характеризовать числа a и b.

Задание. Охарактеризуйте геометрически коэффициенты остальных евклидово-канонических уравнений.

Замечание 6. Аналогичные результаты справедливы, конечно, и в вещественной (евклидовой или аффинной) плоскости.

Единственное отличие состоит в том, что линии [3] интерпретируются теперь как точки, а линии [2] и [8] объединяются в один класс нулевых линий (пустых множеств).

Обратим внимание на то, что, оставаясь на вещественной плоскости, мы не можем доказать (геометрически), что, например, уравнение [2] нельзя преобразовать в уравнение [8]. Поэтому теорему 2 мы можем доказать на вещественной плоскости только для линий второго порядка, содержащих более одной точки.

Полученные результаты позволяют без особого труда доказать обещанную в п. 2 § 2 гл. 2 теорему единственности для линий второго порядка:

Теорема 3. *Если на вещественно-комплексной плоскости (евклидовой или аффинной) два уравнения второй степени*

$$F(x, y) = 0, \quad G(x, y) = 0$$

определяют (в одной и той же координатной системе) одну и ту же линию, то существует такое число $k \neq 0$, *что*

$$G(x, y) = kF(x, y).$$

На вещественной плоскости аналогичное утверждение справедливо для линий, содержащих более одной точки.

Доказательство. Приведем уравнение

$$F(x, y) = 0 \tag{2}$$

к каноническому виду (евклидову — на евклидовой плоскости и аффинному — на аффинной). Это приведение состоит в переходе к некоторым каноническим координатам x', y' и в умножении получающегося уравнения на некоторое число. Следовательно, если

$$f(x', y') = 0 \tag{3}$$

— каноническое уравнение, а

$$x' = \varphi_1(x, y),$$
$$y' = \varphi_2(x, y)$$

— формулы перехода от координат x, y к координатам x', y', то тождественно по x и y имеет место равенство

$$F(x, y) = k_1 f(\varphi_1(x, y), \varphi_2(x, y)),$$

где k_1 — некоторое число.

Аналогично, приводя к каноническому виду уравнение

$$G(x, y) = 0, \tag{4}$$

мы получим тождество вида

$$G(x, y) = k_2 g(\psi_1(x, y), \psi_2(x, y)),$$

где k_2 — некоторое число,

$$x'' = \psi_1(x, y),$$
$$y'' = \psi_2(x, y)$$

— канонические координаты для уравнения (3), а

$$g(x'', y'') = 0 \qquad (5)$$

— соответствующее каноническое уравнение.

Но согласно теореме 2 канонические уравнения (3) и (5) совпадают (с точностью до обозначений координат) и, более того, сохраняются при переходе

$$x'' = h_1(x', y'),$$
$$y'' = h_2(x', y')$$

от канонических координат x', y' к каноническим координатам x'', y''.

Это означает, что

$$f(x', y') = g(h_1(x', y'), h_2(x', y')).$$

Кроме того, по определению

$$\psi_1(x, y) = h_1(\varphi_1(x, y), \varphi_2(x, y)),$$
$$\psi_2(x, y) = h_2(\varphi_1(x, y), \varphi_2(x, y)).$$

Поэтому

$$g(\psi_1(x, y), \psi_2(x, y)) = f(\varphi_1(x, y), \varphi_2(x, y)),$$

т. е.

$$\frac{G(x, y)}{k_2} = \frac{F(x, y)}{k_1},$$

что и требовалось доказать.

На вещественной плоскости теорема 2 справедлива только для линий, имеющих более одной точки, и потому теорема 3 справедлива тоже только для таких линий.

Теоремы 1 и 2 позволяют также полностью выяснить вопрос об аффинной эквивалентности линий второго порядка.

Согласно общему определению (см. определение 3 п. 3 гл. 2) две линии второго порядка называются *аффинно эквивалентными* («одинаковыми с аффинной точки зрения»), если существует аффинный автоморфизм (преобразование по равенству координат в двух аффинных координатных системах; см. п. 3 § 3 гл. 2), переводящий одну линию в другую.

Пусть

$$F(x, y) = 0 \qquad (6)$$

— уравнение некоторой линии в аффинной координатной системе Oxy. Перейдем к новой координатной системе $O'x'y'$. Преобразование по равенству координат в системах Oxy и $O'x'y'$ переводит линию (6) в линию, имеющую в координатах x', y' то же уравнение

$$F(x', y') = 0, \qquad (7)$$

что и линия (6) в координатах x, y. Чтобы получить уравнение

$$G(x, y) = 0 \tag{8}$$

линии (7) в исходных координатах x, y, нужно, следовательно, в уравнении (7) координаты x', y' выразить согласно формулам перехода через координаты x, y.

Обратно, пусть нам дана некоторая линия, имеющая в координатах x, y уравнение (8). Перейдя к координатам x', y', мы получим уравнение (7) той же линии в координатах x', y'. Обозначив в этом уравнении координаты x', y' через x, y, мы получим уравнение (6) аффинно эквивалентной линии в исходных координатах.

Все это можно выразить короче, сказав, что

два уравнения

$$F(x, y) = 0 \quad и \quad G(x, y) = 0$$

тогда и только тогда выражают в одних и тех же аффинных координатах аффинно эквивалентные линии, когда одно уравнение получается из другого некоторой заменой аффинных координат (и переобозначением координат).

В применении к линиям второго порядка это означает, что если некоторая линия, имеющая (в произвольных аффинных координатах x, y) уравнение второй степени

$$F(x, y) = 0,$$

имеет (в канонических для этой линии координатах x, y), скажем, уравнение

$$x^2 + y^2 = 1, \tag{9}$$

то она аффинно эквивалентна линии, имеющей в исходных координатах x, y уравнение (9).

Тем самым из теорем 1 и 2 мы получаем следующую теорему аффинной классификации линий второго порядка:

Теорема 4. *Любая линия второго порядка аффинно эквивалентна одной и только одной из девяти линий, выражающихся в произвольной (но фиксированной) системе аффинных координат уравнениями* [1]—[9].

Таким образом, например, *все эллипсы аффинно эквивалентны.*

Замечание 7. Последнее утверждение иногда формулируют следующим образом: «любой эллипс аффинно эквивалентен окружности». Такая формулировка [1]) совершенно правильна, но

[1]) Заметим, что она нам по существу известна из элементарной теории эллипса, поскольку рассмотренное в п. 2 § 1 гл. 5 сжатие является, очевидно, аффинным автоморфизмом.

принадлежит евклидовой геометрии, поскольку в аффинной геометрии понятие окружности отсутствует.

Замечание 8. Из того, что любой эллипс аффинно эквивалентен окружности, вытекает, что любым аффинно инвариантным свойством окружности обладают все эллипсы. Это позволяет очень просто получать (методами евклидовой геометрии) разнообразные свойства эллипсов. Например, любая окружность обладает центром симметрии, что является свойством аффинно инвариантным. Следовательно, центром симметрии обладает и любой эллипс.

Однако этот способ методологически невыдержан (аффинные свойства получаются евклидовыми методами!), и потому мы им пользоваться не будем. Он и опасен: можно ошибиться и счесть аффинно инвариантным свойством свойство, на самом деле таковыми не являющееся.

Конечно, об эллипсе мы говорили только для примера. Все сказанное с равным правом применимо, скажем, к гиперболе (здесь роль окружности играет, естественно, равнобочная гипербола).

Совершенно аналогичные результаты имеют место, конечно, и в евклидовом случае. В частности, имеет место следующая теорема евклидовой классификации линий второго порядка:

Теорема 5. *Любая линия второго порядка евклидово эквивалентна (конгруэнтна) одной и только одной линии, выражающейся в произвольной (но фиксированной) системе евклидовых координат евклидово каноническим уравнением.*

Таким образом, имеется бесконечное множество различных евклидово не эквивалентных линий второго порядка, в то время как различных аффинно не эквивалентных линий второго порядка имеется только девять.

В применении, скажем, к эллипсам (и гиперболам) теорема 5 утверждает, что два эллипса (две гиперболы) тогда и только тогда конгруэнтны, когда они имеют одни и те же полуоси. Аналогично, две параболы тогда и только тогда конгруэнтны, когда их параметры одинаковы.

5. Центры линий второго порядка

Мы теперь приступим к систематическому изучению аффинных свойств линий второго порядка с целью найти подходы к «аффинному» доказательству теоремы 1 п. 4.

При этом мы не будем ограничиваться вещественно-комплексным случаем и будем рассматривать линии второго порядка в аффинной плоскости *над произвольным полем* $\mathbb{K}$ *характеристики, отличной от двух* (т. е. таким, что $2a \neq 0$ при $a \neq 0$). Это не вызовет (по крайней мере на первых порах) никаких

существенных осложнений, но позволит применить полученные результаты, во-первых, к линиям на вещественной плоскости (случай, геометрически наиболее наглядный) и, во-вторых, к линиям на комплексной плоскости (случай, алгебраически наиболее простой).

Пусть

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0, \qquad (1)$$

— уравнение некоторой линии второго порядка в данной аффинной координатной системе Oxy.

Согласно формулам (6) п. 1 (справедливыми, очевидно, и в рассматриваемом сейчас общем случае), если мы хотим переносом начала координат уничтожить члены с x и y, то новое начало координат мы должны выбрать среди решений системы уравнений

$$\begin{aligned} a_{11}x + a_{12}y + a_{13} &= 0, \\ a_{21}x + a_{22}y + a_{23} &= 0. \end{aligned} \qquad (2)$$

Определение 1. Каждая точка (x_0, y_0), координаты которой удовлетворяют уравнениям (2), называется *центром* линии второго порядка (1).

Таким образом, точка (x_0, y_0) является центром линии второго порядка, если в координатной системе с началом в этой точке уравнение линии не содержит членов с x и y, т. е. имеет вид

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + a_{33} = 0. \qquad (3)$$

Поскольку это уравнение не меняется при замене x, y на $-x$, $-y$, начало координат является центром симметрии [1]) линии (3). Но, по условию, это начало координат представляет собой центр линии. Следовательно,

каждый центр линии второго порядка является ее центром симметрии.

Конечно, здесь предполагается, что говорить о центре симметрии имеет смысл, т. е. что кривая (3) содержит хотя бы одну точку (не является нулевой линией) [2]).

Обратное утверждение мы докажем в следующем виде:

Предложение 1. *На вещественной или вещественно-комплексной плоскости каждый центр симметрии линии второго порядка является ее центром.*

Доказательство. Пусть начало координат (0, 0) является центром симметрии линии (1). Мы должны показать, что

[1]) Ясно, что понятие центра симметрии имеет смысл в аффинной геометрии над любым полем (поскольку имеет смысл понятие «середина отрезка»).

[2]) В вещественно-комплексной (или комплексной) плоскости это условие выполнено автоматически.

тогда $a_{13} = 0$ и $a_{23} = 0$. Но поскольку точка (0, 0) является центром симметрии линии (1), вместе с некоторой точкой (x_0, y_0) уравнению (1) удовлетворяет и точка $(-x_0, -y_0)$, т. е. вместе с соотношением

$$a_{11}x_0^2 + 2a_{12}x_0y_0 + a_{22}y_0^2 + 2a_{13}x_0 + 2a_{23}y_0 + a_{33} = 0$$

имеет место и соотношение

$$a_{11}x_0^2 + 2a_{12}x_0y_0 + a_{22}y_0^2 - 2a_{13}x_0 - 2a_{23}y_0 + a_{33} = 0.$$

Вычитая эти соотношения (и сокращая на 2), мы получим равенство

$$a_{13}x_0 + a_{23}y_0 = 0.$$

При $a_{13} \neq 0$ или $a_{23} \neq 0$ это показывает, что любая точка рассматриваемой линии второго порядка принадлежит прямой

$$a_{13}x + a_{23}y = 0.$$

Примером линии второго порядка, расположенной на прямой, является дважды взятая прямая. Наиболее общее уравнение такой линии имеет вид

$$(Ax + By + C)^2 = 0.$$

Ясно, что все центры симметрии такой линии ей принадлежат. Поэтому если точка (0, 0) является ее центром симметрии, то $C = 0$ и, следовательно,

$$a_{13} = AC = 0, \quad a_{23} = BC = 0.$$

Для случая вещественно-комплексной плоскости это доказывает предложение 1, поскольку дважды взятая прямая является, как мы знаем из теоремы 2 п. 4, единственной линией второго порядка, расположенной на прямой.

На вещественной плоскости, кроме дважды взятых прямых, имеется еще лишь один класс ненулевых линий второго порядка, располагающихся на прямой, а именно — точки. Поскольку центром симметрии множества, состоящего из одной точки, является сама эта точка, уравнение (1), изображающее линию второго порядка, являющуюся точкой и имеющую центр симметрии в начале координат, должно удовлетворяться координатами точки (0, 0) и только ее координатами.

Следовательно, если линия второго порядка (1) является точкой (0, 0), то, во-первых,

$$a_{33} = 0$$

и, во-вторых, уравнение

$$a_{11}x^2 + 2a_{13}x = 0,$$

получающееся из уравнения (1) подстановкой $y = 0$, должно удовлетворяться только при $x = 0$, что возможно лишь при $a_{11} = 0$ и $a_{13} \neq 0$ или при $a_{11} \neq 0$ и $a_{13} = 0$.

Аналогично показывается, что либо $a_{22} = 0$ и $a_{23} \neq 0$, либо $a_{22} \neq 0$ и $a_{23} = 0$.

Если $a_{13} = 0$ и $a_{23} \neq 0$, то $a_{11} \neq 0$ и $a_{22} = 0$, так что уравнение (1) имеет вид

$$a_{11}x^2 + 2a_{12}xy + 2a_{23}y = 0,$$

$$a_{11} \neq 0, \qquad a_{23} \neq 0$$

и поэтому имеет ненулевые решения (x_0, y_0) (достаточно взять любое $x_0 \neq 0$, для которого $a_{12}x_0 + a_{23} \neq 0$). Следовательно, случай $a_{13} = 0$, $a_{23} \neq 0$ невозможен.

Аналогично показывается, что случай $a_{13} \neq 0$, $a_{23} = 0$ также невозможен.

Пусть теперь $a_{13} \neq 0$ и $a_{23} \neq 0$. Тогда $a_{11} = 0$ и $a_{22} = 0$, и уравнение (1) имеет вид

$$2a_{12}xy + 2a_{13}x + 2a_{23} = 0,$$

$$a_{12} \neq 0, \quad a_{13} \neq 0, \quad a_{23} \neq 0$$

и потому также имеет ненулевые решения. Следовательно, этот случай также невозможен. Поэтому $a_{13} = a_{23} = 0$.

Тем самым, предложение 1 полностью доказано.

Для случая произвольного поля $\mathbb{K}$ (характеристики, отличной от двух) изложенное доказательство показывает лишь, что *если линия второго порядка либо не расположена на прямой, либо является дважды взятой прямой или точкой, то любой центр симметрии этой линии служит ее центром.*

Замечание 1. Определение 2 центра линии второго порядка привязано к определенному уравнению линии и определенной координатной системе, и потому неясно, не зависит ли оно от выбора этой координатной системы? Теперь мы охарактеризовали центры линии второго порядка (по крайней мере, на вещественной или вещественно-комплексной плоскости) как ее центры симметрии, т. е. способом, не зависящим от системы координат. Поэтому центры линии второго порядка на вещественной или вещественно-комплексной плоскости определены корректно.

В общем случае любого поля $\mathbb{K}$ для доказательства корректности определения центра следует найти, как преобразуются коэффициенты уравнения (1) при произвольном преобразовании

координат, и на этой основе показать, что для преобразованного уравнения центры будут те же самые, что и для исходного уравнения.

Упражнение. Докажите в общем случае корректность определения центра линий второго порядка.

Для системы уравнений (2) возможны следующие три случая:

I. Уравнения (2) имеют единственное решение.

II. Уравнения (2) не имеют ни одного решения.

III. Уравнения (2) имеют бесконечно много решений.

Определение 2. В случае I линия второго порядка называется *центральной*. Она имеет единственный центр.

В случаях II и III линия (1) либо вообще не имеет центров, либо они заполняют некоторую прямую, называемую *прямой центров* (уравнения (2) не могут удовлетворяться тождественно, ибо, по условию, хотя бы одно из чисел a_{11}, a_{12}, a_{22} отлично от нуля). Такие линии второго порядка называются *нецентральными*.

Примерами центральных линий являются эллипсы, гиперболы и пары пересекающихся прямых, а нецентральных — параболы (центров нет) и пары параллельных прямых (имеется прямая центров).

Из алгебры известно, что система двух уравнений с двумя неизвестными тогда и только тогда имеет единственное решение, когда ее определитель отличен от нуля. Но определитель системы (2) — это знакомый нам определитель

$$\delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

Следовательно,

центральные линии второго порядка характеризуются условием $\delta \neq 0$, *а нецентральные — условием* $\delta = 0$.

Далее, система двух уравнений с двумя неизвестными тогда и только тогда имеет семейство решений, зависящих от одного параметра, когда уравнения пропорциональны. Но пропорциональность уравнений (2) означает пропорциональность первых двух строк определителя

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Следовательно, если имеет место случай III, то $\Delta = 0$.

Обратно, пусть $\Delta = 0$ (и $\delta = 0$). Покажем, что тогда первые две строки определителя Δ пропорциональны (т. е. что имеет место случай III).

Действительно, так как $\delta = 0$, то строки определителя δ пропорциональны. Хотя бы одна строка этого определителя по условию отлична от нуля. Пусть, для определенности, отлична от нуля первая строка. Тогда

$$a_{21} = ka_{11}, \quad a_{22} = ka_{12},$$

где k — некоторое число.

Разложив определитель Δ по элементам последней строки, мы, следовательно, получим, что

$$\Delta = a_{31}\begin{vmatrix} a_{11} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} - a_{32}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} + a_{33}\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} =$$

$$= (ka_{31} - a_{32})\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}.$$

Поскольку $\Delta = 0$, отсюда вытекает, что либо $ka_{31} - a_{32} = 0$, либо

$$\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} = 0$$

и, следовательно, существует такое число l, что

$$a_{13} = la_{11}, \quad a_{23} = la_{21},$$

и потому $a_{23} = lka_{11} = ka_{13}$. В обоих случаях $a_{32} = ka_{31}$, так что первые две строки определителя Δ пропорциональны.

Тем самым доказано, что

случай II *характеризуется условиями* $\delta = 0$, $\Delta \neq 0$, *а случай* III — *условиями* $\delta = 0$, $\Delta = 0$.

Замечание 2. Сравнивая полученные утверждения с результатами п. 1, мы немедленно получаем следующую геометрическую интерпретацию типа линии второго порядка (введенного в п. 1 довольно формальным образом):

линии типа I — *это центральные линии,*

линии типа II — *это линии, не имеющие центра,*

линии типа III — *это линии, обладающие прямой центров.*

Таким образом, на вещественно-комплексной плоскости

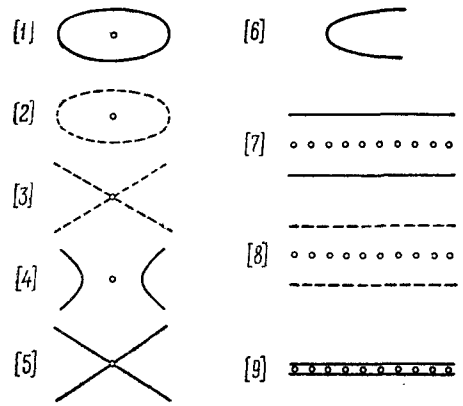
центральные линии — это эллипсы (действительные или мнимые), гиперболы и пары пересекающихся прямых (вещественных или мнимых),

линии, не имеющие центра, — это параболы (и только они),

линии, обладающие прямой центров — это пары параллельных (или совпадающих) прямых (вещественных или мнимых).

Для эллипса (действительного) и гиперболы центр совпадает с центром в смысле пп. 2 и 3 § 1 гл. 5. Для пары пересекающихся (вещественных или мнимых) прямых центром является их точка пересечения. Для параллельных (вещественных) прямых прямой центров является прямая, им параллельная и находящаяся от них на половинном расстоянии. Прямой центров дважды взятой прямой является сама эта прямая.

Схематически расположение центра (центров) изображено на следующих диаграммах:



6. Асимптоты и диаметры линий второго порядка

В соответствии с определением 1 п. 6 § 1 гл. 5, направление $l:m$ на плоскости называется *асимптотическим* для линии второго порядка

$$a_{11}x^2+2a_{12}xy+a_{22}y^2+2a_{13}x+2a_{23}y+a_{33}=0, \tag{1}$$

если

$$a_{11}l^2+2a_{12}lm+a_{22}m^2=0. \tag{2}$$

Пусть, как и выше,

$$\delta=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

Легко видеть, что

если основное поле $\mathbb{K}$ *содержит элемент* $\sqrt{-\delta}$, *т. е. элемент, квадрат которого равен* $-\delta$, *то решения уравнения* (2) *могут быть выражены любой из двух формул*:

$$\begin{aligned} l:m&=(-a_{12}\pm\sqrt{-\delta}):a_{11},\\ l:m&=a_{22}:(-a_{12}\mp\sqrt{-\delta}); \end{aligned} \tag{3}$$

в противном случае уравнение (2) *решений не имеет.*

Замечание 1. При $a_{11}\neq 0$ и $a_{22}\neq 0$ формулы (3) равносильны. В случае же, когда $a_{11}=0$ или $a_{22}=0$, следует пользоваться той из формул (3), которая имеет смысл.

В частности, на вещественной плоскости (т. е. при $\mathbb{K}=\mathbb{R}$)

эллиптические линии (*с* $\delta>0$) *не имеют асимптотических направлений*; *параболические линии* (*с* $\delta=0$) *имеют одно асимптотическое направление* (*характеризующееся отношением*

$-a_{12}:a_{11}=-a_{22}:a_{12}$), *а гиперболические линии* ($c\ \delta<0$) *имеют два асимптотических направления.*

Для гиперболы асимптотическими направлениями являются направления ее асимптот (этим и объясняется термин «асимптотическое направление»), а для пары пересекающихся прямых — направления этих прямых.

Для параболы единственным асимптотическим направлением является направление ее оси, а для пары параллельных (или совпадающих) прямых — направления этих прямых.

На вещественно-комплексной плоскости (когда $\mathbb{K}=\mathbb{C}$),

эллиптические линии имеют два комплексно-сопряженных асимптотических направления, параболические линии имеют одно (вещественное) асимптотическое направление, а гиперболические линии — два вещественных асимптотических направления.

Заметим, что асимптотическое направление параболической линии вещественно и тогда, когда линия является парой мнимых параллельных прямых (и совпадает с направлением этих прямых, которое, следовательно, вещественно).

На комплексной плоскости (т. е. при $\mathbb{K}=\mathbb{C}$) различение эллиптических и гиперболических линий смысла не имеет. На этой плоскости можно говорить только о центральных линиях ($c\ \delta\neq 0$) и нецентральных (параболических) линиях (с $\delta=0$). При этом

центральные линии имеют два асимптотических направления, а нецентральные — только одно.

Замечание 1. Мы видим, в частности, что асимптотические направления линии второго порядка могут быть охарактеризованы (по крайней мере на вещественной плоскости) чисто геометрически. Это показывает, что хотя в определении асимптотического направления и участвует система координат, но на самом деле асимптотические направления от выбора координатной системы не зависят (определены корректно).

В общем случае корректность определения асимптотических направлений следует (подобно корректности определения центра; см. замечание 1 из п. 5) доказывать вычислением, выяснив предварительно, как меняются коэффициенты уравнения (1) при преобразовании координат.

Упражнение. Докажите корректность определения асимптотических направлений.

На вещественной, комплексной или вещественно-комплексной плоскости асимптотами (в смысле п. 2 § 2 гл. 5) обладают из всех линий второго порядка только гиперболы и пары параллельных (или совпадающих) прямых (в последнем случае асимптотой является любая прямая, параллельная этим прямым, но от них отличная). Напротив, асимптотами в смысле п. 3 § 1 гл. 5 обладают (на вещественной плоскости) кроме

гипербол также пары (необязательно различных) прямых (безразлично, параллельных или пересекающихся), причем теперь асимптотами будут уже сами эти прямые и только они (ибо постоянная величина, равная нулю, имеет предел 0). Впрочем, допуская определенную непоследовательность, обычно говорят об асимптотах (в смысле п. 3 § 1 гл. 5) только тогда, когда «уходящая в бесконечность» ветвь линии, к которой приближается асимптота, не является прямой. В этом смысле, из всех линий второго порядка асимптотами обладают лишь гиперболы.

Пусть теперь $l:m$ — произвольное неасимптотическое направление.

Поскольку

$$a_{11}l^2+2a_{12}lm+a_{22}m^2=(a_{11}l+a_{12}m)\,l+(a_{12}l+a_{22}m)\,m\neq 0, \quad (4)$$

то хотя бы одно из чисел $a_{11}l+a_{12}m$ и $a_{12}l+a_{22}m$ отлично от нуля. Поэтому уравнение

$$(a_{11}l+a_{12}m)\,x+(a_{12}l+a_{22}m)\,y+(a_{13}l+a_{23}m)=0 \quad (5)$$

определяет на плоскости некоторую прямую.

Определение 1. Прямая (5) называется *диаметром* линии второго порядка, сопряженным с неасимптотическим направлением $l:m$.

Прямые направления $l:m$ имеют параметрические уравнения вида

$$x=x_0+lt,$$
$$y=y_0+mt.$$

Предположим, что одна из таких прямых пересекает линию (1) в двух различных точках $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$. Будем считать, что за точку (x_0, y_0) на такой прямой выбрана середина отрезка $\overline{M_1M_2}$. Тогда

$$x_0=\frac{x_1+x_2}{2}, \quad y_0=\frac{y_1+y_2}{2}$$

и, одновременно,

$$x_1=x_0+lt_1, \quad x_2=x_0+lt_2,$$
$$y_1=y_0+mt_1, \quad y_2=y_0+mt_2,$$

где t_1 и t_2 — корни уравнения

$$(a_{11}l^2+2a_{12}lm+a_{22}m^2)\,t^2+$$
$$+2\,[a_{11}lx_0+a_{12}(ly_0+mx_0)+a_{22}my_0+a_{13}l+a_{23}m]\,t+$$
$$+a_{11}x_0^2+2a_{12}x_0y_0+a_{22}y_0^2+2a_{13}x_0+2a_{23}y_0+a_{33}=0$$

(см. п. 2 § 2 гл. 5).

Следовательно,

$$x_0=\frac{x_0+lt_1+x_0+lt_2}{2}=x_0+\frac{l}{2}(t_1+t_2)$$

и, аналогично,

$$y_0 = y_0 + \frac{m}{2}(t_1 + t_2).$$

Поскольку числа l и m одновременно нулю не равны, эти равенства возможны тогда и только тогда, когда

$$t_1 + t_2 = 0.$$

Но если сумма корней квадратного уравнения равна нулю, его средний коэффициент также равен нулю. Поэтому при указанном выборе точки (x_0, y_0) имеет место равенство

$$a_{11}lx_0 + a_{12}(ly_0 + mx_0) + a_{22}my_0 + a_{13}l + a_{23}m = 0,$$

т. е. равенство

$$(a_{11}l + a_{12}m)x_0 + (a_{12}l + a_{22}m)y_0 + (a_{13}l + a_{23}m) = 0.$$

Тем самым доказано следующее

Предложение 1. *Если среди прямых неасимптотического направления $l : m$ имеются прямые, пересекающие линию* (1) *в двух точках, то диаметр* (5), *сопряженный с направлением $l : m$, является прямой, содержащей середины хорд, высекаемых на прямых этого направления линией* (1).

Определение 2. Направление $l : m$ мы назовем *хордальным* (по отношению к данной линии второго порядка (1)), если существуют по крайней мере две прямые этого направления, каждая из которых пересекает линию (1) в двух различных точках.

В этой терминологии предложение 1 утверждает, что

диаметр, сопряженный с произвольным хордальным неасимптотическим направлением, однозначно характеризуется как прямая, содержащая середины хорд этого направления.

На вещественной (а также на вещественно-комплексной и комплексной) плоскости каждое неасимптотическое направление эллипса, параболы, гиперболы и пары несовпадающих прямых является, очевидно, хордальным. Поэтому для этих линий предложение 1 характеризует все диаметры. При этом диаметры эллипса, параболы и гиперболы оказываются совпадающими с их диаметрами в смысле п. 5 § 2 гл. 5, а диаметрами пары пересекающихся прямых являются все прямые, проходящие через их точку пересечения. Что же касается пары параллельных (различных) прямых, то они имеют единственный диаметр — прямую, им параллельную и находящуюся от них на половинном расстоянии (т. е. прямую центров).

Найдем теперь диаметры линии второго порядка, являющейся дважды взятой прямой:

$$(Ax + By + C)^2 = 0.$$

В развернутом виде уравнение такой линии имеет вид

$$A^2x^2 + 2ABxy + B^2y^2 + 2ACx + 2BCy + C^2 = 0.$$

Следовательно, уравнение (5) имеет для такой линии вид

$$(A^2l + ABm)x + (ABl + B^2m)y + (ACl + BCm) = 0.$$

Сокращая на $Al + Bm$ (что возможно, ибо в силу условия неасимптотичности $Al + Bm \neq 0$), мы получаем уравнение

$$Ax + By + C = 0.$$

Следовательно,

любой диаметр линии второго порядка, являющейся дважды взятой прямой, совпадает с этой прямой.

Таким образом, и для этих кривых диаметр может быть охарактеризован инвариантным геометрическим образом.

Естественно считать, что дважды взятую прямую каждая прямая (от нее отличная) пересекает в двух совпадающих точках. Поскольку середина вырожденного отрезка совпадает с его концами, отсюда следует, что и для дважды взятой прямой диаметр может быть охарактеризован как прямая, содержащая середины хорд данного направления.

Поскольку определение 1 диаметра использует координатные системы, возникает вопрос о его корректности (независимости от координатной системы). Для дважды взятой прямой ответ на этот вопрос — утвердительный, поскольку такая линия имеет единственный диаметр, совпадающий с этой прямой. Для других линий второго порядка утвердительный ответ обеспечивается предложением 1 для диаметров, сопряженных с хордальными направлениями. В частности, для всех диаметров ответ утвердительный (на плоскости над любым полем) для пар прямых (пересекающихся или параллельных), а на вещественной, вещественно-комплексной или комплексной плоскости — для эллипса, гиперболы и параболы.

На вещественно-комплексной плоскости это полностью решает вопрос, а на вещественной плоскости корректность определения диаметров остается открытой только для нулевых линий и линий, состоящих из одной точки.

Общее доказательство корректности определения диаметра можно получить (так же как и общее доказательство корректности определения центра; см. п. 5), исследовав, как меняются коэффициенты уравнения (1) при преобразовании координат.

Упражнение. Проведите доказательство корректности определения диаметра.

Пусть (x_0, y_0) — произвольный центр линии второго порядка (1). Тогда, по определению,

$$a_{11}x_0 + a_{12}y_0 + a_{13} = 0,$$
$$a_{21}x_0 + a_{22}y_0 + a_{23} = 0.$$

Умножая первое уравнение на l, а второе на m и складывая, мы получаем, что

$$(a_{11}l + a_{12}m)\,x_0 + (a_{12}l + a_{22}m)\,y_0 + (a_{13}l + a_{23}m) = 0.$$

Этим доказано, что

каждый диаметр линии второго порядка проходит через любой ее центр.

В частности, отсюда вытекает, что

если линия второго порядка обладает прямой центров, то эта линия имеет единственный диаметр, совпадающий с этой прямой.

Таким образом, для линии, являющейся парой параллельных (или совпадающих) прямых, имеется единственный диаметр — ее прямая центров. Впрочем, этот факт мы выше уже доказали другим способом.

Определение 3. Направления $l : m$ и $l' : m'$ называются *сопряженными* (относительно данной кривой второго порядка (1)), если

$$(a_{11}l + a_{12}m)\,l' + (a_{21}l + a_{22}m)\,m' = 0,$$

т. е. если

$$a_{11}ll' + a_{12}(lm' + l'm) + a_{22}mm' = 0. \tag{6}$$

Поскольку равенство (6) симметрично относительно $l : m$ и $l' : m'$, то *отношение сопряженности симметрично*, т. е.

если направление $l : m$ сопряжено с направлением $l' : m'$, то направление $l' : m'$ сопряжено с направлением $l : m$.

Направление $l : m$ тогда и только тогда *самосопряжено*, т. е. сопряжено с самим собой, когда

$$a_{11}l^2 + 2a_{12}lm + a_{22}m^2 = 0,$$

т. е. когда оно является асимптотическим направлением.

Определение 4. Направление $l : m$ называется *особым*, если оно сопряжено с к а ж д ы м направлением $l' : m'$.

Ясно, что для того чтобы направление $l : m$ было особым, необходимо и достаточно, чтобы

$$\begin{aligned} a_{11}l + a_{12}m &= 0, \\ a_{21}l + a_{22}m &= 0. \end{aligned} \tag{7}$$

Поскольку система двух уравнений с двумя неизвестными тогда и только тогда имеет нетривиальное решение, когда ее определитель равен нулю, и поскольку определителем системы (7) служит определитель δ, мы видим, что

особые направления могут быть только у нецентральных (параболических) линий.

При $\delta = 0$ каждое из уравнений (7) является следствием другого, и потому эти уравнения имеют единственное (с точностью до пропорциональности) нетривиальное решение $l : m =$

$= -a_{12} : a_{11} = -a_{22} : a_{12}$. Поскольку отношение $-a_{12} : a_{11} = -a_{22} : a_{12}$ характеризует, как мы уже знаем, асимптотическое направление параболической линии, тем самым доказано, что

любая параболическая линия имеет единственное особое направление, совпадающее с ее асимптотическим направлением.

Направление $l' : m'$, сопряженное с неособым направлением $l : m$, однозначно характеризуется равенством

$$l' : m' = -(a_{21}l + a_{22}m) : (a_{11}l + a_{12}m).$$

Поэтому, если направление $l : m$ не асимптотично (и, в частности, не особо), то сопряженное с ним направление $l' : m'$ может быть описано как направление диаметра, сопряженного с направлением $l : m$, поскольку, как показывает уравнение (5), направление этого диаметра как раз и характеризуется отношением $-(a_{12}l + a_{22}m) : (a_{11}l + a_{12}m)$.

Определение 5. Диаметры линии второго порядка называются *сопряженными*, если они оба имеют неасимптотические направления (ситуация, возможная только для центральных линий) и эти направления сопряжены. Согласно только что сказанному,

диаметры (центральной) линии второго порядка тогда и только тогда сопряжены, когда каждый из них сопряжен с направлением другого.

Для центральных линий второго порядка каждый из сопряженных диаметров, имеющих хордальные направления, может быть описан как прямая, содержащая середины хорд, которые данная линия высекает на прямых, параллельных другому диаметру.

Эти утверждения геометрически характеризуют сопряженные направления, что, в частности, доказывает корректность определения сопряженности направлений (и диаметров).

Замечание 2. На евклидово-проективной плоскости направления (собственных) прямых находятся в биективном соответствии с точками несобственной прямой. Будем называть несобственные точки *сопряженными*, если соответствующие им направления сопряжены. Предположим, что рассматриваемая линия второго порядка центральна. Тогда для любой несобственной точки M существует единственная несобственная точка M', сопряженная с точкой M. Следовательно, сопоставив каждой точке M сопряженную точку M', мы получим некоторое однозначно определенное отображение T несобственной прямой на себя. Поскольку отношение сопряженности симметрично, это отображение является инволюцией. Для эллиптических линий эта инволюция не имеет неподвижных точек, а для гиперболических линий она обладает двумя неподвижными точками (соответствующими асимптотическим направлениям). В параболическом случае отображение T также определено и представляет собой отображение, переводящее всю несобственную прямую в одну «особую» точку (соответствующую особому направлению). Если рассматривать отношение $x = l : m$ как «координату» на несобственной прямой, то отображение T будет задаваться дробно-линейной функцией

$$x' = -\frac{a_{21}x + a_{22}}{a_{11}x + a_{12}}.$$

На основании этой связи между линиями второго порядка и инволюциями на несобственной прямой любая инволюция на произвольной прямой, выражающаяся дробно-линейной функцией, называется *эллиптической*, если она не имеет неподвижных точек, *гиперболической*, если она имеет две неподвижные точки, и *параболической*, если она имеет одну неподвижную точку (и представляет собой отображение всей прямой в эту точку). Ср. замечание 3 п. 6 § 3 гл. 4.

7. Приведение уравнений линий второго порядка к простейшему виду

Направления координатных осей характеризуются отношениями 1 : 0 и 0 : 1. Отсюда и из формулы (6) п. 6 непосредственно вытекает, что

направления координатных осей тогда и только тогда сопряжены, когда

$$a_{12} = 0.$$

Кроме того, мы знаем (см. п. 5), что

начало координат (0, 0) *тогда и только тогда является центром линии второго порядка, когда*

$$a_{13} = a_{23} = 0.$$

Тот факт, что начало координат является центром линии второго порядка, а направления координатных осей сопряжены, означает для центральных линий (тип. I), что *координатные оси являются сопряженными диаметрами.* В этом случае уравнение линии имеет вид

$$a_{11}x^2 + a_{22}y^2 + a_{33} = 0, \qquad (1)$$

причем $a_{11} \neq 0$ и $a_{22} \neq 0$ (ибо в противном случае $\delta = 0$).

Для нецентральных линий условие сопряженности осей однозначно определяет направление одной из них (оно должно быть особым, т. е. асимптотическим). Для определенности мы будем считать, что *особое направление имеет ось абсцисс.* Направление оси ординат мы тогда можем выбрать произвольно (поскольку любое направление сопряжено с особым). Так как координатные оси сопряжены, то $a_{12} = 0$, а так как направление оси абсцисс является особым направлением, то $a_{11} = 0$. Поэтому, если рассматриваемая нецентральная линия обладает прямой центров (принадлежит типу III), то, выбирая указанным способом направления координатных осей и помещая начало координат в произвольный центр линии, мы получим для нее уравнение вида

$$a_{22}y^2 + a_{33} = 0, \quad a_{22} \neq 0. \qquad (2)$$

Заметим, что при таком выборе начала координат ось абсцисс будет не просто прямой асимптотического направления, а будет диаметром, сопряженным с направлением оси ординат (прямой центров).

Для нецентральных линий, не имеющих центра (тип. II), уничтожить линейные члены невозможно. Поэтому в этом случае мы поступим несколько иначе. Именно, так же как для линий типа III, мы направление оси ординат выберем произвольно, а за ось абсцисс примем диаметр, сопряженный с этим направлением. Поскольку направление этого диаметра является асимптотическим, то $a_{11} = 0$. С другой стороны, согласно формуле (5) п. 6 диаметр, сопряженный с направлением 0 : 1 оси ординат, имеет уравнение

$$a_{12}x + a_{22}y + a_{23} = 0.$$

Так как, по условию, этим диаметром является ось абсцисс, коэффициенты a_{12} и a_{23} этого уравнения должны быть равны нулю (а коэффициент a_{22} отличен от нуля). Таким образом, в описанной координатной системе (с пока еще точно не определенным положением оси ординат) уравнение рассматриваемой линии второго порядка имеет вид

$$a_{22}y^2 + 2a_{13}x + a_{33} = 0, \quad a_{22} \neq 0.$$

Полагая в этом уравнении $y = 0$, мы немедленно получим, что ось абсцисс пересекает нашу линию в некоторой точке (а именно, в точке с абсциссой $-\frac{a_{33}}{2a_{13}}$; заметим, что $a_{13} \neq 0$, поскольку в противном случае наша линия принадлежала бы не типу II, а типу III). Поэтому мы можем окончательно выбрать систему координат, потребовав, чтобы начало координат было точкой пересечения оси абсцисс с линией. При таком выборе начала координат будет иметь место равенство $a_{33} = 0$ и потому уравнение линии будет иметь вид

$$a_{22}y^2 + 2a_{13}x = 0, \tag{3}$$

где $a_{22} \neq 0$ и $a_{13} \neq 0$.

Подводя итоги, мы получаем следующую теорему приведения:

Теорема 1. *Для любой линии второго порядка существует аффинная координатная система, в которой ее уравнение имеет либо вид* (1), *либо вид* (2), *либо вид* (3).

К виду (1) *могут быть приведены уравнения центральных линий, причем уравнение центральной линии тогда и только тогда имеет вид* (1), *когда координатные оси являются сопряженными диаметрами.*

К виду (2) *могут быть приведены уравнения нецентральных линий, не имеющих центров, причем уравнение такой линии тогда и только тогда имеет вид* (2), *когда ось абсцисс является диаметром, сопряженным с направлением оси ординат, а начало координат является точкой пересечения линии с осью абсцисс.*

К виду (3) *могут быть приведены уравнения нецентральных линий, обладающих прямой центров, причем уравнение такой линии тогда и только тогда имеет вид* (3), *когда ось абсцисс является прямой центров этой линии.*

Замечание 1. Подчеркнем, что эту теорему мы доказали для линий второго порядка над любым полем $\mathbb{K}$ (характеристики, отличной от двух). Более того, коэффициенты приведенных уравнений (1), (2) и (3) (так же как и соответствующие координаты) мы можем найти (по коэффициентам исходного уравнения) лишь рациональными действиями (сложением и вычитанием, умножением и делением). Например, если коэффициенты исходного уравнения линии второго порядка были вещественными числами, то коэффициенты приведенных уравнений (и коэффициенты формул перехода от исходных координат к новым) также будут вещественными числами.

Возможность дальнейшего упрощения уравнений (1), (2) и (3) зависит от арифметических свойств поля $\mathbb{K}$. Например, пусть поле $\mathbb{K}$ является полем $\mathbb{R}$ вещественных чисел (т. е. мы находимся в вещественной или вещественно-комплексной плоскости). Тогда для любой центральной линии мы от координат x, y, в которых ее уравнение имеет приведенный вид (1), можем перейти к координатам x', y', связанным с координатами x, y формулами

$$\left.\begin{aligned} x' &= \sqrt{\left|\frac{a_{11}}{a_{33}}\right|}\,x, \\ y' &= \sqrt{\left|\frac{a_{22}}{a_{33}}\right|}\,y, \end{aligned}\right\} \quad \text{если} \quad a_{33} \neq 0,$$

$$\left.\begin{aligned} x' &= \sqrt{|a_{11}|}\,x, \\ y' &= \sqrt{|a_{22}|}\,y, \end{aligned}\right\} \quad \text{если} \quad a_{33} = 0.$$

Для линий с уравнением (3) мы положим

$$\begin{aligned} x' &= a_{13}x, \\ y' &= \sqrt{|a_{22}|}\,y, \end{aligned}$$

а для линий с уравнением (2) положим

$$\left.\begin{aligned} x' &= x, \\ y' &= \sqrt{\frac{a_{11}}{a_{33}}}\,y, \end{aligned}\right\} \quad \text{если} \quad a_{33} \neq 0,$$

$$\left.\begin{aligned} x' &= x, \\ y' &= \sqrt{|a_{11}|}\,y, \end{aligned}\right\} \quad \text{если} \quad a_{33} = 0.$$

В результате мы получим следующие уравнения (штрихи в обозначениях координат мы опускаем):

$$\pm x^2 \pm y^2 + 1 = 0,$$
$$\pm x^2 \pm y^2 = 0,$$
$$\pm y^2 + 2x = 0,$$
$$\pm y^2 + 1 = 0,$$
$$\pm y^2 = 0.$$

Меняя (если нужно) ориентации координатных осей и умножая уравнения на ± 1, мы немедленно получаем из этих уравнений знакомые нам девять уравнений, перечисленные в теореме 1 п. 4:

$$[1]\ x^2 + y^2 = 1, \quad [4]\ x^2 - y^2 = 1, \quad [7]\ y^2 - 1 = 0,$$
$$[2]\ x^2 + y^2 = -1, \quad [5]\ x^2 - y^2 = 0, \quad [8]\ y^2 + 1 = 0,$$
$$[3]\ x^2 + y^2 = 0, \quad [6]\ y^2 = 2x, \quad [9]\ y^2 = 0.$$

Тем самым мы заново (и чисто «аффинно») доказали теорему 1 п. 4 (одновременно выяснив геометрический смысл процесса приведения).

Пусть теперь поле $\mathbb{K}$ является полем $\mathbb{C}$ комплексных чисел, т. е. мы рассматриваем комплексную аффинную плоскость и на ней линии второго порядка, выражающиеся уравнениями с комплексными коэффициентами (напомним, что на комплексной плоскости — в отличие от вещественно-комплексной плоскости — понятие линии, выражающейся уравнением с вещественными коэффициентами, бессмысленно). Тогда совершенно аналогичные (лишь несколько более простые) рассуждения доказывают следующую теорему об аффинно канонических уравнениях линий второго порядка на комплексной плоскости:

Теорема 2. *Для любой линии второго порядка на комплексной плоскости существует аффинная координатная система, в которой уравнением линии является одно из следующих пяти уравнений*:

$$[1]\ x^2 + y^2 = 1, \quad [3]\ y^2 = 2x,$$
$$[2]\ x^2 + y^2 = 0, \quad [4]\ y^2 - 1 = 0,$$
$$[5]\ y^2 = 0.$$

При этом

никаким преобразованием аффинных координат (на комплексной плоскости) ни одно из уравнений [1]—[5] *нельзя перевести в другое.*

Задание. Докажите это утверждение (ср. п. 4).

Линии, имеющие каноническое уравнение (1), являются невырожденными центральными линиями (их с равным правом можно называть «эллипсами» или «гиперболами»).

Линии, имеющие каноническое уравнение [3], являются невырожденными параболическими (нецентральными) линиями (их можно называть «параболами»).

Остальные линии ([2], [4] и [5]) являются парами прямых (пересекающихся, параллельных, но не совпадающих, и совпадающих).

Чтобы понять, каким образом эллипсы и гиперболы объединились на комплексной плоскости в одну линию, напомним, что любая линия в комплексной плоскости геометрически представляет собой поверхность и что вещественно-комплексная плоскость получается из комплексной плоскости некоторым выбором вещественной плоскости $\mathfrak{M}^{\text{вещ}}$. Поверхность, изображающая линию [1], обладает тем свойством, что при одном выборе вещественной плоскости $\mathfrak{M}^{\text{вещ}}$ в пересечении получается эллипс, а при другом — гипербола.

Наглядно представить себе эту поверхность нельзя.

Непосредственным следствием теоремы 2 является (ср. п. 4) следующая теорема об аффинной классификации линий второго порядка на комплексной плоскости:

Теорема 3. *Любая линия второго порядка на комплексной плоскости аффинно эквивалентна одной и только одной из пяти линий, имеющих в произвольной (но фиксированной) системе аффинных координат уравнения* [1]—[5].

В частности,

на комплексной плоскости все невырожденные центральные линии аффинно эквивалентны.

8. Главные направления и диаметры линий второго порядка

Вернемся снова к евклидовой (вещественной или вещественно-комплексной) геометрии и в соответствии с этим будем опять считать координаты x, y прямоугольными.

Определение 1. Направление $l : m$ называется *главным* по отношению к данной линии второго порядка

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0, \qquad (1)$$

если перпендикулярное к нему направление $l' : m'$ с ним сопряжено.

Таким образом, в частности,

для любой параболической линии особое направление является главным.

Главным является и направление, перпендикулярное особому.

Если направление $l : m$ не особо, то с ним сопряжено единственное направление

$$l' : m' = -(a_{21}l + a_{22}m) : (a_{11}l + a_{12}m).$$

Так как направления $l:m$ и $l':m'$ тогда и только тогда перпендикулярны, когда

$$ll' + mm' = 0,$$

отсюда вытекает, что неособое направление $l:m$ тогда и только тогда является главным, когда

$$l(a_{21}l + a_{22}m) - m(a_{11}l + a_{12}m) = 0,$$

т. е. когда

$$a_{21}l^2 + (a_{22} - a_{11})lm - a_{12}m^2 = 0. \qquad (2)$$

Поскольку для особого направления это соотношение также справедливо, тем самым доказано, что

направление $l:m$ тогда и только тогда является главным, когда для него имеет место соотношение (2).

При $a_{12} = 0$ и $a_{11} = a_{22}$, т. е. (см. конец п. 3) в случае, когда линия (1) является окружностью (быть может, мнимого или нулевого радиусов), соотношение (2) удовлетворяется тождественно. В противном случае оно представляет собой уравнение, которому должны удовлетворять главные направления. Решения этого уравнения могут быть, очевидно, выражены любой из формул

$$\begin{aligned} l:m &= \left(a_{11} - a_{22} \pm \sqrt{(a_{11} - a_{22})^2 + 4a_{12}^2}\right) : 2a_{12}, \\ l:m &= -2a_{12} : \left(a_{11} - a_{12} \mp \sqrt{(a_{11} - a_{22})^2 + 4a_{12}^2}\right). \end{aligned} \qquad (3)$$

(При $a_{12} \neq 0$ эти формулы равносильны. В случае же, когда $a_{12} = 0$, следует пользоваться той из формул (3), которая имеет смысл; именно, первой формулой, когда корень берется с положительным знаком, и второй — в противном случае.)

Поскольку подкоренное выражение

$$(a_{11} - a_{22})^2 + 4a_{12}^2$$

в формулах (3) положительно, мы видим, что даваемые этими формулами значения отношения $l:m$ всегда вещественны (и различны).

Тем самым доказано следующее

Предложение 1. *Для любой линии второго порядка, не являющейся окружностью, существует точно два различных вещественных главных направления.*

Для окружности (возможно, мнимого или нулевого радиусов) все направления являются главными.

Поскольку направление, перпендикулярное главному, также, очевидно, является главным, найденные главные направления перпендикулярны друг другу.

Замечание 1. Согласно формулам (3) нахождение главных направлений требует, вообще говоря, извлечения квадратных корней. Замечательно, что *для параболических* (*нецентральных*)

линий главные направления можно найти, и не извлекая квадратных корней, поскольку для параболических линий главные направления характеризуются отношениями

$$a_{11} : a_{12} = a_{12} : a_{22} \quad и \quad -a_{12} : a_{11} = -a_{22} : a_{12}.$$

Действительно, отношение $-a_{12} : a_{11} = -a_{22} : a_{12}$ характеризует, как мы знаем, асимптотическое (особое) направление параболических линий, а отношение $a_{11} : a_{12} = a_{12} : a_{22}$ — направление, ему перпендикулярное.

Определение 2. Диаметр линии второго порядка называется *главным*, если он сопряжен с перпендикулярным ему направлением, т. е. если он перпендикулярен некоторому главному направлению (ср. определение 2 п. 5 § 2 гл. 5).

Так как любой диаметр проходит через середины хорд сопряженного направления (см. п. 6), то

любой главный диаметр линии второго порядка является ее осью симметрии.

Обратное, как показывает пример двух перпендикулярных прямых, вообще говоря, неверно.

Далее, ясно, что

направление любого главного диаметра является главным направлением.

Обратное также, вообще говоря, неверно. Действительно, для параболических линий все диаметры имеют особое (асимптотическое) направление, но существует главное направление, которое неасимптотично и потому не является направлением никакого диаметра.

Для вырожденных параболических линий (линий типа III) существует только один диаметр (прямая центров) и поскольку перпендикулярное к нему направление представляет собой главное направление, он является главным диаметром.

Для парабол (линий типа II) также существует только один главный диаметр, а именно, диаметр, сопряженный с главным неасимптотическим направлением.

Для центральных линий (линий типа I), отличных от окружности, существует два главных диаметра, каждый из которых сопряжен с направлением другого (и потому ему перпендикулярен).

Для окружностей все диаметры являются главными.

Для нецентральных линий (линий типа II и III) главный диаметр сопряжен с направлением, характеризующимся отношением $a_{11} : a_{12} = a_{12} : a_{22}$. Следовательно, чтобы получить уравнение главного диаметра, следует в уравнении (5) п. 6 положить либо $l = a_{11}$, $m = a_{12}$, либо $l = a_{12}$, $m = a_{22}$.

Полагая, например, $l = a_{11}$, $m = a_{12}$, мы получим уравнение

$$(a_{11}^2 + a_{12}^2)x + a_{12}(a_{11} + a_{12})y + a_{11}a_{13} + a_{12}a_{23} = 0.$$

Но, по условию, $a_{11}a_{22} = a_{12}^2$, и потому

$$a_{11}^2 + a_{12}^2 = a_{11}(a_{11} + a_{22}).$$

Кроме того,

$$a_{11} + a_{22} \neq 0,$$

поскольку, как мы знаем (см. п. 3), равенство $a_{11} + a_{22} = 0$ характеризует равнобочные гиперболы и пары перпендикулярных прямых.

Следовательно, мы можем полученное уравнение сократить на $a_{11} + a_{22}$. В результате получим уравнение

$$a_{11}x + a_{12}y + \frac{a_{11}a_{13} + a_{12}a_{23}}{a_{11} + a_{22}} = 0. \qquad (4)$$

Аналогично, полагая в уравнении (5) п. 6 $l = a_{21}$, $m = a_{22}$, мы (после сокращения на $a_{11} + a_{12}$) получим уравнение

$$a_{21}x + a_{22}y + \frac{a_{21}a_{13} + a_{22}a_{23}}{a_{11} + a_{22}} = 0. \qquad (5)$$

Таким образом,

уравнение главного диаметра параболической линии второго порядка может быть записано либо в виде (4), *либо в виде* (5).

Уравнение (4) имеет смысл (не сводится к тождеству $0 = 0$) тогда и только тогда, когда либо $a_{11} \neq 0$, либо $a_{12} \neq 0$. Это не удивительно, поскольку при его выводе мы принимали за l и m числа a_{11} и a_{12}, что имеет смысл тогда и только тогда, когда хотя бы одно из этих чисел отлично от нуля.

Аналогично, уравнение (5) имеет смысл тогда и только тогда, когда либо $a_{21} \neq 0$, либо $a_{22} \neq 0$.

Когда оба уравнения (4) и (5) имеют смысл, они равносильны.

В случае, когда рассматриваемая линия является параболой (имеет тип II), ее главный диаметр является ее осью. Следовательно,

для параболы уравнения (4) *и* (5) *являются уравнениями ее оси.*

В случае вырожденной параболической линии (линии типа III) из условий $\delta = 0$ и $\Delta = 0$ легко вытекает, что

$$a_{12}a_{23} = a_{22}a_{13}; \quad a_{21}a_{13} = a_{11}a_{23}.$$

Поэтому

$$\frac{a_{11}a_{13} + a_{12}a_{23}}{a_{11} + a_{22}} = a_{13} \quad \text{и} \quad \frac{a_{21}a_{13} + a_{22}a_{23}}{a_{11} + a_{22}} = a_{23},$$

так что уравнения (4) и (5) приобретают вид

$$a_{11}x + a_{12}y + a_{13} = 0 \quad \text{и} \quad a_{21}x + a_{22}y + a_{23} = 0.$$

Каждое из этих уравнений (когда оно имеет смысл) является уравнением прямой центров вырожденной параболической

линии, как это и должно быть, поскольку главным диаметром такой линии является ее прямая центров.

Уравнения главных диаметров центральных линий второго порядка рационально через коэффициенты уравнения кривой не записываются.

Теперь мы можем без труда охарактеризовать прямоугольную систему координат, в которой уравнение данной линии второго порядка имеет приведенный вид:

Осью абсцисс этой системы является главный диаметр (единственный главный диаметр для нецентральных линий и любой из двух главных диаметров для центральных линий).

Осью ординат этой системы является для центральных линий другой главный диаметр, а для нецентральных линий — прямая, перпендикулярная главному диаметру (т. е. прямая, имеющая неасимптотическое главное направление) и проходящая (в случае параболы) через точку пересечения линий с главным диаметром.

Замечание 2. Обратим внимание на то, что мы заново доказали и существование прямоугольных координат, в которых линия второго порядка имеет приведенное уравнение.

Дополнение. **Классификация поверхностей второго порядка**

Совершенно аналогичные результаты имеют место и для поверхностей второго порядка в пространстве (аффинном или евклидовом, вещественном, вещественно-комплексном или комплексном).

Мы приведем здесь формулировки соответствующих теорем для евклидова случая. Доказывать их мы не будем.

Во-первых, оказывается, что в вещественном евклидовом пространстве все поверхности второго порядка исчерпываются поверхностями, рассмотренными в § 3 гл. 5, и парами плоскостей, т. е.

любая поверхность второго порядка в вещественном евклидовом пространстве представляет собой поверхность одного из следующих шести видов:

1) *эллипсоид,*
2) *гиперболоид (однополостный или двуполостный),*
3) *параболоид (эллиптический или гиперболический),*
4) *конус,*
5) *цилиндр (эллиптический, параболический или гиперболический),*
6) *пара плоскостей (различных или совпадающих).*

Впрочем, пары плоскостей мы можем считать цилиндрами (а пары пересекающихся или совпадающих плоскостей — и конусами); см. пп. 7—8 § 3 гл. 5.

В вещественно-комплексном евклидовом пространстве возможны еще *мнимые эллипсоиды* с каноническим уравнением вида

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=-1, \qquad a\geqslant b\geqslant c>0,$$

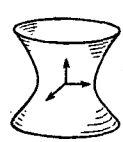
(в вещественном пространстве эти эллипсоиды являются пустыми множествами), *мнимые конусы* с каноническим уравнением вида

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=0 \qquad a\geqslant b\geqslant c>0$$

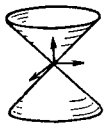
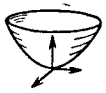
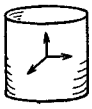
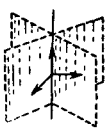
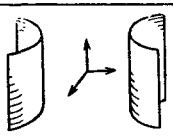
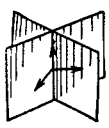
(являющиеся в вещественном пространстве точками) и *мнимые эллиптические цилиндры* (являющиеся в вещественном пространстве пустыми множествами). Кроме того, конечно, возможны *пары мнимых* (комплексно-сопряженных) *плоскостей* (в вещественном пространстве являющиеся либо прямыми, либо пустыми множествами).

Всего мы получаем, таким образом, семнадцать классов различных поверхностей второго порядка в евклидовом вещественно-комплексном пространстве:

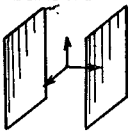
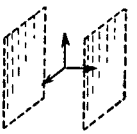
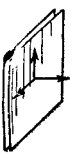
Таблица вещественных поверхностей второго порядка в евклидовом вещественно-комплексном пространстве

№№ пп	Название	Схематический вид поверхности	Каноническое уравнение поверхности
[1]	Эллипсоид (действительный)		$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1,$ $a\geqslant b\geqslant c>0$
[2]	Мнимый эллипсоид		$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=-1,$ $a\geqslant b\geqslant c>0$
[3]	Мнимый конус второго порядка		$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=0$ $a\geqslant b\geqslant c>0,$ $\frac{1}{a^2}+\frac{1}{b^2}+\frac{1}{c^2}=1$
[4]	Двуполостный гиперболоид		$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1,$ $a\geqslant b\geqslant 0, \quad c>0$
[5]	Однополостный гиперболоид		$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1,$ $a\geqslant b>0, \quad c>0$

Продолжение

№№ п/п	Название	Схематический вид поверхности	Каноническое уравнение поверхности
[6]	Действительный конус второго порядка		$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0,$ $a\geqslant b>0, \quad c>0,$ $\frac{1}{a^2}+\frac{1}{b^2}+\frac{1}{c^2}=1$
[7]	Эллиптический параболоид		$\frac{x^2}{p}+\frac{y^2}{q}=2z,$ $p\geqslant q>0$
[8]	Гиперболический параболоид		$\frac{x^2}{p}-\frac{y^2}{q}=2z,$ $p>0, \quad q>0$
[9]	Эллиптический цилиндр (действительный)		$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1,$ $a\geqslant b>0$
[10]	Мнимый эллиптический цилиндр		$\frac{x^2}{a^2}+\frac{y^2}{b^2}=-1,$ $a\geqslant b>0$
[11]	Пара мнимых пересекающихся плоскостей		$\frac{x^2}{a^2}+\frac{y^2}{b^2}=0,$ $a\geqslant b>0,$ $\frac{1}{a^2}+\frac{1}{b^2}=1$
[12]	Гиперболический цилиндр		$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1,$ $a>0, \quad b>0$
[13]	Пара пересекающихся плоскостей (вещественных)		$\frac{x^2}{a^2}-\frac{y^2}{b^2}=0$ $a>0, \quad b>0,$ $\frac{1}{a^2}+\frac{1}{b^2}=1$
[14]	Параболический цилиндр		$y^2=2px, \quad p>0$

Продолжение

№№ пп	Название	Схематический вид поверхности	Каноническое уравнение поверхности
[15]	Пара параллельных (не совпадающих) плоскостей (вещественных)		$y^2 - b^2 = 0, \quad b > 0$
[16]	Пара мнимых параллельных плоскостей		$y^2 + b^2 = 0, \quad b > 0$
[17]	Пара совпадающих плоскостей		$y^2 = 0$

Самое общее уравнение поверхности второго порядка в прямоугольных (или аффинных) координатах имеет вид

$$\begin{aligned} a_{11}x^2 + 2a_{12}xy + 2a_{13}xz + 2a_{14}x + \\ + a_{22}y^2 + 2a_{23}yz + 2a_{24}y + \\ + a_{33}z^2 + 2a_{34}z + \\ + a_{44} = 0. \end{aligned} \tag{1}$$

Считая, что

$$a_{ij} = a_{ji}, \qquad i, j = 1, 2, 3, 4,$$

мы сопоставим этому уравнению симметрическую матрицу

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}. \tag{2}$$

Пусть

$$S = a_{11} + a_{22} + a_{33} + a_{44},$$

$$\delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix},$$

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix},$$

$$D = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}.$$

Оказывается, что

числа S, δ, Δ, D не меняются при переходе к любой другой системе прямоугольных координат.

На этом основании они называются *инвариантами* уравнения (1).

Нам понадобятся также числа

$$K=\begin{vmatrix} a_{11} & a_{14} \\ a_{41} & a_{44} \end{vmatrix}+\begin{vmatrix} a_{22} & a_{24} \\ a_{42} & a_{44} \end{vmatrix}+\begin{vmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{vmatrix},$$

$$L=\begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{21} & a_{22} & a_{24} \\ a_{41} & a_{42} & a_{44} \end{vmatrix}+\begin{vmatrix} a_{11} & a_{13} & a_{14} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{vmatrix}+\begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix}.$$

Оказывается, что

числа K и L не меняются при повороте координатных осей (при переносе начала координат они, вообще говоря, меняются).

На этом основании числа K и L называются *семиинвариантами.*

Оказывается, что класс (вещественной) поверхности второго порядка может быть определен по инвариантам и семиинвариантам его уравнения:

№№ пп	Поверхности	Характеристика по инвариантам и семиинвариантам
[1]	Эллипсоид	$S\Delta>0,\ \delta>0,\ D<0$
[2]	Мнимый эллипсоид	$S\Delta>0,\ \delta>0,\ D>0$
[3]	Мнимый конус	$S\Delta>0,\ \delta>0,\ D=0$
[4]	Двуполостный гиперболоид	$S\Delta\leqslant 0,\ \Delta\neq 0,\ \delta\leqslant 0,\ D>0$
[5]	Однополостный гиперболоид	$S\Delta\leqslant 0,\ \Delta\neq 0,\ \delta\leqslant 0,\ D<0$
[6]	Конус второго порядка	$S\Delta\leqslant 0,\ \Delta\neq 0,\ \delta\leqslant 0,\ D=0$
[7]	Эллиптический параболоид	$\Delta=0,\ D<0$
[8]	Гиперболический параболоид	$\Delta=0,\ D>0$
[9]	Эллиптический цилиндр	$\Delta=0,\ D=0,\ \delta>0,\ SL<0$
[10]	Мнимый эллиптический цилиндр	$\Delta=0,\ D=0,\ \delta>0,\ SL>0$
[11]	Пара мнимых пересекающихся плоскостей	$\Delta=0,\ D=0,\ \delta>0,\ L=0$
[12]	Гиперболический цилиндр	$\Delta=0,\ D=0,\ \delta<0,\ L\neq 0$
[13]	Пара пересекающихся (вещественных) плоскостей	$\Delta=0,\ D=0,\ \delta<0,\ L=0$
[14]	Параболический цилиндр	$\Delta=0,\ D=0,\ \delta=0,\ L\neq 0$
[15]	Пара параллельных (вещественных) плоскостей	$\Delta=0,\ D=0,\ \delta=0.\ L=0,\ K<0$
[16]	Пара мнимых параллельных плоскостей	$\Delta=0,\ D=0,\ \delta=0,\ L=0,\ K>0$
[17]	Пара совпадающих (вещественных) плоскостей	$\Delta=0,\ D=0,\ \delta=0,\ L=0,\ K=0$

Таким образом, с помощью инвариантов и семиинвариантов можно определить класс поверхности второго порядка, не производя фактического приведения к каноническому виду (возможность использования семиинвариантов для выделения поверхностей классов [9]—[17] объясняется тем, что для этих поверхностей они являются инвариантами).

Более того, зная инварианты и семиинварианты уравнения поверхности, можно написать даже ее каноническое уравнение (по-прежнему не производя приведения, т. е. не находя канонической системы координат).

Для этого следует ввести в рассмотрение кубическое уравнение

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & a_{13} \\ a_{21} & a_{22}-\lambda & a_{23} \\ a_{31} & a_{32} & a_{33}-\lambda \end{vmatrix} = 0,$$

в развернутом виде имеющее вид

$$\lambda^3 - S\lambda^2 + \delta\lambda - \Delta = 0$$

(это уравнение называется *характеристическим уравнением* для уравнения (1)). Можно показать, что

корни λ_1, λ_2, λ_3 *характеристического уравнения всегда вещественны* [1]).

При $\Delta \neq 0$ все корни λ_1, λ_2, λ_3 отличны от нуля, при $\Delta = 0$ и $\delta \neq 0$ один (и только один) из этих корней равен нулю (пусть это будет корень λ_3), при $\Delta = 0$ и $\delta = 0$ равны нулю два из корней (скажем, корни λ_2 и λ_3), а третий корень λ_1 равен S (и заведомо отличен от нуля, ибо в противном случае, как можно показать, все коэффициенты уравнения (1) равны нулю).

Оказывается, что

преобразованием прямоугольных координат уравнение произвольной поверхности второго порядка можно привести к одному из следующих пяти типов:

(I) $\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 + \dfrac{D}{\Delta} = 0$ (*при* $\Delta \neq 0$)

(II) $\lambda_1 x^2 + \lambda_2 y^2 \pm 2\sqrt{-\dfrac{D}{\delta}}\, z = 0$ (*при* $\Delta = 0$, $D \neq 0$, $\delta \neq 0$) [2]),

(III) $\lambda_1 x^2 + \lambda_2 y^2 + \dfrac{L}{\delta} = 0$ (*при* $\Delta = 0$, $D = 0$, $\delta \neq 0$),

(IV) $Sx^2 \pm 2\sqrt{-\dfrac{L}{S}}\, y = 0$ (*при* $\Delta = 0$, $D = 0$, $\delta = 0$, $L \neq 0$) [3]),

(V) $Sx^2 + \dfrac{K}{S} = 0$ (*при* $\Delta = 0$, $D = 0$, $\delta = 0$, $L = 0$).

Эти уравнения называются *приведенными уравнениями поверхностей второго порядка.*

---

[1]) Алгебраической причиной этого является симметричность матрицы (2).

[2]) При $\Delta = 0$ величины D и δ имеют противоположные знаки.

[3]) При $\Delta = D = \delta = 0$ величины L и S имеют противоположные знаки.

Подчеркнем, что написать приведенное уравнение произвольной поверхности второго порядка мы можем непосредственно по ее инвариантам и семиинвариантам.

Переход от приведенного уравнения к каноническому сводится к тривиальным алгебраическим преобразованиям и никакого труда не представляет.

Приведенные уравнения типа (I) имеют поверхности [1]—[6], типа (II)— поверхности [7]—[8], типа (III) — поверхности [9]—[13], типа (IV) — поверхности [14] и типа (V) — поверхности [15]—[17].

§ 2. ПРОЕКТИВНАЯ ТЕОРИЯ ЛИНИЙ ВТОРОГО ПОРЯДКА

1. Линии второго порядка на аффинно-проективной и проективной плоскостях

Определение 1. *Линией второго порядка* на аффинно-проективной плоскости называется множество точек, однородные аффинные координаты $X : Y : Z$ которых удовлетворяют уравнению вида

$$a_{11}X^2 + 2a_{12}XY + a_{22}Y^2 + 2a_{13}XZ + 2_{23}YZ + a_{33}Z^2 = 0. \qquad (1)$$

Замечание 1. Ясно, что это определение корректно, т. е. в любой другой однородной аффинной координатной системе линия второго порядка выражается аналогичным уравнением (но, конечно, с другими коэффициентами).

Замечание 2. Определение 1 имеет смысл в аффинно-проективной плоскости под любым полем $\mathbb{K}$ (а в случае поля $\mathbb{C}$ — как в комплексной, так и в вещественно-комплексной плоскости). Для определенности мы в дальнейшем ограничимся лишь аффинно-проективной вещественно-комплексной плоскостью (случай, для нас наиболее интересный), хотя почти все результаты будут справедливы и над любым полем $\mathbb{K}$ характеристики, отличной от двух (конечно, с соответствующими тривиальными и само собой разумеющимися изменениями).

Аффинные координаты $x = \frac{X}{Z}$, $y = \frac{Y}{Z}$ собственных точек линии (1) удовлетворяют уравнению

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0 \qquad (2)$$

и, следовательно, составляют (если хотя бы один из коэффициентов a_{11}, a_{12}, a_{22} отличен от нуля) некоторую линию второго порядка на аффинной плоскости.

Если $a_{11} = a_{12} = a_{22} = 0$, но $a_{13} \neq 0$ или $a_{23} \neq 0$, то уравнение (2) выражает прямую, а если $a_{11} = a_{12} = a_{22} = a_{13} = a_{23} = 0$ (но, конечно, $a_{33} \neq 0$), то ему не удовлетворяют координаты никакой собственной точки.

Координаты $X:Y:Z=l:m:0$ несобственных точек линии (1) определяются из уравнения

$$a_{11}l^2+2a_{12}lm+a_{22}m^2=0,$$

удовлетворяющегося тождественно при $a_{11}=a_{12}=a_{22}=0$ и определяющего асимптотические направления линии (2) в противном случае.

Таким образом,

если хотя бы один из коэффициентов a_{11}, a_{12}, a_{22} *отличен от нуля, то линия второго порядка* (1) *на аффинно-проективной плоскости получается из линии второго порядка* (2) *на аффинной плоскости присоединением несобственных точек ее асимптотических направлений,*

если же $a_{11}=a_{12}=a_{22}=0$, *то линия второго порядка* (1) *распадается на две прямые, одна из которых является несобственной прямой* $Z=0$ (*а другая — либо некоторой собственной прямой, либо также несобственной прямой* $Z=0$).

Если линия (2) является эллипсом, параболой или гиперболой, то те же названия сохраняются и для соответствующей линии (1) на аффинно-проективной плоскости.

Поскольку эллипс и гипербола имеют два различных асимптотических направления (эллипс — мнимые, а гипербола — вещественные), а парабола — одно асимптотическое направление, то, следовательно,

на аффинно-проективной вещественно-комплексной плоскости эллипс имеет две дополнительные мнимые несобственные точки, гипербола — две вещественные несобственные точки, а парабола — одну (*вещественную*) *несобственную точку.*

Если линия (2) является парой прямых, то ее асимптотическими направлениями являются направления этих прямых. Поэтому присоединение несобственных точек этих направлений превращает эти прямые в прямые на аффинно-проективной плоскости.

Таким образом, мы видим, что, так же как и на аффинной плоскости,

любая (*вещественная*) *линия второго порядка на аффинно-проективной* (*вещественно-комплексной*) *плоскости представляет собой линию одного из следующих четырех видов*:

1) *эллипс* (*действительный или мнимый*);
2) *парабола*;
3) *гипербола*;
4) *пара прямых* (*вещественных или мнимых, различных или совпадающих, собственных или несобственных*).

Поскольку мнимые прямые (составляющие вещественную линию второго порядка) обязательно различны, а совпадающие прямые обязательно вещественны, мы, следовательно, получаем о д и н н а д ц а т ь различных классов вещественных линий второго порядка на аффинно-проективной вещественно-комплексной плоскости:

Таблица вещественных линий второго порядка на аффинно-проективной вещественно-комплексной плоскости

№№ пп	Линия	Каноническое уравнение
[1]	Действительный эллипс	$X^2+Y^2-Z^2=0$
[2]	Мнимый эллипс	$X^2+Y^2+Z^2=0$
[3]	Пара различных мнимых (комплексно-сопряженных) прямых, пересекающихся в собственной точке	$X^2+Y^2=0$
[4]	Гипербола	$X^2-Y^2-Z^2=0$
[5]	Пара различных вещественных собственных прямых, пересекающихся в собственной точке	$X^2-Y^2=0$
[6]	Парабола	$Y^2=2XZ$
[7]	Пара различных собственных вещественных прямых, пересекающихся в несобственной точке	$Y^2-Z^2=0$
[8]	Пара различных мнимых (комплексно-сопряженных) прямых, пересекающихся в несобственной точке	$Y^2+Z^2=0$
[9]	Пара совпадающих вещественных собственных прямых	$Y^2=0$
[10]	Пара прямых, состоящая из вещественной собственной прямой и несобственной прямой	$XZ=0$
[11]	Пара совпадающих несобственных прямых	$Z^2=0$

Аналогично, на аффинно-проективной комплексной плоскости получается следующая таблица:

Таблица линий второго порядка на аффинно-проективной комплексной плоскости

№№ пп	Линия	Каноническое уравнение
[1]	Центральная невырожденная линия (эллипс или гипербола)	$X^2+Y^2+Z^2=0$
[2]	Пара различных собственных прямых, пересекающихся в собственной точке	$X^2+Y^2=0$
[3]	Нецентральная невырожденная линия (парабола)	$Y^2=2XZ$
[4]	Пара различных собственных прямых, пересекающихся в несобственной точке	$Y^2+Z^2=0$
[5]	Пара совпадающих собственных прямых	$Y^2=0$
[6]	Пара прямых, состоящая из собственной и несобственной прямой	$XZ=0$
[7]	Пара совпадающих несобственных прямых	$Z^2=0$

Таким образом, в этом случае имеется семь классов линий второго порядка.

Аналогичные результаты имеют место и для линий второго порядка на проективной плоскости.

Определение 2. *Линией второго порядка* на проективной плоскости называется множество всех точек, однородные проективные координаты $X : Y : Z$ которых удовлетворяют уравнению вида

$$a_{11}X^2 + 2a_{12}XY + a_{22}Y^2 + 2a_{13}XZ + 2a_{23}YZ + a_{33}Z^2 = 0, \quad (3)$$

Ясно, что это определение корректно.

На аффинно-проективной (вещественно-комплексной) плоскости эллипсы, параболы и гиперболы различились по характеру их пересечения с несобственной прямой. Поскольку на проективной плоскости несобственной прямой нет, различие между эллипсами, параболами и гиперболами пропадает. Аналогично пропадает различие между парами прямых, состоящих из собственных прямых, и парами прямых, одна из которых (или обе) являются несобственными прямыми. Поэтому на проективной вещественно-комплексной плоскости имеется только пять классов (вещественных) линий второго порядка:

Таблица вещественных линий второго порядка на проективной вещественно-комплексной плоскости

№№ пп	Линия	Каноническое уравнение
[1]	Действительная невырожденная линия	$X^2 + Y^2 - Z^2 = 0$
[2]	Мнимая невырожденная линия	$X^2 + Y^2 + Z^2 = 0$
[3]	Пара вещественных различных прямых	$X^2 - Y^2 = 0$
[4]	Пара мнимых (комплексно-сопряженных) прямых	$X^2 + Y^2 = 0$
[5]	Пара совпадающих прямых	$X^2 = 0$

На проективной комплексной плоскости остается только три класса линий второго порядка:

Таблица линий второго порядка на проективной комплексной плоскости

№№ пп	Линия	Каноническое уравнение
[1]	Невырожденная линия	$X^2 + Y^2 + Z^2 = 0$
[2]	Пара различных прямых	$X^2 + Y^2 = 0$
[3]	Пара совпадающих прямых	$X^2 = 0$

2. Пересечение прямой и линии второго порядка

Пересечение прямой и линии второго порядка мы исследовали до сих пор (см., например, п. 2 § 2 гл. 5) на аффинной (или евклидовой) плоскости. Рассмотрим теперь этот вопрос на аффинно-проективной и проективной плоскостях. Естественно, что мы получим аналогичные результаты, но несколько более простые (поскольку нам не придется различать асимптотических и неасимптотических направлений).

Исследование мы будем вести одновременно на аффинно-проективной и проективной плоскостях. В первом случае координаты $X:Y:Z$ будут предполагаться, естественно, однородными аффинными координатами, а во втором — произвольными проективными координатами.

Для определенности мы будем предполагать плоскость вещественно-комплексной.

В аффинно-проективной (и проективной) плоскости прямая, проходящая через (фиксированные) точки $M_0(X_0:Y_0:Z_0)$ и $M_1(X_1:Y_1:Z_1)$, имеет параметрические уравнения вида

$$\begin{aligned} X &= \mu X_0 + \nu X_1, \\ Y &= \mu Y_0 + \nu Y_1, \\ Z &= \mu Z_0 + \nu Z_1, \end{aligned} \tag{1}$$

где μ, ν — произвольные параметры, одновременно не равные нулю.

Пусть

$$a_{11}X^2 + 2a_{12}XY + a_{22}Y^2 + 2a_{13}XZ + 2a_{23}YZ + a_{33}Z^2 = 0 \tag{2}$$

— произвольная линия второго порядка.

Обозначения. Для упрощения формул введем следующие сокращенные обозначения:

$$\begin{aligned} F(M) &= a_{11}X^2 + 2a_{12}XY + a_{22}Y^2 + 2a_{13}XZ + 2a_{23}YZ + a_{33}Z^2, \\ F_1(M) &= a_{11}X + a_{12}Y + a_{13}Z, \\ F_2(M) &= a_{21}X + a_{22}Y + a_{23}Z, \\ F_3(M) &= a_{31}X + a_{32}Y + a_{33}Z, \end{aligned} \tag{3}$$

где M — точка с координатами $X:Y:Z$. Аналогично,

$$\begin{aligned} F(M_0, M_1) = {} & a_{11}X_0X_1 + a_{12}X_0Y_1 + a_{13}X_0Z_1 + \\ & + a_{21}Y_0X_1 + a_{22}Y_0Y_1 + a_{23}Y_0Z_1 + \\ & + a_{31}Z_0X_1 + a_{32}Z_0Y_1 + a_{33}Z_0Z_1, \end{aligned}$$

где M_0 и M_1 — точки с координатами $X_0 : Y_0 : Z_0$ и $X_1 : Y_1 : Z_1$. Таким образом,

$$F(M) = F(M, M) = F_1(M)\,X + F_2(M)\,Y + F_3(M)\,Z,$$

и

$$F(M_0, M_1) = F_1(M_0)\,X_1 + F_2(M_0)\,Y_1 + F_3(M_0)\,Z_1 = \\ = F_1(M_1)\,X_0 + F_2(M_1)\,Y_0 + F_3(M_1)\,Z_0.$$

Обратим внимание на то, что, скажем, число $F_1(M)$ зависит от выбора координат точки M (при переходе от координат X, Y, Z к пропорциональным координатам ρX, ρY, ρZ оно умножается на ρ).

Замечание 1. (Для читателей, знакомых с дифференциальным исчислением функций многих переменных.) Ясно, что

$$F_1 = \frac{1}{2}\frac{\partial F}{\partial X}, \quad F_2 = \frac{1}{2}\frac{\partial F}{\partial Y}, \quad F_3 = \frac{1}{2}\frac{\partial F}{\partial Z}.$$

Чтобы найти точки пересечения прямой (1) с линией (2), следует, как мы знаем, подставить выражения (1) в уравнение (2) и решить получающееся уравнение относительно $\mu : \nu$.

В обозначениях (3) и (3′) это уравнение имеет, очевидно, вид

$$\mu^2 F(M_0) + 2\mu\nu F(M_0, M_1) + \nu^2 F(M_1) = 0. \qquad (4)$$

Если

$$F(M_0) = F(M_0, M_1) = F(M_1) = 0,$$

то уравнение (4) удовлетворяется тождественно, т. е. прямая (1) целиком содержится в линии (2).

В противном случае уравнение (4) имеет два решения, выражающиеся формулами

$$\mu : \nu = \left(-F(M_0, M_1) \pm \sqrt{F(M_0, M_1)^2 - F(M_0)\,F(M_1)}\right) : F(M_0),$$

$$\mu : \nu = F(M_1) : \left(-F(M_0, M_1) \mp \sqrt{F(M_0, M_1)^2 - F(M_0)\,F(M_1)}\right).$$

(Если $F(M_0) \neq 0$ и $F(M_1) \neq 0$, то эти формулы равносильны; при $F(M_0) = 0$ или $F(M_1) = 0$ следует пользоваться той из этих формул — в зависимости от знака корня,— которая имеет смысл.)

Таким образом,

в аффинно-проективной (проективной) вещественно-комплексной плоскости каждая линия второго порядка, не содержащая данную прямую, пересекается с этой прямой не более чем в двух точках.

Определение 1. В случае, когда имеются две точки пересечения, они, по определению, считаются *точками пересечения кратности* 1, а в случае, когда точка пересечения только одна, она, по определению, считается *точкой пересечения кратности* 2.

Точки кратности 1 называются также *простыми точками пересечения*, а точки кратности 2 — *двойными точками пересечения*.

Прямая, имеющая с линией второго порядка двойную точку пересечения (или целиком принадлежащая линии) называется *касательной* к линии в этой точке (соответственно — в любой своей точке). Для прямой (1) это имеет место тогда и только тогда, когда

$$F(M_0,\ M_1)^2 - F(M_0)\,F(M_1) = 0. \tag{5}$$

Если данные прямая и линия второго порядка вещественны, то их точки пересечения либо вещественны, либо комплексно-сопряжены.

Если точка $M_0(X_0 : Y_0 : Z_0)$ принадлежит линии (2), т. е. если $F(M_0) = 0$, то условие (5) выполнено тогда и только тогда, когда

$$F(M_0,\ M_1) = 0,$$

т. е. когда

$$F_1(M_0)\,X_1 + F_2(M_0)\,Y_1 + F_3(M_0)\,Z_1 = 0. \tag{6}$$

Таким образом, соотношению (6) удовлетворяют координаты $X_1 : Y_1 : Z_1$ любой точки M_1 каждой касательной, проходящей через точку M_0.

Определение 2. Если хотя бы одно из чисел $F_1(M_0)$, $F_2(M_0)$, $F_3(M_0)$ отлично от нуля, то точка M_0 называется *простой точкой* линии (2). Если же

$$F_1(M_0) = F_2(M_0) = F_3(M_0) = 0, \tag{7}$$

то точка M_0 называется *двойной точкой* линии (2).

Если (фиксированная) точка M_0 — простая, то уравнение

$$F_1(M_0)\,X + F_2(M_0)\,Y + F_3(M_0)\,Z = 0 \tag{8}$$

определяет на плоскости некоторую прямую. Поскольку уравнению (8) должны удовлетворять, согласно сказанному выше, координаты любой точки каждой касательной в точке M_0, мы получаем, что

в каждой простой точке M_0 линия второго порядка (2) *имеет единственную касательную, выражающуюся уравнением* (8).

В обозначениях (3) уравнение (8) имеет вид

$$F(M_0,\ M) = 0 \tag{8'}$$

(где точка M_0 — фиксированная, а точка M — переменная), т. е. вид

$$(a_{11}X_0 + a_{12}Y_0 + a_{13}Z_0)\,X + (a_{21}X_0 + a_{22}Y_0 + a_{23}Z_0)\,Y + \\ + (a_{31}X_0 + a_{32}Y_0 + a_{33}Z_0)\,Z = 0. \tag{8''}$$

Если точка M_0 — двойная, условие (6) выполнено для любой точки M_1. Это означает, что

каждая прямая, проходящая через двойную точку линии (2), *является касательной к этой линии.*

Таким образом,

точка линии второго порядка тогда и только тогда является ее простой точкой, когда в этой точке существует единственная касательная.

Уравнения (7), определяющие двойные точки, т. е. уравнения

$$\begin{aligned} a_{11}X_0 + a_{12}Y_0 + a_{13}Z_0 &= 0, \\ a_{21}X_0 + a_{22}Y_0 + a_{23}Z_0 &= 0, \\ a_{31}X_0 + a_{32}Y_0 + a_{33}Z_0 &= 0 \end{aligned} \tag{9}$$

тогда и только тогда имеют нетривиальное решение $(X_0, Y_0, Z_0) \neq (0, 0, 0)$, когда определитель

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

равен нулю.

Определение 3. Линия второго порядка (2) называется *вырожденной*, если $\Delta = 0$, и *невырожденной*, если $\Delta \neq 0$.

Умножив равенства (9) соответственно на X_0, Y_0, Z_0 и сложив, мы получим соотношение

$$F(X_0, Y_0, Z_0) = 0,$$

показывающее, что

любая точка, координаты $X_0 : Y_0 : Z_0$ *которой удовлетворяют уравнениям* (9), *принадлежит линии* (2) (*и потому является ее двойной точкой*).

Следовательно,

линия второго порядка тогда и только тогда вырождена, когда она имеет двойные точки.

Это, в частности, показывает, что определение 3 корректно.

Пусть M_0 — двойная точка линии (2) и M_1 — другая ее точка (безразлично, простая или двойная). Прямая M_0M_1 является касательной в точке M_0 и проходит через точку M_1. Но если касательная не содержится в линии (2), то она имеет с ней только одну общую точку. Поэтому прямая M_0M_1 целиком содержится в линии (2), что, как мы знаем, возможно тогда и только тогда, когда линия (2) является парой прямых (быть может, совпадающих).

Обратно, пусть линия (2) является парой прямых, т. е. имеет уравнение вида

$$(A_1X + B_1Y + C_1Z)(A_2X + B_2Y + C_2Z) = 0.$$

Тогда

$$a_{11}=A_1A_2,\quad a_{12}=\frac{A_1B_2+A_2B_1}{2},\quad a_{13}=\frac{A_1C_2+A_2C_1}{2},$$

$$a_{22}=B_1B_2,\quad a_{23}=\frac{B_1C_2+B_2C_1}{2},$$

$$a_{33}=C_1C_2$$

и

$$\Delta=\begin{vmatrix} A_1A_2 & \frac{A_1B_2+A_2B_1}{2} & \frac{A_1C_2+A_2C_1}{2} \\ \frac{A_1B_2+A_2B_1}{2} & B_1B_2 & \frac{B_1C_2+B_2C_1}{2} \\ \frac{A_1C_2+A_2C_1}{2} & \frac{B_1C_2+B_2C_1}{2} & C_1C_2 \end{vmatrix}=0.$$

Таким образом,

линия второго порядка тогда и только тогда вырождена (обладает двойными точками), когда она является парой прямых.

Задание. Покажите, что если прямые, составляющие линию (2), различны, то эта линия обладает единственной двойной точкой и этой точкой будет точка пересечения этих прямых. Покажите также, что если линия (2) является двойной прямой, то любая ее точка будет двойной точкой.

Для линии (2), являющейся парой прямых, касательной в любой ее простой точке является та из прямых, составляющих линию, которая проходит через эту точку. Касательной в двойной точке является (согласно сказанному выше) любая прямая, проходящая через эту точку.

В частности, для двойной прямой каждая прямая на плоскости является касательной.

На аффинно-проективной плоскости эллипс и гипербола пересекают, как мы знаем, несобственную прямую в двух различных точках, а парабола — в одной. Следовательно,

на аффинно-проективной плоскости параболы могут быть охарактеризованы как невырожденные линии второго порядка, касающиеся несобственной прямой.

Точкой касания является, очевидно, несобственная точка оси параболы.

Далее,

собственная прямая тогда и только тогда касается эллипса, параболы или гиперболы, когда она является либо асимптотой (если точка касания — несобственная), либо касательной в смысле п. 6 § 1 гл. 5 *(если точка касания — собственная).*

Задание. Докажите это утверждение.

Таким образом, в частности,

асимптоты линии второго порядка могут быть охарактеризованы (на аффинно-проективной плоскости) как прямые (собственные), касающиеся линий в несобственной точке.

3. Поляры и полюсы

Обратим внимание на то, что уравнение (8) п. 2, т. е. уравнение

$$F(M_0, M) = 0 \qquad (1)$$

(с фиксированной точкой M_0 и переменной точкой M), определяет некоторую прямую и тогда, когда точка M_0 не принадлежит линии второго порядка.

Определение 1. Прямая (1) называется *полярой* точки M_0 (по отношению к линии $F = 0$), а точка M_0 называется ее *полюсом.*

Замечание 1. Так как $F_1(M_0) = F_2(M_0) = F_3(M_0) = 0$ тогда и только тогда, когда точка M_0 является двойной точкой линии $F = 0$ (см. п. 2), то для любой точки M плоскости, не являющейся двойной точкой линии $F = 0$, уравнение (1) действительно определяет некоторую прямую.

Таким образом,

поляра существует (и однозначно определена) для любой точки, не являющейся двойной точкой линии $F = 0$.

При этом

полярой любой простой точки линии $F = 0$ *является касательная к линии* $F = 0$ *в этой точке.*

Короче: *касательная есть поляра точки касания.*

В частности, мы видим, что

касательная является полярой, содержащей свой полюс.

Обратно,

если поляра содержит свой полюс, то она является касательной (в полюсе).

Действительно, так как

$$F(M_0) = F(M_0, M_0),$$

точка M_0, принадлежащая своей поляре, принадлежит линии $F = 0$.

Пусть

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0 \qquad (2)$$

— произвольная линия второго порядка на аффинной плоскости. Согласно определению 1 п. 6 § 1 диаметром, сопряженным с неасимптотическим направлением $l : m$, называется прямая

$$(a_{11}l + a_{12}m)x + (a_{12}l + a_{22}m)y + (a_{13}l + a_{23}m) = 0.$$

В однородных аффинных координатах $X : Y : Z = x : y : 1$ это уравнение имеет вид

$$(a_{11}l + a_{12}m)X + (a_{12}l + a_{22}m)Y + (a_{13}l + a_{23}m)Z = 0.$$

Сравнив это уравнение с уравнением поляры несобственной точки $X_0 : Y_0 : Z_0 = l : m : 0$, мы немедленно получим, что эти два уравнения совпадают.

Следовательно,

каждый диаметр линии (2) (*пополненный, конечно, своей несобственной точкой*) *является полярой несобственной точки сопряженного направления.*

Таким образом, поляры можно рассматривать как обобщения диаметров.

Предложение 1. *Если точка* M_1 *принадлежит поляре точки* M_0, *то точка* M_0 *принадлежит поляре точки* M_1.

Доказательство. Точка M_1 тогда и только тогда принадлежит поляре точки M_0, когда

$$F(M_0,\ M_1) = 0,$$

а точка M_0 тогда и только тогда принадлежит поляре точки M_1, когда

$$F(M_1,\ M_0) = 0.$$

Остается заметить, что

$$F(M_0,\ M_1) = F(M_1,\ M_0)$$

для любых точек M_0 и M_1.

Конечно, в предложении 1 предполагается, что ни точка M_0, ни точка M_1 не являются двойными точками линии $F = 0$.

Чтобы избежать подобного рода оговорок, будем далее предполагать, что линия $F = 0$ невырождена (не является парой прямых). Тогда поляра будет существовать для любой точки плоскости.

Более того, легко видеть, что

относительно невырожденной линии второго порядка $F = 0$ *любая прямая является полярой некоторой однозначно определенной точки.*

Действительно, пусть

$$AX + BY + CZ = 0$$

— произвольная прямая. Рассмотрим систему уравнений

$$\begin{aligned} a_{11}X + a_{12}Y + a_{13}Z &= A,\\ a_{21}X + a_{22}Y + a_{23}Z &= B,\\ a_{31}X + a_{32}Y + a_{33}Z &= C. \end{aligned}$$

Поскольку определитель Δ этой системы отличен, по условию, от нуля, она имеет единственное решение $(X_0,\ Y_0,\ Z_0) \neq (0, 0, 0)$. Тогда полярой точки $M_0(X_0 : Y_0 : Z_0)$ и будет, очевидно, данная прямая.

Найдем, в частности, на аффинно-проективной плоскости полюс несобственной прямой. Согласно предложению 1 через этот полюс проходят поляры всех несобственных точек, т. е. (см. выше) диаметры рассматриваемой линии второго порядка. Следовательно,

полюсом несобственной прямой относительно некоторой (*невырожденной*) *линии второго порядка является* (*в случае, когда этот полюс представляет собой собственную точку*) *центр данной линии.*

Мы видим, что
отображение

$$\text{«полюс»} \longmapsto \text{«поляра»} \tag{3}$$

определяет биективное соответствие между множеством всех точек и множеством всех прямых, сохраняющее отношение инцидентности точек и прямых.

Определение 2. Соответствие (3) называется *поляритетом* (относительно данной невырожденной линии второго порядка $F = 0$).

В терминологии, введенной определением 3 п. 7 § 1 гл. 4, любой поляритет представляет собой *антиавтоморфизм* (коррелятивное преобразование) проективной плоскости.

Заметим, что построенный в п. 7 § 1 гл. 4 антиизоморфизм проективной плоскости над произвольным полем $\mathbb{K}$ является не чем иным, как поляритетом относительно линии второго порядка

$$X^2 + Y^2 + Z^2 = 0.$$

(Ясно, что понятия поляры, полюса и поляритета имеют смысл в проективной геометрии над л ю б ы м полем.)

Не нужно думать, что любой антиизоморфизм проективной плоскости является поляритетом.

Упражнение. Докажите, что
антиавтоморфизм проективной (вещественно-комплексной) плоскости тогда и только тогда является поляритетом относительно некоторой невырожденной линии второго порядка, когда он инволютивен, т. е. обладает тем свойством, что если точке M сопоставлена прямая λ, то и, наоборот, прямой λ сопоставлена точка M.

Существует красивый геометрический способ отыскания полюсов (доказывающий еще раз их существование). Так как для любой прямой λ_0, являющейся касательной, полюсом служит точка касания, то достаточно рассмотреть случай, когда прямая λ_0 пересекает (невырожденную) линию $F = 0$ в двух различных точках M_1 и M_2.

Пусть λ_1 — касательная в точке M_1, а λ_2 — касательная в точке M_2. Эти касательные различны и потому пересекаются в единственной точке M_0. Оказывается, что
точка M_0 является полюсом прямой λ_0.

Действительно, пусть μ_0 — поляра точки M_0. Так как касательная λ_1 является полярой точки M_1 и содержит точку M_0, то согласно предложению 1 поляра μ_0 точки M_0 содержит точку M_1. По аналогичным соображениям поляра μ_0 содержит и точку M_2. Поэтому $\mu_0 = M_1M_2 = \lambda_0$.

Следствие 1. *Через любую точку M_0 плоскости, не принадлежащую невырожденной линии второго порядка $F = 0$, проходят две и только две касательные к этой линии.*

Действительно, поскольку каждая прямая имеет единственный полюс, точка M_0 является точкой пересечения касательных

к линии $F = 0$, касающихся этой линии в точках ее пересечения с полярой λ_0 точки M_0. Никакой третьей касательной через точку M_0 проходить не может, поскольку ее точка касания была бы третьей точкой пересечения прямой λ_0 с линией $F = 0$.

Следствие 2. *Полярой произвольной точки M_0, не принадлежащей невырожденной линии второго порядка $F = 0$, является прямая, проходящая через точки касания двух касательных к линии $F = 0$, проходящих через точку M_0.*

Замечание 1. Предусмотренные следствием 1 касательные могут быть мнимыми (комплексно-сопряженными), даже если точка M_0 вещественна. Это позволяет разбить совокупность всех вещественных точек проективной (или аффинно-проективной) вещественно-комплексной плоскости, не принадлежащих данной невырожденной (вещественной) линии второго порядка, на две области: *внешнюю*, для которой эти касательные вещественны, и *внутреннюю*, для которой эти касательные мнимы (для нулевых линий последняя область пуста).

Замечание 2. Следствие 2 показывает, что понятие поляры не зависит от выбора системы координат (ее определение корректно).

Рассмотрим в заключение вопрос: как найти предусмотренные следствием 1 касательные?

Пусть невырожденная линия второго порядка

$$F(X, Y, Z) \equiv a_{11}X^2 + 2a_{12}XY + a_{22}Y^2 + 2a_{13}XZ + 2a_{23}YZ + a_{33}Z^2 = 0$$

касается в некоторой (заранее не фиксированной) точке $(X_0 : Y_0 : Z_0)$ прямой

$$AX + BY + CZ = 0. \tag{4}$$

Сравнивая уравнение (4) с уравнением касательной (уравнение (8) п. 2), мы немедленно получаем, что это возможно тогда и только тогда, когда

$$\begin{aligned} a_{11}X_0 + a_{12}Y_0 + a_{13}Z_0 &= \rho_0 A, \\ a_{21}X_0 + a_{22}Y_0 + a_{23}Z_0 &= \rho_0 B, \\ a_{31}X_0 + a_{32}Y_0 + a_{33}Z_0 &= \rho_0 C, \end{aligned} \tag{5}$$

где ρ_0 — некоторый множитель пропорциональности (отличный от нуля).

Кроме соотношений (5), должно иметь место также равенство

$$AX_0 + BY_0 + CZ_0 = 0,$$

означающее, что точка $(X_0 : Y_0 : Z_0)$ принадлежит прямой (4).

Таким образом, мы видим, что если прямая (4) касается линии второго порядка $F = 0$, то система однородных уравнений

$$\begin{aligned} a_{11}X + a_{12}Y + a_{13}Z - A\rho &= 0, \\ a_{21}X + a_{22}Y + a_{23}Z - B\rho &= 0, \\ a_{31}X + a_{32}Y + a_{33}Z - C\rho &= 0, \\ AX + BY + CZ &= 0 \end{aligned} \tag{6}$$

имеет нетривиальное решение (X_0, Y_0, Z_0, ρ_0) и потому ее определитель

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & A \\ a_{21} & a_{22} & a_{23} & B \\ a_{31} & a_{32} & a_{33} & C \\ A & B & C & 0 \end{vmatrix} \tag{7}$$

равен нулю.

Обратно, если определитель (7) равен нулю и потому система уравнений (6) имеет нетривиальное решение

$$(X_0, Y_0, Z_0, \rho_0),$$

то, во-первых, $\rho_0 \neq 0$, поскольку в противном случае определитель

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

был бы равен нулю и линия $F = 0$ была бы вырожденной, и, во-вторых, хотя бы одно из чисел X_0, Y_0, Z_0 также не равно нулю, ибо в противном случае имели бы место равенства $A = B = C = 0$. Таким образом, числа X_0, Y_0, Z_0 являются однородными координатами некоторой точки M_0, принадлежащей прямой (4) и удовлетворяющей соотношениям (5). Умножая эти соотношения на X_0, Y_0, Z_0 и складывая, мы немедленно получим, что

$$\begin{aligned} F(M_0) = F_1(M_0)X_0 + F_2(M_0)Y_0 + F_3(M_0)Z_0 = \\ = \rho_0(AX_0 + BY_0 + CZ_0) = 0, \end{aligned}$$

т. е. что точка M_0 принадлежит линии $F = 0$. Поэтому в этой точке существует единственная касательная, совпадающая в силу соотношений (5) с данной прямой (4).

Тем самым доказано, что

линия $F=0$ тогда и только тогда касается прямой (4) *в некоторой (заранее не фиксированной) точке, когда*

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & A \\ a_{21} & a_{22} & a_{23} & B \\ a_{31} & a_{32} & a_{33} & C \\ A & B & C & 0 \end{vmatrix} = 0. \tag{8}$$

Задание. Проверьте, что это утверждение верно для любых (а не только невырожденных) линий второго порядка.

Теперь уже ясно, как искать касательную, проходящую через точку $M_0(X_0, Y_0, Z_0)$ (безразлично, лежащую или не лежащую на линии $F=0$). Для этого нужно решить (относительно $A:B:C$) систему двух уравнений, состоящую из уравнения (8) и уравнения

$$AX_0+BY_0+CZ_0=0,$$

выражающего тот факт, что касательная проходит через точку M_0. Соответствующие числа A, B, C и будут коэффициентами уравнения искомой касательной.

Дополнение. Поляры в пространстве

Совершенно аналогичным образом можно определить относительно произвольной поверхности второго порядка *поляру точки* в пространстве, являющуюся некоторой плоскостью. Если поверхность невырождена ($D\neq 0$), то эта поляра определена для любой точки (проективного пространства) и отображение

«точка» $\longmapsto$ «ее поляра»

определяет биективное соответствие, сохраняющее инцидентность.

Наглядно, поляру можно определить, рассмотрев для любой точки M_0 *конус касательных* с вершиной в точке M_0, образующими которого являются касательные к данной (невырожденной) поверхности второго порядка, проходящие через точку M_0. Если точка M_0 не принадлежит поверхности, то, как можно показать, этот конус касается поверхности по некоторой плоской линии второго порядка. Плоскость, содержащая эту линию, и является полярой точки M_0.

Далее, оказывается, что поляры всех точек, принадлежащих некоторой прямой λ, проходят через вполне определенную прямую μ, называемую *полярой прямой* λ. При этом полярой прямой μ оказывается исходная прямая λ.

Упражнение. Покажите (см. п. 6 § 3 гл. 4), что

если пучок окружностей изображается в (евклидово-проективном) пространстве прямой λ, то сопряженный пучок окружностей изображается полярой прямой λ относительно параболоида

$$X^2+Y^2+2ZT=0$$

Таким образом, понятие поляры прямой в пространстве обобщает понятие сопряженного пучка окружностей.

4. Теорема Безу

Пусть

$$F=0 \quad \text{и} \quad G=0$$

— две линии второго порядка. Отыскание точек пересечения этих линий на аффинной плоскости сводится к решению системы двух уравнений второй степени с двумя неизвестными x, y. Вообще говоря, такая система имеет четыре решения, так что данные линии пересекаются в четырех точках. Поскольку решения могут оказаться комплексными, мы при этом с неизбежностью должны выйти в вещественно-комплексную плоскость (или просто в комплексную, если мы не хотим ограничиваться вещественными линиями). Однако и в комплексной (аффинной) плоскости число точек пересечения может оказаться меньше четырех (например, пара сопряженных гипербол вообще не имеет общих точек). Суть дела здесь в том, что точки пересечения «могут уходить в бесконечность». Это показывает, что для того, чтобы добиться «правильного» числа точек пересечения, необходимо перейти в аффинно-проективную (или просто проективную) плоскость.

Все это объясняет, почему мы будем рассматривать задачу о точках пересечения двух линий второго порядка *на аффинно-проективной вещественно-комплексной плоскости.* Конечно, все результаты будут справедливы (с соответствующими упрощениями) и для линий второго порядка на проективной, или комплексной, плоскости.

Однако и в аффинно-проективной вещественно-комплексной плоскости число точек пересечения двух линий второго порядка может оказаться меньшим четырех (примером является та же пара сопряженных гипербол, имеющих только две общие точки — несобственные точки их общих асимптот). Здесь уже дело в том, что некоторые точки пересечения (аналогично корням алгебраических уравнений) могут иметь кратность и такие точки нужно считать столько раз, какова их кратность.

Оказывается, что для двух алгебраических линий можно так определить кратность точки их пересечения (собственной или несобственной, вещественной или мнимой), что будет справедлива следующая теорема:

Теорема Безу. *Число точек пересечения двух алгебраических линий соответственно порядков m и n либо бесконечно (и в этом случае данные линии имеют общую часть, также являющуюся алгебраической линией*[1]*), либо это число, подсчитанное с учетом кратностей всех точек (иными словами — сумма кратностей всех общих точек наших линий), равно mn.*

---

[1]) Порядка, не превосходящего меньшего из чисел m и n.

В частности,

число точек пересечения произвольной прямой с данной алгебраической линией (не содержащей этой прямой) равно порядку этой линии.

Таким образом, теорема Безу позволяет, в частности, дать *геометрическую характеристику понятия порядка алгебраической линии.*

Для линии второго порядка мы получаем из теоремы Безу правильное число точек их пересечения (четыре).

Доказать теорему Безу не очень сложно после того, как дано определение кратности точки пересечения. Однако дать такое определение, пригодное во всех случаях, оказывается очень деликатной и трудной задачей[1]). Мы ограничимся здесь тем, что определим кратности точек пересечения *только для линий порядка не выше второго* (и докажем в этом случае теорему Безу).

В первую очередь мы рассмотрим случай, когда одна из рассматриваемых линий является прямой.

Если и вторая линия является прямой, то никакой проблемы нет: кратность точки пересечения двух (различных) прямых по определению считается равной единице. Теорема Безу при этом, конечно, автоматически справедлива (на аффинно-проективной плоскости).

Кратности точек пересечения прямой и линии второго порядка мы уже определили в п. 2 (определение 1). Ясно, что согласно этому определению теорема Безу справедлива и в этом случае (число точек пересечения, с учетом кратности, всегда равно двум).

Рассмотрим теперь точки пересечения двух вырожденных линий второго порядка:

$$f_1 g_1 = 0 \quad \text{и} \quad f_2 g_2 = 0,$$

т. е. линий, распадающихся в пары прямых

$$f_1 = 0, \quad g_1 = 0 \tag{1}$$

и

$$f_2 = 0, \quad g_2 = 0. \tag{2}$$

Мы будем предполагать, что *эти линии не имеют общей прямой*, т. е. что ни одна из прямых (1) не совпадает ни с одной из прямых (2).

Тогда возможны следующие четыре случая:

Случай 1. *Никакие три из прямых* (1) *и* (2) *не имеют общей точки.*

[1]) Ее решение использует, в частности, известное из алгебры понятие результанта двух многочленов.

Тогда они пересекаются в четырех различных точках. Мы, по определению, будем считать эти точки *простыми* точками пересечения.

Случай 2. *Три и только три из прямых* (1) *и* (2) *имеют общую точку* M_0.

Пусть, для определенности, эта точка принадлежит прямым $f_1 = 0$, $f_2 = 0$, $g_1 = 0$ (и потому является двойной точкой линии $f_1 g_1 = 0$). Тогда прямая $g_1 = 0$ пересекает кривую $f_2 g_2 = 0$ в двух различных точках. Эти точки мы, по определению, будем считать *простыми* точками пересечения данных линий второго порядка, а точку M_0 — их *двойной* точкой пересечения.

Случай 3. *Одна из данных линий* (*скажем, первая*) *является двойной прямой и эта прямая не проходит через двойную точку другой линии.*

В этом случае наши линии пересекаются в двух точках (каждая из которых является двойной точкой первой линии). Мы, по определению, будем считать эти точки *двойными* точками пересечения.

Случай 4. *Все четыре прямые* (1) *и* (2) *проходят через одну точку.*

Тогда эта точка является единственной точкой пересечения данных кривых (и на каждой из них является двойной точкой). Мы, по определению, будем считать ее точкой пересечения *кратности* 4.

Мы видим, что в каждом из этих четырех случаев теорема Безу справедлива (общее число точек пересечения с учетом кратности всегда равно четырем).

Сопоставим каждой точке первой кривой число $p = 1$, если эта точка — простая, и число $p = 2$, если эта точка — двойная. Аналогично, любой точке второй кривой сопоставим число $q = 1$, если эта точка — простая, и число $q = 2$, если эта точка — двойная.

Тогда все четыре случая можно объединить следующим общим определением:

Определение 1. *Кратностью* произвольной точки пересечения двух вырожденных линий второго порядка называется произведение pq.

Для точек пересечения вырожденной и невырожденной линий второго порядка дело обстоит сложнее.

Пусть

$$fg = 0$$

— произвольная вырожденная линия второго порядка, и пусть

$$F = 0$$

— произвольная невырожденная линия второго порядка.

Возможны следующие шесть случаев:

Случай 1. *Общая точка M_0 прямых $f=0$ и $g=0$ не принадлежит линии $F=0$ и ни одна из прямых $f=0$ и $g=0$ не касается линии $F=0$.*

В этом случае имеется четыре точки пересечения, все, по определению, *простые*.

Случай 2. *Точка M_0 не принадлежит линии $F=0$, но одна (и только одна) из прямых $f=0$ и $g=0$ касается линии $F=0$.*

В этом случае имеется три точки пересечения: две *простые* и одна *двойная* (точка касания).

Случай 3. *Точка M_0 не принадлежит линии $F=0$ и обе прямые $f=0$ и $g=0$ касаются линии $F=0$.*

Имеется только две точки пересечения, обе, по определению, *двойные*.

Случай 4. *Точка M_0 принадлежит линии $F=0$, но прямые $f=0$ и $g=0$ не касаются линии $F=0$ в этой точке.*

В этом случае каждая из прямых $f=0$ и $g=0$ пересекает линию $F=0$ еще в одной точке. Эти точки, по определению, считаются *простыми* точками пересечения обеих линий второго порядка, а точка M_0 — их *двойной* точкой пересечения.

Случай 5. *Одна (и только одна) из прямых $f=0$ и $g=0$ касается линии $F=0$ в точке M_0.*

Тогда другая прямая пересекает линию $F=0$ еще в одной точке. Эта точка, по определению, считается *простой* точкой пересечения, а точка M_0 — *тройной* точкой пересечения (*кратности* 3).

Случай 6. *Обе прямые $f=0$ и $g=0$ касаются линии $F=0$ в точке M_0 (и, следовательно, совпадают).*

В этом случае никаких других точек пересечения нет и точка M_0, по определению, считается точкой пересечения *кратности* 4.

Снова теорема Безу во всех случаях автоматически выполнена.

Условно считая, что точка, не являющаяся общей точкой двух кривых второго порядка, является их точкой пересечения *нулевой кратности*, мы можем все шесть случаев объединить в следующей формулировке:

Определение 2. Пусть p — кратность точки M как точки пересечения прямой $f=0$ и линии $F=0$, и пусть q — кратность точки M как точки пересечения прямой $g=0$ и линии $F=0$. Тогда *кратностью* точки M как точки пересечения линий $fg=0$ и $F=0$ называется сумма $p+q$.

Определим теперь кратности точек пересечения двух невырожденных линий второго порядка

$$F=0, \quad G=0. \tag{3}$$

Случай 1. *В точке пересечения M_0 данные линии имеют различные касательные.*

В этом случае мы, по определению, будем считать точку M_0 *простой* точкой пересечения.

Пусть линии $F = 0$ и $G = 0$ имеют в точке M_0 совпадающие касательные.

Лемма 1. *Линии* (3) *тогда и только тогда имеют общую касательную в их точке пересечения* M_0, *когда существуют такие числа* μ_0 *и* ν_0 (*не равные одновременно нулю*), *что линия второго порядка*

$$\overset{\circ}{H} \equiv \mu_0 F + \nu_0 G = 0 \tag{4}$$

имеет точку M_0 *своей двойной точкой.*

Этим условием линия $\overset{\circ}{H} = 0$ *однозначно определена.*

Доказательство. Если касательные к линиям (3) в точке M_0 совпадают, то

$$\frac{F_1(M_0)}{G_1(M_0)} = \frac{F_2(M_0)}{G_2(M_0)} = \frac{F_3(M_0)}{G_3(M_0)} \tag{5}$$

(см. в п. 2 уравнение касательной (8)), и потому существуют такие числа μ_0 и ν_0 (одновременно не равные нулю), что

$$\begin{aligned} \mu_0 F_1(M_0) + \nu_0 G_1(M_0) &= 0, \\ \mu_0 F_2(M_0) + \nu_0 G_2(M_0) &= 0, \\ \mu_0 F_3(M_0) + \nu_0 G_3(M_0) &= 0. \end{aligned} \tag{6}$$

Тогда для линии $\overset{\circ}{H} = 0$ будут иметь место равенства

$$\overset{\circ}{H}_1(M_0) = \overset{\circ}{H}_2(M_0) = \overset{\circ}{H}_3(M_0) = 0, \tag{7}$$

т. е. точка M_0 будет двойной точкой этой линии.

Обратно, если линия вида (4) с двойной точкой M_0 существует, то имеют место равенства (7), т. е. равенства (6), означающие справедливость пропорции (5). Поэтому касательные к линиям (3) в точке M_0 совпадают.

Наконец, ясно, что поскольку точка M_0 не является двойной точкой линий $F = 0$ и $G = 0$ (по условию, эти линии вообще не имеют двойных точек), соотношения (6) однозначно определяют отношение $\mu_0 : \nu_0$, а значит, и линию (4).

Поскольку линия $\overset{\circ}{H} = 0$ обладает двойной точкой M_0, она распадается в пару прямых, пересекающихся в точке M_0.

Случай 2. *Ни одна из прямых, составляющих линию* $\overset{\circ}{H} = 0$, *не касается в точке* M_0 *линий* $F = 0$ *и* $G = 0$ (*т. е. не совпадает с их общей касательной*).

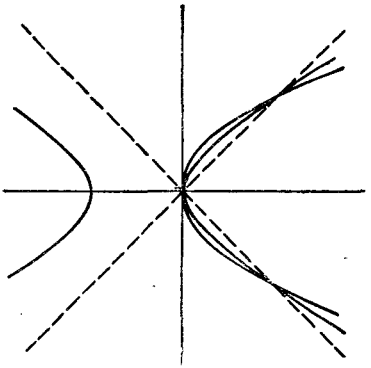
В этом случае, мы, по определению, будем считать точку M_0 *двойной* точкой пересечения линий $F = 0$ и $G = 0$.

Случай 3. *Одна и только одна прямая из прямых, составляющих линию* $\overset{\circ}{H} = 0$, *совпадает с общей касательной линий* $F = 0$ *и* $G = 0$.

В этом случае мы, по определению, будем считать точку M_0 *тройной* точкой пересечения линий $F = 0$ и $G = 0$.

Случай 4. *Обе прямые, составляющие линию* $\overset{\circ}{H} = 0$, *совпадают с общей касательной линий* $F = 0$ *и* $G = 0$.

В этом случае мы, по определению, будем считать точку M_0 точкой пересечения кратности 4.



Например, парабола

$$Y^2 - 2XZ = 0$$

и гипербола

$$X^2 - 2Y^2 + 2XZ = 0$$

имеют в собственной точке $(0:0:1)$ общую касательную

$$X = 0.$$

Кривая $\overset{\circ}{H} = 0$ имеет вид

$$(Y^2 - 2XZ) + (X^2 - 2Y^2 + 2XZ) = 0,$$

т. е. вид

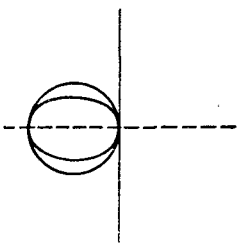
$$X^2 - Y^2 = 0,$$

и потому является парой прямых

$$X + Y = 0 \quad \text{и} \quad X - Y = 0.$$

Поскольку эти прямые отличны от общей касательной $X = 0$, точка $(0:0:1)$ является двойной точкой пересечения.

Аналогично, два эллипса



$$X^2 + Y^2 + 2XZ = 0$$

и

$$X^2 + 2Y^2 + 2XZ = 0$$

имеют в точке $(0:0:1)$ общую касательную

$$X = 0.$$

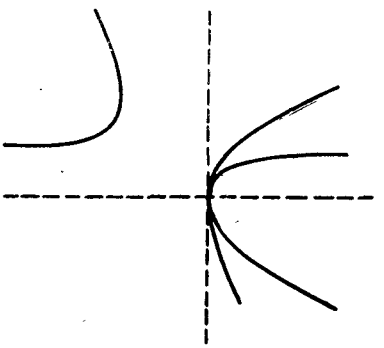
Кривая $\overset{\circ}{H} = 0$ теперь имеет вид

$$(X^2 + Y^2 + 2XZ) - (X^2 + 2Y^2 + 2XZ) = 0,$$

т. е. вид

$$Y^2 = 0,$$

и потому является двойной прямой $Y = 0$. Поскольку эта прямая не совпадает с касательной $X = 0$, точка $(0:0:1)$ также является двойной точкой пересечения.



Парабола

$$Y^2 - 2XZ = 0$$

и гипербола

$$2XY + Y^2 - 2XZ = 0$$

также имеют в точке $(0:0:1)$ общую касательную

$$X = 0.$$

Но кривая $\overset{\circ}{H} = 0$ для них имеет вид

$$(Y^2 - 2XZ) - (2XY + Y^2 - 2XZ) = 0$$

или

$$XY = 0,$$

т. е. представляет собой пару различных прямых, одной из которых является прямая $X = 0$, а потому точка $(0:0:1)$ является тройной точкой пересечения.

Наконец, эллипс

$$X^2 + Y^2 - 2XZ = 0$$

и парабола

$$Y^2 - 2XZ = 0$$

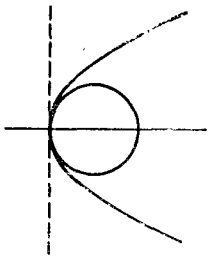
обладают в точке $(0:0:1)$ общей касательной

$$X = 0;$$

эта точка является их точкой пересечения кратности 4, поскольку соответствующая кривая $\overset{\circ}{H} = 0$ имеет вид

$$(X^2 + Y^2 - 2XZ) - (Y^2 - 2XZ) = 0$$

и потому является дважды взятой общей касательной $X = 0$.



Сравнивая изложенные определения с определением 2, мы немедленно получаем, что в каждом из случаев 2—4 кратность точки M_0 как точки пересечения линий $F = 0$ и $G = 0$ равна ее кратности как точки пересечения линий $F = 0$ и $\overset{\circ}{H} = 0$ (или, что то же самое, — линий $G = 0$ и $\overset{\circ}{H} = 0$). Поэтому все сказанное можно объединить в следующем общем определении:

Определение 3. Пусть для общей точки M_0 двух невырожденных линий второго порядка $F = 0$ и $G = 0$ существует линия $\overset{\circ}{H} = 0$ вида (4), имеющая точку M_0 своей двойной точкой. Тогда *кратностью* точки M_0 как точки пересечения линий $F = 0$ и $G = 0$ называется ее кратность как точки пересечения линий $F = 0$ и $\overset{\circ}{H} = 0$ (см. определение 2). Если такой линии $\overset{\circ}{H} = 0$ не существует, точка M_0, по определению, считается *простой* точкой пересечения (кратности 1).

Докажем теперь для линий (3) теорему Безу. Для этого нам понадобятся три вспомогательные леммы.

Лемма 2. *Если линии $F = 0$ и $G = 0$ имеют в их общей точке M_0 общую касательную, то эта касательная будет касательной в точке M_0 и для любой линии вида*

$$H \equiv \mu F + \nu G = 0,$$

где μ и ν — произвольные числа, не равные одновременно нулю.

Доказательство. Тот факт, что линии $F = 0$ и $G = 0$ имеют в точке M_0 общую касательную, означает (см. доказательство леммы 1), что имеет место пропорция

$$\frac{F_1(M_0)}{G_1(M_0)} = \frac{F_2(M_0)}{G_2(M_0)} = \frac{F_3(M_0)}{G_3(M_0)}.$$

Но тогда ясно, что для любых μ и ν будет иметь место и пропорция

$$\frac{H_1(M_0)}{G_1(M_0)} = \frac{H_2(M_0)}{G_2(M_0)} = \frac{H_3(M_0)}{G_3(M_0)}$$

(ибо, например, $H_1 = \mu F_1 + \nu G_1$).

Лемма 3. *Кратность точки пересечения M_0 любых двух невырожденных линий второго порядка $F=0$ и $G=0$ равна ее кратности и как точки пересечения линии $F=0$ и любой линии вида*

$$H \equiv \mu F + \nu G = 0,$$

где μ и ν — произвольные числа (причем $\nu \neq 0$).

Доказательство. Пусть M_0 является простой точкой пересечения, т. е. пусть касательные к линиям $F=0$ и $G=0$ в точке M_0 различны. Применим лемму 2 к линиям $F=0$ и $H=0$. Так как

$$G = -\frac{\mu}{\nu} F + \frac{1}{\nu} H$$

(напомним, что, по условию, $\nu \neq 0$), то согласно этой лемме касательные к линиям $F=0$ и $H=0$ различны. Поэтому точка M_0 является их простой точкой пересечения.

Пусть кратность точки пересечения M_0 больше единицы. Тогда эта кратность, по определению, равна ее кратности как точки пересечения линий $F=0$ и $\overset{\circ}{H}=0$, где $\overset{\circ}{H}=0$ — линия вида (4), являющаяся парой прямых. Но ясно (в силу единственности линии $\overset{\circ}{H}=0$), что если мы составим аналогичную линию для линий $F=0$ и $H=0$ (напомним, что согласно лемме 2 эти линии также имеют в точке M_0 общую касательную), то она совпадет с линией $\overset{\circ}{H}=0$. Поэтому кратность точки M_0 как точки пересечения линий $F=0$ и $H=0$ и в этом случае равна ее кратности как точки пересечения линий $F=0$ и $G=0$.

Лемма 4. *Для любых двух линий второго порядка $F=0$ и $G=0$ существуют такие числа μ и ν (не равные одновременно нулю), что линия*

$$H \equiv \mu F + \nu G = 0 \tag{8}$$

вырождена (является парой прямых).

Доказательство. Пусть

$$F(X, Y, Z) = a_{11}X^2 + 2a_{12}XY + a_{22}Y^2 + 2a_{13}XZ + 2a_{23}YZ + a_{33}Z^2,$$
$$G(X, Y, Z) = b_{11}X^2 + 2b_{12}XY + b_{22}Y^2 + 2b_{13}XZ + 2b_{23}YZ + b_{33}Z^2.$$

Тогда условие вырожденности линии

$$\mu F + \nu G = 0$$

имеет вид

$$\begin{vmatrix} \mu a_{11} + \nu b_{11} & \mu a_{12} + \nu b_{12} & \mu a_{13} + \nu b_{13} \\ \mu a_{21} + \nu b_{21} & \mu a_{22} + \nu b_{22} & \mu a_{23} + \nu b_{23} \\ \mu a_{31} + \nu b_{31} & \mu a_{32} + \nu b_{32} & \mu a_{33} + \nu b_{33} \end{vmatrix} = 0. \tag{9}$$

Это уравнение (относительно $\mu : \nu$) либо удовлетворяется тождественно, либо имеет три (возможно, совпадающих) корня,

каждый из которых отвечает некоторой вырожденной линии вида (8).

В случае комплексной плоскости этим все доказано. В случае вещественно-комплексной плоскости достаточно заметить, что уравнение (9), являясь (для вещественных линий $F = 0$ и $G = 0$) уравнением третьей степени с вещественными коэффициентами, обязательно имеет вещественный корень.

Теперь у нас уже все готово для доказательства теоремы Безу (в оставшемся случае, когда линии $F = 0$ и $G = 0$ невырождены). Действительно, для этих линий согласно лемме 4 существует вырожденная линия $H = 0$ вида (8), а согласно лемме 3 кратность любой точки пересечения линий $F = 0$ и $G = 0$ равна ее кратности как точки пересечения линий $F = 0$ и $H = 0$. Поэтому сумма кратностей всех точек пересечения линий $F = 0$ и $G = 0$ равна сумме кратностей всех точек пересечения линий $F = 0$ и $H = 0$. Но, поскольку линия $H = 0$ вырождена, последняя сумма, как мы знаем, равна четырем.

Тем самым теорема Безу (для линий второго порядка) полностью доказана.

Легко видеть, что

если вещественные линии $F = 0$ и $G = 0$ (на вещественно-комплексной плоскости) имеют две мнимые (комплексно-сопряженные) точки пересечения, то кратности этих точек совпадают.

Задание. Докажите это утверждение.

Отсюда (и из теоремы Безу) вытекает, что

любая мнимая точка пересечения двух (вещественных) линий второго порядка (на вещественно-комплексной плоскости) является либо простой, либо двойной точкой пересечения (в последнем случае линии пересекаются только в двух точках: в данной и в комплексно-сопряженной).

Дополнение. Дифференциально-геометрическое истолкование кратности точки пресечения

Для читателей, знакомых с основными понятиями дифференциального исчисления, мы напомним, что для любой линии вида

$$y = f(x)$$

ее касательная в точке $(x_0, y_0) = (x_0, f(x_0))$ имеет уравнение

$$y - y_0 = f'(x_0)(x - x_0).$$

Если функция $y = f(x)$ задана неявным уравнением

$$F(x, y) = 0,$$

то по правилу дифференцирования неявной функции $y' = -\frac{F_x}{F_y}$, и потому

уравнение касательной может быть переписано в виде

$$F_x(x_0, y_0)(x - x_0) + F_y(x_0, y_0)(y - y_0) = 0.$$

Для линии второго порядка

$$F(x, y) \equiv a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0$$

это вполне согласуется с полученными выше результатами, поскольку уравнение касательной (уравнение (8) п. 2) в неоднородных координатах

$$x = \frac{X}{Z}, \quad y = \frac{Y}{Z}$$

имеет вид

$$(a_{11}x_0 + a_{12}y_0 + a_{13})x + (a_{21}x_0 + a_{22}y_0 + a_{23})y + (a_{31}x_0 + a_{32}y_0 + a_{33}) = 0$$

и, очевидно, равносильно уравнению

$$2(a_{11}x_0 + a_{12}y_0 + a_{13})(x - x_0) + 2(a_{21}x_0 + a_{22}y_0 + a_{23})(y - y_0) = 0.$$

По определению, две линии

$$y = f(x), \quad y = g(x),$$

проходящие через точку (x_0, y_0), *касаются* друг друга в этой точке, если их касательные совпадают, т. е. если

$$f'(x_0) = g'(x_0).$$

Поэтому линии

$$F(x, y) = 0, \quad G(x, y) = 0,$$

заданные неявными уравнениями, касаются друг друга в точке (x_0, y_0), если

$$\frac{F_x(x_0, y_0)}{G_x(x_0, y_0)} = \frac{F_y(x_0, y_0)}{G_y(x_0, y_0)}.$$

Линии

$$y = f(x), \quad y = g(x)$$

называются *соприкасающимися* (в точке (x_0, y_0)), если они касаются и, кроме того,

$$f''(x_0) = g''(x_0).$$

По правилу дифференцирования неявных функций вторая производная $f''(x)$ функции $y = f(x)$, заданной неявно уравнением

$$F(x, y) = 0,$$

выражается формулой

$$f''(x) = -\frac{F_{xx}(F_y)^2 - 2F_{xy}F_xF_y + F_{yy}(F_x)^2}{F_y^3}.$$

Поэтому, полагая

$$l = F_y(x_0, y_0), \quad m = -F_x(x_0, y_0), \quad k = \frac{F_y(x_0, y_0)}{G_y(x_0, y_0)},$$

мы можем для функций, заданных неявно, условие $f''(x_0) = g''(x_0)$ записать (при выполнении условия $f'(x_0) = g'(x_0)$, т. е. при $F_x(x_0, y_0) = kG_x(x_0, y_0)$) в следующем виде:

$$\mathring{H}_{xx}(x_0, y_0)\, l^2 + 2\mathring{H}_{xy}(x_0, y_0)\, lm + \mathring{H}_{yy}(x_0, y_0)\, m^2 = 0,$$

где

$$\overset{\circ}{H} = F - kG.$$

При

$$F(x, y) = a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33},$$
$$G(x, y) = b_{11}x^2 + 2b_{12}xy + b_{22}y^2 + 2b_{13}x + 2b_{23}y + b_{33}$$

условие соприкасания приобретает вид

$$(a_{11} - kb_{11})\, l^2 + 2(a_{12} - kb_{12})\, lm + (a_{22} - kb_{22})\, m^2 = 0,$$

а это означает, что направление $l : m$ является асимптотическим для линии $\overset{\circ}{H} = 0$. Таким образом, мы установили, что линии второго порядка $F = 0$ и $G = 0$, имеющие в точке (x_0, y_0) общую касательную, направление которой характеризуется отношением $l : m$, тогда и только тогда соприкасаются в этой точке, когда линия второго порядка $\overset{\circ}{H} = 0$ содержит несобственную точку $(l : m : 0)$.

Но ясно, что линия $\overset{\circ}{H} = 0$ как раз и является указанной в лемме 1 линией $\overset{\circ}{H} = 0$, имеющей в точке пересечения двойную точку. С другой стороны, мы знаем, что если некоторая прямая проходит через двойную точку линии второго порядка и еще через одну ее точку, то эта прямая целиком содержится в линии. Поэтому утверждение, что линия $\overset{\circ}{H} = 0$ содержит несобственную точку $(l : m : 0)$ касательной, равносильно утверждению, что эта касательная целиком принадлежит линии $\overset{\circ}{H} = 0$.

Сопоставляя все эти обстоятельства, мы немедленно получаем, что

собственная точка пересечения двух (вещественных) линий второго порядка является

а) *простой точкой пересечения, если линии в этой точке не касаются*;

б) *двойной точкой пересечения, если линии в этой точке касаются, но не соприкасаются*;

в) *точкой пересечения кратности* 3 *или* 4, *если линии в этой точке соприкасаются.*

Вычисляя третьи производные, можно аналогичным образом различнть точки пересечения кратности 3 н 4.

Упражнение. Докажите, что для точки пересечения кратности 3 производные $f'''(x_0)$ и $g'''(x_0)$ различны, а для точки пересечения кратности 4 они совпадают.

§ 3. ГЕОМЕТРИЯ ЛИНИЙ ВТОРОГО ПОРЯДКА

1. Пучки линий второго порядка

Определение 1. Пусть

$$F = 0 \quad \text{и} \quad G = 0$$

—две различные линии второго порядка. *Пучком линий второго порядка* называется множество всех линий вида

$$\mu F + \nu G = 0, \qquad (1)$$

где μ и ν — произвольные числа, одновременно не равные нулю.

Замечание 1. Это определение пригодно в любой плоскости (аффинной, аффинно-проективной или проективной, вещественной, вещественно-комплексной, комплексной или над любым полем). В дальнейшем мы, как правило, будем предполагать плоскость аффинно-проективной и вещественно-комплексной (а все линии — вещественными). Конечно, все результаты будут справедливы (с соответствующими упрощениями) и на проективной или комплексной плоскости.

Замечание 2. Линия второго порядка полностью определена (в данной системе координат) шестью коэффициентами a_{ij} ее уравнения и, обратно, однозначно (с точностью до пропорциональности) их определяет (теорема единственности). Следовательно, эти коэффициенты можно рассматривать как *однородные координаты линии* (ср. определение 1 п. 1 § 1 гл. 4). Однако шесть однородных координат определяют точки «пятимерного» пространства. Поскольку концепцией многомерных пространств мы еще не владеем, возникающим отсюда изображением линий второго порядка как точек пятимерного (проективного) пространства мы пользоваться не можем. Наоборот, геометрию линий второго порядка мы можем истолковать как конкретное воплощение абстрактной пятимерной геометрии.

В частности, пучки линий второго порядка являются, с этой точки зрения, не чем иным, как *прямыми в пятимерном пространстве.*

Подобно точкам (проективной) прямой каждая линия пучка определяется отношением $\mu : \nu$. При $\mu : \nu = 1 : 0$ получается линия $F = 0$, а при $\mu : \nu = 0 : 1$ — линия $G = 0$.

Ясно, что (подобно прямой) *пучок однозначно определяется любыми двумя (различными) линиями, ему принадлежащими.*

Если $F(M_0) = 0$ и $G(M_0) = 0$, то

$$\mu F(M_0) + \nu G(M_0) = 0.$$

Это показывает, что

любая линия пучка $\mu F + \nu G = 0$ *проходит через каждую точку пересечения линий* $F = 0$ *и* $G = 0$.

Рассмотрим, например, случай, когда линии $F = 0$ и $G = 0$ имеют общую прямую $h = 0$ (здесь h — сокращенное обозначение для левой части $Ax + By + C$ уравнения прямой). В этом случае обе линии $F = 0$ и $G = 0$ являются вырожденными линиями, т. е. распадаются в пару прямых. Пусть

$$h = 0, \quad f = 0$$

— прямые, составляющие линию $F = 0$, а

$$h = 0, \quad g = 0$$

— прямые, составляющие линию $G = 0$. Без ограничения общности мы можем, очевидно, считать, что

$$F = fh, \quad G = gh.$$

Поэтому уравнение любой линии пучка $\mu F + \nu G = 0$ имеет вид

$$(\mu f + \nu g) h = 0,$$

и потому эта линия также является вырожденной линией, состоящей из прямой $h = 0$ и некоторой прямой $\mu f + \nu g = 0$, принадлежащей пучку прямых, определенному прямыми $f = 0$ и $g = 0$.

Определение 2. Пучок линий второго порядка, получающийся из некоторого пучка прямых присоединением к каждой прямой пучка одной и той же прямой, называется *вырожденным.*

При выбранной дополнительной прямой $h = 0$ вырожденные пучки находятся в биективном соответствии с пучками прямых.

Ясно, что точки пересечения любой пары линий невырожденного пучка — одни и те же для всех пар.

Определение 3. Общие точки всех линий невырожденного пучка называются *фундаментальными точками* этого пучка.

Ряд элементарных свойств (невырожденных) пучков мы уже фактически установили в леммах п. 4 § 2. Например, согласно лемме 3 п. 4 § 2,

кратность точки пересечения любой пары линий пучка — одна и та же для всех пар.

Определение 4. *Кратностью* фундаментальной точки пучка называется ее кратность как точки пересечения любой пары линий пучка.

Согласно только что сказанному, это определение корректно.

В силу теоремы Безу сумма кратностей всех фундаментальных точек пучка равна четырем. Поэтому возможны следующие типы пучков:

а) пучки с четырьмя простыми фундаментальными точками;

б) пучки с тремя фундаментальными точками — одной двойной и двумя простыми;

в) пучки с двумя двойными фундаментальными точками;

г) пучки с двумя фундаментальными точками — одной тройной и одной простой;

д) пучки с одной фундаментальной точкой кратности 4.

Определение 5. Пучки а) мы будем называть пучками *типа* [1111] (или пучками *общего типа*), пучки б) — пучками *типа* [211], пучки в) — пучками *типа* [22], пучки г) — пучками *типа* [31] и, наконец, пучки д) — пучками *типа* [4].

Согласно лемме 2 п. 4,

если две линии пучка имеют в некоторой (фундаментальной) точке общую касательную, то эта касательная является касательной (в той же точке) и любой другой линии пучка.

Определение 6. Общие касательные всех линий невырожденного пучка называются *фундаментальными прямыми* этого пучка.

Фундаментальные точки и фундаментальные прямые пучков не могут быть выбраны совершенно произвольно. Простое необходимое условие указывает следующее

Предложение 1. *Никакие три фундаментальные точки невырожденного пучка не коллинеарны. Никакая фундаментальная прямая не содержит двух фундаментальных точек.*

Доказательство. Прямая, проходящая через три (с учетом кратностей) фундаментальные точки, пересекаясь с любой линией пучка по трем точкам, должна этой линии принадлежать. Таким образом, все линии пучка содержат одну и ту же прямую, т. е. пучок является, вопреки предположению, вырожденным пучком.

Пучок общего типа не имеет фундаментальных прямых, пучок типа [211] имеет одну фундаментальную прямую (проходящую через двойную фундаментальную точку и не содержащую простых), пучок типа [22] имеет две (различные) фундаментальные прямые, и, наконец, пучок типа [31] имеет одну фундаментальную прямую (проходящую через тройную точку).

Что же касается пучков типа [4] то, как мы увидим ниже, они имеют либо одну фундаментальную прямую, либо любая прямая, проходящая через фундаментальную точку, является фундаментальной прямой.

Все сказанное выше с равным правом применимо как к комплексной, так и к вещественно-комплексной плоскостям. Рассмотрим теперь свойства пучков, имеющие место только в вещественно-комплексной плоскости (проективной или аффинно-проективной).

Определение 7. Пучок линий второго порядка на вещественно-комплексной плоскости называется *вещественным*, если он содержит по крайней мере две (а потому и бесконечно много) различных вещественных кривых.

Вещественные пучки общего типа (типа [1111]) распадаются на следующие классы:

а) класс пучков, все фундаментальные точки которых вещественны;

б) класс пучков, две фундаментальные точки которых вещественны, а другие две — мнимые комплексно-сопряженные;

в) класс пучков, все фундаментальные точки которых мнимы (и попарно комплексно-сопряжены).

Вещественные пучки типа [211] распадаются на два класса:

а) класс пучков, все фундаментальные точки которых вещественны;

б) класс пучков, простые фундаментальные точки которых — мнимые комплексно-сопряженные.

Заметим, что двойная фундаментальная точка вещественного пучка типа [211] всегда вещественна. Вещественна и его фундаментальная прямая.

Вещественные пучки типа [22] также распадаются на два класса:

а) класс пучков, обе фундаментальные точки которых вещественны;

б) класс пучков, фундаментальные точки которых комплексно-сопряжены.

В случае а) фундаментальные прямые вещественны, а в случае б) — комплексно-сопряжены.

Вещественные пучки типов [31] и [4] имеют только вещественные фундаментальные точки и прямые.

Предложение 2. *В произвольном пучке*

$$\mu F + \nu G = 0$$

линий второго порядка для любой точки M_0 *плоскости, отличной от фундаментальных точек пучка, существует единственная линия*

$$H \equiv \mu_0 F + \nu_0 G = 0,$$

проходящая через точку M_0.

Если пучок и точка вещественны, то линия $H = 0$ *также вещественна.*

Доказательство. Условие $H(M_0) = 0$ означает, что

$$\mu_0 F(M_0) + \nu_0 G(M_0) = 0.$$

Поскольку хотя бы одно из чисел $F(M_0)$ и $G(M_0)$ отлично от нуля (точка M_0 не фундаментальна), это соотношение однозначно определяет отношение $\mu_0 : \nu_0$, т. е. однозначно определяет линию $H = 0$. Если числа $F(M_0)$ и $G(M_0)$ вещественны, то отношение $\mu_0 : \nu_0$ также вещественно.

2. Описание пучков линий второго порядка

Теорема 1. *Каждый пучок общего типа* [1111] *состоит из всех линий второго порядка, проходящих через его фундаментальные точки.*

Доказательство. Пусть $F = 0$ — произвольная линия второго порядка, проходящая через фундаментальные точки M_1, M_2, M_3, M_4 данного пучка, и пусть M_0 — произвольная точка этой линии, отличная от точек M_1, M_2, M_3, M_4. Согласно предложению 2 п. 1, в пучке существует линия $H = 0$, проходящая через точку M_0. Линии $F = 0$ и $H = 0$ имеют, таким образом, пять общих точек, что согласно теореме Безу возможно тогда и только тогда, когда эти линии либо совпадают (и в этом случае все доказано), либо имеют общую прямую (и, следовательно, обе являются парами прямых). В последнем случае, мы, без ограничения общности, можем предполагать, что линия $H = 0$ состоит из прямых M_1M_2 и M_3M_4 и что прямая M_1M_2 является общей прямой линий $H = 0$ и $F = 0$. Но тогда линия $F = 0$, состоящая из двух прямых, одной из которых является прямая

M_1M_2, и проходящая через четыре точки M_1, M_2, M_3, M_4, никакие три из которых не коллинеарны (предложение 1 п. 1), обязательно должна содержать прямую M_3M_4. Следовательно, линии $F = 0$ и $H = 0$ совпадают и в этом случае.

Следствие. *Пучок общего типа однозначно определяется его фундаментальными точками.*

Чтобы описать все линии, принадлежащие пучку с фундаментальными точками

$$M_1,\ M_2,\ M_3,\ M_4,$$

рассмотрим прямые M_1M_2, M_3M_4, M_1M_3 и M_2M_4. Пусть

$$f_1 = 0,\quad g_1 = 0,\quad f_2 = 0,\quad g_2 = 0$$

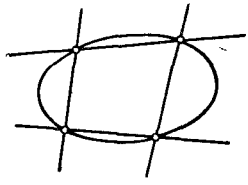
— уравнения этих прямых. Тогда линии второго порядка

$$f_1g_1 = 0 \quad \text{и} \quad f_2g_2 = 0$$

как линии, проходящие через все четыре точки, будут согласно теореме 1 принадлежать рассматриваемому пучку.

Этим доказано следующее

Предложение 1. *Любой пучок общего типа состоит из всех линий второго порядка, имеющих уравнения вида*



$$\mu f_1g_1 + \nu f_2g_2 = 0, \tag{1}$$

где μ *и* ν *— произвольные числа, а*

$$f_1 = 0,\quad g_1 = 0,\quad f_2 = 0,\quad g_2 = 0 \tag{2}$$

— уравнения прямых, никакие три из которых не имеют общей точки.

Фундаментальными точками пучка являются точки пересечения прямых $f_1 = 0$, $g_1 = 0$ *с прямыми* $f_2 = 0$, $g_2 = 0$.

Заметим, что

если пучок (1) *— вещественный, то прямые* (2) *можно выбрать вещественными или комплексно-сопряженными.*

Действительно, если все точки M_1, M_2, M_3, M_4 — вещественные, то все прямые (2) также будут вещественными. Если, скажем, точки M_1 и M_2 комплексно-сопряжены (а точки M_3 и M_4 вещественны или комплексно-сопряжены), то прямые $f_1 = 0$ и $g_1 = 0$ будут вещественными, а прямые $f_2 = 0$ и $g_2 = 0$ — комплексно-сопряженными.

Ясно, что формула (1) определяет пучок общего типа для любых точек M_1, M_2, M_3, M_4, никакие три из которых не лежат на одной прямой (причем этот пучок — вещественный тогда и только тогда, когда мнимые из точек M_1, M_2, M_3, M_4 попарно комплексно-сопряжены). Таким образом, *указанное в предложении* 1 п. 1 *необходимое условие для фундаментальных точек является* (*для пучков общего типа*) *также и достаточным.*

Пучки типа [211] изучаются совершенно аналогично.

Теорема 2. *Каждый пучок типа* [211] *состоит из всех линий второго порядка, проходящих через его фундаментальные точки и касающихся его фундаментальной прямой.*

Доказательство. Пусть M_1 — двойная фундаментальная точка, а M_2 и M_3 — простые фундаментальные точки, и пусть $h = 0$ — фундаментальная прямая (проходящая через точку M_1 и не проходящая через точки M_2 и M_3). Рассмотрим произвольную линию второго порядка $F = 0$, проходящую через точки M_1, M_2, M_3 и касающуюся (в точке M_1) прямой $h = 0$. Пусть M_0 — произвольная точка линии $F = 0$, отличная от точек M_1, M_2, M_3, и пусть $H = 0$ — линия пучка, проходящая через точку M_0 (предложение 2 п. 1). Линии второго порядка $F = 0$ и $H = 0$ имеют (с учетом кратностей) по крайней мере пять общих точек (двойную точку M_1 и еще три точки M_2, M_3, M_0 неизвестных кратностей); поэтому согласно теореме Безу эти линии либо совпадают (и тогда все доказано), либо имеют общую прямую (и, следовательно, обе являются парами прямых). В последнем случае точка M_1 должна быть единственной двойной точкой линии $H = 0$ (ибо согласно предложению 1 п. 1 точки M_1, M_2, M_3 не коллинеарны), и следовательно, эта линия состоит из различных прямых M_1M_2 и M_1M_3. При этом без ограничения общности мы можем считать, что в состав линии $F = 0$ входит прямая M_1M_2. С другой стороны, поскольку прямая $h = 0$ является, по условию, касательной к линии $F = 0$, она должна либо совпадать с прямой M_1M_2, либо точка M_1 должна быть двойной точкой линии $F = 0$. Поскольку первый случай невозможен (фундаментальная прямая $h = 0$ не проходит через фундаментальную точку M_2), остается только второй случай. Но тогда второй прямой, входящей в состав линии $F = 0$, должна быть обязательно прямая M_1M_3. Поэтому и в этом случае линии $F = 0$ и $H = 0$ совпадают.

Следствие. *Пучок типа* [211] *однозначно определяется его фундаментальными точками и фундаментальной прямой.*

Пусть

$$f_1 = 0, \quad g = 0 \quad \text{и} \quad f_2 = 0$$

— уравнения прямых M_1M_2, M_1M_3 и M_2M_3. Тогда линии второго порядка

$$f_1 g = 0 \quad \text{и} \quad f_2 h = 0$$

принадлежат, согласно теореме 2, рассматриваемому пучку.

Этим доказано следующее

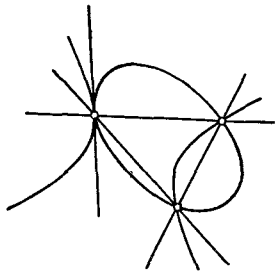
Предложение 2. *Каждый пучок типа* [211] *состоит из всех линий второго порядка, имеющих уравнения вида*

$$\mu f_1 g + \nu f_2 h = 0, \tag{3}$$

где μ *и* ν *— произвольные числа, а*

$$f_1 = 0, \quad g = 0, \quad f_2 = 0, \quad h = 0$$

— уравнения таких (различных) прямых, что прямые $f_1 = 0$, $g = 0$ *и* $h = 0$ *проходят через одну точку, через которую не проходит прямая* $f_2 = 0$.



Фундаментальными точками пучка являются точки попарного пересечения прямых $f_1 = 0$, $g = 0$ *и* $f_2 = 0$, *а фундаментальной прямой — прямая* $h = 0$.

При этом

если пучок (3) *— вещественный, то прямые* $f_2 = 0$ *и* $h = 0$ *обязательно вещественны, а прямые* $f_1 = 0$ *и* $g = 0$ *либо вещественны, либо комплексно-сопряжены.*

Действительно, как мы знаем, прямая $h = 0$ и двойная фундаментальная точка M_1 обязательно вещественны, а простые фундаментальные точки M_2 и M_3 либо вещественны, либо комплексно-сопряжены.

Ясно, что формула (3) определяет некоторый пучок типа [211] для любых неколлинеарных точек M_1, M_2, M_3 и любой прямой $h = 0$, проходящей через точку M_1 и не проходящей через точки M_2 и M_3 (причем этот пучок — вещественный тогда и только тогда, когда точка M_1 и прямая $h = 0$ вещественны, а точки M_2 и M_3 либо вещественны, либо комплексно-сопряжены). Таким образом, и *для пучков типа* [211] *необходимые условия из предложения* 1 п. 2 *оказываются достаточными.*

Рассмотрим теперь пучки типа [22].

Теорема 3. *Каждый пучок типа* [22] *состоит из всех линий второго порядка, проходящих через его фундаментальные точки и касающихся в этих точках его фундаментальных прямых.*

Доказательство этой теоремы вполне аналогично доказательству теорем 1 и 2 и потому мы его опустим.

Задание. Докажите теорему 3.

Следствие. *Пучок типа* [22] *однозначно определяется его фундаментальными точками и прямыми.*

Пусть

$$h_1 = 0, \quad h_2 = 0$$

— фундаментальные прямые пучка типа [22] и пусть

$$f = 0$$

— прямая, соединяющая его фундаментальные точки. Поскольку линии второго порядка

$$f^2 = 0 \quad \text{и} \quad h_1 h_2 = 0$$

принадлежат, согласно теореме 3, пучку, справедливо следующее

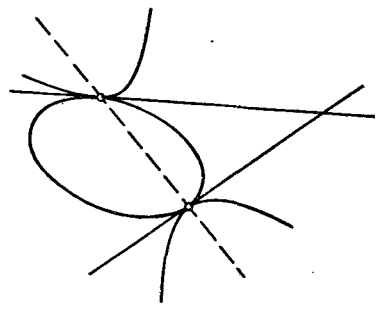
Предложение 3. *Каждый пучок типа* [22] *состоит из всех линий второго порядка, имеющих уравнения вида*

$$\mu f^2 + \nu h_1 h_2 = 0, \qquad (4)$$

где μ *и* ν *— произвольные числа, а*

$$f = 0, \quad h_1 = 0, \quad h_2 = 0 \qquad (5)$$

— уравнения трех прямых, не проходящих через одну точку.



Фундаментальными прямыми этого пучка являются прямые

$$h_1 = 0, \quad h_2 = 0,$$

а фундаментальными точками — точки пересечения этих прямых с прямой $f = 0$.

При этом,

если пучок (4) *— вещественный, то прямая* $f = 0$ *вещественна, а прямые* $h_1 = 0$ *и* $h_2 = 0$ *либо вещественны, либо комплексно-сопряжены.*

Формула (4) определяет, очевидно, некоторый пучок типа [22] для любых прямых (5), не проходящих через одну точку (причем этот пучок — вещественный тогда и только тогда, когда прямая $f = 0$ вещественна, а прямые $h_1 = 0$ и $h_2 = 0$ либо вещественны, либо комплексно-сопряжены). Таким образом, и *для пучков типа* [22] *необходимые условия из предложения* 1 п. 1 *оказываются достаточными.*

Для пучков типа [31] и [4] ситуация значительно сложнее.

Рассмотрим сначала пучки типа [31]. Каждый такой пучок имеет две фундаментальные точки M_1 и M_2 (тройную и простую) и одну фундаментальную прямую $h = 0$, проходящую через тройную точку M_1. Пусть $g = 0$ — прямая M_1M_2 и пусть $f = 0$ — произвольная прямая, проходящая через точку M_2 и отличная от прямой $g = 0$.

Будем пока предполагать, что мы находимся на проективной (вещественно-комплексной или комплексной) плоскости. Тогда без ограничения общности мы можем считать, что прямые $h = 0$, $g = 0$ и $f = 0$ имеют соответственно уравнения

$$X = 0, \quad Y = 0 \quad \text{и} \quad Z = 0.$$

Пусть

$$a_{11}X^2 + 2a_{12}XY + a_{22}Y^2 + 2a_{13}XZ + 2a_{23}YZ + a_{33}Z^2 = 0$$

— уравнение произвольной линии рассматриваемого пучка в координатах $X : Y : Z$. Так как эта линия проходит через точку

M_1 с координатами $X_1 : Y_1 : Z_1 = 0 : 0 : 1$, то $a_{33} = 0$, а так как она проходит через точку M_2 с координатами $X_2 : Y_2 : Z_2 = 1 : 0 : 0$, то $a_{11} = 0$. Кроме того, эта линия касается, по условию, прямой $X = 0$ в точке $M_1(0 : 0 : 1)$, откуда непосредственно вытекает (см. уравнение касательной (8) п. 2 § 2), что $a_{23} = 0$.

Таким образом, уравнение произвольной линии нашего пучка имеет на самом деле вид

$$2a_{12}XY + a_{22}Y^2 + 2a_{13}XZ = 0.$$

Заметим теперь, что в силу нашего выбора координат принадлежащая пучку линия второго порядка $\overset{\circ}{H} = 0$, имеющая в точке M_1 двойную точку (см. лемму 1 п. 4 § 2), задается уравнением $XY = 0$. Поэтому наш пучок содержит некоторую линию с уравнением

$$a_{22}Y^2 + 2a_{13}XZ = 0.$$

Легко видеть, что коэффициенты a_{22} и a_{13} этого уравнения отличны от нуля. Действительно, если, например, $a_{13} = 0$, то пучок содержит две распадающиеся линии $XY = 0$ и $Y^2 = 0$ с общей прямой $Y = 0$ и потому является вырожденным пучком. Аналогично, если $a_{22} = 0$, то пучок также содержит две линии $XY = 0$ и $XZ = 0$ с общей прямой и потому снова является вырожденным пучком.

Следовательно, деля на a_{22} и полагая $p = -\frac{a_{13}}{a_{22}}$, мы окончательно получаем, что

рассматриваемый пучок типа [31] *содержит линию вида*

$$Y^2 = 2pXZ \tag{6}$$

с $p \neq 0$.

В однородных аффинных координатах линия (6) является параболой. Однако в рассматриваемых сейчас произвольных однородных проективных координатах она может быть любой (невырожденной) линией второго порядка.

Легко видеть, что

двух различных линий вида (6) *данный пучок содержать не может.*

Действительно, в противном случае уравнение любой линии пучка было бы линейной комбинацией двух уравнений вида (6). Но ясно, что уравнение $XY = 0$ вырожденной линии пучка получить таким образом невозможно.

Доказанное утверждение означает, что

число p *однозначно определяется пучком* (*при выбранной системе однородных проективных координат*).

Таким образом, мы видим, что в отличие от пучков предыдущих типов,

пучков типа [31] *с данными фундаментальными точками и данной фундаментальной прямой существует целое семейство,*

зависящее от одного параметра p (подчиненного единственному условию: быть не равным нулю).

Поскольку пучок содержит линию (6) и линию $XY = 0$, любая принадлежащая ему линия второго порядка имеет уравнение вида

$$\mu XY + \nu(Y^2 - 2pXY) = 0.$$

Возвращаясь к произвольным координатам и учитывая, что прямые $Z = 0$ и $2pZ = 0$ совпадают, мы получаем, таким образом, следующее

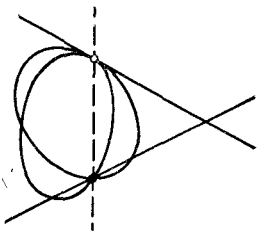
Предложение 4. *Каждый пучок типа* [31] *состоит из всех линий второго порядка, имеющих уравнение вида*

$$\mu gh + \nu(g^2 - fh) = 0, \qquad (7)$$

где μ *и* ν *— произвольные числа, а*

$$f = 0, \quad g = 0, \quad h = 0 \qquad (8)$$

— уравнения трех прямых, не проходящих через одну точку.



Фундаментальной прямой этого пучка является прямая

$$h = 0,$$

а фундаментальными точками — точки пересечения прямых $h = 0$ *и* $f = 0$ *с прямой* $g = 0$.

При этом,

если пучок (7) *— вещественный, то прямые* (8) *также вещественны.*

Формула (7) определяет, очевидно, некоторый пучок типа [31] для любых прямых (8), не проходящих через одну точку (причем этот пучок — вещественный тогда и только тогда, когда все прямые (8) вещественны). Таким образом, и *для пучков типа* [31] *необходимые условия из предложения* 1 п. 1 *оказываются достаточными.*

Итак, мы видим, что *условия предложения* 1 п. 1 *достаточны во всех случаях* (для пучков типа [4] они выполнены тривиально).

Замечание 1. Чтобы от пучка (7) перейти к другому пучку типа [31] с теми же фундаментальными точками и той же фундаментальной прямой, достаточно под $h = 0$ понимать другое (пропорциональное) уравнение фундаментальной прямой.

Замечание 2. Хотя в доказательстве предложения 4 мы пользовались проективными координатами, ясно, что *предложение 4 справедливо и в аффинно-проективной плоскости.*

Рассмотрим, наконец, пучки типа [4].

Каждый такой пучок имеет единственную фундаментальную точку M_1. Если в пучке содержится линия $H = 0$, состоящая из

двух различных прямых, то, во-первых, двойная точка этой линии должна совпадать с точкой M_1, а, во-вторых, любая другая линия пучка должна быть вырожденной линией с той же двойной точкой. Действительно, только при этих условиях каждая линия пучка пересекается с линией $\overset{\circ}{H} = 0$ в точке M_1 кратности 4.

Таким образом, в этом случае

каждая линия пучка является парой прямых (пересекающихся в его фундаментальной точке).

Определение 1. Пучки типа [4], содержащие линии, являющиеся парами различных прямых, мы будем называть *пучками вырожденных линий* (не путать с вырожденными пучками!).

Выбрав в пучке вырожденных линий две линии, мы немедленно получаем следующее

Предложение 5. *Каждый пучок вырожденных линий состоит из всех линий второго порядка, имеющих уравнения вида*

$$\mu f_1 g_1 + \nu f_2 g_2 = 0, \tag{9}$$

где μ *и* ν *— произвольные числа, а*

$$f_1 = 0, \quad g_1 = 0, \quad f_2 = 0, \quad g_2 = 0 \tag{10}$$

— уравнения четырех прямых, проходящих через одну точку и обладающих тем свойством, что каждая из прямых $f_1 = 0$ *и* $g_1 = 0$ *отлична от прямых* $f_2 = 0$ *и* $g_2 = 0$ *(совпадение прямых* $f_1 = 0$ *и* $g_1 = 0$ *или прямых* $f_2 = 0$ *и* $g_2 = 0$ *не исключается).*

Фундаментальной точкой этого пучка является точка M_1 *пересечения прямых* (10), *а фундаментальной прямой — произвольная прямая, проходящая через точку* M_1.

Пучок (9) *тогда и только тогда — вещественный, когда либо все прямые* (10) *вещественны, либо прямые* $f_1 = 0$ *и* $g_1 = 0$ *или прямые* $f_2 = 0$ *и* $g_2 = 0$ *комплексно-сопряжены (а остальные прямые вещественны).*

Замечание 3. Не нужно думать, что пучок (9) содержит все вырожденные линии второго порядка с двойной точкой M_1. Напротив, из предложения 2 п. 1 непосредственно вытекает, что

каждая прямая, проходящая через точку M_1, *входит в состав одной и только одной линии пучка* (9).

Рассмотрим теперь пучки типа [4], содержащие невырожденные линии второго порядка. Согласно сказанному выше, содержащаяся в таком пучке вырожденная линия $\overset{\circ}{H} = 0$ (см. лемму 3 п. 4 § 2) обязана быть двойной прямой, и эта прямая должна касаться в точке M_1 каждой линии пучка, т. е. должна быть фундаментальной прямой пучка. Поэтому, в частности, линия $\overset{\circ}{H} = 0$ будет единственной вырожденной линией, содержащейся в пучке.

Мы выберем систему однородных проективных координат $X:Y:Z$ так, чтобы точка M_1 имела координаты $(1:0:0)$, а линия $\mathring{H}=0$ — уравнение $Z^2=0$. Тогда фундаментальная прямая пучка будет иметь уравнение

$$Z=0.$$

Пусть

$$a_{11}X^2+2a_{12}XY+a_{22}Y^2+2a_{13}XZ+2a_{23}YZ+a_{33}Z^2=0$$

— уравнение произвольной линии пучка. Так как эта линия проходит через точку $M_1(1:0:0)$, то $a_{11}=0$, а так как она касается в этой точке прямой $Z=0$, то $a_{12}=0$.

Таким образом, уравнение произвольной линии рассматриваемого пучка имеет на самом деле вид

$$a_{22}Y^2+2a_{13}XZ+2a_{23}YZ+a_{33}Z^2=0.$$

Поскольку линия $Z^2=0$ принадлежит, по условию, пучку, отсюда вытекает, что в пучке содержится линия с уравнением вида

$$a_{22}Y^2+2a_{13}XZ+2a_{23}YZ=0.$$

Ясно, что коэффициенты a_{22} и a_{13} этого уравнения отличны от нуля, поскольку в противном случае пучок будет содержать две вырожденные линии, что согласно сказанному выше невозможно.

Следовательно, деля на a_{22} и полагая

$$p=-\frac{a_{13}}{a_{22}},\quad q=-\frac{a_{23}}{a_{22}},$$

мы получаем, что

рассматриваемый пучок типа [4] *содержит линию вида*

$$Y^2=2pXZ+2qYZ, \tag{11}$$

где $p\neq 0$.

Так же как для пучков типа [31], немедленно доказывается, что при выбранной системе однородных проективных координат *числа* $p\neq 0$ *и* q *однозначно определяются пучком.*

Таким образом,

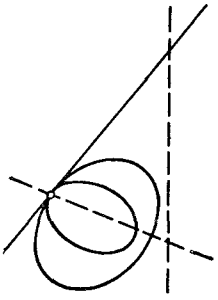
пучков типа [4] *с данной фундаментальной точкой и данной фундаментальной прямой существует целое семейство, зависящее от двух параметров* p *и* q (*подчиненных единственному условию*: $p\neq 0$).

Поскольку пучок содержит линию (11) и линию $Z^2=0$, любая принадлежащая ему линия второго порядка имеет уравнение вида

$$\mu Z^2+\nu(Y^2-2pXZ-2qYZ)=0.$$

Возвращаясь к произвольным координатам и обозначая через $f=0$, $g=0$ и $h=0$ прямые, имеющие в координатах $X:Y:Z$ уравнения $2pX+2qY=0$, $Y=0$ и $Z=0$, мы немедленно получаем отсюда следующее

Предложение 6. *Каждый пучок типа* [4], *не являющийся пучком вырожденных линий, состоит из всех линий второго порядка, имеющих уравнение вида*



$$\mu h^2+\nu(g^2-fh)=0, \tag{12}$$

где μ *и* ν *— произвольные числа, а*

$$f=0,\quad g=0,\quad h=0 \tag{13}$$

— уравнения трех прямых, не проходящих через одну точку.

Фундаментальной прямой пучка (12) *является прямая* $h=0$, *а его фундаментальной точкой — точка пересечения прямых* $h=0$ *и* $g=0$.

Пучок (12) *тогда и только тогда — вещественный, когда все прямые* (13) *вещественны.*

Замечание 4. Подобно предложению 4 *предложение* 6 *справедливо и в аффинно-проективной плоскости.*

Замечание 5. Сопоставляя все доказанные предложения, мы видим, что

каждый пучок линий второго порядка может быть представлен в одном из следующих трех видов:

$$\mu f_1 g_1+\nu f_2 g_2=0, \tag{14}$$

$$\mu gh+\nu(g^2-fh)=0, \tag{15}$$

$$\mu h^2+\nu(g^2-fh)=0, \tag{16}$$

где

$$f_1=0,\quad g_1=0,\quad f_2=0,\quad g_2=0 \tag{17}$$

и

$$f=0,\quad g=0,\quad h=0 \tag{18}$$

— уравнения некоторых прямых.

При этом, если из прямых (17) никакие три не проходят через одну точку, то пучок (14) является пучком общего типа [1111]; если три (и только три) из прямых (17) проходят через одну точку, то пучок (14) является пучком типа [211], когда все прямые (17) различны, и пучком типа [22], когда либо $f_1=g_1$, либо $f_2=g_2$; если все четыре прямые (17) проходят через одну точку (и ни одна из прямых $f_1=0$, $g_1=0$ не совпадает ни с одной из прямых $f_2=0$, $g_2=0$), то пучок (14) является пучком типа [4], состоящим из вырожденных линий; и, наконец, во всех остальных случаях пучок (14) вырожден.

Пучки (15) и (16) являются (при прямых (18), не проходящих через одну точку) соответственно пучками типа [31] и

пучками типа [4], не являющимися пучками вырожденных линий.

Упражнение. Докажите, что

если прямые (18) *проходят через одну точку, то пучки* (15) *и* (16) *являются либо пучками вырожденных линий, либо вырожденными пучками*

Дополнение. Еще раз о пучках окружностей

Все сказанное выше справедливо и для линий второго порядка на евклидово-проективной (вещественно-комплексной или комплексной) плоскости, поскольку такую плоскость мы можем единственным образом превратить в аффинно-проективную плоскость (см. п. 4 § 1 гл. 4).

Как мы знаем (см. п. 6 § 3 гл. 4), любая окружность на евклидово-проективной вещественно-комплексной плоскости проходит через циклические точки $(\pm i : 1 : 0)$ (конечно, мы здесь пользуемся однородными евклидовыми координатами). Покажем, что и, обратно,

любая вещественная линия второго порядка на евклидово-проективной вещественно-комплексной плоскости, проходящая через циклические точки, является окружностью.

Действительно, если линия

$$a_{11}X^2 + 2a_{12}XY + a_{22}Y^2 + 2a_{13}XZ + 2a_{23}YZ + a_{33}Z^2 = 0 \qquad (1)$$

проходит через точки $(\pm i : 1 : 0)$, то

$$-a_{11} + 2ia_{12} + a_{22} = 0$$

и, следовательно (поскольку числа a_{11}, a_{12}, a_{22}, по условию, вещественны),

$$a_{11} = a_{22}, \quad a_{12} = 0.$$

Поэтому при $a_{11} = a_{22} \neq 0$ линия (1) — настоящая окружность (см. п. 6 § 3 гл. 4), а при $a_{11} = a_{22} = 0$ — прямая (плюс несобственная прямая $Z = 0$).

Отсюда вытекает, что *любой вещественный пучок линий второго порядка, среди фундаментальных точек которого содержатся обе циклические точки, состоит только из окружностей* (точнее, из окружностей состоит его подмножество вещественных линий).

Если этот пучок имеет еще две собственные фундаментальные точки (и потому является пучком общего типа), то он будет эллиптическим пучком окружностей, если эти точки вещественны, и гиперболическим пучком, если эти точки комплексно-сопряжены. Если этот пучок имеет одну двойную собственную фундаментальную точку, то он будет параболическим пучком окружностей. Если среди его фундаментальных точек имеется (кроме циклических) несобственная точка, то он будет вырожденным пучком с общей несобственной прямой, так что, удалив эту прямую, мы получим пучок прямых. Пучок типа [22], фундаментальными точками которого являются циклические точки (а фундаментальные прямые комплексно-сопряжены), является несобственным пучком концентрических окружностей (имеющих центры в точке пересечения фундаментальных прямых).

Таким образом, теория пучков окружностей естественным образом вкладывается в общую теорию пучков линий второго порядка.

3. Линии второго порядка, проходящие через пять точек

В этом и следующих пунктах мы применим теорию пучков линий второго порядка к доказательству некоторых важных теорем о линиях второго порядка. Конечно, эти теоремы можно доказывать, и не пользуясь пучками, но тогда их доказательства получаются гораздо более сложными.

Пусть нам даны пять различных точек

$$1, 2, 3, 4, 5$$

аффинно-проективной вещественно-комплексной плоскости, из которых никакие четыре не коллинеарны. Покажем, что

из этих пяти точек можно выбрать четыре, из которых никакие три не коллинеарны.

Действительно, рассмотрим точки **1**, **2**, **3**, **4**. Если никакие три из них не коллинеарны, то все доказано. Пусть среди этих точек имеются три точки (скажем, точки **1**, **2**, **3**), принадлежащие одной прямой. По условию точка **4** не принадлежит прямой **123** и потому определены три прямые **14**, **24** и **34**. Точка **5** может принадлежать не более чем одной из этих прямых. Пусть она не принадлежит прямым **14** и **24**. Тогда ясно, что никакие три из четырех точек **1**, **2**, **4** и **5** не коллинеарны.

Из этого утверждения немедленно вытекает следующее

Предложение 1. *Через любые пять точек* **1**, **2**, **3**, **4**, **5** *аффинно-проективной вещественно-комплексной плоскости, из которых никакие четыре не коллинеарны, проходит единственная линия второго порядка.*

Доказательство. Пусть **1**, **2**, **3**, **4** — те из данных точек, никакие три из которых не коллинеарны. Тогда существует единственный пучок линий второго порядка общего типа, фундаментальными точками которого являются эти точки, а в этом пучке существует единственная линия, проходящая через точку **5**. Поскольку этот пучок состоит из всех линий второго порядка, проходящих через точки **1**, **2**, **3** и **4**, никакой другой линии второго порядка, проходящей через точки **1**, **2**, **3**, **4** и **5**, существовать не может.

Чтобы найти эту линию, следует составить уравнения

$$f_{12}=0, \quad f_{34}=0, \quad f_{13}=0, \quad f_{24}=0 \tag{1}$$

прямых, проходящих соответственно через точки **1**, **2**, через точки **3**, **4**, через точки **1**, **3** и через точки **2**, **4**, а затем найти отношение $\mu:\nu$ из условия, чтобы линия

$$\mu f_{12}f_{34}+\nu f_{13}f_{24}=0$$

проходила через точку **5**.

Если точки **1**, **2**, **3**, **4** — вещественные, то прямые (1) — также вещественные. Если точки **1**, **4** — вещественные, а точки **2**, **3** — комплексно-сопряженные, то прямые $f_{12}=0$ и $f_{13}=0$, а также

прямые $f_{34}=0$ и $f_{24}=0$ комплексно-сопряжены. Наконец, если точки **1**, **2**, а также точки **3**, **4** — комплексно-сопряженные, то прямые $f_{12}=0$ и $f_{34}=0$ — вещественные, а прямые $f_{13}=0$ и $f_{23}=0$ — комплексно-сопряженные. Таким образом, во всех случаях пучок содержит две вещественные линии (линии $f_{12}f_{34}=0$, $f_{13}f_{24}=0$ — в первом и третьем случае, и линии $f_{12}f_{34}+f_{13}f_{24}=0$, $i(f_{12}f_{34}-f_{13}f_{24})=0$ — во втором случае) и потому является вещественным пучком.

Поэтому,

если среди пяти точек **1**, **2**, **3**, **4** *и* **5** *каждая мнимая содержится вместе со своей комплексно-сопряженной, то проходящая через них линия второго порядка вещественна.*

Если нам даны пять точек **1**, **2**, **3**, **4** и **5**, из которых четыре (скажем, точки **1**, **2**, **3** и **4**) принадлежат одной прямой, то линия второго порядка, проходящая через эти точки, все равно будет существовать (ею будет, например, пара прямых, состоящая из прямой **1234** и произвольной прямой, проходящей через точку **5**), но эта линия уже не будет единственной.

Замечание 1. Предложение 1 можно легко доказать, и не пользуясь понятием пучка (но зато используя более серьезные алгебраические соображения, чем те, которыми мы пользовались до сих пор). Действительно, линия второго порядка с уравнением

$$a_{11}X^2+2a_{12}XY+a_{22}Y^2+2a_{13}XZ+2a_{23}YZ+a_{33}Z^2=0$$

тогда и только тогда проходит через точку $(X_0:Y_0:Z_0)$, когда

$$a_{11}X_0^2+2a_{12}X_0Y_0+a_{22}Y_0^2+2a_{13}X_0Z_0+2a_{23}Y_0Z_0+a_{33}Z_0^2=0.$$

Но это равенство является линейным однородным уравнением относительно шести неизвестных

$$a_{11},\ a_{12},\ a_{22},\ a_{13},\ a_{23},\ a_{33}.$$

Поэтому требование, чтобы линия проходила через пять точек, алгебраически выражается в виде системы пяти таких уравнений. А так как из алгебры известно, что система однородных уравнений, число уравнений которой меньше числа неизвестных, всегда имеет нетривиальное решение, то всегда существует линия второго порядка, проходящая через пять данных точек.

Чтобы на этом пути получить единственность линии, проходящей через пять точек, из которых никакие четыре не лежат на одной прямой, необходимо воспользоваться более точным алгебраическим фактом, а именно тем, что система линейно независимых однородных линейных уравнений от $n+1$ неизвестных, состоящая из n уравнений, имеет единственное (с точностью до пропорциональности) нетривиальное решение. Чтобы мы могли эту теорему использовать, нам нужно только доказать, что наши пять уравнений линейно независимы. Но это легко доказывается от противного.

Действительно, если, скажем, пятое уравнение является следствием четырех других уравнений, то любая линия второго порядка, проходящая через точки **1**, **2**, **3** и **4**, будет проходить и через точку **5**. В частности, точка **5** должна принадлежать линии, состоящей из прямых **12** и **34**, а также линии, состоящей из прямых **13** и **24**. Но если никакие три из точек **1**, **2**, **3** и **4** не коллинеарны, то эти две линии имеют общими только точки **1**, **2**, **3** и **4**, так что в этом случае мы приходим к противоречию. Если же, скажем, точки **1**, **2**, **3** коллинеарны, то чтобы прийти к противоречию, достаточно рассмотреть линию второго порядка, состоящую из прямой **123** и произвольной прямой, проходящей через точку **4** и не содержащей точку **5**.

Замечание 2. Обратим внимание на то, что последним способом мы доказали не только единственность линии, проходящей через точки **1**—**5**, но даже единственность (с точностью до пропорциональности) ее уравнения. Это позволяет нам немедленно получить теорему единственности для линий второго порядка в следующей формулировке:

Любые два уравнения произвольной линии второго порядка на аффинно-проективной вещественно-комплексной плоскости пропорциональны.

Действительно, если линия не является дважды взятой прямой, то на ней существует пять точек, из которых никакие четыре не коллинеарны, а согласно только что сказанному уравнение линии второго порядка, проходящей через эти точки, определено однозначно (с точностью до пропорциональности).

Если же линия является дважды взятой прямой, то ее уравнение имеет (с точностью до числового множителя) вид

$$(AX + BY + CZ)^2 = 0,$$

и его единственность немедленно вытекает из единственности (с точностью до пропорциональности) уравнения прямой.

Теорема единственности для линий второго порядка на аффинной (а значит, и на евклидовой) вещественно-комплексной плоскости (см. п. 4 § 1) вытекает отсюда немедленно, поскольку любая линия второго порядка на аффинной плоскости однозначно дополняется не более чем двумя точками до линии на аффинно-проективной плоскости.

Совершенно так же может быть доказана и теорема единственности для линий второго порядка (содержащих более одной точки) на вещественной плоскости.

Дополнение. Поверхности второго порядка, проходящие через девять точек

Уравнение поверхности второго порядка в пространстве имеет десять коэффициентов a_{ij}. Условие, что поверхность проходит через данную точку, накладывает на эти коэффициенты одно линейное соотношение. Следова-

тельно, поскольку система десяти линейных однородных уравнений с десятью неизвестными всегда имеет хотя бы одно нетривиальное решение.

через любые девять точек (аффинно-проективного или проективного) пространства проходит хотя бы одна поверхность второго порядка.

Упражнение. Найдите необходимые и достаточные условия, которым должны удовлетворять девять точек пространства, чтобы через них проходила единственная поверхность второго порядка.

Из доказанного утверждения непосредственно вытекает, что

через любые три прямые проходит хотя бы одна поверхность второго порядка.

Действительно, достаточно выбрать на каждой из данных прямых по три точки и провести через получающиеся девять точек поверхность второго порядка. Эта поверхность, имея с каждой из данных прямых три общие точки, будет эти прямые целиком содержать.

Построенная поверхность второго порядка не может быть ни эллипсоидом, ни двуполостным гиперболоидом, ни эллиптическим параболоидом, поскольку ни одна из этих поверхностей не может содержать ни одной прямой. Аналогично, если данные прямые скрещиваются, то она не может быть цилиндром (и, в частности, парой плоскостей). Поэтому для скрещивающихся прямых содержащая их поверхность второго порядка обязана быть либо однополостным гиперболоидом, либо гиперболическим параболоидом (заметим, что здесь мы существенно используем оставленный нами в дополнении к § 1 без доказательства факт, что перечисленными в этом дополнении семнадцатью классами поверхностей исчерпываются все поверхности второго порядка в пространстве). Но мы знаем, что прямые, лежащие на однополостном гиперболоиде или гиперболическом параболоиде, являются его прямолинейными образующими, причем три прямолинейные образующие тогда и только тогда скрещиваются, когда они принадлежат одному семейству образующих. Таким образом, наши три прямые являются прямолинейными образующими либо однополостного гиперболоида, либо гиперболического параболоида, принадлежащими одному семейству образующих. Поскольку три образующих одного семейства однозначно определяют поверхность (потому что эта поверхность может быть охарактеризована как геометрическое место всех точек прямых, проходящих через данные три образующие; см. пп. 4—5 § 2 гл. 5), тем самым доказано, что

через любые три скрещивающиеся прямые проходит единственная поверхность второго порядка (являющаяся однополостным гиперболоидом, если эти прямые не параллельны одной плоскости, и гиперболическим параболоидом — в противном случае).

Упражнение. Существуют ли не скрещивающиеся прямые, для которых содержащая их поверхность второго порядка все же единственна?

Одновременно мы также доказали, что

для любых трех скрещивающихся прямых существует линейчатая поверхность второго порядка (однополостный гиперболоид или гиперболический параболоид), прямолинейными образующими которой, принадлежащими одному семейству, они являются.

Это означает (см. пп. 4—5 § 2 гл. 5), что

какие бы три скрещивающиеся прямые мы ни взяли, геометрическое место точек всех прямых, пересекающих эти прямые, является либо однополостным гиперболоидом (если данные прямые не параллельны одной плоскости), либо гиперболическим параболоидом (если они параллельны некоторой плоскости).

4. Теорема Штурма

Пусть нам даны три линии второго порядка

$$F_1 = 0, \quad F_2 = 0, \quad F_3 = 0,$$

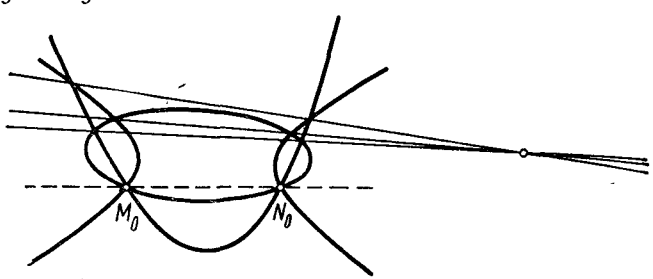
имеющие попарно лишь простые точки пересечения, и пусть эти кривые проходят через две данные точки M_0 и N_0. Тогда любые две из этих линий пересекаются еще в двух точках (отличных от точек M_0 и N_0). Пусть

M_1, N_1 — дополнительные точки пересечения линий $F_2 = 0$ и $F_3 = 0$;

M_2, N_2 — дополнительные точки пересечения линий $F_1 = 0$ и $F_3 = 0$;

M_3, N_3 — дополнительные точки пересечения линий $F_1 = 0$ и $F_2 = 0$.

Теорема Штурма. *Прямые M_1N_1, M_2N_2 и M_3N_3 проходят через одну точку.*



Доказательство. Пусть

$g_0 = 0$ — уравнение прямой M_0N_0;
$g_1 = 0$ — уравнение прямой M_1N_1;
$g_2 = 0$ — уравнение прямой M_2N_2;
$g_3 = 0$ — уравнение прямой M_3N_3.

Если одна из данных линий (скажем, линия $F_3 = 0$) является парой прямых, одной из которых служит прямая $g_0 = 0$, то другой прямой, входящей в состав линии, будет прямая $g_1 = 0$ (поскольку линия $F_3 = 0$ содержит точки M_1 и N_1), а также прямая $g_2 = 0$ (поскольку линия $F_3 = 0$ содержит точки M_2 и N_2). Следовательно, прямые $g_1 = 0$ и $g_2 = 0$ совпадают, и потому в этом случае утверждение теоремы очевидно.

Пусть теперь ни одна из данных линий не является парой прямых, одной из которых служит прямая $g_0 = 0$. Рассмотрим

пучок линий второго порядка

$$\mu F_1 + \nu F_2 = 0.$$

По условию, он является пучком общего типа с фундаментальными точками M_0, N_0, M_3, N_3 и потому содержит любую линию второго порядка, проходящую через эти точки. В частности, этот пучок содержит пару прямых

$$g_0 g_3 = 0.$$

Так как линии $F_1 = 0$ и $g_0 g_3 = 0$, по условию, различны, существуют такие числа μ и ν (оба отличные от нуля), что

$$F_2 = \mu F_1 + \nu g_0 g_3.$$

Поскольку все уравнения линий нам заданы только с точностью до пропорциональности, мы можем (заменяя, если нужно, F_2 на $\frac{1}{\mu} F_2$, а g_3 — на $\frac{\nu}{\mu} g_3$) без ограничения общности считать, что

$$F_2 = F_1 + g_0 g_3.$$

Аналогично показывается, что

$$F_3 = F_1 + g_0 g_2$$

(при соответствующей нормировке уравнений $F_3 = 0$ и $g_2 = 0$).

Поэтому

$$F_2 - F_3 = g_0 (g_3 - g_2),$$

так что линия $F_2 - F_3 = 0$ является парой прямых $g_0 = 0$ и $g_3 - g_2 = 0$.

Но линия $F_2 - F_3 = 0$ принадлежит пучку $\mu F_2 + \nu F_3 = 0$ и потому проходит через точки M_0, N_0, M_1, N_1. Поскольку прямой $M_0 N_0$ является прямая $g_0 = 0$, это показывает, что прямая $M_1 N_1$ (т. е. прямая $g_1 = 0$) имеет уравнение

$$g_3 - g_2 = 0.$$

Следовательно, прямая $g_1 = 0$ принадлежит пучку, определенному прямыми $g_2 = 0$, $g_3 = 0$, и потому проходит через точку их пересечения.

Тем самым теорема Штурма полностью доказана.

Рассмотрим частный случай, когда точки M_0 и N_0 являются циклическими точками. Тогда линии второго порядка

$$F_1 = 0, \quad F_2 = 0, \quad F_3 = 0$$

будут окружностями, а прямые $g_1 = 0$, $g_2 = 0$, $g_3 = 0$ — общими хордами этих окружностей, т. е. их радикальными осями (см. п. 1 § 3 гл. 4).

В этом случае теорема Штурма переходит в уже известную нам теорему о том, что для любых трех окружностей их радикальные оси пересекаются в одной точке (радикальном центре этих трех окружностей; см. определение 3 п. 1 § 3 гл. 4).

Другие варианты теоремы Штурма получаются, когда некоторые из рассмотренных выше точек совпадают.

Пусть, например, совпадают точки M_0 и N_0, т. е. пусть линии второго порядка

$$F_1 = 0, \quad F_2 = 0, \quad F_3 = 0$$

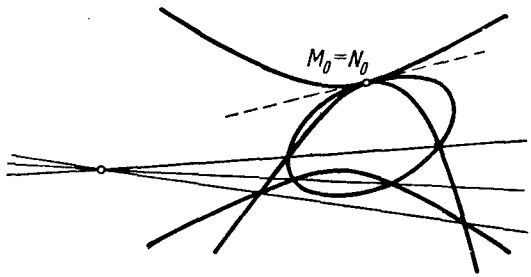
проходят через точку M_0, причем эта точка является двойной точкой пересечения любых двух из этих линий. Другими словами, пусть в точке M_0 все три линии имеют одну и ту же касательную

$$g_0 = 0.$$

Предполагая по-прежнему, что все другие точки пересечения этих линий различны (и отличны от точки M_0), мы можем повторить все изложенные выше рассуждения. Единственное отличие будет теперь в том, что пучок $\mu F_1 + \nu F_2 = 0$ будет теперь пучком типа [211] с фундаментальной прямой $g_0 = 0$, но распадающаяся на прямые линия $g_0 g_3 = 0$ будет, как и раньше, ему принадлежать.

Таким образом,

для трех линий второго порядка, касающихся друг друга в некоторой точке, прямые, проходящие через другие точки их пересечения, проходят через одну точку.



Точно так же, теорема Штурма сохраняется и тогда, когда, например, точки M_1 и N_1 совпадают (т. е. когда в точке M_1 линии $F_2 = 0$ и $F_3 = 0$ касаются друг друга). Естественно, что в этом случае под прямой $g_1 = 0$ следует понимать общую касательную линий $F_2 = 0$ и $F_3 = 0$ в точке M_1.

Задание. Разберите все возможные здесь случаи (когда совпадают точки пересечения одной пары линий, двух пар и т. д.) и сформулируйте соответствующие теоремы.

5. Теорема Паскаля

Пусть на некоторой линии второго порядка

$$F = 0$$

дано шесть ее простых точек. Обозначим их (в произвольном порядке) символами

1, 2, 3, 4, 5, 6.

Определение 1. О шести прямых

12, 23, 34, 45, 56, 61

мы будем говорить, что они образуют *шестиугольник, вписанный в линию* $F = 0$.

Обратим внимание на то, что термин «шестиугольник» здесь используется в смысле, отличном от элементарно-геометрического («сторонами» шестиугольника являются не отрезки, а прямые, и эти стороны могут как угодно пересекаться).

Стороны, между которыми лежит по одной «свободной вершине», т. е. стороны

12 и 45, 23 и 56, 34 и 61,

мы будем называть *противоположными* сторонами нашего шестиугольника.

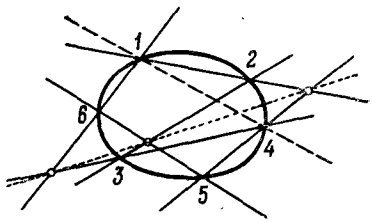
Теорема Паскаля. *Точки пересечения противоположных сторон шестиугольника, вписанного в линию второго порядка, коллинеарны.*

Доказательство. Рассмотрим три линии второго порядка:

1) данную линию,

2) вырожденную линию, состоящую из прямых **12** и **34**,

3) вырожденную линию, состоящую из прямых **16** и **45**.



Эти линии пересекаются в двух точках **1** и **4** (имеют общую хорду **14**), и потому к ним применима теорема Штурма.

Соответствующими прямыми

$$g_1 = 0, \quad g_2 = 0, \quad g_3 = 0$$

являются здесь, как легко видеть, прямые **23**, **65** и прямая, соединяющая точку **12·45** пересечения прямых **12** и **45** с точкой пересечения **34·61** прямых **34** и **61**. Поскольку эти прямые пересекаются в одной точке, мы получаем, что три точки:

а) точка **12·45** пересечения прямых **12** и **45**,

б) точка **23·56** пересечения прямых **23** и **56**,

в) точка **34·61** пересечения прямых **34** и **61**,

действительно лежат на одной прямой.

Теорема Паскаля позволяет по пяти данным точкам (невырожденной) линии второго порядка строить (с помощью одной только линейки) любое число других ее точек. Действительно, пусть

$$\mathbf{1,\ 2,\ 3,\ 4,\ 5}$$

— данные точки (из которых никакие четыре не коллинеарны). Построим точку $\mathbf{12\cdot 45}$ пересечения прямых **12** и **45** и через эту точку проведем произвольную прямую $f = 0$. Пусть эта прямая пересекает прямую **23** в точке M, а прямую **34** — в точке N. Соединив прямыми точку M с точкой **5**, а точку N — с точкой **1**, рассмотрим точку **6** пересечения этих прямых. Легко видеть, что

точка **6** *принадлежит линии второго порядка* $F = 0$, *проходящей через точки* **1—5**.

Действительно, пусть $\mathbf{6'}$ — точка пересечения прямой $M\mathbf{5}$ с линией $F = 0$, отличная от точки **5**. Применим к точкам

$$\mathbf{1,\ 2,\ 3,\ 4,\ 5,\ 6'}$$

теорему Паскаля. Согласно этой теореме три точки пересечения

$$\mathbf{12\cdot 45,\ 23\cdot 56',\ 34\cdot 6'1}$$

лежат на одной прямой. Но поскольку прямая $\mathbf{56'}$ совпадает с прямой $M\mathbf{5}$, то $M = \mathbf{23\cdot 56'}$. Следовательно, указанная прямая совпадает с прямой $f = 0$ (проходящей, по построению, через точки $\mathbf{12\cdot 45}$ и M), и потому точка N (являясь точкой пересечения прямой $f = 0$ с прямой **34**) совпадает с точкой $\mathbf{34\cdot 6'1}$. Это показывает, что точка $\mathbf{6'}$ расположена на прямой $N\mathbf{1}$. Будучи точкой прямой $M\mathbf{5}$, она является, следовательно, точкой пересечения этих прямых, т. е. совпадает с точкой **6**. Поэтому точка **6** принадлежит линии $F = 0$.

Правду сказать, в этом доказательстве допущена небольшая неточность, поскольку не рассмотрены различные «вырожденные случаи», например, случай, когда прямая $M\mathbf{5}$ касается линии $F = 0$ и потому точка $\mathbf{6'}$ совпадает с точкой **5**. Малоприятный труд рассмотрения всех возможных здесь вырождений мы предоставим читателю. Для этого необходимы, конечно, соответствующие «вырожденные варианты» теоремы Паскаля, простейшим из которых является следующее утверждение (получающееся, когда две вершины сливаются в одну):

две пары несмежных сторон вписанного в линию второго порядка пятиугольника пересекаются в двух точках, коллинеарных с точкой пересечения пятой стороны и касательной к линии в противоположной вершине.

Для доказательства этого утверждения достаточно повторить изложенное выше доказательство теоремы Паскаля, только вместо теоремы Штурма следует воспользоваться соответствующим ее вариантом.

Аналогично доказывается, что

в четырехугольнике, вписанном в линию второго порядка, противоположные стороны пересекаются в двух точках, коллинеарных с точкой пересечения касательных к линии в любой паре противоположных вершин,

а также, что

в треугольнике, вписанном в линию второго порядка, точки пересечения каждой из сторон с касательной к линии в противоположной вершине коллинеарны.

Задание. Докажите эти утверждения.

Замечание 1. В справедливости последних утверждений можно непосредственно убедиться, заметив, что они получаются из теоремы Паскаля о шестиугольнике предельным переходом, когда некоторые вершины шестиугольника сливаются в одну точку, поскольку предельным положением секущей является касательная. Однако такого рода соображения «по непрерывности» носят эвристический характер и не могут рассматриваться как доставляющие строгое доказательство.

Рассмотрим теорему Паскаля в частном случае, когда линия $F = 0$ является парой различных прямых (парой совпадающих прямых она быть не может, поскольку на такой линии нет простых точек). Если на одной из прямых, составляющих линию $F = 0$, находятся вершины **1**—**6** с номерами разной четности, то, как легко убедиться, по крайней мере две из рассматриваемых в теореме Паскаля трех точек пересечения совпадают, и потому эта теорема сводится к тавтологии. Напротив, если точки **1**, **3**, **5** расположены на одной прямой, а точки **2**, **4**, **6** — на другой, то, сравнивая формулировки, мы немедленно получим, что в этом случае теорема Паскаля превращается в уже известную нам теорему Паппа — Паскаля (см. п. 6 § 1 гл. 4). Таким образом, общую теорему Паскаля можно рассматривать как обобщение теоремы Паппа — Паскаля на любые линии второго порядка.

6. Квадратичные пучки прямых

Принцип двойственности на проективной плоскости (см. п. 6 § 1 гл. 4) мотивирует следующее

Определение 1. Множество всех прямых $(A : B : C)$, координаты которых удовлетворяют некоторому однородному алгебраическому уравнению

$$\Phi(A, B, C) = 0,$$

называется *алгебраическим пучком прямых.* Степень n этого уравнения называется *порядком* пучка.

В этой терминологии пучки прямых в смысле п. 1 § 1 гл. 4 являются не чем иным, как *пучками первого порядка.*

Для *пучков второго порядка* (которые будем также называть *квадратичными пучками*) может быть развита теория, полностью параллельная (или, лучше сказать, двойственная) развитой выше теории кривых второго порядка.

Например, через любую точку (аффинно-проективной вещественно-комплексной) плоскости проходят, вообще говоря, две прямые данного пучка второго порядка, определяемые корнями некоторого квадратного уравнения. Если это уравнение имеет двойной корень, то точка называется *касательной* к пучку, а единственная прямая пучка, проходящая через эту точку, называется *прямой касания* (о точке говорят также, что она *касается* пучка в этой прямой).

Прямая пучка второго порядка называется *двойной*, если любая ее точка касается пучка в этой прямой. Остальные прямые пучка называются *простыми*. На каждой простой прямой пучка существует единственная точка, касающаяся пучка в этой прямой.

Пучок второго порядка называется *вырожденным*, если он имеет двойную прямую. Каждый такой пучок состоит из двух пучков первого порядка. Если эти пучки различны, то единственной двойной прямой является прямая, соединяющая их центры. Если же они совпадают, то любая прямая данного пучка второго порядка является его двойной прямой.

Если пучок второго порядка имеет уравнение

$$b_{11}A^2 + 2b_{12}AB + b_{22}B^2 + 2b_{13}AC + 2b_{23}BC + b_{33}C^2 = 0, \qquad (1)$$

то он вырожден тогда и только тогда, когда

$$\begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} = 0$$

(как и в аналогичной ситуации для линий, мы считаем, что $b_{ij} = b_{ji}$ для любых i, j). Его двойные прямые $(A : B : C)$ определяются в этом случае из уравнений

$$\begin{aligned} b_{11}A + b_{12}B + b_{13}C &= 0, \\ b_{21}A + b_{22}B + b_{23}C &= 0, \\ b_{31}A + b_{32}B + b_{33}C &= 0. \end{aligned}$$

В любой простой прямой $(A_0 : B_0 : C_0)$ пучка (1) касательная является центром пучка первого порядка, состоящего из прямых $(A : B : C)$, координаты которых удовлетворяют уравнению

$$(b_{11}A_0 + b_{12}B_0 + b_{13}C_0)\,A + (b_{21}A_0 + b_{22}B_0 + b_{23}C_0)\,B + \\ + (b_{31}A_0 + b_{32}B_0 + b_{33}C_0)\,C = 0,$$

т. е. является точкой с координатами

$$(b_{11}A_0 + b_{12}B_0 + b_{13}C_0) : (b_{21}A_0 + b_{22}B_0 + b_{23}C_0) : (b_{31}A_0 + b_{32}B_0 + b_{33}C_0).$$

Любые (не совпадающие и не имеющие общего пучка первого порядка) пучки второго порядка

$$\Phi = 0, \quad \Psi = 0$$

имеют не более четырех общих прямых. Каждой такой общей прямой можно приписать некоторую *кратность* так, чтобы сумма всех их кратностей была равна четырем. Общие прямые кратности 1 называются *простыми прямыми пересечения* двух пучков второго порядка. Если все общие прямые двух пучков второго порядка (не имеющих общего пучка первого порядка) являются простыми прямыми пересечения, то их ровно четыре.

Пусть нам даны три пучка второго порядка

$$\Phi_1 = 0, \quad \Phi_2 = 0, \quad \Phi_3 = 0,$$

имеющие попарно лишь простые прямые пересечения, и пусть эти пучки содержат две данные прямые α_0 и β_0. Тогда любые два из этих пучков содержат еще по паре прямых (отличных от прямых α_0 и β_0). Пусть

α_1, β_1 — дополнительные прямые пересечения пучков $\Phi_2 = 0$ и $\Phi_3 = 0$,

α_2, β_2 — дополнительные прямые пересечения пучков $\Phi_1 = 0$ и $\Phi_3 = 0$,

α_3, β_3 — дополнительные прямые пересечения пучков $\Phi_1 = 0$ и $\Phi_2 = 0$,

и пусть

M_1 — точка пересечения прямых α_1 и β_1,

M_2 — точка пересечения прямых α_2 и β_2,

M_3 — точка пересечения прямых α_3 и β_3.

Теорема, двойственная теореме Штурма, утверждает, что *точки* M_1, M_2, M_3 *расположены на одной прямой.*

Аналогично, пусть в произвольном пучке второго порядка

$$\Phi = 0$$

даны шесть простых прямых

1, 2, 3, 4, 5, 6,

и пусть

12, 23, 34, 45, 56, 61

— их точки пересечения. Рассмотрим прямые

12 · 45, 23 · 56, 34 · 61,

соединяющие точку **12** с точкой **45**, точку **23** — с точкой **56** и точку **34** — с точкой **61**. Тогда теорема, двойственная теореме Паскаля, утверждает, что

прямые **12·34, 23·56** *и* **34·61** *проходят через одну точку.*

Иногда эту теорему называют теоремой Брианшона для пучков второго порядка.

Доказанная в п. 6 § 1 гл. 4 теорема Паппа — Брианшона является, очевидно, частным случаем теоремы Брианшона, получающимся, когда пучок $\Phi = 0$ вырожден, а прямые **1—6**, имеющие номера разной четности, не принадлежат одному из пучков первого порядка, составляющих пучок $\Phi = 0$.

Подчеркнем, что все эти результаты немедленно вытекают из соответствующих результатов для линий второго порядка и того факта, что для проективной плоскости существует хотя бы один поляритет (см. п. 3 § 2). Тем не менее, прямые доказательства этих результатов очень полезны, поскольку только таким образом можно полностью осознать их геометрический смысл.

Задание. Дайте прямые доказательства всех перечисленных выше утверждений.

Между невырожденными линиями и пучками второго порядка существует замечательное взаимоотношение, позволяющее интерпретировать результаты о пучках как результаты о линиях. Поэтому теория пучков второго порядка является не просто теорией, двойственной к теории линий второго порядка, но и представляет собой мощное орудие доказательства новых теорем об этих линиях.

Пусть нам дан произвольный невырожденный квадратичный пучок прямых

$$\Phi(A, B, C) \equiv b_{11}A^2 + 2b_{12}AB + b_{22}B^2 + 2b_{13}AC + \\ + 2b_{23}BC + b_{33}C^2 = 0. \quad (2)$$

Поскольку этот пучок невырожден, в любой его прямой $(A : B : C)$ существует единственная касательная, имеющая координаты

$$\begin{aligned} X &= b_{11}A + b_{12}B + b_{13}C, \\ Y &= b_{21}A + b_{22}B + b_{23}C, \\ Z &= b_{31}A + b_{32}B + b_{33}C. \end{aligned} \quad (3)$$

Определение 2. Множество всех касательных к невырожденному пучку второго порядка называется *огибающей* этого пучка.

Предложение 1. *Огибающая произвольного невырожденного пучка второго порядка является невырожденной линией второго порядка.*

Доказательство. Чтобы найти уравнение, которому удовлетворяют координаты касательной, следует из соотношений (2) и (3) исключить переменные A, B, C. Это делается следующим образом.

Поскольку пучок (2) невырожден, соотношения (3) можно единственным образом разрешить относительно A, B, C.

В результате мы получим формулы вида

$$\begin{aligned} A &= a_{11}X + a_{12}Y + a_{13}Z, \\ B &= a_{21}X + a_{22}Y + a_{23}Z, \\ C &= a_{31}X + a_{32}Y + a_{33}Z, \end{aligned} \tag{4}$$

где

$$U = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

— матрица, обратная к матрице

$$V = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}.$$

Подставив эти выражения в уравнение (2), мы получим уравнение

$$F(X, Y, Z) = 0, \tag{5}$$

которому должны удовлетворять точки, являющиеся касательными. При этом ясно, что любая точка $X : Y : Z$, удовлетворяющая этому уравнению, является касательной к пучку (2) в некоторой его прямой (а именно, в прямой $(A : B : C)$, определяемой соотношениями (4)).

Поскольку уравнение (5) является, очевидно, однородным уравнением второй степени, для завершения доказательства осталось лишь доказать, что линия (5) невырождена, т. е. что невырождена матрица коэффициентов многочлена $F(X, Y, Z)$.

Мы докажем даже большее, а именно, что

матрицей коэффициентов многочлена $F(X, Y, Z)$ *является матрица* U, *обратная матрице* V *коэффициентов многочлена* $\Phi(A, B, C)$.

Это утверждение можно, конечно, проверить прямым вычислением. Однако поскольку такого рода вычисление весьма длинно и утомительно, мы предпочтем обходной путь, использующий символику матричного исчисления.

Введем в рассмотрение столбец x, составленный из координат X, Y, Z, и столбец a, составленный из координат A, B, C:

$$x = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}, \quad a = \begin{pmatrix} A \\ B \\ C \end{pmatrix}.$$

Тогда соотношения (3) мы можем записать в виде одного матричного равенства

$$x = Va,$$

где, как и выше, V — матрица коэффициентов однородного многочлена Φ.

Умножая это равенство слева на $U = V^{-1}$, мы немедленно получим соотношения (4), записанные в виде одного матричного равенства:

$$a = Ux. \tag{6}$$

Рассмотрим теперь матрицу

$$a^{\top} V a.$$

По правилам матричного умножения это должна быть некоторая 1×1-матрица, т. е. число. Произведя умножение, мы немедленно получим, что это число равно $\Phi(A, B, C)$. Таким образом,

$$\Phi(A, B, C) = a^{\top} V a.$$

По определению, если мы выразим здесь столбец a через столбец x, т. е. положим $a = Ux$ (см. формулу (6)), то получим многочлен $F(X, Y, Z)$. Следовательно,

$$F(X, Y, Z) = (Ux)^{\top} V (Ux) = x^{\top} U^{\top} V U x = x^{\top} U x,$$

поскольку, как известно из алгебры,

$$U^{\top} = (V^{-1})^{\top} = (V^{\top})^{-1} = V^{-1} = U$$

(матрица V — симметрическая, и потому $V^{\top} = V$).

Так как однородный многочлен $F(X, Y, Z)$ второй степени единственным образом записывается в виде

$$F(X, Y, Z) = x^{\top} U x,$$

с симметрической матрицей U, то тем самым наше утверждение полностью доказано.

Вместе с тем полностью доказано и предложение 1.

Пусть теперь, обратно,

$$F(X, Y, Z) \equiv a_{11}X^2 + 2a_{12}XY + a_{22}Y^2 + \\ + 2a_{13}XZ + 2a_{23}YZ + a_{33}Z^2 = 0 \tag{7}$$

— произвольная невырожденная линия второго порядка. В каждой точке $(X : Y : Z)$ этой линии существует единственная касательная $(A : B : C)$ с координатами

$$\begin{aligned} A &= a_{11}X + a_{12}Y + a_{13}Z, \\ B &= a_{21}X + a_{22}Y + a_{23}Z, \\ C &= a_{31}X + a_{32}Y + a_{33}Z. \end{aligned} \tag{8}$$

Определение 3. Множество всех этих касательных мы будем называть *пучком касательных* к линии (7).

Предложение 2. *Для произвольной невырожденной линии второго порядка пучок касательных является невырожденным пучком второго порядка.*

Доказательство. Пусть, как и выше, x — столбец координат X, Y, Z, а a — столбец координат A, B, C. Тогда

$$F(X, Y, Z) = x^{\top} U x$$

и (см. формулы (8))

$$a = V x.$$

Поэтому при подстановке в многочлен $F(X, Y, Z)$ вместо X, Y, Z их выражений через A, B, C, получающихся из соотношений (8), мы получим однородный многочлен второй степени

$$\Phi(A, B, C) = a^{\top} (U^{-1})^{\top} U U^{-1} a = a^{\top} U^{-1} a,$$

матрицей коэффициентов которого является матрица U^{-1}.

Поскольку прямая $(A : B : C)$ тогда и только тогда является касательной к линии (7), когда

$$\Phi(A, B, C) = 0$$

(в одну сторону это следует из определения многочлена $\Phi(A, B, C)$, а в другую — из того, что если $\Phi(A, B, C) = 0$, то числа X, Y, Z, определенные уравнениями (8), удовлетворяют уравнению (7) и имеет место равенство $AX + BY + CZ = 0$), тем самым предложение 2 полностью доказано.

На самом деле мы доказали даже большее, а именно, что *соответствия*

$$\text{«пучок»} \longmapsto \text{«огибающая»},$$

$$\text{«линия»} \longmapsto \text{«пучок касательных»}$$

являются взаимно обратными биективными соответствиями между множеством всех невырожденных линий второго порядка и множеством всех невырожденных пучков второго порядка.

Действительно, в силу полной симметричности полученных выше формул, пучок касательных к огибающей является исходным пучком, а огибающая пучка касательных является исходной линией.

Таким образом,

каждый невырожденный пучок второго порядка является пучком касательных к некоторой невырожденной линии второго порядка,

и обратно,

каждая невырожденная линия второго порядка является огибающей некоторого невырожденного пучка второго порядка.

Замечание 1. В п. 3 § 2 мы вывели формулу (см. п. 3 § 2, формула (6)), выражающую необходимое и достаточное условие того, что данная прямая касается данной линии второго порядка. С нашей теперешней точки зрения эта формула является не чем иным, как уравнением пучка касательных. Таким образом,

пучок касательных к линии второго порядка (7) *задается уравнением*

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & A \\ a_{21} & a_{22} & a_{23} & B \\ a_{31} & a_{32} & a_{33} & C \\ A & B & C & 0 \end{vmatrix} = 0. \tag{9}$$

Сравнив это уравнение с полученным выше уравнением $\Phi(A, B, C) = 0$, мы немедленно получаем тождество

$$a^{\top}U^{-1}a = k \begin{vmatrix} a_{11} & a_{12} & a_{13} & A \\ a_{21} & a_{22} & a_{23} & B \\ a_{31} & a_{32} & a_{33} & C \\ A & B & C & 0 \end{vmatrix}, \tag{10}$$

где k — некоторый (отличный от нуля) множитель.

Задание. Проверьте тождество (10) прямым вычислением (и, в частности, найдите множитель k).

Ясно, что уравнение, аналогичное уравнению (9), имеет место и для огибающих, т. е.

огибающая к пучку второго порядка (2) *задается уравнением*

$$\begin{vmatrix} b_{11} & b_{12} & b_{13} & X \\ b_{21} & b_{22} & b_{23} & Y \\ b_{31} & b_{32} & b_{33} & Z \\ X & Y & Z & 0 \end{vmatrix} = 0.$$

Упражнение. Докажите это утверждение тем же методом, которым была доказана формула (8) п. 3 § 2.

Установленные биективные соответствия между линиями и пучками второго порядка позволяют доказывать новые теоремы о линиях второго порядка (или очень просто доказывать уже известные).

Например (см. следствие 1 в п. 3 § 2), мы немедленно получаем, что

через каждую точку (*вещественно-комплексной*) *плоскости, не принадлежащую данной невырожденной линии второго порядка, проходит точно две различных касательных к этой линии.*

Действительно, так как пучок касательных является пучком второго порядка, то через каждую точку плоскости проходят две (возможно, совпадающие) касательные. Если эти касательные совпадают, то рассматриваемая точка является, по определению, касательной к пучку, т. е. принадлежит его оги-

бающей. Для завершения доказательства остается заметить, что огибающей пучка касательных как раз и является данная линия.

Аналогично, применив теорему, двойственную теореме Штурма, к пучкам касательных линий второго порядка, мы немедленно получим, что

если три невырожденные линии второго порядка имеют две общие касательные, то три точки пересечения трех пар прямых, являющихся касательными к двум из данных линий, лежат на одной прямой.

Эта теорема известна как теорема Штурма для касательных.

Чтобы сформулировать теорему, получающуюся подобным образом из теоремы Брианшона, мы для невырожденной линии второго порядка будем называть *описанным вокруг нее шестисторонником* произвольную систему из шести ее касательных («сторон» шестисторонника). Считая эти касательные занумерованными в произвольном порядке, мы будем называть две его «вершины» (точки пересечения соседних сторон) *противоположными*, если между ними расположена «свободная сторона». Таким образом, если **1**, **2**, **3**, **4**, **5**, **6** — это стороны, то противоположные вершины — это вершины

12 и **45**

23 и **56**

34 и **61**.

Прямые, соединяющие противоположные вершины, мы будем называть *главными диагоналями* описанного шестисторонника.

В этой терминологии теорема, получающаяся применением теоремы Брианшона к пучку касательных, утверждает, что

три главные диагонали произвольного шестисторонника, описанного вокруг невырожденной линии второго порядка, пересекаются в одной точке.

Эта теорема обычно называется теоремой Брианшона для линий второго порядка.

Подобно теореме Паскаля, теорема Брианшона обладает рядом «предельных случаев» (теоремы Брианшона для пятиугольника, для четырехсторонника и т. д.).

Задание. Сформулируйте и докажите эти теоремы.

Аналогично тому как теорема Паскаля позволила нам указать построение любого числа точек линии второго порядка по ее пяти точкам, из теоремы Брианшона непосредственно вытекает способ построения (одной только линейкой) любого числа касательных к (невырожденной) линии второго порядка по ее пяти касательным.

Задание. Сформулируйте этот способ и докажите его правильность.

7. Фокусы линий второго порядка

Вернемся к формуле (9) п. 6 (т. е. к формуле (6) п. 2 § 2). При фиксированных a_{ij} эта формула является уравнением пучка касательных к данной линии второго порядка. Однако ее можно читать и «наоборот», предполагая коэффициенты A, B, C фиксированными, а коэффициенты a_{ij} — переменными. Тогда она будет уравнением совокупности линий второго порядка, касающихся данной прямой.

Обратим внимание на то, что уравнение (9) п. 6 является (относительно a_{ij}) уравнением второй степени[1]), тогда как условие, что линия второго порядка проходит через данную точку, линейно (см. п. 3). Это кажется странным, ибо условие, что линия второго порядка касается данной прямой, означает, что ее пучок касательных содержит данную прямую, а по соображениям двойственности условие, что квадратичный пучок прямых содержит данную прямую, линейно. Это кажущееся противоречие разрешается легко: коэффициенты уравнения пучка касательных квадратично выражаются через коэффициенты уравнения линии, и поэтому линейное условие на коэффициенты уравнения пучка оказывается квадратичным условием на коэффициенты уравнения линии.

Можно сказать, что совокупность всех линий второго порядка, проходящих через данную точку, изображается в пятимерном пространстве коэффициентов (см. п. 1) некоторой *гиперплоскостью*, а совокупность всех линий второго порядка, касающихся данной прямой, изображается *гиперповерхностью второго порядка*.

Поэтому, в то время как семейство всех линий второго порядка, проходящих через четыре фиксированные точки, изображается линией пересечения четырех гиперплоскостей, т. е. некоторой прямой (и потому является пучком), семейство всех линий второго порядка, касающихся четырех данных прямых, изображается линией пересечения четырех гиперповерхностей второго порядка, и потому представляет собой весьма сложную кривую.

По этой причине изучение семейств линий второго порядка, касающихся четырех фиксированных прямых, несравненно сложнее изучения пучков. Поэтому мы их рассматривать не будем.

Любопытно все же отметить, что одно такое семейство (правда, очень специального вида) выше мы уже изучали. Именно, оказывается, что рассмотренное в п. 4 § 2 гл. 5 *семейство софокусных эллипсов и гипербол может быть охарактеризовано (на евклидовой вещественно-комплексной плоскости) как семейство линий второго порядка, касающихся четырех прямых.*

Чтобы показать это, мы напомним (см. пп. 1—3 § 1 гл. 5), что фокусом невырожденной линии второго порядка (эллипса, гиперболы или параболы) является точка, обладающая тем свойством, что для любой точки линии отношение ее расстояния до

[1]) На первый взгляд даже третьей! Однако, ввиду нуля в правом нижнем углу определителя, члены третьей степени (по a_{ij}) в его разложении отсутствуют.

этой точки к ее расстоянию до некоторой фиксированной прямой (директрисы) постоянно (равно эксцентриситету e линии).

Если в некоторой системе прямоугольных координат фокус имеет координаты (x_0, y_0), а директриса — уравнение

$$Ax + By + C = 0, \tag{1}$$

то это характеристическое свойство фокуса может быть записано в виде соотношения

$$\sqrt{(x - x_0)^2 + (y - y_0)^2} = \frac{e}{\sqrt{A^2 + B^2}} | Ax + By + C |.$$

Предполагая, что уравнение директрисы (1) нормировано так, что

$$\sqrt{A^2 + B^2} = e,$$

мы (после возведения в квадрат) можем это уравнение переписать в следующем виде:

$$(x - x_0)^2 + (y - y_0)^2 = (Ax + By + C)^2. \tag{2}$$

Поскольку это уравнение является алгебраическим эквивалентом описания линии второго порядка как геометрического места точек, отношение расстояния которых до точки (x_0, y_0) к расстоянию до прямой (1) равно e, мы видим, что *уравнение (2) является уравнением невырожденной линии второго порядка с фокусом в точке (x_0, y_0) и директрисой (1), причем любое уравнение такого вида (для которого прямая (1) не проходит через точку (x_0, y_0)) является уравнением невырожденной линии второго порядка с фокусом в точке (x_0, y_0), директрисой (1) и эксцентриситетом $e = \sqrt{A^2 + B^2}$.*

Задание. Покажите, что если прямая (1) проходит через точку (x_0, y_0), то линия (2) является парой прямых, проходящих через точку (x_0, y_0).

Уравнение (2) показывает, что выражаемая этим уравнением линия принадлежит пучку линий второго порядка, определенному окружностью нулевого радиуса

$$(x - x_0)^2 + (y - y_0)^2 = 0 \tag{3}$$

и двойной прямой

$$(Ax + By + C)^2 = 0. \tag{4}$$

Фундаментальными точками этого пучка являются точки пересечения (имеющие, очевидно, кратность 2) окружности (3) и двойной прямой (4). (Таким образом, рассматриваемый пучок имеет тип [22].)

Но тогда, согласно общей теории пучков линий второго порядка, каждая линия пучка (и, в частности, данная линия (2)) будет иметь эти точки двойными точками пересечения с окружностью (3). Поскольку тот факт, что точка пересечения — двой-

ная, равносилен, по определению, тому, что линии в этой точке касаются (имеют общую касательную), тем самым доказано, что

если точка (x_0, y_0) является фокусом линии второго порядка, то окружность нулевого радиуса с центром в этой точке касается линии в двух точках.

Задание. Могут ли эти точки касания совпадать?

Обратно,

если окружность нулевого радиуса с центром в точке (x_0, y_0) касается в двух точках некоторой невырожденной линии второго порядка, то точка (x_0, y_0) является фокусом этой линии.

Действительно, пусть

$$F = 0 \tag{5}$$

— уравнение данной линии. По условию пучок линий второго порядка, определенный линией (5) и окружностью (3), имеет две двойные фундаментальные точки. Поэтому содержащаяся в нем вырожденная линия второго порядка является двойной прямой, проходящей через эти точки. Следовательно, уравнение (5) является линейной комбинацией уравнения окружности (3) и уравнения некоторой линии вида (4). Пользуясь тем, что уравнения (5) и (4) нам заданы только с точностью до пропорциональности (а также учитывая, что по условию линия (5) невырождена), мы, следовательно, без ограничения общности можем считать, что

$$F = (x - x_0)^2 + (y - y_0)^2 - (Ax + By + C)^2,$$

т. е. что уравнение линии имеет вид (2). Поэтому точка (x_0, y_0) является фокусом (а прямая $Ax + By + C = 0$ — директрисой).

Таким образом, фокус невырожденной линии второго порядка мы можем определить как точку, обладающую тем свойством, что окружность нулевого радиуса с центром в этой точке касается кривой в двух точках. При этом прямая, проходящая через точки касания, является соответствующей директрисой.

Но мы знаем (см. п. 6 § 3 гл. 4), что окружность нулевого радиуса является парой изотропных прямых, проходящих через ее центр. С другой стороны, если пара прямых касается линии второго порядка в двух точках, то это значит, что каждая из этих прямых касается линии (в соответствующей точке).

Все эти соображения мотивируют следующее общее

Определение 1. Точка вещественно-комплексной (евклидовой или евклидово-проективной) плоскости называется *фокусом* линии второго порядка, если проходящие через эту точку изотропные прямые касаются этой линии (вообще говоря, в точках, отличных от фокуса).

В частности, мы видим, что

изотропные прямые, проходящие через центр произвольной окружности, касаются этой окружности в циклических точках.

Как мы знаем, через любую точку плоскости (не принадлежащую данной линии второго порядка) проходят две касательные к этой линии. В частности, это верно для циклических точек.

Определение 2. Касательные, проходящие через циклические точки, называются *изотропными касательными.*

Вообще говоря, линия второго порядка имеет четыре изотропные касательные (по две для каждой циклической точки). Если же линия проходит через циклические точки, то число ее изотропных касательных соответственно уменьшается. Например, окружность имеет только две изотропные касательные (пересекающиеся в ее центре).

Пользуясь понятием изотропной касательной, мы можем сказать, что

фокус линии второго порядка является точкой пересечения изотропных касательных этой линии.

Следовательно,

линия второго порядка имеет, вообще говоря, четыре фокуса.

Действительно, две пары изотропных касательных пересекаются, вообще говоря, в четырех точках.

Если линия второго порядка вещественна, то ее изотропные касательные попарно комплексно-сопряжены и потому пересекаются в двух вещественных и двух мнимых (комплексно-сопряженных) точках. Таким образом,

вещественная линия второго порядка имеет, вообще говоря, два вещественных и два комплексно-сопряженных фокуса.

Найдем, например, фокусы эллипса

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1.$$

Пучок его касательных имеет уравнение

$$\begin{vmatrix} \frac{1}{a^2} & 0 & 0 & A \\ 0 & \frac{1}{b^2} & 0 & B \\ 0 & 0 & -1 & C \\ A & B & C & 0 \end{vmatrix}=0,$$

т. е. уравнение

$$a^2A^2+b^2B^2=C^2.$$

Чтобы найти прямые этого пучка, проходящие через циклическую точку $(1:i:0)$, надо решить это уравнение совместно с уравнением

$$A+iB=0.$$

Сделав это, мы получим, что

$$B=iA,$$

$$C=\pm cA,$$

где, как всегда, $c=\sqrt{a^2-b^2}$.

Следовательно, изотропные касательные эллипса, проходящие через циклическую точку $(1 : i : 0)$, имеют уравнения

$$x + iy \pm c = 0.$$

Аналогично показывается, что изотропными касательными эллипса, проходящими через циклическую точку $(1 : -i : 0)$, являются прямые

$$x - iy \pm c = 0.$$

Из четырех точек попарного пересечения этих четырех прямых имеем две вещественные:

$$(-c, 0), \quad (c, 0)$$

(это — известные из элементарной теории фокусы эллипса) и две мнимые:

$$(0, -ic), \quad (0, ic)$$

(эти фокусы в элементарной теории не рассматриваются).

Аналогично находятся фокусы гиперболы и параболы.

В частности, мнимыми фокусами любой параболы являются циклические точки, а из вещественных фокусов один является известным элементарным фокусом $(p/2, 0)$, а другой — несобственной точкой ее оси.

Для окружности имеется один («четырехкратный») фокус — ее центр.

Теперь мы видим, что

семейство всех (вещественных) линий второго порядка, касающихся четырех (попарно комплексно-сопряженных) изотропных прямых, является не чем иным, как семейством софокусных эллипсов и гипербол.

Понятие фокуса (и директрисы) имеет красивую интерпретацию и в терминах теории поляр.

Например, легко понять, что

директриса является полярой (одноименного) фокуса.

Упражнение. Докажите это утверждение.

Назовем две прямые *взаимно полярными* (по отношению к данной невырожденной линии второго порядка), если каждая из них проходит через полюс другой прямой (см. предложение 1 п. 3 § 2). Оказывается, что

точка (проективно-евклидовой вещественно-комплексной или вещественной плоскости) тогда и только тогда является фокусом невырожденной линии второго порядка, когда любые две взаимно полярные прямые, проходящие через эту точку, перпендикулярны.

Упражнение. Докажите это утверждение.

Заметим, что это утверждение дает, в частности, проективно-евклидову интерпретацию фокуса на вещественной плоскости (без выхода в комплексную область).

Упражнение. Докажите директориальное свойство эллипса, гиперболы и параболы (предложение 2 п. 1 § 1 гл. 5, предложение 3 п. 2 § 1 гл. 5 и предложение 3 п. 3 § 1 гл. 5), основываясь на материале этого пункта.

Глава 7

ГЕОМЕТРИЧЕСКИЕ ПРЕОБРАЗОВАНИЯ

§ 1. АФФИННЫЕ, ПРОЕКТИВНЫЕ И ОРТОГОНАЛЬНЫЕ ПРЕОБРАЗОВАНИЯ

1. Аффинные преобразования

Основным понятием аффинной геометрии является, конечно, понятие прямой. Поэтому естественно возникает вопрос о наиболее общих преобразованиях (вещественной) аффинной плоскости или (вещественного) аффинного пространства, переводящих прямые в прямые. Такого рода преобразования будут в этой главе изучены в рамках элементарного содержательного понимания геометрии. Уточнения аксиоматического плана предоставляются инициативе читателя.

Определение 1. Преобразование (биективное отображение на себя; см. п. 1 § 7 гл. 1) аффинной плоскости или аффинного пространства называется *аффинным*, если любые три коллинеарные точки оно переводит в три коллинеарные точки («сохраняет отношение коллинеарности»).

Замечание 1. Для «плоскостного случая» можно рассматривать также биективные отображения одной плоскости на другую, сохраняющие отношение коллинеарности точек. Такие отображения называются *аффинными отображениями.* Читателю предлагается самостоятельно продумать, что́ из доказываемых ниже утверждений об аффинных преобразованиях сохраняется для аффинных отображений (и в какой форме).

Как и для произвольного преобразования, для любого аффинного преобразования Φ существует обратное преобразование Φ^{-1}.

Оказывается, что

преобразование Φ^{-1}, обратное к аффинному преобразованию Φ, также является аффинным преобразованием.

Рассуждая «от противного», предположим, что существует коллинеарная тройка точек M_1', M_2', M_3', которую преобразование Φ^{-1} переводит в неколлинеарную тройку точек M_1, M_2, M_3. Пусть α' — прямая $M_1'M_2'M_3'$, P — плоскость $M_1M_2M_3$, а α_{12}, α_{13} и α_{23} — прямые M_1M_2, M_1M_3 и M_2M_3 соответственно. По условию

все три прямые α_{12}, α_{13} и α_{23} попарно различны (и не проходят через одну точку).

Пусть M — произвольная точка прямой α_{12} и пусть M' — ее образ при преобразовании Ф. Поскольку тройка M_1, M_2, M коллинеарна, тройка M_1', M_2', M' также коллинеарна, т. е. точка M' принадлежит прямой $\alpha' = M_1'M_2'$. Таким образом, всю прямую α_{12} преобразование Ф переводит в прямую α'. Аналогично доказывается, что преобразование Ф переводит в прямую α' также и прямые α_{13} и α_{23}.

Пусть теперь M — произвольная точка плоскости P, отличная от точек прямых α_{12}, α_{13} и α_{23}. Проведем через эту точку произвольную прямую, пересекающую, скажем, прямые α_{12} и α_{13} в двух различных точках N_2 и N_3. По только что доказанному, образы N_2' и N_3' точек N_2 и N_3 при преобразовании Ф принадлежат прямой α'. С другой стороны, так как точки N_2, N_3, M по построению коллинеарны, то точки N_2', N_3', M' также коллинеарны. Следовательно, точка $M' = \Phi(M)$ принадлежит прямой $\alpha' = N_2'N_3'$. Этим доказано, что преобразование Ф всю плоскость P отображает в прямую α'.

В «плоскостном случае» (Ф является преобразованием плоскости) мы тем самым уже пришли к противоречию (всю плоскость преобразование Ф отображает в ее часть — прямую α', что в силу условия биективности преобразования невозможно).

В «пространственном случае» (Ф является преобразованием пространства) нам следует сделать еще один шаг. Пусть M_0 — произвольная точка пространства, не принадлежащая плоскости P. Рассмотрим плоскость P', содержащую прямую α' и образ M_0' точки M_0 при преобразовании Ф (если точка M_0' не принадлежит прямой α', эта плоскость однозначно определена; в противном случае мы за нее принимаем произвольную плоскость, проходящую через прямую α').

Пусть M — произвольная точка пространства, отличная от точки M_0. Рассмотрим сначала случай, когда прямая M_0M пересекает плоскость P. Пусть N — соответствующая точка пересечения. Поскольку точки M_0, M, N коллинеарны, их образы M_0', M', N' также коллинеарны. Но, по построению, точки M_0' и N' принадлежат плоскости P'. Поэтому точка M' также принадлежит этой плоскости.

Пусть, наконец, прямая M_0M параллельна плоскости P, т. е. пусть точка M принадлежит плоскости P_0, проходящей через точку M_0 параллельно плоскости P. Проведем через точку M произвольную прямую, не лежащую в плоскости P_0, и выберем на ней две точки N_1, N_2, отличные от точки M (и потому не принадлежащие плоскости P_0). По уже доказанному образы N_1' и N_2' точек N_1 и N_2 принадлежат плоскости P'. Следовательно, поскольку точки N_1, N_2, M, а значит, и точки N_1', N_2', M' коллинеарны, точка M' также принадлежит плоскости P'.

Таким образом, преобразование Φ переводит все пространство в плоскость P', что невозможно. Следовательно, преобразование Φ^{-1} не может не быть аффинным.

Ясно, что композиция двух аффинных преобразований также является аффинным преобразованием. Поэтому (см. п. 1 § 7 гл. 1) справедливо следующее

Предложение 1. *Совокупность всех аффинных преобразований плоскости или пространства является группой преобразований.*

Эту группу мы будем обозначать символом Aff(2) в случае плоскости и символом Aff(3) — в случае пространства.

Докажем теперь несколько простейших свойств аффинных преобразований.

Пусть Φ — произвольное аффинное преобразование (плоскости или пространства) и пусть α — произвольная прямая. Выбрав на прямой α две точки M_1 и M_2, рассмотрим прямую α', проходящую через точки M_1' и M_2', являющиеся образами точек M_1 и M_2 при преобразовании Φ. Ясно, что прямая α' содержит образ M' любой точки M прямой α (и потому, в частности, не зависит от выбора точек M_1 и M_2).

Определение 2. Мы будем говорить, что прямая α' *соответствует* прямой α в аффинном преобразовании Φ.

Свойство 1. *Аффинное преобразование* Φ *каждую прямую* α *отображает на соответствующую прямую* α'.

Доказательство. В обратном преобразовании Φ^{-1} прямой α' соответствует, очевидно, прямая α (ибо точки M_1' и M_2' преобразование Φ^{-1} переводит в точки M_1 и M_2). Это означает, что каждая точка M' прямой α' преобразованием Φ^{-1} переводится в некоторую точку M прямой α (напомним, что преобразование Φ^{-1} аффинно). Поскольку $M' = \Phi(M)$, то, следовательно, образом прямой α при преобразовании Φ является вся прямая α'.

Предположим теперь, что Φ является аффинным преобразованием пространства, и рассмотрим произвольную плоскость P. Выбрав на этой плоскости три неколлинеарные точки M_1, M_2, M_3, рассмотрим их образы M_1', M_2', M_3' при преобразовании Φ. По доказанному выше точки M_1', M_2', M_3' не коллинеарны, и потому определяют единственную плоскость P'.

Пусть M — произвольная точка плоскости P и пусть M' — ее образ при преобразовании Φ. Если точка M принадлежит прямой M_1M_2, то точка M' принадлежит прямой $M_1'M_2'$, а значит, и плоскости P'. Аналогично доказывается, что точка M' принадлежит плоскости P' и тогда, когда точка M расположена на прямых M_1M_3 и M_2M_3.

Пусть точка M не принадлежит ни одной из прямых M_1M_2, M_1M_3, M_2M_3. Проведем через эту точку произвольную прямую, пересекающую, скажем, прямые M_1M_2 и M_1M_3. Пусть N_2 и

N_3 — соответствующие точки пересечения. По уже доказанному, образы N_2' и N_3' этих точек принадлежат плоскости P′. Поэтому точка M', будучи коллинеарной с точками N_2' и N_3', также принадлежит плоскости P′. Таким образом, мы доказали, что

плоскость P′ *содержит образ* M' *произвольной точки* M *плоскости* P.

Это показывает, в частности, что плоскость P′ зависит только от плоскости P и преобразования Ф (но не зависит от выбора точек M_1, M_2, M_3).

Определение 2′. О плоскости P′ мы будем говорить, что она *соответствует* плоскости P в аффинном преобразовании Ф.

Свойство 1′. *Аффинное преобразование* Ф *каждую плоскость* P *отображает на соответствующую плоскость* P′.

Доказательство по существу дословно совпадает с доказательством свойства 1, и потому мы его опустим.

Задание. Докажите свойство 1′.

Замечание 2. Определим отображение $Ф_P$ плоскости P на плоскость P′, считая, что каждую точку M оно переводит в ту же точку M', что и исходное преобразование Ф. (Таким образом, отображения Ф и $Ф_P$ действуют на точках одинаковым образом; их различие состоит только в том, что отображение $Ф_P$ определено лишь для точек M плоскости P, а соответствующие точки M' считаются точками плоскости P′). Свойство 1′ утверждает, что отображение $Ф_P$ является биективным отображением плоскости P на плоскость P′. Кроме того, это отображение сохраняет, очевидно, отношение коллинеарности точек (плоскости P). Таким образом,

отображение $Ф_P$ *является аффинным отображением плоскости* P *на плоскость* P′.

Определение 3. Мы будем говорить, что отображение $Ф_P$ *индуцировано* преобразованием Ф.

Согласно сказанному выше каждое аффинное преобразование Ф определяет некоторое биективное отображение (преобразование) множества всех прямых (плоскости или пространства) на себя, а в «пространственном случае» — и множества всех плоскостей. Мы будем эти преобразования обозначать тем же самым символом Ф.

Свойство 2. *Преобразование* Ф *переводит параллельные прямые в параллельные.*

Доказательство. Параллельные прямые — это непересекающиеся прямые, расположенные в одной плоскости. Поскольку любое аффинное преобразование переводит прямые, расположенные в одной плоскости, в прямые, также расположенные в одной плоскости (для аффинных преобразований плоскости это верно автоматически, а для аффинных преобразований пространства вытекает из свойства 1′), для доказательства свойства 2 достаточно поэтому доказать, что любое аф-

финное преобразование Φ переводит непересекающиеся прямые α и β в прямые α' и β', также не пересекающиеся.

Рассуждая «от противного», предположим, что прямые α' и β' имеют общую точку M'. Пусть $M = \Phi^{-1}(M')$. Поскольку (свойство 1) преобразование Φ отображает прямую α на прямую α', точка M принадлежит прямой α. Но по тем же соображениям точка M принадлежит и прямой β. Таким образом, вопреки предположению, прямые α и β пересекаются.

Свойство 2'. *Преобразование Φ переводит параллельные плоскости в параллельные.*

Задание. Докажите свойство 2'.

Чтобы сформулировать следующее свойство аффинных преобразований, мы рассмотрим две различные точки M_1, M_2 и проходящую через них прямую α. Пусть k — отношение, в котором некоторая точка M прямой α делит направленный отрезок $\overrightarrow{M_1M_2}$ (см. п. 3 § 1 гл. 2). При произвольном аффинном преобразовании Φ коллинеарные точки M_1, M_2, M перейдут в коллинеарные точки M_1', M_2', M', и потому будет определено отношение k', в котором точка M' делит отрезок $\overrightarrow{M_1'M_2'}$.

Свойство 3. *Имеет место равенство*

$$k' = k,$$

т. е. аффинное преобразование сохраняет отношение, в котором данная точка делит данный отрезок.

Поскольку внутренние точки отрезка $\overline{M_1M_2}$ характеризуются условием $0 < k < \infty$, из этого свойства вытекает

Следствие. *Преобразование Φ каждую внутреннюю точку отрезка $\overline{M_1M_2}$ переводит во внутреннюю точку отрезка $\overline{M_1'M_2'}$.*

В отличие от первых двух свойств свойство 3 требует для своего доказательства довольно глубоких соображений. Мы докажем его в следующем пункте, связав аффинные преобразования с линейными операторами.

2. Линейный оператор, индуцированный аффинным преобразованием

Определение 1. Оператором (см. определение 1 п. 1 § 7 гл. 1), *индуцированным* аффинным преобразованием Φ, называется оператор $\boldsymbol{\Phi}$, определенный формулой

$$\boldsymbol{\Phi}(\overrightarrow{M_1M_2}) = \overrightarrow{M_1'M_2'},$$

где $M_1' = \Phi(M_1)$ и $M_2' = \Phi(M_2)$.

Это определение нуждается, конечно, в проверке корректности, т. е. нужно доказать, что

если направленные отрезки $\overrightarrow{M_1M_2}$ *и* $\overrightarrow{N_1N_2}$ *эквиполлентны, то направленные отрезки* $\overrightarrow{M_1'M_2'}$ *и* $\overrightarrow{N_1'N_2'}$ *также эквиполлентны.*

Если бы свойство 3 из п. 1 нами уже было доказано, доказательство этого факта не составляло бы никакого труда. Действительно, по определению, отрезки $\overrightarrow{M_1M_2}$ и $\overrightarrow{N_1N_2}$ эквиполлентны, если середина отрезка $\overline{M_1N_2}$ совпадает с серединой отрезка $\overline{M_2N_1}$. Поскольку в силу свойства 3 (примененного к случаю $k = 1/2$) середина отрезка переходит при преобразовании Ф в середину отрезка, то середина отрезка $\overline{M_1'N_2'}$ совпадает с серединой отрезка $\overline{M_2'N_1'}$ и потому направленные отрезки $\overrightarrow{M_1'M_2'}$ и $\overrightarrow{N_1'N_2'}$ эквиполлентны.

Однако к доказательству свойства 3 мы по существу и стремимся. Чтобы разорвать получающийся порочный круг, мы заметим, что в изложенном доказательстве свойство 3 использовалось не в полном объеме, а лишь при $k = 1/2$. Поэтому все будет в порядке, если мы этот частный случай свойства 3 независимо докажем.

Таким образом, все свелось к доказательству того, что

при любом аффинном преобразовании Ф (*плоскости или пространства*) *середина* C *произвольного отрезка* $\overline{M_1M_2}$ *переходит в середину* C' *отрезка* $\overline{M_1'M_2'}$.

Мы сначала докажем это утверждение для случая, когда Ф является аффинным преобразованием плоскости.

С этой целью мы проведем в рассматриваемой плоскости через точки M_1 и M_2 две различные пары параллельных прямых, отличных от прямой M_1M_2. Эти прямые образуют параллелограмм с вершинами в точках M_1, M_2 и еще в двух точках N_1 и N_2. Согласно известному свойству диагоналей параллелограмма, прямые M_1M_2 и N_1N_2 пересекаются в середине C отрезка $\overline{M_1M_2}$. При аффинном преобразовании Ф точки M_1, M_2, N_1 и N_2 перейдут в некоторые точки M_1', M_2', N_1', N_2', причем согласно свойству 2 из п. 1 прямая $M_1'N_1'$ останется параллельна прямой $M_2'N_2'$, а прямая $M_1'N_2'$—прямой $M_2'N_1'$. Другими словами, четырехугольник с вершинами M_1', M_2', N_1', N_2' по-прежнему останется параллелограммом. Следовательно, прямые $M_1'M_2'$ и $N_1'N_2'$ будут пересекаться в точке C', являющейся серединой отрезка $\overrightarrow{M_1'M_2'}$.

Но согласно свойству 1 из п. 1 прямая M_1M_2 отображается на прямую $M_1'M_2'$, а прямая N_1N_2 — на прямую $N_1'N_2'$. Следовательно, точка C пересечения этих прямых перейдет в точку C' пересечения прямых $M_1'M_2'$ и $N_1'N_2'$. Другими словами, середина отрезка $\overline{M_1M_2}$ перейдет в середину отрезка $\overline{M_1'M_2'}$.

В случае, когда Ф представляет собой аффинное преобразование пространства, достаточно рассмотреть произвольную

плоскость, содержащую точки M_1, M_2, и применить уже доказанное утверждение к аффинному отображению этой плоскости, индуцированному преобразованием Φ.

Тем самым корректность определения 1 полностью доказана.

Предложение 1. *Оператор* Φ, *индуцированный аффинным преобразованием* Φ, *является линейным оператором.*

Доказательство. Свойство аддитивности

$$\Phi(x+y)=\Phi x+\Phi y$$

означает, что для любых трех точек M_1, M_2, M_3 имеет место равенство

$$\overrightarrow{M_1'M_2'}+\overrightarrow{M_2'M_3'}=\overrightarrow{M_1'M_3'}.$$

Но это очевидно.

Труднее доказывается свойство однородности

$$\Phi(kx)=k\Phi x. \tag{1}$$

В первую очередь мы заметим, что

если $x=0$, *то* $\Phi x=0$.

Действительно, если $x=\overrightarrow{MM}$, то $\Phi x=\overrightarrow{M'M'}$.

Поэтому соотношение (1) нам нужно доказать только при $x\neq 0$ и $k\neq 0$.

Это соотношение при $x\neq 0$ означает, во-первых, что

векторы Φx *и* $\Phi(kx)$ *коллинеарны,*

а во-вторых, что

отношение $\Phi(kx):\Phi x$ *коллинеарных векторов* $\Phi(kx)$ *и* Φx *равно* k.

Первое утверждение доказывается без труда. Действительно, пусть $x=\overrightarrow{OM}$ и $kx=\overrightarrow{ON}$. Тогда $\Phi x=\overrightarrow{O'M'}$ и $\Phi(kx)=\overrightarrow{O'N'}$. Поскольку векторы x и kx коллинеарны, точки O, M, N также коллинеарны. Но тогда коллинеарны и точки O', M', N', что и означает, что векторы $\Phi x=\overrightarrow{O'M'}$ и $\Phi(kx)=\overrightarrow{O'M'}$ коллинеарны.

Таким образом, нам нужно только доказать, что отношение векторов $\Phi(kx)$ и Φx равно k. Для этого мы предварительно докажем, что

отношение $\Phi(kx):\Phi x$ *не зависит от вектора* x, *т. е. для любых двух векторов* $x\neq 0$ *и* $y\neq 0$ *имеет место равенство*

$$\Phi(kx):\Phi x=\Phi(ky):\Phi y.$$

Предположим сначала, что векторы x и y не коллинеарны. Пусть, как и выше,

$$x=\overrightarrow{OM},\quad kx=\overrightarrow{ON}$$

и пусть

$$y=\overrightarrow{OM_1},\quad ky=\overrightarrow{ON_1}.$$

По условию, прямые $OM = ON$ и $OM_1 = ON_1$ (пересекающиеся в точке O) различны. В этом случае тот факт, что отношение $k\boldsymbol{y} : \boldsymbol{y}$ вектора $k\boldsymbol{y}$ к вектору $\boldsymbol{y}$ равно отношению $k\boldsymbol{x} : \boldsymbol{x}$ вектора $k\boldsymbol{x}$ к вектору $\boldsymbol{x}$, означает, что прямые MM_1 и NN_1 параллельны. Но тогда (свойство 2 из п. 1) прямые $M'M_1'$ и $N'N_1'$ также параллельны. Следовательно, отношение вектора $\Phi(k\boldsymbol{y}) = \overrightarrow{O'N'}$ к вектору $\Phi\boldsymbol{y} = \overrightarrow{O'M_1'}$ равно отношению вектора $\Phi(k\boldsymbol{x}) = \overrightarrow{O'N'}$ к вектору $\Phi\boldsymbol{x} = \overrightarrow{O'M'}$.

В случае, когда векторы $\boldsymbol{x}$ и $\boldsymbol{y}$ коллинеарны, мы введем в рассмотрение третий вспомогательный вектор $\boldsymbol{z}$, им не коллинеарный. Поскольку векторы $\boldsymbol{x}$ и $\boldsymbol{z}$ не коллинеарны, то по уже доказанному

$$\Phi(k\boldsymbol{x}) : \Phi\boldsymbol{x} = \Phi(k\boldsymbol{z}) : \Phi\boldsymbol{z}.$$

Аналогично,

$$\Phi(k\boldsymbol{y}) : \Phi\boldsymbol{y} = \Phi(k\boldsymbol{z}) : \Phi\boldsymbol{z}.$$

Следовательно,

$$\Phi(k\boldsymbol{x}) : \Phi\boldsymbol{x} = \Phi(k\boldsymbol{y}) : \Phi\boldsymbol{y}.$$

Тем самым независимость отношения $\Phi(k\boldsymbol{x}) : \Phi\boldsymbol{x}$ от выбора вектора $\boldsymbol{x} \neq 0$ полностью доказана.

Это отношение является, таким образом, некоторой функцией только числа k:

$$\Phi(k\boldsymbol{x}) : \Phi\boldsymbol{x} = f(k),$$

и наша задача состоит в том, чтобы доказать, что эта функция тождественно равна k.

Пусть $\boldsymbol{e}$ — произвольный отличный от нуля вектор. Так как для любых чисел k и l

$$(k + l)\boldsymbol{e} = k\boldsymbol{e} + l\boldsymbol{e},$$

то в силу уже доказанного свойства аддитивности

$$\Phi((k + l)\boldsymbol{e}) = \Phi(k\boldsymbol{e}) + \Phi(l\boldsymbol{e}),$$

и потому

$$\Phi((k + l)\boldsymbol{e}) : \Phi\boldsymbol{e} = \Phi(k\boldsymbol{e}) : \Phi\boldsymbol{e} + \Phi(l\boldsymbol{e}) : \Phi\boldsymbol{e},$$

т. е.

$$f(k + l) = f(k) + f(l).$$

Далее,

$$\Phi((kl)\boldsymbol{e}) = \Phi(k(l\boldsymbol{e})) = f(k)\Phi(l\boldsymbol{e}) = f(k)f(l)\Phi(l\boldsymbol{e}),$$

т. е.

$$f(kl) = f(k)f(l).$$

Наконец, поскольку аффинное преобразование Φ переводит равные векторы в равные, то

$$f(1) = 1.$$

Таким образом, мы доказали, что
функция $f(k)$ обладает следующими тремя свойствами:

$$f(k+l)=f(k)+f(l),$$
$$f(kl)=f(k)f(l),$$
$$f(1)=1.$$

Любая числовая функция $f(k)$, обладающая этими свойствами, называется *автоморфизмом поля* $\mathbb{R}$ вещественных чисел. Пользуясь этой терминологией, мы можем, следовательно, сказать, что

функция $f(k)$ является автоморфизмом поля $\mathbb{R}$.

Функция $f(k)=k$ является, конечно, автоморфизмом. Этот автоморфизм называется *тождественным.*

Для доказательства предложения 1 нам достаточно теперь доказать следующую лемму:

Лемма Дарбу. *Поле вещественных чисел не имеет автоморфизмов, отличных от тождественного.*

Доказательство. Пусть f — произвольный автоморфизм. Нам нужно доказать, что для любого числа $x\in\mathbb{R}$ имеет место равенство

$$f(x)=x.$$

Мы это сделаем в семь этапов.

1°. *Для любых чисел x_1, x_2 имеет место равенство*

$$f(x_2-x_1)=f(x_2)-f(x_1).$$

Действительно, $x_1+(x_2-x_1)=x_2$, и потому

$$f(x_1)+f(x_2-x_1)=f(x_2).$$

2°. *Если $x_1\leqslant x_2$, то $f(x_1)\leqslant f(x_2)$.*

Действительно,

$$f(x_2)-f(x_1)=f(x_2-x_1)=f\left(\left(\sqrt{x_2-x_1}\right)^2\right)=\left(f\left(\sqrt{x_2-x_1}\right)\right)^2\geqslant 0.$$

3°. $f(0)=0$.

Действительно,

$$f(0)=f(1-1)=f(1)-f(1)=0.$$

4°. $f(-x)=-f(x)$.

Действительно,

$$f(-x)=f(0-x)=f(0)-f(x)=0-f(x)=-f(x).$$

5°. $f(x)=x$, *если x — целое.*

Согласно свойству 4° мы без ограничения общности можем предполагать, что целое число x положительно, т. е. является натуральным числом n. Тогда

$$f(n)=f(\underbrace{1+\ldots+1}_{n\text{ раз}})=\underbrace{f(1)+\ldots+f(1)}_{n\text{ раз}}=\underbrace{1+\ldots+1}_{n\text{ раз}}=n.$$

6°. $f(x)=x$, *если* x *рационально.*

Для рационального числа x существует такое натуральное n, что число nx—целое. По уже доказанному $f(nx)=nx$ и $f(n)=n$, а по условию

$$f(nx)=f(n)f(x).$$

Следовательно,

$$f(x)=\frac{f(nx)}{f(x)}=\frac{nx}{n}=x.$$

7°. $f(x)=x$ *для любого вещественного* x.

Действительно, пусть для некоторого x имеет место неравенство $f(x)\neq x$. Предположим для определенности, что $x<f(x)$. Тогда существует такое рациональное число r, что

$$x<r<f(x).$$

В частности, $x<r$, и потому согласно свойствам 2° и 6°

$$f(x)\leqslant f(r)=r,$$

что противоречит неравенству $r<f(x)$. Следовательно, неравенство $x<f(x)$ невозможно.

Аналогично показывается, что невозможно и неравенство $f(x)<x$. Следовательно, $f(x)=x$ для всех x.

Тем самым лемма Дарбу, а вместе с ней и предложение 1, полностью доказаны.

Ясно, что соотношение (1) в точности равносильно сформулированному в конце предыдущего пункта свойству 3 аффинных преобразований. Таким образом, это свойство мы также можем считать доказанным.

Замечание 1. Поскольку аффинное преобразование Φ биективно, оператор $\mathbf{\Phi}$ также биективен. Другими словами (см. п. 4 § 7 гл. 1), *оператор, индуцированный аффинным преобразованием, обратим.*

В п. 3 § 3 гл. 2 мы подробно рассмотрели автоморфизмы аффинной плоскости и показали, в частности, что они совпадают с преобразованиями по равенству координат в двух аффинных координатных системах (для понимания дальнейшего текста знакомство с § 3 гл. 2 необязательно; термин «автоморфизм» можно далее понимать просто как синоним термина «преобразование по равенству координат в двух аффинных координатных системах»).

Аналогичным образом можно, конечно, говорить об автоморфизмах аффинного пространства.

Тот факт, что понятие коллинеарности троек точек имеет смысл в аффинной геометрии, означает, что

любой автоморфизм аффинной плоскости (аффинного пространства) является аффинным преобразованием.

Замечательно, что обратное тоже верно.

Теорема 1. *Любое аффинное преобразование* Φ *аффинной (вещественной) плоскости (аффинного пространства) является ее автоморфизмом.*

Доказательство. Пусть $Oe_1 \ldots e_n$ — произвольный аффинный координатный репер (где, как всегда, $n = 3$ в случае пространства и $n = 2$ в случае плоскости) и пусть

$$O' = \Phi(O), \quad e'_1 = \Phi e_1, \ \ldots, \ e'_n = \Phi e_n.$$

Поскольку оператор Φ обратим, векторы $e'_1, \ldots, e'_n$ составляют базис, т. е. мы имеем аффинный координатный репер $O'e'_1 \ldots e'_n$.

Пусть M — произвольная точка и $M' = \Phi(M)$ — ее образ при преобразовании Φ. По определению, координаты $x^1, \ldots, x^n$ точки M в репере $Oe_1 \ldots e_n$ определяются из разложения

$$\overrightarrow{OM} = x^1 e_1 + \ \ldots \ + x^n e_n$$

ее радиус-вектора $\overrightarrow{OM}$ по векторам базиса $e_1, \ldots, e_n$. Применив к обеим частям этого соотношения оператор Φ и воспользовавшись его линейностью, мы получим соотношение

$$\overrightarrow{O'M'} = x^1 e'_1 + \ \ldots \ + x^n e'_n,$$

показывающее, что точка M имеет в репере $O'e'_1 \ldots e'_n$ те же самые координаты $x^1, \ldots, x^n$.

Следовательно, преобразование Φ действительно является преобразованием по равенству аффинных координат, т. е. является автоморфизмом.

Теорема 1 объясняет, почему аффинная геометрия играет такую роль во всех вопросах, связанных с прямыми линиями: она является самой общей геометрией (на обычной плоскости), в которой еще имеет смысл понятие прямой линии.

По аналогии, автоморфизмы аффинной прямой, а также аффинной плоскости (аффинного пространства) над любым полем $\mathbb{K}$ также называются «аффинными преобразованиями».

Подчеркнем, что в случае произвольного поля $\mathbb{K}$ так понимаемые аффинные преобразования уже не будут самыми общими преобразованиями, сохраняющими коллинеарность точек. Например, в случае поля $\mathbb{C}$ преобразование, переводящее каждую точку в точку, имеющую (в данной координатной системе) комплексно-сопряженные координаты, сохраняет, очевидно, отношение коллинеарности, но автоморфизмом (преобразованием по равенству аффинных координат) не является.

Суть дела здесь в том, что для произвольного поля лемма Дарбу неверна.

Задание. Покажите, что преобразование (вещественной) прямой тогда и только тогда является аффинным преобразованием, когда оно сохраняет отношение, в котором данная точка делит данный отрезок.

Поскольку преобразование по равенству координат вполне определено, когда известны реперы $Oe_1 \ldots e_n$ и $O'e'_1 \ldots e'_n$, мы видим, что

аффинное преобразование однозначно определяется репером $O'e'_1 \dots e'_n$, *в который оно переводит данный репер* $Oe_1 \dots e_n$.

Другими словами,

аффинные преобразования, одинаково преобразующие некоторый репер, совпадают.

При этом, поскольку для любых двух реперов $Oe_1 \dots e_n$ и $O'e'_1 \dots e'_n$ существует (единственное) преобразование по равенству координат, мы видим также, что

для любых двух реперов $Oe_1 \dots e_n$ и $O'e'_1 \dots e'_n$ *существует единственное аффинное преобразование, переводящее первый репер во второй.*

Вместо реперов $Oe_1 \dots e_n$ мы можем рассматривать соответствующие системы точек общего положения, состоящие из $n+1$ точек (см. определение 3 п. 1 § 4 гл. 1). Поэтому справедливо следующее

Предложение 2. *Для любых двух троек точек плоскости общего положения (пар точек прямой и четверок точек пространства) существует единственное аффинное преобразование, переводящее первую тройку (пару или четверку) во вторую.*

Замечание 2. Не любую коллинеарную тройку точек плоскости (компланарную четверку точек пространства) можно аффинным преобразованием перевести в данную коллинеарную тройку (компланарную четверку). Когда это все же возможно, соответствующее аффинное преобразование не единственно.

3. Общий вид аффинных преобразований

Предположим теперь, что нами выбрана и зафиксирована некоторая точка O («начало»). Тогда точкам M будут взаимно однозначно соответствовать их радиус-векторы $\boldsymbol{x} = \overrightarrow{OM}$, что позволит нам отождествлять точки и векторы.

В силу этого отождествления каждое аффинное преобразование Φ (прямой, плоскости или пространства) мы можем рассматривать как преобразование соответствующих радиус-векторов. Иными словами, вместо того, чтобы рассматривать точки M и $M' = \Phi(M)$, мы можем рассматривать векторы $\boldsymbol{x} = \overrightarrow{OM}$ и $\boldsymbol{x}' = \overrightarrow{OM'}$. При этом последний вектор естественно обозначать также через $\Phi(\boldsymbol{x})$:

$$\Phi(\boldsymbol{x}) = \overrightarrow{OM'}.$$

Замечание 1. Обратим внимание на различие между векторами $\Phi(\boldsymbol{x})$ и $\boldsymbol{\Phi x}$. Первый зависит от выбора точки O, а второй не зависит.

По определению,

$$\boldsymbol{\Phi x} = \overrightarrow{O'M'}, \quad \text{где} \quad O' = \Phi(O).$$

Поэтому

$$\Phi(x) = \Phi x + b, \tag{1}$$

где

$$b = \overrightarrow{OO'}.$$

Таким образом, мы видим, что

аффинное преобразование Φ *полностью известно* (*при выбранной точке* O), *если известны линейный оператор* Φ *и вектор* b.

Обратно, пусть нам дан произвольный линейный оператор A и произвольный вектор b. Определим отображение Φ (прямой, плоскости или пространства) в себя формулой

$$\Phi(x) = Ax + b, \tag{2}$$

где x — радиус-вектор произвольной точки относительно выбранной точки O.

В соответствии со сказанным выше эта формула означает, что отображение Φ точку M с радиус-вектором $x = \overrightarrow{OM}$ переводит в точку M' с радиус-вектором $OM' = Ax + b$.

Легко видеть, что

если точки с радиус-векторами x_1, x_2, x_3 *коллинеарны, то точки с радиус-векторами* $\Phi(x_1)$, $\Phi(x_2)$, $\Phi(x_3)$ *также коллинеарны.*

Задание. Докажите это утверждение.

Таким образом, мы видим, что рассматриваемое отображение Φ сохраняет отношение коллинеарности точек.

Однако утверждать, что отображение Φ будет аффинным преобразованием, было бы неосторожно, поскольку для этого нужно, чтобы оно было биективным.

Очевидно, что отображение Φ обладает последним свойством (и, следовательно, является аффинным преобразованием) тогда и только тогда, когда линейный оператор A обратим.

Заметим еще, что для аффинного преобразования Φ, определенного формулой (2), индуцированный линейный оператор Φ совпадает с исходным оператором A, а вектор $\overrightarrow{OO'}$ совпадает с вектором b.

Резюмируя все сказанное, мы получаем следующее

Предложение 1. *Выбор точки* O *определяет биективное соответствие между аффинными преобразованиями* Φ *и парами* (A, b), *состоящими из обратимого линейного оператора* A *и произвольного вектора* b. *Оператор* A *и вектор* b, *соответствующие преобразованию* Φ, *определяются формулами*

$$\begin{aligned} A &= \Phi, \\ b &= \overrightarrow{OO'}. \end{aligned} \tag{3}$$

Аффинное преобразование Φ, *соответствующее паре* $(\boldsymbol{A}, \boldsymbol{b})$, *определяется формулой*

$$\Phi(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b},$$

где $\boldsymbol{x}$ — *радиус-вектор произвольной точки* (*относительно выбранной точки* O).

Замечание 2. Иногда полезно рассматривать отображения Φ, задаваемые формулой (2), и в случае, когда оператор необратим. Такие отображения называются *вырожденными аффинными преобразованиями.*

Замечание 3. В случае геометрии над произвольным полем $\mathbb{K}$ предложение 1 описывает лишь аффинные автоморфизмы.

Предложение 1 по существу полностью сводит теорию аффинных преобразований к теории линейных операторов. Например, вспомнив формулы (2) п. 3 § 7 гл. 1, выражающие линейный оператор в координатах, мы немедленно получаем, что

в произвольной аффинной координатной системе $Ox^1 \dots x^n$ *каждое аффинное преобразование* Φ *записывается в координатах формулами вида*

$$\begin{aligned} y^1 &= a^1_1 x^1 + \dots + a^1_n x^n + b^1, \\ &\dots\dots\dots\dots\dots \\ y^n &= a^n_1 x^1 + \dots + a^n_n x^n + b^n, \end{aligned} \tag{4}$$

где $x^1, \dots, x^n$ — *координаты произвольной точки* M (*т. е. координаты ее радиус-вектора* $\boldsymbol{x} = \overrightarrow{OM}$), $y^1, \dots, y^n$ — *координаты преобразованной точки* $M' = \Phi(M)$ (*т. е. координаты радиус-вектора* $\Phi(\boldsymbol{x}) = \overrightarrow{OM'}$),

$$\begin{pmatrix} a^1_1 & \dots & a^1_n \\ \dots & \dots & \dots \\ a^n_1 & \dots & a^n_n \end{pmatrix}$$

— *матрица обратимого линейного оператора* $\boldsymbol{\Phi}$, *индуцированного аффинным преобразованием* Φ, $b^1, \dots, b^n$ — *координаты точки* $O' = \Phi(O)$ (*т. е. координаты радиус-вектора* $\boldsymbol{b} = \overrightarrow{OO'}$).

Без ссылки на линейные операторы матрица

$$A = \begin{pmatrix} a^1_1 & \dots & a^1_n \\ \dots & \dots & \dots \\ a^n_1 & \dots & a^n_n \end{pmatrix}$$

может быть описана как матрица, столбцы которой состоят из координат точек $E'_1, \dots, E'_n$, являющихся образами при преобразовании Φ единичных точек $E_1, \dots, E_n$ координатных осей (т. е. при $n = 3$ — точек $E_1(1, 0, 0)$, $E_2(0, 1, 0)$, $E_3(0, 0, 1)$; при $n = 2$ — точек $E_1(1, 0)$, $E_2(0, 1)$ и при $n = 1$ — точки $E_1(1)$).

Обратно,
любое неоднородное линейное преобразование

$$y^1 = a_1^1 x^1 + \ldots + a_n^1 x^n + b^1,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$y^n = a_1^n x^1 + \ldots + a_n^n x^n + b^n$$

с невырожденной матрицей

$$A = \begin{pmatrix} a_1^1 & \ldots & a_n^1 \\ \ldots & \ldots & \ldots \\ a_1^n & \ldots & a_n^n \end{pmatrix}$$

задает (в данной аффинной координатной системе) некоторое аффинное преобразование Φ.

Действительно, преобразование Φ можно записать в виде

$$\Phi(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b},$$

где $\boldsymbol{A}$ — линейный оператор с матрицей A, а $\boldsymbol{b}$ — вектор с координатами $b^1, \ldots, b^n$. Поэтому это преобразование аффинно.

Замечание 4. Установленные результаты утверждают изоморфизм группы аффинных преобразований $\mathrm{Aff}(n)$ с группой $\mathrm{Aff}(\mathbb{R}^n)$ неоднородных линейных преобразований. Очевидно, что это в точности изоморфизм

$$\Phi \longmapsto \alpha \circ \Phi \circ \alpha^{-1},$$

рассмотренный в п. 3 § 3 гл. 2.

Определение 1. Аффинное преобразование Φ называется *сохраняющим ориентации*, если индуцированный им линейный оператор $\boldsymbol{\Phi}$ сохраняет ориентации (см. п. 6 § 7 гл. 1), т. е. если оно является преобразованием по равенству координат в двух одноименных координатных системах.

Ясно, что
совокупность $\mathrm{Aff}^+(n)$ *всех сохраняющих ориентации аффинных преобразований является подгруппой (даже нормальным делителем) группы* $\mathrm{Aff}(n)$.

Аналогично,
совокупность $\mathrm{Aff}^-(n)$ *аффинных преобразований, обращающих ориентации, является смежным классом группы* $\mathrm{Aff}(n)$ *по подгруппе* $\mathrm{Aff}^+(n)$.

Определение 2. Аффинное преобразование называется *параллельным переносом*, если оно индуцирует тождественный линейный оператор $\boldsymbol{E}$.

Согласно сказанному выше любой параллельный перенос Φ записывается (в радиус-векторах относительно некоторой точки O) формулой вида

$$\Phi(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{b},$$

где $\boldsymbol{b} = \overrightarrow{OO'}$. Но, если точка M имеет радиус-вектор $\boldsymbol{x}$, а точка $M' = \Phi(M)$ — радиус-вектор $\boldsymbol{x} + \boldsymbol{b}$, то вектор $\overrightarrow{MM'}$ равен $\boldsymbol{b}$. Следовательно, параллельный перенос Φ произвольную точку M переводит в такую точку M', что

$$\overrightarrow{MM'} = \boldsymbol{b}.$$

Обратно, ясно, что если преобразование Φ таково, что для любой точки M имеет место равенство

$$\overrightarrow{MM'} = \boldsymbol{b},$$

где $\boldsymbol{b}$ — некоторый фиксированный вектор, то это преобразование является параллельным переносом. Таким образом, мы могли бы о п р е д е л и т ь параллельный перенос как преобразование, обладающее этим свойством.

С последним определением читатель, без сомнения, уже знаком из элементарного курса геометрии.

Ясно, что, сопоставив параллельному переносу $\Phi(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{b}$ вектор $\boldsymbol{b}$, мы получим биективное соответствие между переносами и векторами из $\mathrm{Vect}(n)$. Более того, легко видеть, что композиция $\Phi_2 \circ \Phi_1$ двух параллельных переносов Φ_1 и Φ_2 также является параллельным переносом, причем, если переносам Φ_1 и Φ_2 соответствуют векторы $\boldsymbol{b}_1$ и $\boldsymbol{b}_2$, то переносу $\Phi_2 \circ \Phi_1$ соответствует вектор $\boldsymbol{b}_1 + \boldsymbol{b}_2$. Это означает, во-первых, что

совокупность $\mathrm{Trans}(n)$ *всех параллельных переносов является подгруппой группы* $\mathrm{Aff}(n)$ (*и даже подгруппы* $\mathrm{Aff}^+(n)$), и, во-вторых, что

соответствие

$$\text{«параллельный перенос»} \longmapsto \text{«вектор } \overrightarrow{OO'}\text{»}$$

определяет изоморфизм группы $\mathrm{Trans}(n)$ *на линеал* $\mathrm{Vect}(n)$, *рассматриваемый как группа относительно операции сложения векторов.*

В частности, мы видим, что

группа $\mathrm{Trans}(n)$ — *абелева, т. е. композиция любых двух параллельных переносов* Φ_1 *и* Φ_2 *не зависит от их порядка*:

$$\Phi_1 \circ \Phi_2 = \Phi_2 \circ \Phi_1.$$

Замечание 5. Очевидно, что отображение

$$\Phi \longmapsto \boldsymbol{\Phi}$$

является эпиморфизмом группы $\mathrm{Aff}(n)$ на группу $\mathrm{Op}(n)$ (отображающим подгруппу $\mathrm{Aff}^+(n)$ на подгруппу $\mathrm{Op}^+(n)$, а смежный класс $\mathrm{Aff}^-(n)$ — на смежный класс $\mathrm{Op}^-(n)$).

Поскольку ядром этого эпиморфизма является, по определению, группа $\mathrm{Trans}(n)$, мы видим, что

группа $\mathrm{Trans}(n)$ *является не только подгруппой, но даже нормальным делителем группы* $\mathrm{Aff}(n)$ (*а также группы* $\mathrm{Aff}^+(n)$), *причем имеют место изоморфизмы*:

$$\mathrm{Aff}(n)/\mathrm{Trans}(n) \approx \mathrm{Op}(n), \quad \mathrm{Aff}^+(n)/\mathrm{Trans}(n) \approx \mathrm{Op}^+(n).$$

4. Проективные преобразования

По аналогии с определением 1 п. 1 мы принимаем следующее

Определение 1. Преобразование вещественной проективной плоскости (или вещественного проективного пространства) называется *проективным*, если каждую тройку коллинеарных точек оно переводит в тройку коллинеарных точек.

Таким образом, различие между аффинными и проективными преобразованиями состоит формально только в том, что первые действуют на аффинной плоскости (аффинном пространстве), а вторые — на проективной плоскости (проективном пространстве).

Точно так же, как и для аффинных преобразований, показывается, что

преобразование, обратное к проективному, проективно.

Задание. Докажите это утверждение.

Отсюда непосредственно вытекает, что

совокупность всех проективных преобразований плоскости или пространства является группой.

Эту группу мы будем обозначать символом $\mathrm{Proj}(2)$ в случае плоскости и символом $\mathrm{Proj}(3)$ в случае пространства.

Пусть Π — произвольное проективное преобразование проективной плоскости $\mathfrak{M}$ и пусть α — произвольная прямая плоскости $\mathfrak{M}$. Эту прямую преобразование Π отображает на некоторую прямую α', а множество $\mathfrak{M} \setminus \alpha$ точек плоскости $\mathfrak{M}$, не принадлежащих прямой α, — на множество $\mathfrak{M} \setminus \alpha'$ точек, не принадлежащих прямой α'. Но мы знаем (п. 5 § 1 гл. 4), что при удалении из проективной плоскости некоторой прямой остается аффинная плоскость. Таким образом, проективное преобразование Π индуцирует некоторое отображение аффинной плоскости $\mathfrak{M} \setminus \alpha$ на аффинную плоскость $\mathfrak{M} \setminus \alpha'$. Это отображение биективно и сохраняет коллинеарность троек точек, т. е. является аффинным отображением.

Тем самым доказано, что

проективное преобразование Π *проективной плоскости* $\mathfrak{M}$, *переводящее прямую* α *в прямую* α', *индуцирует некоторое аффинное отображение* Π' *аффинной плоскости* $\mathfrak{M} \setminus \alpha$ *на аффинную плоскость* $\mathfrak{M} \setminus \alpha'$.

Докажем теперь следующее

Предложение 1. *Соответствие*

$$\Pi \longmapsto \Pi' \qquad (1)$$

является биективным соответствием между множеством всех проективных преобразований проективной плоскости $\mathfrak{M}$, *переводящих прямую* α *в прямую* α', *и множеством всех аффинных отображений аффинной плоскости* $\mathfrak{M} \setminus \alpha$ *на аффинную плоскость* $\mathfrak{M} \setminus \alpha'$.

Доказательство. Чтобы доказать инъективность соответствия (1), нужно доказать, что аффинное отображение Π' однозначно определяет соответствующее проективное преобразование Π.

Но для точек M, не принадлежащих прямой α, это очевидно, ибо для таких точек по определению

$$\Pi(M) = \Pi'(M).$$

Поэтому вопрос может стоять только для точек M прямой α.

Пусть M — произвольная такая точка. Проведем через эту точку произвольную прямую β, отличную от прямой α, и выберем на прямой β две точки N_1 и N_2, отличные от точки M. Прямая β' плоскости $\mathfrak{M}$, в которую преобразование Π переводит прямую β, однозначно характеризуется как прямая, проходящая через точки $N_1' = \Pi(N_1)$ и $N_2' = \Pi(N_2)$. С другой стороны, поскольку точки N_1 и N_2 не принадлежат, очевидно, прямой α, то

$$N_1' = \Pi'(N_1), \quad N_2' = \Pi'(N_2).$$

Следовательно, прямая β' однозначно определяется отображением Π', т. е. может быть полностью восстановлена, когда известно это отображение. Но точка $M' = \Pi(M)$ является, очевидно, точкой пересечения прямых α' и β', и потому также однозначно определяется отображением Π'. Таким образом, аффинное отображение Π' однозначно определяет проективное преобразование Π, т. е. соответствие (1) инъективно.

Чтобы доказать его надъективность, мы «обратим» предыдущее рассуждение. Именно для любого аффинного отображения Φ плоскости $\mathfrak{M} \setminus \alpha$ на плоскость $\mathfrak{M} \setminus \alpha'$ мы для точек M, не принадлежащих прямой α, определим преобразование Π формулой

$$\Pi(M) = \Phi(M),$$

а для точек M, принадлежащих этой прямой, — следующим построением: выберем произвольную прямую β, проходящую через точку M, на этой прямой выберем две точки N_1 и N_2, отличные от точки M, рассмотрим прямую β', проходящую через точки $N_1' = \Phi(N_1)$, $N_2' = \Phi(N_2)$, и примем за точку $M' = \Pi(M)$ точку пересечения прямых α' и β'.

Конечно, здесь нужно проверить, во-первых, корректность, т. е. независимость точки M' от выбора прямой β и точек N_1 и N_2, а во-вторых, — что получающееся отображение Π является проективным преобразованием. Все эти проверки совершенно автоматичны (хотя и несколько утомительны), и мы оставим их читателю.

Поскольку построенное проективное преобразование $\Pi\colon \mathfrak{M} \to \mathfrak{M}$ индуцирует, очевидно, данное аффинное отображение:

$$\Pi' = \Phi,$$

предложение 1 тем самым полностью доказано.

Упражнение. Сформулируйте и докажите аналог предложения 1 для проективных преобразований пространства.

Предложение 1 по существу сводит изучение проективных преобразований к изучению аффинных отображений. Например, мы знаем (теорема 1 п. 2), что любое аффинное преобразование является преобразованием по равенству аффинных координат. Ясно, что аналогичное утверждение имеет место и для аффинных отображений. Отсюда уже без особого труда вытекает аналогичное утверждение и для проективных преобразований:

Предложение 2. *Любое проективное преобразование является преобразованием по равенству координат в двух проективных координатых системах (т. е. является проективным автоморфизмом).*

Доказательство. Пусть проективное преобразование Π проективной плоскости $\mathfrak{M}$ переводит прямую α в прямую α' и потому индуцирует аффинное отображение Π' аффинной плоскости $\mathfrak{M}\setminus\alpha$ на аффинную плоскость $\mathfrak{M}\setminus\alpha'$. Согласно сказанному выше отображение Π' является отображением по равенству некоторых аффинных координат. Пусть x, y — эти координаты на плоскости $\mathfrak{M}\setminus\alpha$, а x', y' — координаты на плоскости $\mathfrak{M}\setminus\alpha'$. Рассмотрим соответствующие однородные аффинные координаты $X : Y : Z$ и $X' : Y' : Z'$. Как первые, так и вторые координаты являются проективными координатами на плоскости $\mathfrak{M}$. Соответствующее преобразование по равенству координат (проективный автоморфизм) переводит прямую α (имеющую уравнение $Z = 0$) в прямую α' (имеющую уравнение $Z' = 0$) и индуцирует, очевидно, аффинное отображение Π' аффинной плоскости $\mathfrak{M}\setminus\alpha$ на аффинную плоскость $\mathfrak{M}\setminus\alpha'$. Поэтому этот проективный автоморфизм совпадает с данным проективным преобразованием Π.

Замечание 1. Подчеркнем, что предложение 2 (так же как и аналогичное утверждение для аффинных преобразований) справедливо только для вещественной проективной плоскости. Для плоскостей над другими полями класс преобразований, сохраняющих коллинеарность троек точек, как правило, шире

класса проективных автоморфизмов. Тем не менее, и в случае произвольного поля $\mathbb{K}$ (а также в случае проективной прямой) принято употреблять термин «проективное преобразование» как синоним термина «проективный автоморфизм».

Из предложения 2 немедленно вытекает, что

любое проективное преобразование проективной плоскости в произвольной системе проективных координат $X:Y:Z$ задается формулами вида

$$\rho X' = a_1^1 X + a_2^1 Y + a_3^1 Z,$$
$$\rho Y' = a_1^2 X + a_2^2 Y + a_3^2 Z,$$
$$\rho Z' = a_1^3 X + a_2^3 Y + a_3^3 Z,$$

где

$$\begin{vmatrix} a_1^1 & a_2^1 & a_3^1 \\ a_1^2 & a_2^2 & a_3^2 \\ a_1^3 & a_2^3 & a_3^3 \end{vmatrix} \neq 0,$$

и, обратно, любые формулы такого вида задают некоторое проективное преобразование.

В этих формулах ρ — произвольный (отличный от нуля) множитель пропорциональности.

В более «инвариантных» терминах это утверждение означает, что соответствие

$$\Pi \longmapsto \alpha \circ \Pi \circ \alpha^{-1},$$

где $\Pi\colon \mathfrak{M} \to \mathfrak{M}$ — произвольное проективное преобразование, а $\alpha\colon \mathfrak{M} \to \mathbb{R}P^2$ — данная проективная координатная система, устанавливает изоморфизм группы $\mathrm{Proj}(2)$ проективных преобразований на группу $\mathrm{Proj}(\mathbb{R}P^2)$ проективных автоморфизмов стандартной проективной плоскости $\mathbb{R}P^2$ (ср. замечание 4 п. 3).

Предложение 3. *Для любых двух четверок E_0, E_1, E_2, E_3 и E_0', E_1', E_2', E_3' точек общего положения проективной плоскости существует одно и только одно проективное преобразование, переводящее первую четверку во вторую.*

Доказательство этого предложения вполне аналогично доказательству соответствующего предложения для аффинных преобразований (см. предложение 2 п. 2): требуемым проективным преобразованием является преобразование по равенству проективных координат с реперами E_0, E_1, E_2, E_3 и E_0', E_1' E_2' E_3' соответственно (см. предложение 1 п. 5 § 1 гл. 4).

Задание. Сформулируйте и докажите аналог предложения 3 для проективных преобразований пространства.

Наглядное описание проективных преобразований плоскости можно получить, воспользовавшись проективной плоскостью $\mathfrak{M}_0$ (см. п. 5 § 1 гл. 4), точками которой являются прямые про-

странства (аффинного), проходящие через фиксированную точку O. Каждое аффинное преобразование Φ пространства, оставляющее точку O на месте, переводит прямые в прямые и потому индуцирует некоторое преобразование Π_Φ проективной плоскости $\mathfrak{M}_O$, сохраняющее коллинеарность троек точек, т. е. являющееся проективным преобразованием. Ясно, что соответствие

$$\Phi \longmapsto \Pi_\Phi$$

является гомоморфизмом группы $\mathrm{Aff}_O(3)$ всех аффинных преобразований пространства, оставляющих на месте точку O, в группу $\mathrm{Proj}(2) = \mathrm{Proj}(\mathfrak{M}_O)$.

Упражнение. Докажите, что отображение $\Phi \longmapsto \Pi_\Phi$ является *эпиморфизмом*, т. е. что

любое проективное преобразование плоскости индуцируется некоторым аффинным преобразованием пространства.

К числу аффинных преобразований из группы $\mathrm{Aff}_O(3)$ принадлежат *гомотетии* пространства с центром в точке O. Ясно, что эти гомотетии образуют подгруппу $\mathrm{Hom}_O(3)$ группы $\mathrm{Aff}_O(3)$.

Упражнение. Докажите, что

ядром гомоморфизма $\Phi \longmapsto \Pi_\Phi$ *является группа* $\mathrm{Hom}_O(3)$.

Следовательно,

группа $\mathrm{Proj}(2)$ *изоморфна факторгруппе группы* $\mathrm{Aff}_O(3)$ *по подгруппе* $\mathrm{Hom}_O(3)$.

Замечание 2. В частности, мы видим, что группа $\mathrm{Hom}_O(3)$ является нормальным делителем группы $\mathrm{Aff}_O(3)$.

Выражаясь несколько вольно, мы можем, таким образом, сказать, что проективные преобразования проективной плоскости являются не чем иным, как рассматриваемыми с точностью до гомотетии аффинными преобразованиями аффинного пространства, оставляющими на месте данную точку.

5. Ортогональные преобразования

Основным понятием евклидовой геометрии является понятие расстояния. Естественно поэтому, что важную роль в евклидовой геометрии должны играть преобразования, сохраняющие расстояния. Такие преобразования называются «ортогональными».

Дадим точное определение.

Определение 1. Отображение Φ прямой, плоскости или пространства в себя называется *ортогональным преобразованием*, если для любых точек M_1 и M_2 длина $|M_1M_2|$ отрезка $\overline{M_1M_2}$ равна длине $|M_1'M_2'|$ отрезка $\overline{M_1'M_2'}$:

$$|M_1'M_2'| = |M_1M_2|, \qquad M_1' = \Phi(M_1), \quad M_2' = \Phi(M_2).$$

Чтобы оправдать эту терминологию, следует, конечно, показать, что ортогональные преобразования действительно являются преобразованиями, т. е. биективными отображениями прямой, плоскости или пространства на себя.

Инъективность ортогонального преобразования Φ очевидна: если $M_1' = M_2'$, то $|M_1'M_2'| = 0$, и потому $|M_1M_2| = 0$, т. е. $M_1 = M_2$.

Чтобы показать, что Φ надъективно, нам понадобятся некоторые предварительные рассмотрения.

Пусть нам даны три различные точки M_1, M_2 и M_3. Для упрощения формул положим

$$r_1 = |M_2M_3|, \quad r_2 = |M_1M_3|, \quad r_3 = |M_1M_2|.$$

Лемма 1. *Точки M_1, M_2, M_3 тогда и только тогда не коллинеарны, когда одновременно выполнены три неравенства*

$$r_1 < r_2 + r_3, \quad r_2 < r_1 + r_3, \quad r_3 < r_1 + r_2.$$

Для коллинеарных (различных) точек M_1, M_2, M_3 одно и только одно из этих неравенств обращается в равенство. Именно,

$$r_1 = r_2 + r_3$$

тогда и только тогда, когда точка M_1 лежит между точками M_2 и M_3;

$$r_2 = r_1 + r_3$$

тогда и только тогда, когда точка M_2 лежит между точками M_1 и M_3;

$$r_3 = r_1 + r_2$$

тогда и только тогда, когда точка M_3 лежит между точками M_1 и M_2.

Доказательство. Первое утверждение непосредственно вытекает из известной теоремы о сумме длин двух сторон треугольника. Остальные утверждения очевидны.

Эта лемма показывает, что

коллинеарность или неколлинеарность тройки точек, а в случае коллинеарности — характер их взаимного расположения (какая точка находится между какими), полностью определяется взаимными расстояниями между этими точками.

Но, по определению, ортогональное преобразование сохраняет расстояние между точками. Поэтому

каждое ортогональное преобразование переводит неколлинеарные точки в неколлинеарные, а коллинеарные — в коллинеарные, причем в последнем случае характер взаимного расположения точек не меняется.

Более того, так как отношение k, в котором, скажем, точка M_2 делит отрезок $\overrightarrow{M_1M_3}$, определяется характером взаимного расположения этих точек и расстояниями между ними (расстоя-

ния определяют абсолютную величину отношения, а характер взаимного расположения — его знак), то

каждое ортогональное преобразование сохраняет отношение, в котором точка делит отрезок,

т. е., более распространенно,

если точка M_2 делит отрезок $\overrightarrow{M_1M_3}$ в отношении k, то для любого ортогонального преобразования Φ точка $M_2' = \Phi(M_2)$ делит отрезок $\overrightarrow{M_1'M_3'}$, где $M_1' = \Phi(M_1)$ и $M_3' = \Phi(M_3)$, в том же отношении k.

Так как ортогональные преобразования сохраняют коллинеарность точек, то для любых двух различных точек M_1 и M_2 ортогональное преобразование Φ отображает прямую M_1M_2 в прямую $M_1'M_2'$. Более того, на самом деле,

преобразование Φ отображает прямую M_1M_2 на прямую $M_1'M_2'$.

Действительно, пусть M' — произвольная точка прямой $M_1'M_2'$ и пусть k — отношение, в котором эта точка делит отрезок $\overrightarrow{M_1'M_2'}$. Рассмотрим на прямой M_1M_2 точку M, делящую отрезок $\overrightarrow{M_1M_2}$ в отношении k. По доказанному выше, ортогональное преобразование Φ переводит точку M в точку M'', делящую отрезок $\overrightarrow{M_1'M_2'}$ в том же отношении k. Но если две точки делят один и тот же отрезок в одном и том же отношении, то они совпадают. Следовательно, $M'' = M'$, так что любая точка прямой $M_1'M_2'$ действительно является образом при преобразовании Φ некоторой точки прямой M_1M_2.

Теперь мы уже можем доказать, что

каждое ортогональное преобразование Φ (прямой, плоскости или пространства) надъективно.

Для случая прямой это непосредственно вытекает из предыдущего утверждения.

Рассмотрим случай плоскости. Пусть M_1, M_2, M_3 — три неколлинеарные точки. По доказанному выше, преобразование Φ переводит эти точки в три неколлинеарные точки M_1', M_2', M_3' и отображает прямые M_1M_2, M_1M_3, M_2M_3 на прямые $M_1'M_2'$, $M_1'M_3'$, $M_2'M_3'$. Мы должны доказать, что для любой точки M' существует такая точка M, что $M' = \Phi(M)$. При этом без ограничения общности мы можем предполагать, что точка M' не принадлежит прямым $M_1'M_2'$, $M_1'M_3'$ и $M_2'M_3'$. Проведем через точку M' произвольную прямую, пересекающую, скажем, прямые $M_1'M_2'$ и $M_1'M_3'$ в некоторых точках N_2' и N_3'. По доказанному выше, точки N_2' и N_3' являются образами некоторых точек N_2 и N_3 (лежащих соответственно на прямых M_1M_2 и M_1M_3). Рассмотрим прямую N_2N_3. Как мы знаем, эта прямая отображается на прямую $N_2'N_3'$. Но точка M' принадлежит, по построению, прямой $N_2'N_3'$. Поэтому на прямой N_2N_3 существует

такая точка M, что $M' = \Phi(M)$. Тем самым для случая плоскости наше утверждение полностью доказано.

Для случая пространства изложенное рассуждение показывает, что ортогональное преобразование Φ отображает плоскость, проходящую через три неколлинеарные точки M_1, M_2, M_3, на плоскость, проходящую через точки M_1', M_2', M_3'. Выберем в пространстве две различные непараллельные плоскости P_1 и P_2. Согласно только что сказанному, преобразование Φ отображает эти плоскости на некоторые плоскости P_1' и P_2'. Поскольку линия пересечения плоскостей P_1 и P_2 переходит при преобразовании Φ в линию пересечения плоскостей P_1' и P_2', последние плоскости не параллельны (и различны). Рассмотрим теперь в пространстве произвольную точку M'. Без ограничения общности мы можем считать, что эта точка не принадлежит плоскостям P_1' и P_2', так что через нее можно провести прямую, пересекающую эти плоскости в двух различных точках N_1' и N_2'. Пусть N_1 и N_2 — точки, переходящие при преобразовании Φ в точки N_1' и N_2' (на основании только что сказанного такие точки существуют). Как мы знаем, преобразование Φ отображает прямую N_1N_2 на прямую $N_1'N_2'$. Следовательно, на прямой N_1N_2 найдется точка M, обладающая тем свойством, что $M' = \Phi(M)$. Тем самым наше утверждение доказано и в случае пространства.

Тот факт, что ортогональное преобразование сохраняет коллинеарность точек (и отношение, в котором точка делит отрезок), означает теперь, что оно является аффинным преобразованием (в том числе и в случае прямой). Таким образом, в качестве окончательного результата (покрывающего все доказанные в этом пункте утверждения) мы получаем следующее

Предложение 1. *Любое ортогональное преобразование является аффинным преобразованием.*

Это означает, что совокупность $\mathrm{Ort}(n)$ всех ортогональных преобразований прямой ($n = 1$), плоскости ($n = 2$) или пространства ($n = 3$), являющаяся, очевидно, группой, представляет собой подгруппу группы $\mathrm{Aff}(n)$.

Будучи аффинным преобразованием, каждое ортогональное преобразование Φ индуцирует некоторый линейный оператор $\boldsymbol{\Phi}$. По определению, этот оператор переводит произвольный вектор $\boldsymbol{x} = \overrightarrow{M_1M_2}$ в вектор $\boldsymbol{\Phi x} = \overrightarrow{M_1'M_2'}$, где $M_1' = \Phi(M_1)$ и $M_2' = \Phi(M_2)$. Но, по условию,

$$|M_1'M_2'| = |M_1M_2|,$$

т. е.

$$|\boldsymbol{\Phi x}| = |\boldsymbol{x}|.$$

Тем самым доказано (см. п. 7 §7 гл. 1), что

для любого ортогонального преобразования Φ индуцированный линейный оператор $\boldsymbol{\Phi}$ является изометричным оператором.

Легко видеть, что и обратно,

аффинное преобразование Φ, *индуцирующее изометричный линейный оператор* $\boldsymbol{\Phi}$, *является ортогональным преобразованием.*

Действительно, как мы знаем, в радиус-векторах относительно некоторой точки O аффинное преобразование Φ записывается формулой

$$\Phi(\boldsymbol{x}) = \boldsymbol{\Phi x} + \boldsymbol{b},$$

где $\boldsymbol{b}$ — некоторый вектор. С другой стороны, для любых точек $M_1(\boldsymbol{x}_1)$ и $M_2(\boldsymbol{x}_2)$

$$|M_1M_2| = |\boldsymbol{x}_2 - \boldsymbol{x}_1|.$$

Поэтому, если линейный оператор $\boldsymbol{\Phi}$ изометричен, то

$$|M_1'M_2'| = |\Phi(\boldsymbol{x}_2) - \Phi(\boldsymbol{x}_1)| = |\boldsymbol{\Phi x}_2 - \boldsymbol{\Phi x}_1| = |\boldsymbol{\Phi}(\boldsymbol{x}_2 - \boldsymbol{x}_1)| = \\ = |\boldsymbol{x}_2 - \boldsymbol{x}_1| = |M_1M_2|,$$

так что преобразование Φ ортогонально.

На языке теории групп эти утверждения означают, что

эпиморфизм $\Phi \longmapsto \boldsymbol{\Phi}$ *группы* $\mathrm{Aff}(n)$ *на группу* $\mathrm{Op}(n)$ *отображает подгруппу* $\mathrm{Ort}(n)$ *на подгруппу* $\mathrm{Iso}(n)$.

Отсюда и из известных свойств изометричных линейных операторов (см. п. 7 § 7 гл. 1) немедленно вытекают следующие свойства ортогональных преобразований:

Свойство 1. *Каждое ортогональное преобразование* Φ *сохраняет не только длины, но и углы, т. е. для любых трех неколлинеарных точек* O, A, B *угол* $\angle OAB$ *равен углу* $\angle O'A'B'$.

Замечание 1. Это свойство легко доказывается и непосредственно: поскольку преобразование Φ сохраняет длины, стороны треугольника OAB равны соответствующим сторонам треугольника $O'A'B'$, так что эти треугольники конгруэнтны и потому их соответственные углы равны.

Свойство 2. *Преобразование (прямой, плоскости или пространства) тогда и только тогда ортогонально, когда оно действует по равенству координат в двух прямоугольных реперах (т. е. является евклидовым автоморфизмом*; см. п. 2 § 3 гл. 2).

В частности, это означает, что

любое ортогональное преобразование переводит произвольный прямоугольный репер $O\boldsymbol{i}_1\boldsymbol{i}_2 \ldots \boldsymbol{i}_n$ *снова в прямоугольный репер* $O'\boldsymbol{i}_1' \ldots \boldsymbol{i}_n'$

и обратно,

для любых двух прямоугольных реперов $O\boldsymbol{i}_1 \ldots \boldsymbol{i}_n$ *и* $O'\boldsymbol{i}_1' \ldots \boldsymbol{i}_n'$ *существует (единственное) ортогональное преобразование, переводящее первый репер во второй.*

Свойство 3. *Аффинное преобразование тогда и только тогда ортогонально, когда в некоторой (а потому и в любой) системе*

прямоугольных координат оно выражается формулами

$$y^1 = a_1^1 x^1 + \ldots + a_n^1 x^n + b^1,$$
$$\ldots\ldots\ldots\ldots$$
$$y^n = a_1^n x^1 + \ldots + a_n^n x^n + b^n,$$

где

$$\begin{pmatrix} a_1^1 & \ldots & a_n^1 \\ a_1^n & \ldots & a_n^n \end{pmatrix}$$

— ортогональная матрица.

Являясь аффинным преобразованием, каждое ортогональное преобразование либо сохраняет, либо обращает ориентации. Ясно, что

совокупность $\mathrm{Ort}^+(n)$ *всех ортогональных преобразований, сохраняющих ориентации, является подгруппой (даже нормальным делителем) группы* $\mathrm{Ort}(n)$, *а совокупность* $\mathrm{Ort}^-(n)$ всех *ортогональных преобразований, обращающих ориентации, является смежным классом группы* $\mathrm{Ort}(n)$ *по подгруппе* $\mathrm{Ort}^+(n)$.

Замечание 2. Обратим внимание на то, что группа $\mathrm{Ort}^+(n)$ не является нормальным делителем группы $\mathrm{Aff}(n)$ (и даже группы $\mathrm{Aff}^+(n)$).

Каждый параллельный перенос является, очевидно, ортогональным преобразованием, сохраняющим ориентации. Следовательно,

группа $\mathrm{Trans}(n)$ *является подгруппой группы* $\mathrm{Ort}^+(n)$.

Более того, так как $\mathrm{Trans}(n)$ является ядром гомоморфизма, сопоставляющего каждому ортогональному преобразованию индуцированный линейный оператор, то

группа $\mathrm{Trans}(n)$ *является нормальным делителем группы* $\mathrm{Ort}^+(n)$ *(и даже группы* $\mathrm{Ort}(n)$*), причем*:

$$\mathrm{Ort}^+(n)/\mathrm{Trans}(n) \approx \mathrm{Iso}^+(n), \quad \mathrm{Ort}(n)/\mathrm{Trans}(n) \approx \mathrm{Iso}(n).$$

Определение 2. Ортогональные преобразования, сохраняющие ориентации (т. е. элементы группы $\mathrm{Ort}^+(n)$), мы будем называть также *движениями.*

Обычно (например, в школьном курсе) движение (скажем, плоскости) определяют как результат непрерывного перемещения плоскости самой по себе. (Обратим внимание на различие между движениями и перемещениями: движение является результатом перемещения.) Поскольку при любом перемещении взаимные расстояния между точками не меняются и поскольку (в силу непрерывности перемещения) ориентация каждого репера при перемещении остается той же самой, движение в этом смысле является движением в смысле определения 2.

Обратно, ясно, что, непрерывно двигая некоторый прямоугольный репер $Oe_1 \ldots e_n$, мы можем его совместить с любым другим одноименным с ним прямоугольным репером $O'e_1' \ldots e_n'$. (Например, на плоскости, чтобы совместить прямоугольный репер Oe_1e_2 с одноименным прямоугольным репером

$O'e_1'e_2'$, достаточно передвинуть точку O в точку O', а затем повернуть плоскость на угол между векторами e_1 и e_1'.) Следовательно, для любого сохраняющего ориентации ортогонального преобразования Φ существует движение Ψ (результат перемещения), переводящее данный репер $Oe_1 \ldots e_n$ в тот же репер, что и преобразование Φ. Но мы знаем, что для любых двух прямоугольных реперов существует только одно ортогональное преобразование, переводящее первый репер во второй. Следовательно, $\Phi = \Psi$.

Таким образом,

движения, как мы их определили, совпадают с движениями в элементарно-геометрическом смысле.

Преимущество нашего определения состоит в том, что оно не использует лежащих вне математики понятий механики.

Определение 3. Движение, оставляющее на месте некоторую точку O, называется *вращением с центром* O.

Ясно, что композиция двух вращений с центром O также является вращением с центром O (ибо если $\Phi(O) = O$ и $\Psi(O) = O$, то $(\Phi \circ \Psi)(O) = O$). Это означает, что

совокупность $\mathrm{Rot}_O(n)$ *всех вращений с данным центром* O *является подгруппой группы движений* $\mathrm{Ort}^+(n)$.

В радиус-векторах относительно точки O каждое вращение Φ с центром O записывается формулой

$$\Phi(x) = \mathbf{\Phi} x,$$

где, как всегда, $\mathbf{\Phi}$ — изометричный линейный оператор, индуцированный преобразованием Φ. Это показывает, что

соответствие

«вращение Φ» $\longmapsto$ *«индуцированный оператор* $\mathbf{\Phi}$»

устанавливает изоморфизм группы $\mathrm{Rot}_O(n)$ *на группу* $\mathrm{Iso}^+(n)$.

Поскольку последняя группа изоморфна группе $\mathrm{SO}(n)$, мы видим также, что

для любой точки O *группа* $\mathrm{Rot}_O(n)$ *изоморфна группе* $\mathrm{SO}(n)$.

Заметим, что последний изоморфизм зависит от выбора репера $Oe_1 \ldots e_n$.

Элемент группы $\mathrm{SO}(n)$ (собственную ортогональную матрицу), соответствующий данному вращению Φ, мы, допуская определенную вольность речи, будем называть *матрицей вращения* Φ в данном репере. Вольность здесь состоит в том, что эта матрица является на самом деле матрицей соответствующего изометричного оператора.

Замечание 3. Подчеркнем, что (при $n > 1$) группы $\mathrm{Rot}_O(n)$ и $\mathrm{Rot}_{O'}(n)$, соответствующие различным точкам O и O', различны (хотя и изоморфны одной и той же группе $\mathrm{Iso}^+(n) \approx \mathrm{SO}(n)$).

Предложение 2. *Любое движение* Φ *(прямой, плоскости или пространства) единственным образом разлагается в композицию некоторого вращения* R *(с заранее заданным центром* O*) и некоторого параллельного переноса* T.

Доказательство. В радиус-векторах относительно точки O движение Φ записывается формулой $\Phi(\boldsymbol{x}) = \boldsymbol{\Phi x} + \boldsymbol{b}$ и потому является композицией вращения $R(\boldsymbol{x}) = \boldsymbol{\Phi x}$ и параллельного переноса $T(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{b}$.

Замечание 4. Обратим внимание на то, что в полученном разложении $\Phi = T \circ R$ порядок множителей существен:

$$T \circ R \neq R \circ T.$$

Тем не менее,

движение Φ допускает также разложение вида

$$\Phi = R \circ T',$$

где R — то же самое вращение, что и выше, а T' — некоторый параллельный перенос (вообще говоря, отличный от переноса T).

Действительно, пусть $\boldsymbol{b}'$ — такой вектор, что

$$\boldsymbol{\Phi b}' = \boldsymbol{b}.$$

Тогда

$$\Phi(\boldsymbol{x}) = \boldsymbol{\Phi}(\boldsymbol{x} + \boldsymbol{b}'),$$

и, следовательно, преобразование Φ является композицией параллельного переноса $T'(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{b}'$ и вращения $R(\boldsymbol{x}) = \boldsymbol{\Phi x}$.

Упражнение. Докажите, что

$$\Phi = R' \circ T,$$

где T — тот же параллельный перенос, что и выше, а R' — вращение с центром $O' = \Phi(O)$, индуцирующее тот же изометричный оператор $\boldsymbol{\Phi}$, что и вращение R.

§ 2. РАЗЛОЖЕНИЕ АФФИННЫХ И ОРТОГОНАЛЬНЫХ ПРЕОБРАЗОВАНИЙ В КОМПОЗИЦИЮ БОЛЕЕ ПРОСТЫХ

В этом параграфе аффинные преобразования будут изучаться средствами евклидовой геометрии. В соответствии с этим все координаты мы будем предполагать прямоугольными (евклидовыми).

1. Аффинные и ортогональные преобразования прямой

Аффинные преобразования прямой устроены очень просто: любое такое преобразование записывается в координатах формулой

$$y = ax + b,$$

где a и b — произвольные числа, причем $a \neq 0$.

Это преобразование ортогонально тогда и только тогда, когда $|a| = 1$, т. е. когда $a = \pm 1$.

Простейшим неортогональным аффинным преобразованием прямой является преобразование, задаваемое формулой

$$y = kx,$$

где $k > 0$. По аналогии со сжатиями плоскости к прямой (см. п. 2 § 1 гл. 5) и пространства к плоскости (см. п. 1 § 2 гл. 5), мы будем называть это преобразование *сжатием* прямой с коэффициентом k *к точке* O. Оно, конечно, сохраняет ориентации (поскольку $k > 0$).

Предложение 1. *Любое аффинное преобразование прямой является композицией сжатия к фиксированной точке O и некоторого ортогонального преобразования.*

Доказательство. Преобразование

$$y = ax + b$$

является композицией преобразования

$$y = |a|x$$

и преобразования

$$y = \varepsilon x + b,$$

где

$$\varepsilon = \begin{cases} +1, & \text{если } a > 0 \\ -1, & \text{если } a < 0. \end{cases}$$

Замечание 1. Любое ортогональное преобразование прямой

$$y = \pm x + b$$

является либо параллельным переносом

$$y = x + b, \tag{1}$$

либо *симметрией*

$$y = -x \tag{2}$$

относительно точки O, либо композицией симметрии (2) и параллельного переноса (1). При этом, если преобразование сохраняет ориентации (является движением, см. определение 2 п. 5 § 1), то оно непременно является параллельным переносом (1) (так что $\mathrm{Ort}^+(1) = \mathrm{Trans}(1)$)[1].

Следовательно,

сохраняющее ориентации аффинное преобразование Φ прямой единственным образом разлагается в композицию некоторого параллельного переноса T и некоторого сжатия K к точке O:

$$\Phi = K \circ T, \tag{3}$$

[1]) Это согласуется с предложением 2 п. 4 § 1, поскольку единственным вращением прямой является тождественное преобразование.

а аффинное преобразование Φ, *обращающее ориентации, разлагается (также единственным образом) в композицию некоторого параллельного переноса* T, *симметрии* S *относительно точки* O *и некоторого сжатия* K *к точке* O:

$$\Phi = K \circ S \circ T. \tag{4}$$

Конечно, в этих разложениях перенос T и сжатие K могут оказаться и тождественными преобразованиями.

Подчеркнем, что в разложениях (3) и (4) точка O предполагается фиксированной. Если ее не фиксировать, то будет справедливо следующее (в некотором смысле более точное)

Предложение 2. *Любое аффинное преобразование прямой является либо параллельным переносом, либо сжатием к некоторой (зависящей от преобразования) точке (если оно сохраняет ориентации), либо композицией сжатия к некоторой точке и симметрии относительно той же точки (если оно обращает ориентации).*

Доказательство. Преобразование

$$y = ax + b$$

при $a \neq 1$ оставляет на месте точку

$$x_0 = \frac{b}{1-a}.$$

Поэтому в координате

$$x' = x - x_0$$

оно записывается формулой

$$y' = ax'$$

и потому при $a > 0$ является сжатием к точке x_0, а при $a < 0$ — композицией такого сжатия и симметрии

$$y' = -x'.$$

Замечание 2. В частности, мы видим, что

любое ортогональное преобразование прямой является либо параллельным переносом, либо симметрией относительно некоторой точки.

Для любой точки O симметрия S_O относительно этой точки составляет вместе с тождественным преобразованием E группу второго порядка

$$\mathrm{Sym}_O(1) = \{E, S_O\}.$$

Поэтому мы получаем, что

группа $\mathrm{Ort}(1)$ *является теоретико-множественным объединением группы* $\mathrm{Trans}(1)$ *(изоморфной группе вещественных чисел по сложению) и групп второго порядка* $\mathrm{Sym}_O(1)$, *отвечающих всевозможным точкам прямой.*

Заметим, что любые две группы, объединением которых является группа Ort(1), пересекаются лишь по тождественному преобразованию.

Конечно, это еще не описывает полностью алгебраического строения группы Ort(1): нужно указать, что получается в результате перемножения (компонирования) элементов различных групп, объединением которых является группа Ort(1).

Задание. Покажите, что

1) *композицией двух симметрий относительно точек* O_1 *и* O_2 *является параллельный перенос на вектор* $2\overrightarrow{O_1O_2}$;

2) *композицией симметрии относительно точки* O *и параллельного переноса на вектор* $\boldsymbol{b}$ *является симметрия относительно точки* O', *имеющей относительно точки* O *радиус-вектор* $\frac{1}{2}\boldsymbol{b}$;

3) *композицией параллельного переноса на вектор* $\boldsymbol{b}$ *и симметрии относительно точки* O *является симметрия относительно точки* O', *имеющей относительно точки* O *радиус-вектор* $-\frac{1}{2}\boldsymbol{b}$.

Таким образом,

любое ортогональное преобразование прямой является либо симметрией, либо композицией двух симметрий.

Наша ближайшая цель — перенести доказанные в этом пункте утверждения на случаи плоскости и пространства.

2. Разложение аффинных преобразований плоскости и пространства

Предложение 1. *Любое аффинное преобразование плоскости (пространства) является композицией некоторого ортогонального преобразования и двух (трех) сжатий к взаимно перпендикулярным прямым (плоскостям).*

Для доказательства этого предложения (аналогичного предложению 1 п. 1) мы заметим, что поскольку аффинное преобразование переводит параллельные прямые в параллельные, оно индуцирует некоторое преобразование направлений (на плоскости или, соответственно, в пространстве).

Определение 1. Пусть Φ — произвольное аффинное преобразование плоскости. Пару перпендикулярных направлений на плоскости мы будем называть *парой главных направлений* аффинного преобразования Φ, если преобразование Φ переводит эти направления снова в перпендикулярные направления. Аналогично, тройку перпендикулярных направлений в пространстве мы будем называть *тройкой главных направлений* аффинного преобразования Φ пространства, если преобразование Φ переводит эти направления снова в перпендикулярные направления.

Если преобразование Φ ортогонально, то любая пара (тройка) перпендикулярных направлений является для него главной.

Теорема 1. *Для любого аффинного преобразования* Φ *плоскости (пространства) существует пара (тройка) главных направлений.*

Из этой теоремы предложение 1 вытекает уже без особого труда. Действительно, пусть, например, Φ — произвольное аффинное преобразование плоскости. Выбрав прямоугольный координатный репер Oe_1e_2, орты e_1 и e_2 которого имеют главные направления преобразования Φ, рассмотрим репер $O'e_1'e_2'$, в который это преобразование переводит репер Oe_1e_2. По условию, векторы $e_1' = \Phi e_1$ и $e_2' = \Phi e_2$ ортогональны (и отличны от нуля). Поэтому, положив

$$f_1 = k_1 e_1', \quad f_2 = k_2 e_2', \quad \text{где} \quad k_1 = \frac{1}{|e_1'|}, \quad k_2 = \frac{1}{|e_2'|},$$

мы получим прямоугольный координатный репер $O'f_1f_2$.

Но, как мы знаем (см. п. 5 § 1), любые два прямоугольных репера можно перевести друг в друга некоторым ортогональным преобразованием. Пусть Φ' — ортогональное преобразование, переводящее репер Oe_1e_2 в репер $O'f_1f_2$.

Пусть, далее, Φ'' — аффинное преобразование, оставляющее точку O' на месте и переводящее векторы f_1, f_2 в векторы e_1', e_2'. Это преобразование является, очевидно, композицией двух сжатий плоскости к взаимно перпендикулярным прямым (координатным осям репера $O'f_1f_2$) с коэффициентами k_1 и k_2. В координатах x, y с репером $O'f_1f_2$ преобразование Φ'' записывается формулами

$$x' = k_1 x,$$
$$y' = k_2 y.$$

Поскольку преобразование Φ' переводит репер Oe_1e_2 в репер $O'f_1f_2$, а преобразование Φ'' переводит репер $O'f_1f_2$ в репер $O'e_1'e_2'$, то композиция $\Phi'' \circ \Phi'$ переводит репер Oe_1e_2 в репер $O'e_1'e_2'$. Но мы знаем, что аффинное преобразование однозначно определяется репером, в который оно переводит данный репер. Следовательно, поскольку преобразования Φ и $\Phi'' \circ \Phi'$ переводят репер Oe_1e_2 в один и тот же репер $O'e_1'e_2'$, эти преобразования совпадают.

Задание. Докажите предложение 1 для случая пространства.

Замечание 1. Обратим внимание на то, что предложение 1 не является полным аналогом предложения 1 п. 1, поскольку в предусмотренных им разложениях сначала делается ортогональное преобразование, а затем сжатия, тогда как в предложении 1 п. 1 дело обстояло как раз наоборот.

Задание. Докажите для случая прямой точный аналог предложения 1, а для случая плоскости и пространства — точный аналог предложения 1 п. 1.

Таким образом, нам осталось только доказать теорему 1.

Доказательство теоремы 1. Рассмотрим сначала случай плоскости.

Пусть Σ — произвольная окружность с центром в некоторой точке O. При преобразовании Φ точка O перейдет в некоторую точку O', а каждая точка M окружности Σ — в некоторую точку M'. Несмотря на то, что все отрезки $\overline{OM}$ имеют одну и ту же длину, длины отрезков $\overline{O'M'}$, вообще говоря, различны. Пусть M_0 — та из точек M окружности Σ, для которой длина отрезка $\overline{O'M'}$ имеет наименьшее значение, и пусть α — прямая OM_0. Пусть, далее, β — прямая, проходящая через точку M_0 перпендикулярно прямой α. Заметим, что все точки прямой β, отличные от точки M_0, лежат вне окружности Σ.

При преобразовании Φ прямая α переходит в прямую $\alpha' = O'M_0'$, а прямая β — в некоторую прямую β', проходящую через точку M_0'. Для доказательства теоремы 1 достаточно, очевидно, показать, что

прямая β' перпендикулярна прямой α'.

Пусть это не так, т. е. пусть прямая α' отлична от перпендикуляра, опущенного из точки O' на прямую β'. Пусть N' — основание этого перпендикуляра и пусть N — точка, переходящая при преобразовании Φ в точку N. Точка N принадлежит прямой β и отлична от точки M_0. Поэтому точка N лежит вне окружности Σ и, следовательно, прямая ON пересекает эту окружность в некоторой точке M_1, являющейся внутренней точкой отрезка $\overline{ON}$. При аффинном преобразовании Φ отрезок $\overline{ON}$ переходит, как мы знаем, в отрезок $\overline{O'N'}$, причем внутренние точки отрезка $\overline{ON}$ переходят во внутренние точки отрезка $\overline{O'N'}$. В частности, точка M_1 перейдет в некоторую точку M_1', являющуюся внутренней точкой отрезка $\overline{O'N'}$. Следовательно, длина отрезка $\overline{O'M_1'}$ будет меньше длины отрезка $\overline{O'N'}$. Но длина отрезка $\overline{O'N'}$ в свою очередь меньше длины отрезка $\overline{O'M_0'}$ (перпендикуляр короче наклонной). Поэтому длина отрезка $\overline{O'M_1'}$ меньше длины отрезка $\overline{O'M_0'}$. Таким образом, мы нашли на окружности Σ точку M_1, для которой длина отрезка $\overline{O'M_1'}$ меньше длины отрезка $\overline{O'M_0'}$, что противоречит выбору точки M_0. Следовательно, прямая α' перпендикулярна прямой β'.

Рассмотрим теперь случай пространства.

Пусть Σ — произвольная сфера и O — ее центр. Найдем на сфере Σ точку M_0, для которой длина отрезка $\overline{O'M_0'}$, где $O' = \Phi(O)$ и $M_0' = \Phi(M_0)$, имеет наименьшее возможное значение.

Пусть α — прямая OM_1 и пусть P — плоскость, проходящая через точку M_1 перпендикулярно прямой α. Заметим, что все точки плоскости P, отличные от точки M_0, лежат вне сферы Σ.

При преобразовании Ф прямая α переходит в прямую $\alpha' = O'M_0'$, а плоскость P — в некоторую плоскость P′. Покажем, что *плоскость* P′ *перпендикулярна прямой* α'.

Действительно, пусть это не так, т. е. пусть прямая α' отлична от перпендикуляра, опущенного из точки O' на плоскость P′. Пусть N' — основание этого перпендикуляра и пусть N — точка, переходящая при преобразовании Ф в точку N'. Точка N принадлежит плоскости P и отлична от точки M_0. Поэтому точка N лежит вне сферы Σ и, следовательно, прямая ON пересекает сферу Σ в некоторой точке M_1, являющейся внутренней точкой отрезка $\overline{ON}$. Но тогда точка $M_1' = \Phi(M_1)$ будет внутренней точкой отрезка $\overline{O'N'}$ и потому длина отрезка $O'M_1'$ будет меньше длины отрезка $\overline{O'N'}$, а, следовательно, и меньше длины отрезка $\overline{O'M_0'}$. Таким образом, мы нашли на сфере Σ точку M_1, для которой длина отрезка $\overline{O'M_1'}$ меньше длины отрезка $\overline{O'M_0'}$, что невозможно. Следовательно, плоскость P′ перпендикулярна прямой α'.

Рассмотрим теперь аффинное отображение Φ_P плоскости P на плоскость P′, индуцированное аффинным преобразованием Ф. По уже доказанной теореме для аффинных преобразований плоскости (справедливой, очевидно, и для аффинных отображений одной плоскости на другую) на плоскости P существуют две перпендикулярные прямые β_1 и β_2, переходящие при отображении Φ_P (а значит, и при преобразовании Ф) в перпендикулярные прямые β_1' и β_2' плоскости P′. Ясно тогда, что три попарно перпендикулярные прямые α, β_1 и β_2 переходят при преобразовании Ф в три попарно перпендикулярные прямые α', β_1' и β_2'.

Тем самым теорема 1 полностью доказана.

Замечание 2. Внимательный читатель заметит в изложенном доказательстве существенный пробел: почему мы знаем, что точка M_0 с минимальным расстоянием $|O'M_0'|$ существует? Интуитивно существование такой точки очевидно. Для строгого доказательства мы должны воспользоваться известной из курса анализа теоремой Вейерштрасса, утверждающей, что для любой непрерывной функции на компактном (замкнутом и ограниченном) множестве точек плоскости (или пространства) существует точка, в которой эта функция имеет наименьшее значение. В нашем случае этим компактным множеством является окружность (сфера) Σ, а непрерывной функцией — функция $f(M)$, сопоставляющая произвольной точке M окружности (сферы) Σ длину отрезка $\overline{O'M'}$. (Тот факт, что эта функция непрерывна, проще всего усмотреть, заметив, что ввиду известных нам формул, выражающих в координатах аффинные преобразования, квадрат функции $f(M)$ является некоторым многочленом второй степени от координат точки M и потому непрерывен.)

Замечание 3. Для случая плоскости можно дать очень простое доказательство теоремы 1, не использующее теоремы Вейерштрасса.

Рассмотрим на плоскости произвольную окружность Σ. Как мы знаем (см. п. 4 § 1 гл. 5), аффинное преобразование Φ переводит эту окружность в некоторый эллипс Σ'. При этом любая пара сопряженных диаметров эллипса Σ' является образом некоторой пары перпендикулярных диаметров окружности Σ. Но (см. п. 5 § 2 гл. 5 или п. 8 § 1 гл. 6) любой эллипс обладает парой главных, т. е. сопряженных и перпендикулярных, диаметров. Соответствующие диаметры окружности Σ являются, таким образом, парой перпендикулярных прямых, переходящих при преобразовании Φ в пару перпендикулярных прямых. По определению, направления этих прямых и будут главными направлениями аффинного преобразования Φ.

Обратим внимание на то, что главные направления эллипса Σ' отнюдь не являются главными направлениями аффинного преобразования Φ (они будут главными направлениями обратного преобразования Φ^{-1}).

Замечание 4. Если аффинное преобразование Φ оставляет на месте точку O, то соответствующее ортогональное преобразование Φ' также оставляет на месте эту точку, и потому однозначно определяется индуцированным изометричным оператором $\boldsymbol{\Phi}$. Что же касается сжатий к двум (трем) взаимно перпендикулярным прямым (плоскостям), то их композиция Φ'' индуцирует, очевидно, оператор, обладающий тем свойством, что в соответствующем ортонормированном базисе он выражается диагональной матрицей с положительными диагональными элементами (и подобно преобразованию Φ' однозначно этим оператором определяется).

Линейный оператор, имеющий в некотором ортонормированном базисе диагональную матрицу, называется *ортогонально диагонализируемым* или *симметрическим*. (Последнее название объясняется тем, что, как показывается в курсе линейной алгебры, линейный оператор тогда и только тогда симметричен, когда в произвольном ортонормированном базисе его матрица симметрична.) Симметрический оператор, обладающий (в некотором ортонормированном базисе) диагональной матрицей с положительными диагональными элементами, называется *положительным*.

Сопоставляя все сказанное, мы немедленно получаем следующее (очевидное при $n = 1$)

Предложение 2. *Любой линейный обратимый оператор* $\boldsymbol{A}$ *разлагается в произведение некоторого изометричного оператора* $\boldsymbol{U}$ *и некоторого положительного симметрического оператора* $\boldsymbol{P}$:

$$\boldsymbol{A = PU}.$$

Это разложение операторов называется *полярным разложением*. В некоторых отношениях оно аналогично разложению $z = re^{i\alpha}$ произвольного комплексного числа $z \neq 0$ в произведение его модуля $r = |z|$ (являющегося положительным числом) и комплексного числа $e^{i\alpha}$, имеющего модуль, равный единице.

Упражнение. Докажите, что
операторы U и P однозначно определяются оператором A.

Упражнение. Докажите, что полярное разложение возможно и в случае, когда оператор A необратим. Будет ли оно в этом случае однозначно определенным?

3. Разложение ортогональных преобразований плоскости

Ортогональные преобразования прямой полностью описаны в замечаниях 1 и 2 п. 1. В этом и следующем пунктах мы получим аналогичные результаты для ортогональных преобразований соответственно плоскости и пространства.

На плоскости любое вращение $R \in \mathrm{Rot}_O(2)$ задается (в некотором евклидовом репере Oi_1i_2) собственной ортогональной матрицей, т. е. (см. п. 6 § 5 гл. 1) некоторой матрицей вида

$$A = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix},$$

где $-\pi < \varphi \leqslant \pi$.

Вообще говоря, угол φ зависит от выбора репера Oi_1i_2. Однако легко видеть, что

в одноименных реперах Oi_1i_2 и $Oi_1'i_2'$ угол φ — один и тот же.

Действительно, мы знаем (см. предложение 3 п. 3 § 7 гл. 1), что оператор, имеющий в базисе i_1, i_2 матрицу A, будет иметь в базисе i_1', i_2' матрицу $C^{-1}AC$, где C — матрица перехода от базиса i_1, i_2 к базису i_1', i_2', т. е. в нашем случае — некоторая собственная ортогональная матрица. Но (см. п. 6 § 5 гл. 1) группа $\mathrm{SO}(2)$ собственных ортогональных матриц второго порядка — абелева. Поэтому $C^{-1}AC = A$.

Следовательно, если плоскость ориентирована (и мы условились выбирать репер Oi_1i_2 положительно ориентированным), то угол φ определен однозначно.

Определение 1. Угол φ называется *углом вращения* R (на ориентированной плоскости). Говорят также, что R представляет собой *вращение на угол φ вокруг точки O.*

Это вполне согласуется с «наивно-наглядным» представлением о вращении. Действительно, вращение R переводит, например, вектор i_1 в вектор

$$i_1 \cos\varphi - i_1 \sin\varphi,$$

образующий с вектором i_1 как раз угол φ.

Для любого n имеют место изоморфизмы (см. п. 4 § 1 и п. 7 § 7 гл. 1)

$$\mathrm{Rot}_O(n) \approx \mathrm{Iso}^+(n) \approx \mathrm{SO}(n).$$

Но при $n = 2$ имеет место также изоморфизм

$$\mathrm{SO}(2) \approx \mathrm{S}^1,$$

где S^1 — группа по умножению комплексных чисел $z = e^{i\varphi}$, по модулю равных единице (см. п. 6 § 5 гл. 1).

Композиция этих изоморфизмов дает изоморфизм

$$\mathrm{Rot}_O(2) \approx \mathrm{S}^1.$$

В этом изоморфизме вращению на угол φ отвечает число $e^{i\varphi}$.

Это полностью описывает алгебраическое строение группы $\mathrm{Rot}_O(2)$.

Точки плоскости можно характеризовать не прямоугольными координатами x, y, а одной комплексной координатой

$$z = x + iy.$$

Операция умножения комплексного числа $z = x + iy$ на число $e^{i\varphi} = \cos\varphi + i\sin\varphi$ выражается формулой

$$e^{i\varphi}z = (x\cos\varphi - y\sin\varphi) + i(x\sin\varphi + y\cos\varphi).$$

Это показывает, что, сопоставив точке M с комплексной координатой z точку M' с комплексной координатой $e^{i\varphi}z$, мы получим вращение вокруг точки O (имеющей комплексную координату $z = 0$) на угол φ.

Аналогично, сопоставив точке $M(z)$ точку $M'(z+\beta)$, где $\beta = b_1 + ib_2$, мы получим, очевидно, параллельный перенос на вектор $\boldsymbol{b}(b_1, b_2)$.

Это означает, что справедливо следующее предложение, полностью описывающее алгебраическое строение группы $\mathrm{Ort}^+(2)$:

Предложение 1. *В комплексной координате произвольное движение плоскости однозначно записывается формулой вида*

$$z' = \zeta z + \beta, \qquad \zeta, \beta \in \mathbb{C},$$

где $|\zeta| = 1$. *Обратно, любое такое преобразование является движением.*

Обратим внимание на то, что это описание группы $\mathrm{Ort}^+(2)$ полностью аналогично данному в замечании 1 п. 1 описанию группы $\mathrm{Ort}(1)$: единственное отличие состоит в том, что вместо вещественных чисел появляются числа комплексные.

Замечание 1. Напомним (см. замечание 3 п. 4 § 3 гл. 2), что хотя формула

$$x' = ax + b, \quad a \neq 0, \quad a, b \in \mathbb{R},$$

описывает все аффинные преобразования прямой, аналогичная формула для комплексных чисел

$$z' = az + \beta, \qquad a \neq 0, \quad a, \beta \in \mathbb{C}$$

не описывает всех аффинных преобразований плоскости.

Совершенно иное описание группы $\mathrm{Ort}^+(2)$ (аналогичное описанию группы $\mathrm{Ort}(1)$, данному в замечании 2 п. 1) можно получить, введя в рассмотрение вращения с не заданным наперед центром.

Предложение 2. *Каждое движение* Φ *лоскости, не являющееся параллельным переносом, является вращением (вокруг некоторого, зависящего от* Φ*, центра).*

Доказательство. В радиус-векторах относительно произвольной точки O движение Φ записывается формулой

$$\Phi(\boldsymbol{x}) = \boldsymbol{\Phi x} + \boldsymbol{b},$$

где $\boldsymbol{b}$ — некоторый вектор, а $\boldsymbol{\Phi}$ — сохраняющий ориентации линейный изометричный оператор, отличный по условию от тождественного оператора $\boldsymbol{E}$. Чтобы доказать, что это движение является вращением, мы должны показать, что хотя бы одна (на самом деле — только одна) точка плоскости оставляется этим движением на месте, т. е. обладает тем свойством, что ее радиус-вектор $\boldsymbol{x}_0$ удовлетворяет соотношению

$$\Phi(\boldsymbol{x}_0) = \boldsymbol{x}_0.$$

Поскольку $\boldsymbol{x}_0 = \boldsymbol{E}\boldsymbol{x}_0$, это соотношение можно переписать в следующем виде:

$$(\boldsymbol{E} - \boldsymbol{\Phi})\,\boldsymbol{x}_0 = \boldsymbol{b}.$$

Но мы знаем (лемма 1 п. 8 § 7 гл. 1), что при наложенных на оператор $\boldsymbol{\Phi}$ условиях оператор $\boldsymbol{E} - \boldsymbol{\Phi}$ обратим. Следовательно, рассматриваемое уравнение для $\boldsymbol{x}_0$ имеет единственное решение

$$\boldsymbol{x}_0 = (\boldsymbol{E} - \boldsymbol{\Phi}_0)^{-1}\,\boldsymbol{b}_0.$$

Доказанное предложение означает, что

группа $\mathrm{Ort}^+(2)$ *является теоретико-множественным объединением группы параллельных переносов* $\mathrm{Trans}(2)$ *и всевозможных групп вращений* $\mathrm{Rot}_O(2)$*, отвечающих произвольным точкам* O *плоскости.*

Чтобы получить полное описание алгебраического строения группы $\mathrm{Ort}^+(2)$, это предложение необходимо дополнить указанием, что получается в результате умножения (композиции) элементов различных групп, объединением которых является группа $\mathrm{Ort}^+(2)$.

Задание. Покажите, что:

1) *композицией произвольного вращения R (не являющегося тождественным преобразованием E) и параллельного переноса на вектор b является вращение, индуцирующее тот же изометричный оператор $\boldsymbol{R}$, что и вращение R, но имеющее центр в точке, радиус-вектор которой относительно центра O вращения R равен $(\boldsymbol{E}-\boldsymbol{R})^{-1}b$;*

2) *композицией параллельного переноса на вектор b и вращения $R \neq E$ является вращение, индуцирующее оператор $\boldsymbol{R}$, но имеющее центр в точке с радиус-вектором $(\boldsymbol{E}-\boldsymbol{R})^{-1}\boldsymbol{R}b$;*

3) *композицией $R_2 \circ R_1$ двух вращений R_1 и R_2, имеющих центры соответственно в точках O_1 и O_2, является при $\boldsymbol{R}_2 = \boldsymbol{R}_1^{-1}$ параллельный перенос на вектор*

$$(\boldsymbol{E}-\boldsymbol{R}_2)\,b = b - \boldsymbol{R}_2 b,$$

где $b = \overrightarrow{O_1O_2}$, а при $\boldsymbol{R}_2 \neq \boldsymbol{R}_1^{-1}$ — вращение, индуцирующее оператор $\boldsymbol{R}_2\boldsymbol{R}_1$ и имеющее центр в точке, радиус-вектор которой относительно центра O_1 вращения R_1 равен

$$(\boldsymbol{E}-\boldsymbol{R}_2\boldsymbol{R}_1)^{-1}\,(\boldsymbol{E}-\boldsymbol{R}_2)\,b.$$

Поскольку операторы вида $\boldsymbol{E}-\boldsymbol{R}_2$ обратимы, отсюда, в частности, вытекает следующее

Предложение 3. *Каждый элемент группы* $\mathrm{Ort}^+(2)$ *является либо вращением, либо композицией двух вращений.*

Рассмотрим теперь ортогональные преобразования плоскости, не сохраняющие ориентаций.

Простейшим таким преобразованием является *симметрия* относительно произвольной прямой α. В координатах x, y, осью абсцисс которых является прямая α, эта симметрия выражается формулами

$$x' = x,$$
$$y' = -y.$$

В комплексной координате $z = x + iy$ эти формулы записываются в виде одной формулы:

$$z' = \bar{z}.$$

Замечание 2. Таким образом, симметрия плоскости не является аффинным преобразованием (автоморфизмом) комплексной прямой (ср. замечание 3 п. 4 § 3 гл. 2).

Предложение 4. *Каждое обращающее ориентации ортогональное преобразование Ψ плоскости является композицией*

$$\Psi = S \circ \Phi$$

некоторого движения Φ и симметрии S относительно произвольно выбранной (но фиксированной) прямой α.

Доказательство. Преобразование

$$\Phi = S \circ \Psi$$

является ортогональным преобразованием, сохраняющим ориентации, т. е. движением. С другой стороны,

$$\Psi = S \circ \Phi,$$

ибо $S = S^{-1}$.

Замечание 4. Вообще говоря,

$$\Phi \circ S \neq S \circ \Phi.$$

Тем не менее, преобразование Ψ допускает и разложение

$$\Psi = \Phi' \circ S,$$

где S — та же симметрия, что и выше, а Φ' — некоторое движение (вообще говоря, отличное от движения Φ).

Задание. Докажите это утверждение.

Следствие 1. *В комплексной координате z любое обращающее ориентации ортогональное преобразование задается формулой*

$$z' = \zeta \bar{z} + \beta,$$

где $|\zeta| = 1$, и обратно, любая такая формула задает ортогональное преобразование, обращающее ориентации.

Следствие 1 является аналогом предложения 1. Чтобы получить соответствующий аналог предложения 2, нам понадобится следующее

Определение 2. Композиция симметрии относительно прямой α и параллельного переноса на вектор, параллельный этой прямой, называется *скользящей симметрией с осью α.*

Заметим, что согласно этому определению, любая (простая) симметрия является скользящей симметрией.

Предложение 5. *Каждое обращающее ориентации ортогональное преобразование Ψ плоскости является скользящей симметрией.*

Доказательство. Согласно следствию из леммы 1 п. 8 § 7 гл. 1 существует система прямоугольных координат Oxy, в которой преобразование Ψ записывается формулами

$$x' = x + b_1,$$
$$y' = -y + b_2.$$

Поэтому это преобразование является композицией параллельного переноса

$$x' = x + b_1,$$
$$y' = y$$

на вектор $(b_1, 0)$, параллельный оси Ox, и преобразования

$$x' = x,$$
$$y' = -y + b_2,$$

являющегося, как нетрудно видеть, симметрией относительно прямой $y = \frac{b_2}{2}$. (Действительно, в координатах

$$x_1 = x,$$
$$y_1 = y - \frac{b_2}{2}$$

это преобразование записывается формулами $x_1' = x$, $y_1' = -y_1$).

Сопоставляя предложения 2 и 5, мы получаем следующее

Предложение 6. *Каждое ортогональное преобразование плоскости является одним из следующих трех преобразований:*

а) *параллельным переносом,*

б) *вращением,*

в) *скользящей симметрией.*

Рассмотрим, в частности, композицию $R \circ S$ симметрии S относительно оси Oy и вращения R на угол φ вокруг точки O. Эта композиция выражается, очевидно, формулами

$$\begin{aligned} x' &= x\cos\varphi + y\sin\varphi, \\ y' &= x\sin\varphi - y\cos\varphi. \end{aligned} \tag{1}$$

С другой стороны, согласно предложению 5 она является некоторой скользящей симметрией. Но эта скользящая симметрия оставляет на месте точку O. Поэтому

преобразование (1) *является простой симметрией относительно некоторой прямой* α, *проходящей через точку* O.

Упражнение. Покажите, что прямая α образует с осью абсцисс угол $\varphi/2$.

Поскольку

$$R = (R \circ S) \circ S,$$

тем самым доказано, что

каждое вращение является композицией двух симметрий.

Поскольку любая прямая может быть осью абсцисс некоторой системы прямоугольных координат (а любая точка на ней — началом координат), одновременно доказано, что

композиция двух симметрий относительно пересекающихся прямых является вращением (*вокруг точки пересечения этих прямых на удвоенный угол между этими прямыми*).

Задание. Докажите, что аналогичные утверждения справедливы и для параллельных переносов, т. е. что, во-первых,

каждый параллельный перенос является композицией двух симметрий (*относительно параллельных прямых*),

и, во-вторых,

композиция двух симметрий относительно параллельных прямых является параллельным переносом (*на вектор, перпендикулярный этим прямым и имеющий длину, равную удвоенному расстоянию между этими прямыми*).

Отсюда и из предложения 6 мы немедленно получаем следующее

Предложение 7. *Каждое ортогональное преобразование плоскости может быть разложено (не единственным образом) в композицию не более чем трех симметрий.*

4. Разложение ортогональных преобразований пространства

Определение 1. Сохраняющее ориентации ортогональное преобразование пространства, оставляющее на месте все точки некоторой прямой α, называется *вращением вокруг прямой* α, а прямая α называется его *осью*.

Очевидно, что совокупность $\mathrm{Rot}_\alpha(3)$ всех вращений пространства вокруг данной прямой α является группой.

Пусть i_1, i_2, i_3 — такой (ортонормированный) базис пространства, что вектор i_3 параллелен прямой α. Тогда оператор R, индуцированный произвольным вращением $R \in \mathrm{Rot}_\alpha(3)$, будет обладать тем свойством, что

$$Ri_3 = i_3.$$

Кроме того, в силу изометричности оператора R будут иметь место равенства

$$(Ri_1)\, i_3 = (Ri_1)\,(Ri_3) = i_1 i_3 = 0,$$
$$(Ri_2)\, i_3 = (Ri_2)\,(Ri_3) = i_2 i_3 = 0.$$

Следовательно, оператор R в базисе i_1, i_2, i_3 будет иметь матрицу вида

$$\begin{pmatrix} & & 0 \\ & A & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

где A — некоторая собственная ортогональная матрица:

$$A = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}.$$

Этим доказано, что в репере $Oi_1i_2i_3$, где O — точка прямой α (а вектор i_3 по-прежнему параллелен прямой α), вращение R записывается формулами

$$x' = x\cos\varphi - y\sin\varphi,$$
$$y' = x\sin\varphi + y\cos\varphi,$$
$$z' = z.$$

Теперь ясно, что *соответствие*

$$R \longmapsto e^{i\varphi}$$

определяет изоморфизм группы $\mathrm{Rot}_\alpha(3)$ *на группу* S^1.

Тем самым алгебраическое строение группы $\mathrm{Rot}_\alpha(3)$ полностью выяснено.

Кроме того, мы видим, что любую плоскость P, перпендикулярную прямой α, каждое вращение $R \in \mathrm{Rot}_\alpha(3)$ переводит в себя и индуцирует на этой плоскости некоторое вращение вокруг точки пересечения плоскости P с прямой α. Угол φ этого вращения — один и тот же для всех (одинаково ориентированных) плоскостей P.

Определение 2. Этот угол φ называется *углом вращения R*. Подчеркнем, что он зависит от ориентации плоскости P.

Таким образом, наше понятие «вращения пространства вокруг прямой» совпадает с наивно-интуитивным понятием «вращения», известным читателю из школьного курса.

Рассмотрим теперь группы $\mathrm{Rot}_O(3)$ вращений пространства с центром в фиксированной точке O (см. определение 3 п. 5 § 1).

Ясно, что для любой прямой α, проходящей через точку O, имеет место включение

$$\mathrm{Rot}_\alpha(3) \subset \mathrm{Rot}_O(3).$$

Поэтому группа $\mathrm{Rot}_O(3)$ содержит теоретико-множественное объединение

$$\bigcup_{\alpha \ni O} \mathrm{Rot}_\alpha(3)$$

всех групп $\mathrm{Rot}_\alpha(3)$, отвечающих всевозможным прямым α, проходящим через точку O.

Оказывается, что на самом деле группа $\mathrm{Rot}_O(3)$ совпадает с этим объединением:

$$\mathrm{Rot}_O(3) = \bigcup_{\alpha \ni O} \mathrm{Rot}_\alpha(3).$$

Иными словами, справедливо следующее

Предложение 1. *Каждое вращение пространства с центром в точке O является вращением вокруг некоторой прямой, проходящей через эту точку.*

Доказательство. Пусть $R \in \mathrm{Rot}_O(3)$ и пусть $\boldsymbol{R}$ — индуцированный изометричный оператор. Согласно следствию к лемме 4 п. 8 § 7 гл. 1 для оператора $\boldsymbol{R}$ существует ортонормированный базис $\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{i}_3$, в котором он имеет матрицу вида

$$\begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Но тогда ясно, что $R \in \mathrm{Rot}_\alpha(3)$, где α — прямая, проходящая через точку O и имеющая направляющий вектор $\boldsymbol{i}_3$.

Предложение 1, конечно, не определяет полностью алгебраического строения группы $\mathrm{Rot}_O(3)$: нужно еще указать, чему

равны произведения элементов из различных групп $\mathrm{Rot}_\alpha(3)$, объединением которых является группа $\mathrm{Rot}_O(3)$. Мы этого здесь делать не будем.

Полное алгебраическое описание группы $\mathrm{Rot}_O(3)$ (основанное на совершенно иных идеях) мы получим в следующем пункте.

Замечание 1. Обратим внимание на то, что

различные группы $\mathrm{Rot}_\alpha(3)$, *входящие в состав группы* $\mathrm{Rot}_O(3)$, *пересекаются лишь по тождественному преобразованию* E.

Действительно, вращение плоскости, оставляющее на месте две различные точки, является, очевидно, тождественным преобразованием.

Связь между группами $\mathrm{Rot}_O(3)$, отвечающими различным точкам O, описывается следующим предложением:

Предложение 2. *Для любых двух различных точек* O_1 *и* O_2 *имеет место соотношение* [1])

$$\mathrm{Rot}_{O_1}(3) \cap \mathrm{Rot}_{O_2}(3) = \mathrm{Rot}_\alpha(3),$$

где α *— прямая* O_1O_2.

Доказательство. Включение

$$\mathrm{Rot}_{O_1}(3) \cap \mathrm{Rot}_{O_2}(3) \supset \mathrm{Rot}_\alpha(3)$$

очевидно. Обратно, пусть

$$R \in \mathrm{Rot}_{O_1}(3) \cap \mathrm{Rot}_{O_2}(3).$$

Тогда согласно предложению 1

$$R \in \mathrm{Rot}_{\alpha_1}(3), \quad R \in \mathrm{Rot}_{\alpha_2}(3),$$

где α_1 и α_2 — некоторые прямые, проходящие через точки O_1 и O_2 соответственно. Если $R \neq E$, то прямые α_1 и α_2 однозначно определены вращением R и каждая из них состоит из всех точек, остающихся при вращении R на месте. Следовательно, $\alpha_1 = \alpha_2 = O_1O_2$, т. е.

$$R \in \mathrm{Rot}_\alpha(3).$$

Следствие. *Имеет место равенство*

$$\mathrm{Rot}_\alpha(3) = \bigcap_{O \in \alpha} \mathrm{Rot}_O(3),$$

где пересечение распространено на все точки O *прямой* α.

Определение 3. Пусть α — произвольная прямая в пространстве. Движение пространства называется *винтовым движением с осью* α, если оно является композицией некоторого вращения

[1]) Заметим, что пересечение (в отличие от объединения) двух групп преобразований всегда является группой.

вокруг прямой α и параллельного переноса на вектор $\boldsymbol{b}$, параллельный прямой α.

В частности, любое вращение вокруг прямой α и любой параллельный перенос на вектор, параллельный прямой α, являются винтовыми движениями с осью α.

Ясно, что композиция двух винтовых движений с одной и той же осью также является винтовым движением с той же осью, и движение, обратное винтовому движению, также является винтовым движением с той же осью. Это означает, что

совокупность [1]) $\mathrm{Screw}_\alpha(3)$ *всех винтовых движений с осью* α *является подгруппой группы* $\mathrm{Ort}^+(3)$.

Каждое винтовое движение Φ однозначно задается (если выбраны ориентации пространства и оси α) углом вращения φ и величиной b вектора переноса. При этом легко видеть, что если винтовому движению Φ_1 соответствуют угол φ_1 и число b_1, а винтовому движению Φ_2 — угол φ_2 и число b_2, то их композиции $\Phi_2 \circ \Phi_1$ будет соответствовать угол $\varphi_1 + \varphi_2$ и число $b_1 + b_2$. Заменяя углы φ соответствующими комплексными числами $e^{i\varphi}$, мы, следовательно, можем сказать, что

группа $\mathrm{Screw}_\alpha(3)$ *винтовых движений с осью* α *изоморфна группе, элементами которой являются пары вида* $(e^{i\varphi}, b)$ *и умножение в которой определено формулой*

$$(e^{i\varphi_1}, b_1)(e^{i\varphi_2}, b_2) = (e^{i(\varphi_1+\varphi_2)}, b_1 + b_2).$$

Это полностью описывает алгебраическое строение группы $\mathrm{Screw}_\alpha(3)$.

На языке теории групп доказанное утверждение означает, что группа $\mathrm{Screw}_\alpha(3)$ является *прямым произведением* групп S^1 и $\mathbb{R}$.

Предложение 3. *Любое движение пространства является винтовым движением.*

Доказательство. Поскольку любое движение является (см. предложение 2 п. 4 § 1) композицией вращения и параллельного переноса и поскольку любой параллельный перенос является композицией переноса на вектор, параллельный данной прямой, и переноса на вектор, перпендикулярный этой прямой, для доказательства предложения 3 достаточно доказать, что

композиция $T \circ R$ *произвольного вращения* R *и параллельного переноса* T *на вектор* $\boldsymbol{b}$, *перпендикулярный оси* α *вращения* R, *является вращением вокруг прямой* α', *параллельной прямой* α.

С этой целью мы запишем движение $T \circ R$ в координатах x, y, z, осью Oz которых является ось α вращения R:

$$\begin{aligned} x' &= x\cos\varphi - y\sin\varphi + b_1, \\ y' &= x\sin\varphi + y\cos\varphi + b_2, \\ z' &= z. \end{aligned}$$

[1]) От английского «Screw» — винт.

Мы видим, что каждую плоскость $z = \text{const}$ это движение переводит в себя и индуцирует на ней движение

$$x' = x\cos\varphi - y\sin\varphi + b_1,$$
$$y' = x\sin\varphi + y\cos\varphi + b_2.$$

Поскольку $\varphi \neq 0$, мы знаем (предложение 2 п. 2), что последнее движение представляет собой вращение на угол φ вокруг некоторой точки (x_0, y_0). Поэтому и исходное движение $T \circ R$ будет вращением на угол φ вокруг прямой α' с уравнениями

$$x = x_0,$$
$$y = y_0.$$

Тем самым предложение 3 полностью доказано.

Это предложение означает, что

группа $\mathrm{Ort}^+(3)$ *всех движений пространства является теоретико-множественным объединением групп* $\mathrm{Screw}_\alpha(3)$, *соответствующим всевозможным прямым* α.

При этом

любые две группы $\mathrm{Screw}_{\alpha_1}(3)$ *и* $\mathrm{Screw}_{\alpha_2}(3)$ *пересекаются только по тождественному преобразованию, если прямые* α_1 *и* α_2 *не параллельны, и по группе параллельных переносов на векторы, параллельные прямым* α_1 *и* α_2, *если эти прямые параллельны.*

Действительно, для любого движения $\Phi \in \mathrm{Screw}_\alpha(3)$, не являющегося параллельным переносом, прямая α однозначно характеризуется как прямая, на которой это движение индуцирует параллельный перенос.

Предложение 3 еще не дает полного описания алгебраического строения группы $\mathrm{Ort}^+(3)$. Такое описание мы получим в следующем пункте.

Рассмотрим, наконец, группу $\mathrm{Ort}(3)$ всех ортогональных преобразований пространства.

Примером ортогонального преобразования, обращающего ориентации, является симметрия относительно произвольной плоскости P. В прямоугольных координатах x, y, z, для которых плоскость P является плоскостью Oxy, эта симметрия записывается формулами

$$x' = x,$$
$$y' = y,$$
$$z' = -z.$$

Предложение 4. *Каждое обращающее ориентации ортогональное преобразование* Ψ *пространства является композицией*

$$\Psi = S \circ \Phi$$

некоторого движения Φ *и симметрии* S *относительно произвольно выбранной (но фиксированной) плоскости* P.

Доказательство (ср. доказательство предложения 4 п. 2). Движение Φ определяется формулой

$$\Phi = S \circ \Psi.$$

Уточнением этого предложения является следующее предложение (не имеющее аналога на плоскости):

Предложение 5. *Каждое обращающее ориентации ортогональное преобразование* Ψ *пространства является композицией* $S \circ \Phi$ *некоторого винтового движения* Φ *и симметрии* S *относительно плоскости, перпендикулярной оси движения* Φ.

Доказательство. В первую очередь мы покажем, что *для любого обращающего ориентации ортогонального преобразования* Ψ *пространства существует точка* O, *оставляемая этим преобразованием на месте*:

$$\Psi(O) = O.$$

Действительно, если преобразование Ψ является симметрией относительно некоторой плоскости P, то существование точки O очевидно. В противном случае оператор Ψ, индуцированный преобразованием Ψ, обладает тем свойством, что ни в одном ортонормированном базисе он не имеет матрицы вида

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

и потому согласно лемме 5 п. 8 § 7 гл. 1 оператор $\mathbf{E} - \Psi$ обратим. Но тогда уже известное нам рассуждение (см. доказательство предложения 2 п. 2) показывает, что точка O существует (и притом единственная).

Отсюда и из следствия к лемме 4 п. 8 § 7 гл. 1 немедленно вытекает, что в некоторой системе прямоугольных координат преобразование Ψ записывается формулами

$$\begin{aligned} x' &= x\cos\varphi - y\sin\varphi, \\ y' &= x\sin\varphi + y\cos\varphi, \\ z' &= -z. \end{aligned}$$

Поэтому композиция $\Phi = S \circ \Psi$ преобразования Ψ и симметрии S относительно координатной плоскости Oxy будет записываться формулами

$$\begin{aligned} x' &= x\cos\varphi - y\sin\varphi, \\ y' &= x\sin\varphi + y\cos\varphi, \\ z' &= z \end{aligned}$$

и, следовательно, будет винтовым движением с осью Oz.

Поскольку $\Psi = S \circ \Phi$ и поскольку плоскость симметрии S перпендикулярна оси Oz движения Φ, предложение 5 тем самым полностью доказано.

Докажем в заключение, что, аналогично ортогональным преобразованиям прямой и плоскости, любое ортогональное преобразование пространства является композицией некоторого числа симметрий.

Предложение 6. *Каждый элемент группы* $\mathrm{Ort}(3)$ *является композицией не более чем пяти симметрий относительно плоскостей.*

Доказательство. Поскольку любое обращающее ориентации (сохраняющее ориентации) ортогональное преобразование пространства является композицией винтового движения и симметрии (является винтовым движением), а любое винтовое движение — композицией параллельного переноса и вращения, нам достаточно доказать, что

вращение вокруг прямой α *на угол* φ *является композицией двух симметрий относительно плоскостей, пересекающихся по прямой* α *и образующих угол* $\varphi/2$ (*а в остальном произвольных*), и что

параллельный перенос на вектор $\boldsymbol{b}$ *является композицией двух симметрий относительно параллельных плоскостей, перпендикулярных вектору* $\boldsymbol{b}$ *и находящихся друг от друга на расстоянии* $|\boldsymbol{b}|/2$.

Мы докажем только первое утверждение. Второе доказывается аналогично, и мы оставим его читателю.

Пусть R — вращение на угол φ вокруг прямой α и пусть P — плоскость, перпендикулярная прямой α. Как мы знаем, вращение R отображает плоскость P на себя и индуцирует на ней вращение R_{P} на угол φ вокруг точки O пересечения плоскости P и прямой α. Являясь вращением плоскости, преобразование R_{P} разлагается (см. п. 2) в композицию двух симметрий плоскости P относительно некоторых прямых β_1 и β_2, пересекающихся в точке O и образующих угол $\varphi/2$ (а в остальном произвольных). Пусть P_1 — плоскость, проходящая через (пересекающиеся) прямые α и β_1, а P_2 — плоскость, проходящая через (пересекающиеся) прямые α и β_2, и пусть S_1 — симметрия относительно плоскости P_1, а S_2 — симметрия относительно плоскости P_2.

Композиция $S_2 \circ S_1$ симметрий S_1 и S_2 представляет собой сохраняющее ориентации ортогональное преобразование пространства (т. е. винтовое движение) и оставляет на месте каждую точку прямой α. Поэтому оно является вращением вокруг этой прямой на некоторый угол. Но, по построению, вращение $S_2 \circ S_1$ индуцирует на плоскости P то же вращение R_{P}, что и данное вращение R. Следовательно, преобразование $S_2 \circ S_1$ является вращением на угол φ и потому совпадает с R.

5. Представление движений пространства с помощью кватернионов

Как мы видели (см. замечание 1 п. 1), каждое ортогональное преобразование прямой может быть записано формулой

$$x' = ax + b,$$

где $|a| = 1$. Аналогичную запись допускают и сохраняющие ориентации ортогональные преобразования (движения) плоскости, только вместо вещественных чисел необходимо использовать комплексные числа (см. предложение 1 п. 3). Оказывается, что подобными формулами можно задавать и движения в пространстве, если воспользоваться обобщенными комплексными числами, известными под названием кватернионов.

По определению *кватернионом* называется произвольное выражение вида

$$x_0 + x_1 i + x_2 j + x_3 k,$$

где x_0, x_1, x_2, x_3 — вещественные числа, а i, j, k — некоторые символы, называемые *кватернионными единицами*. Коэффициенты x_1, x_2, x_3, равные единице, не пишутся, а члены с коэффициентами, равными нулю, опускаются. Например, вместо $0 + 1i + 0j + 0k$ пишется просто i. Кватернионы вида $x_0 + 0i + 0j + 0k$ отождествляются с вещественными числами x_0.

Кватернионы складываются по «векторному» правилу

$$(x_0 + x_1 i + x_2 j + x_3 k) + (y_0 + y_1 i + y_2 j + y_3 k) = \\ = (x_0 + y_0) + (x_1 + y_1) i + (x_2 + y_2) j + (x_3 + y_3) k,$$

а перемножаются по правилу

$$(x_0 + x_1 i + x_2 j + x_3 k)(y_0 + y_1 i + y_2 j + y_3 k) = \\ = (x_0 y_0 - x_1 y_1 - x_2 y_2 - x_3 y_3) + (x_0 y_1 + x_1 y_0 + x_2 y_3 - x_3 y_2) i + \\ + (x_0 y_2 - x_1 y_3 + x_2 y_0 + x_3 y_1) j + (x_0 y_3 + x_1 y_2 - x_2 y_1 + x_3 y_0) k.$$

В частности,

$$x_0 (y_0 + y_1 i + y_2 j + y_3 k) = (x_0 y_0) + (x_0 y_1) i + (x_0 y_2) j + (x_0 y_3) k.$$

В силу этих правил любой кватернион

$$\xi = x_0 + x_1 i + x_2 j + x_3 k$$

является суммой кватернионов x_0, $x_1 i$, $x_2 j$, $x_3 k$, где вещественные числа x_0, x_1, x_2, x_3 рассматриваются (в соответствии с принятым выше соглашением) как кватернионы.

Операции над кватернионами обладают всеми стандартными свойствами арифметических операций (ассоциативностью, дистрибутивностью и т. п.), за исключением того, что умножение не коммутативно. Например, $ij = k$, но $ji = -k$. Однако, если кватернион является вещественным числом, то он перестановочен с любым кватернионом.

Упражнение. Покажите, что кватернион тогда и только тогда перестановочен с любым кватернионом, когда он является вещественным числом.

Все возможные произведения кватернионных единиц выражаются формулами

$$i^2 = -1, \quad ij = k, \quad ik = -j,$$
$$ji = -k, \quad j^2 = -1, \quad jk = i,$$
$$ki = j, \quad kj = -i, \quad k^2 = -1.$$

Произведение любых двух кватернионов сводится к этим произведениям в силу дистрибутивности сложения относительно умножения.

Для любого кватерниона

$$\xi = x_0 + x_1 i + x_2 j + x_3 k$$

число x_0 называется его *вещественной частью* и обозначается символом $\operatorname{Re}\xi$. Кватернионы вида

$$x_1 i + x_2 j + x_3 k,$$

т. е. кватернионы, вещественная часть которых равна нулю, называются *мнимыми*.

Кватернион

$$x_0 - x_1 i - x_2 j - x_3 k$$

называется *сопряженным* с кватернионом

$$\xi = x_0 + x_1 i + x_2 j + x_3 k.$$

Сопряженный кватернион обозначается символом $\bar{\xi}$.

Ясно, что $\bar{\bar{\xi}} = \xi$ для любого кватерниона ξ.

Столь же ясно, что кватернион ξ тогда и только тогда является вещественным числом, когда $\bar{\xi} = \xi$, и мнимым кватернионом, когда $\bar{\xi} = -\xi$.

Автоматическая проверка показывает, что операция сопряжения удовлетворяет соотношениям

$$\overline{\xi + \eta} = \bar{\xi} + \bar{\eta}, \quad \overline{\xi\eta} = \bar{\eta}\bar{\xi}$$

(обратите внимание на порядок множителей во второй формуле!).

Для любого кватерниона ξ кватернион $\xi\bar{\xi}$ является вещественным числом (ибо $\overline{\xi\bar{\xi}} = \bar{\bar{\xi}}\bar{\xi} = \xi\bar{\xi}$). Более того, вычисляя это число, мы немедленно получим, что

$$\xi\bar{\xi} = x_0^2 + x_1^2 + x_2^2 + x_3^2.$$

Следовательно, число $\xi\bar{\xi}$ неотрицательно (и равно нулю только для кватерниона $\xi = 0$). Арифметический квадратный корень из числа $\xi\bar{\xi}$ обозначается символом $|\xi|$ и называется *модулем* (или *абсолютной величиной*) кватерниона ξ:

$$|\xi| = \sqrt{\xi\bar{\xi}} = \sqrt{x_0^2 + x_1^2 + x_2^2 + x_3^2}.$$

Так как

$$|\xi\eta|^2 = \xi\eta\overline{\xi\eta} = \xi\eta\bar{\eta}\bar{\xi} = \xi|\eta|^2\bar{\xi} = \xi\bar{\xi}|\eta|^2 = |\xi|^2|\eta|^2$$

(напомним, что вещественные числа перестановочны с любыми кватернионами), то, так же как для комплексных чисел, модуль произведения кватернионов равен произведению модулей сомножителей:

$$|\xi\eta| = |\xi| \cdot |\eta|.$$

Поскольку для любого отличного от нуля кватерниона ξ число $|\xi|$ отлично от нуля, то определен кватернион

$$\xi^{-1} = \frac{\bar{\xi}}{|\xi|^2},$$

удовлетворяющий равенствам

$$\xi^{-1}\xi = \xi\xi^{-1} = 1.$$

Это означает, что в миожестве кватернионов возможно деление на любой отличный от нуля кватернион (хотя в силу некоммутативности умножения следует различать левое частное $\xi^{-1}\eta$ от правого частного $\eta\xi^{-1}$).

Множество всех кватернионов η с модулем, равным единице:

$$|\eta| = 1,$$

обозначается символом S^3. Для таких кватернионов

$$\eta^{-1} = \bar{\eta},$$

так что, если $\eta \in S^3$, то $\eta^{-1} \in S^3$. Кроме того, если $\eta_1 \in S^3$ и $\eta_2 \in S^3$, то $\eta_1\eta_2 \in S^3$ (ибо $|\eta_1\eta_2| = |\eta_1||\eta_2| = 1$). Все это означает, что

совокупность S^3 всех кватернионов, равных по модулю единице, является группой по умножению.

Единицей этой группы служит кватернион 1.

Группе S^3 принадлежит, в частности, каждый кватернион вида

$$\cos\varphi + \sin\varphi \cdot i.$$

Такой кватернион мы будем обозначать символом $e^{i\varphi}$:

$$e^{i\varphi} = \cos\varphi + \sin\varphi \cdot i.$$

Поскольку

$$\cos\varphi - \sin\varphi \cdot i = \cos(-\varphi) + \sin(-\varphi) \cdot i,$$

то

$$\overline{e^{i\varphi}} = e^{-i\varphi}.$$

Аналогично показывается, что

$$e^{i\varphi_1}e^{i\varphi_2} = e^{i(\varphi_1+\varphi_2)}.$$

Следовательно, совокупность всех кватернионов вида $e^{i\varphi}$ является группой, изоморфной группе S^1 комплексных чисел, равных по модулю единице.

Аналогично мы положим

$$e^{j\varphi} = \cos\varphi + \sin\varphi \cdot j,$$

$$e^{k\varphi} = \cos\varphi + \sin\varphi \cdot k.$$

Эти кватернионы также принадлежат группе S^3 и для них имеют место аналогичные утверждения (например, совокупность всех кватернионов вида $e^{k\varphi}$ также изоморфна группе S^1).

Замечание 1. Обозначения $e^{i\varphi}$, $e^{j\varphi}$, $e^{k\varphi}$ носят условный характер и вводятся только для сокращения записи. Обращаться с ними нужно осторожно: например, нельзя писать, что $e^{i\varphi}e^{j\psi} = e^{i\varphi+j\psi}$ (хотя бы потому, что обозначение $e^{i\varphi+j\psi}$ смысла не имеет!).

Лемма 1. *Каждый кватернион* $\eta \in S^3$ *допускает разложение вида*

$$\eta = e^{k\varphi}e^{i\psi}e^{k\theta}, \tag{1}$$

где

$$-\pi/2 < \varphi \leqslant \pi/2, \quad 0 \leqslant \psi \leqslant \pi/2, \quad -\pi/2 < \theta \leqslant \pi/2.$$

Если $\psi \neq 0, \pi/2$, *то это разложение единственно.*

Доказательство. Пусть

$$\eta = y_0 + y_1 i + y_2 j + y_3 k.$$

По условию

$$y_0^2 + y_1^2 + y_2^2 + y_3^2 = 1,$$

и потому существует такой угол α, $0 \leqslant \alpha \leqslant \pi/2$, что

$$y_0^2 + y_3^2 = \cos^2\alpha, \quad y_1^2 + y_2^2 = \sin^2\alpha.$$

Если $\alpha \neq \pi/2$, то

$$\left(\frac{y_0}{\cos\alpha}\right)^2 + \left(\frac{y_3}{\cos\alpha}\right)^2 = 1,$$

и потому существует такой угол β, $-\pi/2 < \beta \leqslant \pi/2$, что

$$y_0 = \cos\alpha\cos\beta, \quad y_3 = \cos\alpha\sin\beta.$$

Эти формулы справедливы и при $\alpha = \pi/2$, но только в этом случае угол β может быть любым (а при $\alpha \neq \pi/2$ угол β определен однозначно).

По аналогичным соображениям существует такой угол γ, $-\pi/2 < \gamma \leqslant \pi/2$ (однозначно определенный при $\alpha \neq 0$), что

$$y_1 = \sin\alpha\cos\gamma, \quad y_3 = \sin\alpha\sin\gamma.$$

Таким образом,

$$\eta = \cos\alpha\cos\beta + \sin\alpha\cos\gamma \cdot i + \sin\alpha\sin\gamma \cdot j + \cos\alpha\sin\beta \cdot k.$$

С другой стороны, по правилу умножения кватернионов

$$e^{k\varphi}e^{i\psi} = (\cos\varphi + \sin\varphi \cdot k)(\cos\psi + \sin\psi \cdot i) =$$
$$= \cos\varphi\cos\psi + \cos\varphi\sin\psi \cdot i + \sin\varphi\sin\psi \cdot j + \sin\varphi\cos\psi \cdot k$$

и

$$e^{k\varphi}e^{i\psi}e^{k\theta} = (\cos\varphi\cos\psi + \cos\varphi\sin\psi \cdot i + \sin\varphi\sin\psi \cdot j + \sin\varphi\cos\psi \cdot k) \times$$
$$\times (\cos\theta + \sin\theta \cdot k) = (\cos\varphi\cos\psi\cos\theta - \sin\varphi\cos\psi\sin\theta) +$$
$$+ (\cos\varphi\sin\psi\cos\theta + \sin\varphi\sin\psi\sin\theta)\, i + (\sin\varphi\sin\psi\cos\theta - \cos\varphi\sin\psi\sin\theta)\, j +$$
$$+ (\sin\varphi\cos\psi\cos\theta + \cos\varphi\cos\psi\sin\theta)\, k =$$
$$= \cos\psi\cos(\varphi + \theta) + \sin\psi\cos(\varphi - \theta) \cdot i +$$
$$+ \sin\psi\sin(\varphi - \theta) \cdot j + \cos\psi\sin(\varphi + \theta) \cdot k.$$

Сравнивая полученные формулы, мы немедленно замечаем, что разложение (1) имеет место при

$$\psi = \alpha, \quad \varphi + \theta = \beta, \quad \varphi - \theta = \gamma.$$

Единственность этого разложения непосредственно вытекает из единственности углов α, β, γ.

Заметим в заключение, что, скажем, кватернион $e^{i\varphi}$ тогда и только тогда — мнимый, когда $\varphi = \pi/2$. В этом случае он совпадает с кватернионной единицей i и потому его квадрат равен -1. Оказывается, что это общий факт, т. е.

кватернион $\eta \in S^3$ *тогда и только тогда — мнимый, когда* $\eta^2 = -1$.

Действительно, если кватернион η — мнимый, то $\bar{\eta} = -\eta$, и потому $\eta^2 = -\eta\bar{\eta} = -|\eta|^2 = -1$. Обратно, если $\eta^2 = -1$, то $\eta = \eta \cdot |\eta| = \eta\eta\bar{\eta} = -\bar{\eta}$ и, следовательно, кватернион η — мнимый.

Множество всех мнимых кватернионов, принадлежащих группе S^3, мы будем обозначать символом S^2. Подчеркнем, что группой это множество не является.

Предположим теперь, что в пространстве выбрана (и раз навсегда фиксирована) некоторая система прямоугольных координат x, y, z. Тогда, сопоставив произвольному вектору с координатами x, y, z кватернион

$$\xi = xi + yi + zk,$$

мы получим биективное соответствие между векторами в пространстве и мнимыми кватернионами. Это соответствие сохраняет, очевидно, операции сложения и умножения на вещественное число. Следовательно,

векторы в пространстве можно отождествлять с мнимыми кватернионами.

После такого отождествления мы можем говорить о *скалярном произведении* двух мнимых кватернионов ξ_1 и ξ_2 (понимая под ним скалярное произведение соответствующих векторов). Поскольку символ $\xi_1\xi_2$ у нас уже занят для обозначения произведения кватернионов, их скалярное произведение мы будем обозначать символом (ξ_1, ξ_2).

По правилу умножения кватернионов

$$(x_1i + y_1j + z_1k)(x_2i + y_2j + z_2k) = -(x_1x_2 + y_1y_2 + z_1z_2) + \ldots$$

Это показывает, что

скалярным произведением двух мнимых кватернионов ξ_1 *и* ξ_2 *является взятая с обратным знаком вещественная часть их произведения*:

$$(\xi_1, \xi_2) = -\operatorname{Re}(\xi_1\xi_2).$$

В частности,

мнимые кватернионы тогда и только тогда ортогональны, когда их произведение $\xi_1\xi_2$ *также является мнимым кватернионом.*

Упражнение. Покажите, что «мнимой частью» $\xi_1\xi_2 - \operatorname{Re}(\xi_1\xi_2)$ кватерниона $\xi_1\xi_2$ является векторное произведение векторов ξ_1 и ξ_2.

Для любой точки пространства M ее радиус-вектор (рассматриваемый как мнимый кватернион) мы будем называть *кватернионной координатой* этой точки. Таким образом, если точка имеет координаты x, y, z, то ее кватернионная координата будет выражаться формулой

$$\xi = xi + yj + zk.$$

Для любого кватерниона $\eta \in S^3$ и любой точки M с кватернионной координатой ξ рассмотрим кватернион

$$\xi' = \eta\xi\bar{\eta}.$$

Поскольку

$$\bar{\xi}' = \overline{\eta\xi\bar{\eta}} = \bar{\bar{\eta}}\bar{\xi}\bar{\eta} = -\eta\xi\bar{\eta} = -\xi',$$

кватернион ξ' — мнимый и потому является кватернионной координатой некоторой точки M'. Положив

$$M' = R_\eta(M),$$

мы, таким образом, получим некоторое отображение R_η пространства в себя.

Так как для любых двух точек $M_1(\xi_1)$ и $M_2(\xi_2)$ расстояние $|M_1M_2|$ между этими точками, очевидно, равно модулю кватерниона $\xi_2 - \xi_1$:

$$|M_1M_2| = |\xi_2 - \xi_1|$$

и так как

$$|\xi_2' - \xi_1'| = |\eta\xi_2\bar{\eta} - \eta\xi_1\bar{\eta}| = |\eta(\xi_2 - \xi_1)\bar{\eta}| =$$
$$= |\eta| \cdot |\xi_2 - \xi_1| \cdot |\bar{\eta}| = |\xi_2 - \xi_1|,$$

то, следовательно,

для любого кватерниона $\eta \in S^3$ преобразование R_η является ортогональным преобразованием пространства.

Так как

$$(\eta_1\eta_2)\,\xi\,(\overline{\eta_1\eta_2}) = \eta_1(\eta_2\xi\bar{\eta}_2)\bar{\eta}_1,$$

то для любых кватернионов $\eta_1, \eta_2 \in S^3$ имеет место равенство

$$R_{\eta_1\eta_2} = R_{\eta_1} \circ R_{\eta_2},$$

показывающее, что

соответствие

$$\eta \longmapsto R_\eta$$

является гомоморфизмом группы S^3 в группу $\mathrm{Ort}(3)$.

Найдем преобразование R_η при $\eta = e^{k\alpha}$. По правилу умножения кватернионов

$$e^{k\alpha}\xi = (\cos\alpha + \sin\alpha \cdot k)(xi + yj + zk) =$$
$$= -z\sin\alpha + (x\cos\alpha - y\sin\alpha)\,i +$$
$$+ (x\sin\alpha + y\cos\alpha)\,j + z\cos\alpha \cdot k$$

и, далее,

$$e^{k\alpha}\xi\overline{e^{k\alpha}} = (x\cos 2\alpha - y\sin 2\alpha)\,i + (x\sin 2\alpha + y\cos 2\alpha)\,j + zk.$$

Это показывает, что

преобразование $R_{e^{k\alpha}}$ *задается в координатах* x, y, z *формулами*

$$x' = x\cos 2\alpha - y\sin 2\alpha,$$
$$y' = x\sin 2\alpha + y\cos 2\alpha,$$
$$z' = z$$

и потому представляет собой вращение вокруг оси Oz *на угол* 2α.

Аналогично показывается, что

преобразование $R_{e^{j\alpha}}$ *задается в координатах* x, y, z *формулами*

$$x' = x\cos 2\alpha - z\sin 2\alpha,$$
$$y' = y,$$
$$z' = x\sin 2\alpha + z\cos 2\alpha$$

и потому представляет собой вращение вокруг оси Oy *на угол* 2α, и что

преобразование $R_{e^{i\alpha}}$ *задается в координатах* x, y, z *формулами*

$$x' = x,$$
$$y' = y\cos 2\alpha - z\sin 2\alpha,$$
$$z' = y\sin 2\alpha + z\cos 2\alpha$$

и потому представляет собой вращение вокруг оси Ox *на угол* 2α.

Но согласно доказанной выше лемме 1 любой кватернион $\eta \in S^3$ допускает разложение вида

$$\eta = e^{k\varphi}e^{i\psi}e^{k\theta}.$$

Следовательно,

каждое преобразование R_η, $\eta \in S^3$, *является композицией трех вращений*:

$$R_\eta = R_{e^{k\varphi}} \circ R_{e^{i\psi}} \circ R_{e^{k\theta}}, \qquad (2)$$

и потому само является вращением (*вокруг начала координат* O).

Таким образом,

соответствие

$$\eta \longmapsto R_\eta$$

представляет собой на самом деле гомоморфизм группы S^3 *в группу вращений* $\mathrm{Rot}_O(3)$.

Прежде чем исследовать этот гомоморфизм более подробно, мы введем в рассмотрение еще один класс преобразований пространства, получающихся с помощью кватернионов.

Пусть η — произвольный кватернион, принадлежащий множеству S^2 (т. е. являющийся мнимым кватернионом из S^3). Любой точке $M(\xi)$ мы сопоставим кватернион

$$\xi' = \eta\xi\eta.$$

Поскольку $\bar{\xi} = -\xi$, $\bar{\eta} = -\eta$, то

$$\bar{\xi}' = \bar{\eta}\bar{\xi}\bar{\eta} = -\eta\xi\eta = -\xi$$

и, следовательно, кватернион ξ' — мнимый и потому является кватернионной координатой некоторой точки M'. Сопоставив точке M точку M', мы получим поэтому некоторое отображение S_η пространства в себя:

$$M' = S_\eta(M).$$

Пусть P_η — плоскость, проходящая через точку O перпендикулярно (единичному) вектору η. Кватернионные координаты ξ точек этой плоскости ортогональны, таким образом, кватерниону η.

Покажем, что

отображение S_η представляет собой симметрию пространства относительно плоскости P_η.

Для этого достаточно показать, что если кватернион ξ ортогонален кватерниону η, то

$$\xi' = \xi,$$

а если кватернион ξ коллинеарен кватерниону η (имеет вид $k\eta$, где k — вещественное число), то

$$\xi' = -\xi.$$

Но если кватернион ξ ортогонален кватерниону η, то (см. выше) кватернион $\xi\eta$ — мнимый и потому $\overline{\xi\eta} = -\xi\eta$. Следовательно,

$$\xi' = \eta\xi\eta = -\eta\overline{\xi\eta} = -\eta\bar{\eta}\bar{\xi} = |\eta|\xi = \xi.$$

Аналогично, если $\xi = k\eta$, то

$$\xi' = \eta(k\eta)\eta = k\eta^3 = (k\eta)\eta^2 = -k\eta = -\xi$$

(напомним, что $\eta^2 = -1$ для любого кватерниона $\eta \in S^2$).

Тем самым наше утверждение полностью доказано.

Композиция $S_{\eta_1} \circ S_{\eta_2}$ двух симметрий S_{η_1} и S_{η_2} определяется, очевидно, формулой

$$\xi' = \eta_1\eta_2\xi\eta_2\eta_1.$$

Но, поскольку кватернионы η_1 и η_2 — мнимые, то $\eta_2\eta_1 = \bar{\eta}_2\bar{\eta}_1 = \overline{\eta_1\eta_2}$ и потому

$$\xi' = (\eta_1\eta_2)\,\xi\,(\overline{\eta_1\eta_2}).$$

Это показывает, что

композиция $S_{\eta_1} \circ S_{\eta_2}$ симметрий S_{η_1} и S_{η_2} является вращением $R_{\eta_1\eta_2}$.

Но мы знаем, что (см. п. 3) в виде композиции двух симметрий может быть представлено любое вращение пространства. Следовательно,

каждое вращение пространства вокруг точки O является вращением R_η, соответствующим некоторому кватерниону $\eta \in S^3$.

Другими словами, имеет место следующее

Предложение 1. *Каждое вращение пространства вокруг точки O записывается в кватернионных координатах ξ формулой*

$$\xi' = \eta\xi\bar{\eta},$$

где $|\eta| = 1$.

На языке теории групп это означает, что

гомоморфизм $\eta \to R_\eta$ является эпиморфизмом группы S^3 на группу $\mathrm{Rot}_O(3)$.

Упражнение. Покажите, что мнимая часть $\eta - \mathrm{Re}\,\eta$ кватерниона $\eta \in S^3$ представляет собой вектор, параллельный оси вращения R_η, и выразите его длину через угол вращения R_η.

Эпиморфизм $\eta \longmapsto R_\eta$ изоморфизмом заведомо не является, поскольку кватернионы 1 и -1 оба определяют тождественное преобразование. Другими словами, группа второго порядка

$$S^0 = \{1, -1\}$$

содержится в его ядре.

Оказывается, что *группа S^0 составляет все ядро*, т. е. справедливо следующее

Предложение 2. *Если кватернион $\eta \in S^3$ обладает тем свойством, что для любого мнимого кватерниона ξ имеет место равенство*

$$\eta\xi\bar{\eta} = \xi,$$

то $\eta = \pm 1$.

Доказательство. Пусть

$$\eta = y_0 + y_1 i + y_2 j + y_3 k.$$

Тогда

$$\eta i\bar{\eta} = (-y_1 + y_0 i + y_3 j - y_2 k)(y_0 - y_1 i - y_2 j - y_3 k) =$$
$$= (-y_1 y_0 + y_0 y_1 - y_3^2 - y_2^2) + \ldots$$

Следовательно, равенство $\eta i\bar{\eta} = i$ возможно только тогда, когда $y_3^2 + y_2^2 = 0$, т. е. когда $y_2 = 0$, $y_3 = 0$.

Аналогично, полагая $\xi = j$, мы получим, что $y_1 = 0$, $y_3 = 0$. Следовательно, кватернион η является вещественным числом y_0. Для завершения доказательства остается заметить, что единственными вещественными числами, равными по модулю единице, являются числа ± 1.

Согласно общеалгебраической теореме о гомоморфизмах предложение 2 означает, что справедлива следующая теорема, полностью описывающая алгебраическое строение группы $\mathrm{Rot}_O(3)$:

Теорема 1. *Группа* $\mathrm{Rot}_O(3)$ *изоморфна факторгруппе* S^3/S^0.

Напомним, что элементами факторгруппы S^3/S^0 являются пары $(\eta, -\eta)$, где $\eta \in S^3$. Умножаются эти пары по формуле

$$(\eta_1, -\eta_1)(\eta_2, -\eta_2) = (\eta_1\eta_2, -\eta_1\eta_2).$$

Замечание 2. Совершенно аналогичные формулировки возможны для $n = 1$ и $n = 2$:

группа $\mathrm{Rot}_O(1)$ *изоморфна фактор группе* S^0/S^0, *а группа* $\mathrm{Rot}_O(2)$ — *факторгруппе* S^1/S^0.

Действительно, факторгруппа S^0/S^0 состоит только из одного элемента (единицы), а факторгруппа S^1/S^0 изоморфна группе S^1 (изоморфизм осуществляется соответствием $(e^{i\alpha}, e^{-i\alpha}) \longmapsto e^{2i\alpha}$).

Чтобы получить более «наглядное» представление о многообразии элементов группы $\mathrm{Rot}_O(3)$, мы рассмотрим отображение

$$S^3 \to \mathbb{R}P^3,$$

сопоставляющее каждому кватерниону

$$\eta = y_0 + y_1 i + y_2 j + y_3 k \in S^3$$

точку $(y_0 : y_1 : y_2 : y_3)$ арифметического проективного пространства $\mathbb{R}P^3$ (см. п. 3 § 2 гл. 4).

Пусть $(x_0 : x_1 : x_2 : x_3)$ — произвольная точка пространства $\mathbb{R}P^3$. Положив

$$y_0 = \frac{x_0}{r}, \quad y_1 = \frac{x_1}{r}, \quad y_2 = \frac{x_2}{r}, \quad y_3 = \frac{x_3}{r},$$

где $r = \sqrt{x_0^2 + x_1^2 + x_2^2 + x_3^2}$, мы, очевидно, получим, во-первых, что

$$(x_0 : x_1 : x_2 : x_3) = (y_0 : y_1 : y_2 : y_3),$$

и во-вторых, что

$$y_0^2 + y_1^2 + y_2^2 + y_3^2 = 1.$$

Это показывает, что точка $(x_0 : x_1 : x_2 : x_3)$ получается из кватерниона $\eta = y_0 + y_1 i + y_2 j + y_3 k$, принадлежащего группе S^3. Следовательно, построенное отображение

$$\eta \longmapsto (y_0 : y_1 : y_2 : y_3)$$

является отображением группы S^3 на все пространство $\mathbb{R}P^3$.

Двум кватернионам η_1 и η_2 тогда и только тогда соответствует одна и та же точка пространства $\mathbb{R}P^3$, когда существует такое вещественное число k, что

$$\eta_2 = k\eta_1.$$

Но поскольку $|\eta_2| = |\eta_1| = 1$, это число должно быть равно ± 1.

Это показывает, что отображение

$$S^3 \to \mathbb{R}P^3$$

индуцирует биективное отображение факторгруппы S^3/S^0 на пространство $\mathbb{R}P^3$.

Таким образом, мы получаем, что

группа S^3/S^0, *а следовательно, и группа вращений* $\mathrm{Rot}_O(3)$ *находится в биективном соответствии с проективным пространством* $\mathbb{R}P^3$.

Это утверждение хотя ничего и не говорит об алгебраическом строении группы $\mathrm{Rot}_O(3)$, но зато дает вполне наглядное описание множества ее элементов.

Замечание 3. Обратим внимание на то, что изоморфизм групп S^3/S^0 и $\mathrm{Rot}_O(3)$ зависит от выбора в пространстве некоторой системы прямоугольных координат x, y, z. От этого выбора зависит также и построенный в п. 4 § 1 изоморфизм между группами $\mathrm{Rot}_O(3)$ и $SO(3)$. Композиция этих двух изоморфизмов дает изоморфизм между группами S^3/S^0 и $SO(3)$, уже, как легко видеть, не зависящий от выбора координат (и потому естественный). Это означает, что

группу $SO(3)$ *собственных ортогональных матриц третьего порядка мы можем отождествить с группой* S^3/S^0 (*а следовательно, и с проективным пространством* $\mathbb{R}P^3$).

Замечание 4. Поскольку любое вращение пространства (вокруг точки O) имеет вид R_η, то согласно формуле разложения (1) мы получаем, что

каждое вращение R *пространства вокруг точки* O *может быть разложено в композицию*

$$R = R' \circ R'' \circ R''',$$

где R' *и* R''' *— вращения вокруг оси* Oz, *а* R'' *— вращение вокруг оси* Ox.

Для ортогональных матриц это означает, что

любая собственная ортогональная матрица допускает разложение вида

$$\begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

где $-\pi < \varphi \leqslant \pi$, $0 \leqslant \psi < \pi$, $-\pi < \theta \leqslant \pi$.

Таким образом, мы заново (но «на более высоком уровне») получили найденное в п. 6 § 5 гл. 1 разложение Эйлера.

С другой стороны, заметив, что за оси Oz и Ox мы можем принять любые две перпендикулярные прямые, проходящие через точку O, мы получаем, что

выбрав две перпендикулярные прямые α_1 и α_2, проходящие через точку O, мы каждое вращение R с центром O (элемент группы $\mathrm{Rot}_O(3)$) можем разложить в композицию

$$R = R' \circ R'' \circ R''',$$

где R' и R''' — вращения вокруг прямой α_1 (элементы группы $\mathrm{Rot}_{\alpha_1}(3)$), а R'' — вращение вокруг прямой α_2 (элемент группы $\mathrm{Rot}_{\alpha_2}(3)$).

Замечание 5. Из описания вращений как преобразований, записывающихся формулой

$$\xi' = \eta\xi\bar{\eta},$$

немедленно вытекает следующее предложение, дающее аналогичное описание всех движений:

Предложение 3. *Каждое движение пространства (элемент группы $\mathrm{Ort}^+(3)$) записывается в кватернионной координате ξ формулой*

$$\xi' = \eta\xi\bar{\eta} + \xi_0,$$

где ξ_0 — некоторый мнимый кватернион, а η — некоторый кватернион из S^3 (т. е. такой, что $|\eta| = 1$).

Подчеркнем, что в отличие от аналогичных формул для движений плоскости, это представление движений пространства не единственно (кватернион η может быть заменен кватернионом $-\eta$).

Упражнение. Докажите, что

каждое обращающее ориентации ортогональное преобразование пространства (элемент смежного класса $\mathrm{Ort}^-(3)$) записывается в кватернионной координате ξ формулой

$$\xi' = -\eta\bar{\xi}\bar{\eta} + \xi_0.$$

Это представление также не единственно.

§ 3. КОНФОРМНЫЕ ПРЕОБРАЗОВАНИЯ

1. Инверсия относительно окружности

Точки M и M' плоскости, симметричные относительно некоторой прямой α, обладают, очевидно, тем свойством, что любая окружность, центр которой лежит на прямой α (т. е. любая окружность, ортогональная прямой α) и которая проходит через точку M, проходит также и через точку M'. Обобщением этого утверждения на любые окружности является следующее

Предложение 1. *Для любой точки M, не принадлежащей данной окружности Σ и отличной от ее центра O, существует единственная точка M', обладающая тем свойством, что любая окружность Σ', ортогональная окружности Σ и проходящая через точку M, проходит также и через точку M'.*

Доказательство. Совокупность всех окружностей, ортогональных окружности Σ, является (см. п. 3 § 3 гл. 4) гиперболической связкой окружностей, а совокупность всех окружностей, проходящих через точку M — параболической связкой окружностей. Поэтому совокупность всех окружностей, ортогональных окружности Σ и проходящих через точку M, представляет собой пучок окружностей. Этот пучок не может быть гиперболическим, поскольку все его окружности имеют общую точку M. Он не может быть и параболическим, поскольку тогда ему принадлежала бы окружность нулевого радиуса с центром в точке M, что невозможно (если окружность нулевого радиуса ортогональна окружности Σ, то ее центр должен принадлежать этой окружности, а по условию точка M окружности Σ не принадлежит). Следовательно, этот пучок эллиптичен. Кроме того, он является пучком именно окружностей, а не прямых, поскольку пучок прямых, проходящих через точку M, тогда и только тогда состоит из прямых, ортогональных окружности Σ, когда точка M является центром этой окружности.

Являясь эллиптическим пучком окружностей, рассматриваемый пучок состоит поэтому из всех окружностей, проходящих через две фиксированные точки. Одной из этих точек является данная точка M, а другой — искомая точка M'.

Определение 1. По очевидной аналогии со случаем прямой, точка M' называется точкой, *симметричной точке M относительно окружности* Σ, а преобразование [1])

$$M \longmapsto M'$$

называется *инверсией* (или *симметрией*) относительно окружности Σ.

Центр O окружности Σ называется *центром* инверсии, а ее радиус $R > 0$ — *радиусом* инверсии, квадрат R^2 радиуса инверсии называется *степенью* инверсии.

[1]) Определенное пока лишь для точек M, не принадлежащих окружности Σ и отличных от ее центра.

Одной из окружностей эллиптического пучка, состоящего из окружностей, проходящих через точки M и M', является прямая MM' (прямая центров этого пучка). Поскольку эта прямая ортогональна окружности Σ, она проходит через ее центр. Таким образом,

точки M и M', симметричные относительно окружности Σ, лежат на прямой, проходящей через центр O этой окружности.

Пусть Σ' — произвольная окружность, проходящая через точки M и M'. Степень точки O относительно окружности Σ' равна (см. п. 1 § 3 гл. 4) произведению величин векторов $\overrightarrow{OM}$ и $\overrightarrow{OM'}$. Но поскольку окружность Σ' ортогональна окружности Σ, эта степень равна квадрату R^2 радиуса окружности Σ и, в частности, положительна. Следовательно,

точка M' лежит на прямой OM по ту же сторону от точки O, что и точка M.

Кроме того, мы видим, что

произведение длин отрезков $\overline{OM}$ и $\overline{OM'}$ равно R^2:

$$|OM| \cdot |OM'| = R^2.$$

Ясно, что последние два свойства однозначно определяют на прямой OM точку M'. Поэтому их можно положить в основу определения этой точки. Заметим, что получающееся определение пригодно и тогда, когда точка M принадлежит окружности Σ, и дает в этом случае ту же самую точку M.

Таким образом, мы приходим к следующему окончательному определению инверсии относительно окружности Σ радиуса R с центром в точке O:

точка M' симметрична точке M относительно окружности Σ, если

1) *она расположена на прямой OM по ту же сторону от точки O, что и точка M;*

2) *имеет место равенство*

$$|OM| \cdot |OM'| = R^2.$$

Тем самым инверсия относительно окружности оказывается определенной для всех точек плоскости, отличных от точки O.

При этом $M' = M$ тогда и только тогда, когда M принадлежит окружности Σ.

Ясно, что аналогично симметрии относительно прямых,

каждая инверсия S является инволютивным преобразованием, т. е. обратное преобразование S^{-1} совпадает с S:

$$S^{-1} = S.$$

Пусть x, y — прямоугольные координаты на плоскости. Найдем формулу, выражающую координаты точки $M'(x', y')$ через

координаты точки $M(x, y)$. Для этого оказывается удобным перейти к соответствующей комплексной координате

$$z = x + iy.$$

Рассмотрим сначала случай, когда центр инверсии совпадает с началом координат $z = 0$. Пусть R — радиус инверсии. По определению, произвольную точку z инверсия переводит в точку z', обладающую следующими свойствами:

1) отношение $z' : z$ является положительным числом;

2) имеет место равенство

$$|z| \cdot |z'| = R^2.$$

Ясно, что этим условиям удовлетворяет точка

$$z' = \frac{R^2}{|z|^2} z = \frac{R^2}{\bar{z}}. \tag{1}$$

Поскольку условия 1) и 2) однозначно определяют точку z', тем самым доказано, что

инверсия с центром в точке $z = 0$ задается формулой (1), *где R^2 — степень инверсии.*

В координатах x, y формула (1) превращается в две формулы:

$$x' = \frac{R^2}{x^2 + y^2} x, \quad y' = \frac{R^2}{x^2 + y^2} y.$$

Чтобы получить аналогичную формулу для инверсии с центром в произвольной точке z_0, достаточно перейти к новой координате $\zeta = z - z_0$. Согласно формуле (1) рассматриваемая инверсия выражается в координате ζ формулой

$$\zeta' = \frac{R^2}{\bar{\zeta}}.$$

Возвращаясь к прежней координате z, мы получаем отсюда, что *инверсия с центром в точке z_0 выражается формулой*

$$z' = \frac{R^2}{\bar{z} - \bar{z}_0} + z_0, \tag{2}$$

где R^2 — степень инверсии.

2. Пополненная плоскость

Тот факт, что инверсия с центром O не определена в точке O, очень неприятен. Естественно добиваться того, чтобы этот недостаток был устранен. С этой целью достаточно пополнить плоскость еще одной фиктивной точкой M_∞ и по определению считать, что инверсия с центром в O переводит точку O в точку M_∞ (и, наоборот, точку M_∞ в точку O).

Определение 1. Плоскость, пополненную указанным образом одной точкой M_∞, мы будем называть *пополненной плоскостью*[1]). Добавленную точку M_∞ мы будем называть *бесконечно удаленной точкой*, а все остальные ее точки — *конечными точками.*

По определению, инверсия с произвольным центром является на пополненной плоскости всюду определенным инволютивным преобразованием.

Любое преобразование плоскости (например, ортогональное или аффинное) мы будем считать преобразованием пополненной плоскости, переводящим точку M_∞ в себя.

Следует отчетливо понимать различие между пополненной и расширенной (евклидово-проективной) плоскостями: первая получается добавлением к обычной (евклидовой) плоскости одной точки, а вторая — добавлением бесконечного числа точек (составляющих целую прямую). В том, что к обычной плоскости можно по-разному добавлять фиктивные точки, нет ничего удивительного: все определяется свойствами тех преобразований, которые мы хотим сделать всюду определенными добавлением фиктивных точек. На евклидовой (или аффинной) плоскости проективное преобразование не определено на целой прямой, и поэтому, чтобы сделать любое такое преобразование определенным всюду, надо добавить прямую. Инверсия же не определена только в одной точке, и потому надо добавлять одну точку. Вполне может быть, что для каких-нибудь других преобразований придется пополнять плоскость, скажем, парой прямых или парой точек.

Пусть α — произвольная прямая, проходящая через центр O некоторой инверсии S. По определению, для любой точки M этой прямой (отличной от точки O) точка $M' = S(M)$ также принадлежит прямой α. На этом основании целесообразно считать, что точка M_∞ также принадлежит прямой α. Поскольку точка O может быть любой (конечной) точкой, следует, таким образом, считать, что *точка M_∞ принадлежит всем прямым пополненной плоскости.*

Таким образом, подобно расширенной плоскости пополненная плоскость не имеет непересекающихся прямых. Однако, в отличие от расширенной плоскости, две прямые на пополненной плоскости, вообще говоря, имеют две общие точки (общую конечную точку и точку M_∞), хотя возможен случай, когда они имеют только одну общую точку M_∞ (параллельные прямые). В этом отношении прямые на пополненной плоскости напоминают не столько прямые на расширенной плоскости, сколько окружности на обычной (евклидовой) плоскости. Действительно, две окружности имеют, вообще говоря, две общие (собственные) точки (возможно, мнимые), за исключением того случая, когда они касаются. Это наводит на мысль, что *в геометрии пополненной плоскости прямые следует рассматривать как окружности, проходящие через точку M_∞.*

Это подкрепляется тем, что, как мы ниже покажем, любая инверсия с центром в точке O каждую окружность перево-

[1]) Используется также термин *«круговая плоскость»*.

дит, вообще говоря, в некоторую окружность; однако окружности, проходящие через точку O, эта инверсия переводит в прямые.

Стоит также напомнить, что и в теории пучков окружностей (§ 3 гл. 4) оказалось удобным рассматривать прямые как окружности «бесконечного радиуса».

По всем этим основаниям мы при изучении геометрии пополненной плоскости будем, как правило, избегать термина «прямая», заменяя его оборотом «окружность, проходящая через точку M_∞».

Тогда можно будет уже без всяких исключений утверждать, например, что инверсия переводит окружности в окружности. Делается также совершенно ясно, почему пучок прямых, проходящих через произвольную точку O, следует считать эллиптическим пучком окружностей: действительно, он состоит из окружностей, проходящих через две фиксированные точки O и M_∞.

В этом круге идей следует считать, что параллельные прямые касаются друг друга в точке M_∞. Действительно, при таком соглашении пучок параллельных прямых окажется пучком окружностей, проходящих через точку M_∞ и касающихся в этой точке друг друга, т. е. попадет под общее определение параболических пучков.

Наглядную интерпретацию пополненной плоскости можно получить, рассмотрев в (евклидовом) пространстве произвольную сферу и касающуюся ее в некоторой точке плоскость. Пусть O — точка сферы, диаметрально противоположная точке касания. Тогда для любой точки M плоскости прямая OM пересекает сферу в единственной точке M' (называемой *стереографической проекцией* точки M). Это показывает, что точки плоскости мы можем отождествить с точками сферы, отличными от точки O.

После этого отождествления необходимость в пополненной плоскости делается очевидной: добавление одной и только одной точки нужно, чтобы ликвидировать в сфере получившийся «прокол». Одновременно мы видим, что

пополненную плоскость можно отождествить со сферой.

Эта интерпретация пополненной плоскости как сферы часто бывает полезна. Например, в этой интерпретации совершенно ясно, почему мы должны считать параллельные прямые касающимися в точке M_∞: образы этих прямых при стереографической проекции являются окружностями, касающимися друг друга в точке O.

Комплексную координату

$$z = x + iy$$

мы распространим на пополненую плоскость, считая, что точке M_∞ отвечает «значение» координаты z, равное ∞.

В силу этого соглашения,

формулы (1), (2) п. 1 *будут выражать инверсию и на пополненной плоскости.*

Действительно, при $z = 0$ формула (1) дает $z' = \infty$, а при $z = \infty$ — дает $z' = 0$.

Мы знаем, что любое движение плоскости записывается в комплексной координате формулой

$$z' = az + b, \tag{1}$$

где a, b — комплексные числа, причем $|a| = 1$ (см. п. 3 § 2). При любом $a \neq 0$ преобразование (1) является композицией движения

$$z' = \frac{a}{|a|} z + b$$

и гомотетии

$$z' = |a| z$$

с центром в точке O, т. е. является *преобразованием подобия* (сохраняющим ориентации).

Поскольку в силу формулы (1) $z' = \infty$ при $z = \infty$, эта формула выражает преобразование подобия (а при $|a| = 1$ — движение) и на пополненной плоскости.

Аналогично, формула

$$z' = \bar{z},$$

выражающая, очевидно, симметрию относительно оси Ox, сохраняет силу и на пополненной плоскости.

Определение 2. Преобразование пополненной плоскости называется *дробно-линейным преобразованием*, если оно выражается (в данной координате z) формулой вида

$$z' = \frac{az + b}{cz + d}, \tag{2}$$

и называется *сопряженно дробно-линейным преобразованием*, если оно выражается формулой вида

$$z' = \frac{a\bar{z} + b}{c\bar{z} + d}. \tag{2'}$$

Конечно, здесь следует проверить корректность этого определения, т. е. тот факт, что преобразование, дробно-линейное (сопряженно дробно-линейное) в координате z, остается дробно-линейным (сопряженно дробно-линейным) и в любой другой комплексной координате.

Задание. Проверьте корректность определения 2.

Чтобы формулы (2) и (2′) действительно задавали преобразования (биективные отображения пополненной плоскости на себя), необходимо (и достаточно), чтобы

$$\Delta = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0.$$

При выполнении этого условия преобразованием, обратным, скажем, к преобразованию (2), будет преобразование

$$z' = \frac{dz - b}{cz - a}.$$

Все рассмотренные выше преобразования были либо дробно-линейными, либо сопряженно дробно-линейными преобразованиями.

Например, преобразования подобия (1) являются преобразованиями (2), получающимися при $c = 0$, $d = 1$, а инверсия с центром $z = 0$ — преобразованием (2′), получающимся при $a = 0$, $b = R^2$, $c = 1$, $d = 0$.

Оказывается, что и наоборот,

каждое дробно-линейное или сопряженно дробно-линейное преобразование является композицией преобразований, являющихся преобразованиями подобия или симметриями (относительно окружностей или прямых).

Действительно, каждое преобразование (2) при $c = 0$ является, очевидно, преобразованием подобия. Пусть $c \neq 0$. Тогда

$$\frac{az + b}{cz + d} = \frac{bc - ad}{c(cz + d)} + \frac{a}{c} = -\frac{\Delta}{c^2}\left(\frac{1}{z + \frac{c}{d}}\right) + \frac{a}{c}$$

и потому преобразование (2) является композицией преобразования

$$z' = \frac{1}{z + \frac{c}{d}} \tag{3}$$

и преобразования подобия

$$z' = -\frac{\Delta}{c^2} z + \frac{a}{c}.$$

С другой стороны, преобразование (3) является, очевидно, композицией параллельного переноса

$$z' = z + \frac{c}{d}$$

и преобразования

$$z' = \frac{1}{z},$$

являющегося композицией симметрии $z' = \bar{z}$ и инверсии $z' = \frac{1}{\bar{z}}$.

Подобным же образом может быть разложено и любое преобразование (2′) (с $c \neq 0$): к указанным преобразованиям добавится еще одна симметрия $z' = \bar{z}$.

Определение 3. Преобразование пополненной плоскости называется *конформным*, если оно переводит окружности в окружности.

Замечание 1. Конформные преобразования называются также *круговыми преобразованиями* или *преобразованиями Мёбиуса.*

Ясно, что

все конформные преобразования пополненной плоскости составляют группу.

Мы будем обозначать эту группу символом Conf.

Легко видеть, что

уравнение любой окружности на пополненной плоскости имеет вид

$$az\bar{z} + \bar{b}z + b\bar{z} + c = 0, \tag{4}$$

где b — некоторое комплексное число, а a и c — вещественные числа.

Действительно, если рассматриваемая окружность имеет центр в точке z_0 и радиус R, то ее уравнение может быть записано в виде

$$|z - z_0| = R.$$

Возводя в квадрат и перемножая, мы получаем уравнение

$$z\bar{z} - \bar{z}_0 z - z_0\bar{z} + z_0\bar{z}_0 - R^2 = 0, \tag{5}$$

имеющее вид (4) с $a = 1$, $b = -z_0$ и $c = z_0\bar{z}_0 - R^2$.

Если же рассматриваемая окружность является прямой

$$Ax + By + C = 0,$$

то, положив $a = 0$, $b = \dfrac{A - iB}{2}$, $c = C$, мы снова получим уравнение вида (4).

Обратно, покажем, что

любое уравнение

$$az\bar{z} + \bar{b}z + b\bar{z} + c = 0, \tag{6}$$

где a и c вещественны, является уравнением некоторой окружности.

Действительно, при $a = 0$ это — уравнение прямой

$$Ax + By + C = 0,$$

где $A = \operatorname{Re} b$, $B = -\operatorname{Im} b$ и $C = c$. Пусть $a \neq 0$. Разделив на a и положив

$$z_0 = -\frac{b}{a}, \quad R = \frac{\sqrt{b\bar{b} - ac}}{a},$$

мы получим уравнение вида (5), показывающее, что рассматриваемая линия является окружностью радиуса R с центром в точке z_0.

Замечание 2. Настоящую окружность мы получим только при $b\bar{b} > ac$. При $b\bar{b} < ac$ получается окружность мнимого радиуса, не имеющая вещественных точек, а при $b\bar{b} = ac$ — окружность нулевого радиуса, имеющая только одну вещественную точку.

Инверсия

$$z' = \frac{R^2}{\bar{z}}$$

с центром в точке $z = 0$ переводит окружность

$$az\bar{z} + \bar{b}z + b\bar{z} + c = 0$$

в линию с уравнением

$$a\frac{R^2}{z'}\frac{R^2}{\bar{z}'} + \bar{b}\frac{R^2}{\bar{z}'} + b\frac{R^2}{z'} + c = 0,$$

т. е. с уравнением

$$cz'\bar{z}' + \bar{b}R^2z' + bR^2\bar{z}' + aR^4 = 0.$$

Поскольку это уравнение имеет вид (4), оно является уравнением некоторой окружности.

Таким образом, инверсия переводит произвольную окружность снова в окружность, т. е. является конформным преобразованием. (Ясно, что это справедливо и для инверсии с центром в произвольной точке z_0.)

Поскольку преобразования подобия и симметрии относительно прямых очевидным образом являются конформными преобразованиями, тем самым доказано, что

каждое дробно-линейное или сопряженно дробно-линейное преобразование является конформным преобразованием.

Замечательно, что справедливо и обратное утверждение:

Предложение 1. *Каждое конформное преобразование пополненной плоскости является дробно-линейным или сопряженно дробно-линейным преобразованием.*

Доказательство. Покажем сначала, что

любое конформное преобразование пополненной плоскости, оставляющее на месте точку M_∞, является преобразованием подобия (возможно, не сохраняющим ориентации, т. е. представляющим собой композицию некоторого сохраняющего ориентации преобразования подобия и симметрии).

Действительно, если конформное преобразование оставляет на месте точку M_∞, то любую окружность, проходящую через эту точку, оно переводит снова в окружность, проходящую через точку M_∞. Другими словами, любую прямую оно переводит в прямую. Следовательно, это конформное преобразование является аффинным преобразованием, и потому разлагается в композицию $\Phi_2 \circ \Phi_1$ некоторого ортогонального преобразования Φ_1 и преобразования Φ_2, являющегося композицией двух сжатий к двум взаимно перпендикулярным прямым. Поскольку преобразования $\Phi_2 \circ \Phi_1$ и Φ_1 конформны, преобразование Φ_2 также конформно. Но ясно, что композиция двух сжатий тогда и только тогда переводит окружности в окружности (является

конформным преобразованием), когда коэффициенты этих сжатий одинаковы, т. е. когда эта композиция является гомотетией. Таким образом, данное конформное отображение является композицией некоторого ортогонального преобразования и гомотетии, т. е. представляет собой преобразование подобия.

Пусть теперь Ω — произвольное конформное преобразование плоскости. Пусть это преобразование переводит точку M_∞ в (конечную) точку O. Рассмотрим произвольную инверсию S с центром в точке O. Композиция

$$\Omega' = S \circ \Omega$$

является конформным преобразованием, оставляющим на месте точку M_∞, и потому, по доказанному, представляет собой преобразование подобия. Поскольку $\Omega = S \circ \Omega'$, тем самым доказано, что

любое конформное преобразование пополненной плоскости является либо преобразованием подобия, либо композицией преобразования подобия и некоторой инверсии.

Чтобы завершить доказательство предложения 1, осталось заметить, что если два преобразования выражаются формулами вида (2) или (2′), то их композиция также выражается формулой такого вида.

3. Свойства конформных преобразований

Предложение 1 п. 2 сводит изучение конформных преобразований к простым алгебраическим вычислениям с дробно-линейными и сопряженно дробно-линейными преобразованиями.

Рассмотрим сначала конформные преобразования, являющиеся дробно-линейными преобразованиями:

$$z' = \frac{az + b}{cz + d}. \tag{1}$$

Такое преобразование полностью определено, если известна матрица

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{2}$$

Поскольку эта матрица невырождена, она принадлежит группе $\mathrm{GL}(\mathbb{C}^2)$ всех невырожденных квадратных матриц второго порядка с комплексными элементами.

Таким образом, сопоставив матрице (2) преобразование (1), мы получим некоторое отображение

$$\mathrm{GL}(\mathbb{C}^2) \to \mathrm{Conf} \tag{3}$$

матричной группы $\mathrm{GL}(\mathbb{C}^2)$ в группу Conf.

Легко видеть, что

отображение (3) *является гомоморфизмом, т. е. произведению матриц отвечает композиция соответствующих преобразований* (1).

Задание. Докажите это утверждение.

Отсюда, в частности, вытекает, что

совокупность Conf^+ *всех конформных преобразований, являющихся дробно-линейными преобразованиями* (*т. е. записывающихся формулами вида* (1)), *является подгруппой группы* Conf.

Действительно, эта совокупность является образом группы $\mathrm{GL}(\mathbb{C}^2)$ при гомоморфизме (3).

Было бы неверно утверждать, что гомоморфизм (3) является изоморфизмом группы $\mathrm{GL}(\mathbb{C}^2)$ на группу Conf^+. Действительно, ясно, что матрицы, отличающиеся скалярным множителем, определяют одно и то же преобразование (1). Покажем, что обратное тоже верно, т. е. что

равенство

$$\frac{a_1z+b_1}{c_1z+d_1}=\frac{a_2z+b_2}{c_2z+d_2}$$

тогда и только тогда имеет место для всех z, *когда существует такое число* $k\neq 0$, *что*

$$\begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} = k\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}.$$

Действительно, если

$$\frac{a_1z+b_1}{c_1z+d_1}=\frac{a_2z+b_2}{c_2z+d_2}$$

для всех z, то

$$(a_1z+b_1)(c_2z+d_2)-(c_1z+d_1)(a_2z+b_2)=0,$$

т. е.

$$(a_1c_2-c_1a_2)z^2+[(a_1d_2-d_1a_2)+(b_1c_2-c_1b_2)]z+(b_1d_2-d_1b_2)=0$$

тождественно по z. Но многочлен тогда и только тогда тождественно равен нулю, когда равны нулю все его коэффициенты. Следовательно,

$$a_1c_2-c_1a_2=0,\quad b_1d_2-d_1b_2=0$$

и

$$a_1d_2-d_1a_2=c_1b_2-b_1c_2. \tag{4}$$

Первые два соотношения означают, что существуют такие числа k и l, что

$$a_2=ka_1,\quad c_2=kc_1$$

и

$$b_2=lb_1,\quad d_2=ld_1.$$

Подставляя эти выражения в соотношение (4), мы получаем равенство

$$la_1d_1 - kd_1a_1 = lc_1b_1 - lb_1c_1,$$

т. е. равенство

$$(k - l)(a_1d_1 - c_1b_1) = 0,$$

возможное только при $k = l$ (напомним, что, по условию, $a_1d_1 - c_1b_1 \neq 0$).

На языке теории групп доказанное утверждение означает, что

ядром гомоморфизма (3) *является подгруппа всех диагональных матриц вида*

$$\begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix},$$

где $k \neq 0$.

Можно попытаться уменьшить это ядро, требуя, чтобы определитель матрицы (2) был равен единице:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = 1. \tag{5}$$

Это означает, что вместо группы $GL(\mathbb{C}^2)$ всех невырожденных матриц второго порядка с комплексными элементами мы рассматриваем ее подгруппу $SL(\mathbb{C}^2)$, состоящую из матриц, удовлетворяющих условию (5).

Соответствующий гомоморфизм

$$SL(\mathbb{C}^2) \to \mathrm{Conf}$$

также отображает группу $SL(\mathbb{C}^2)$ на группу Conf^+. Однако он по-прежнему не будет изоморфизмом, поскольку матрицы, отличающиеся знаком, будут определять одно и то же преобразование из группы $SL(\mathbb{C}^2)$. Другими словами, ядром гомоморфизма (3) будет группа второго порядка, состоящая из матриц

$$E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad -E = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Обозначив эту группу символом $\{\pm E\}$, мы можем, таким образом, сказать, что

группа Conf^+ *изоморфна факторгруппе группы* $SL(\mathbb{C}^2)$ *по нормальному делителю* $\{\pm E\}$.

Запись конформного преобразования из группы Conf^+ формулой (1), в которой

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = 1,$$

называется *нормированной*. Подчеркнем, что, согласно сказанному выше, для преобразования (1) существуют д в е нормированные записи. От этой неоднозначности избавиться невозможно, и о ней всегда нужно помнить (в противном случае легко прийти к ошибке).

Обратим внимание на то, что запись (сохраняющего ориентации) преобразования подобия в виде

$$z' = az + b$$

не нормирована. Нормированная запись этого преобразования имеет вид

$$z' = \frac{\sqrt{a}\, z + b/\sqrt{a}}{1/\sqrt{a}}.$$

Рассмотрим теперь конформные преобразования, выражающиеся формулами

$$z' = \frac{a\bar{z} + b}{c\bar{z} + d}, \quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0.$$

Их совокупность мы обозначим символом Conf^-.

Каждое такое преобразование Ω является композицией симметрии

$$S\colon\ z' = \bar{z}$$

и преобразования

$$\Omega^+\colon\ z' = \frac{az + b}{cz + d},$$

принадлежащего группе Conf^+:

$$\Omega = \Omega^+ \circ S.$$

При этом ясно, что

$$\Omega^+ \circ S = S \circ \bar{\Omega}^+,$$

где $\bar{\Omega}^+$ — преобразование из группы Conf^+, выражающееся формулой

$$z' = \frac{\bar{a}z + \bar{b}}{\bar{c}z + \bar{d}}.$$

Из этих формул вытекает, что для любых двух преобразований $\Omega_1^+,\ \Omega_2^+ \in \mathrm{Conf}^+$ имеют место равенства

$$(\Omega_1^+ \circ S) \circ (\Omega_2^+ \circ S) = (\Omega_1^+ \circ S) \circ (S \circ \bar{\Omega}_2^+) = \Omega_1^+ \circ \bar{\Omega}_2^+,$$

$$(\Omega_1^+ \circ S) \circ \Omega_2^+ = (\Omega_1^+ \circ \bar{\Omega}_2^+) \circ S,$$

$$\Omega_1^+ \circ (\Omega_2^+ \circ S) = (\Omega_1^+ \circ \Omega_2^+) \circ S.$$

Поскольку представление преобразования Ω из Conf^- в виде $\Omega^+ \circ S$, очевидно, единственно, мы видим, что произведение любых двух элементов группы

$$\mathrm{Conf} = \mathrm{Conf}^+ \cup \mathrm{Conf}^-$$

мы можем вычислить, если нам известно произведение любых двух элементов группы Conf^+.

Тем самым алгебраическое строение группы Conf полностью выяснено (или, точнее, сведено к алгебраическому строению группы Conf^+).

В частности, мы видим, что

множество Conf^- *является смежным классом группы* Conf *по ее нормальному делителю* Conf^+.

Рассмотрим теперь некоторые «геометрические» свойства конформных преобразований.

Предложение 1. *Для любых двух троек* z_1, z_2, z_3 *и* z_1', z_2', z_3' *точек пополненной плоскости, не содержащих одинаковых точек, существует конформное преобразование из группы* Conf^+, *переводящее точки* z_1, z_2, z_3 *в точки* z_1', z_2', z_3'.

Доказательство. Предполагая сначала, что ни одна из данных шести точек не является бесконечно удаленной точкой $z = \infty$, рассмотрим соотношение

$$\frac{(z' - z_1')(z_2' - z_3')}{(z' - z_3')(z_2' - z_1')} = \frac{(z - z_1)(z_2 - z_3)}{(z - z_3)(z_2 - z_1)}. \tag{6}$$

Выразив из этого соотношения z' через z, мы, очевидно, получим формулу вида

$$z' = \frac{az + b}{cz + d},$$

показывающую, что соответствие $z \longmapsto z'$ является конформным преобразованием из группы Conf^+.

Это преобразование переводит точки z_1, z_2, z_3 в точки z_1', z_2', z_3', так как обе стороны соотношения (6) обращаются в нуль при $z' = z_1'$ и $z = z_1$, в единицу при $z' = z_2'$ и $z = z_2$ и в бесконечность при $z' = z_3'$ и $z = z_3$.

В случае, когда одна из заданных точек является бесконечно удаленной точкой, нужно написать соотношение, аналогичное соотношению (6) и получающееся из него предельным переходом, когда соответствующая точка $z_1, z_2, z_3, z_1', z_2', z_3'$ «уходит в бесконечность». Так, если $z_1 = \infty$, то правую часть соотношения (6) нужно заменить выражением

$$\frac{z_2 - z_3}{z - z_3},$$

если $z_2 = \infty$ — то выражением

$$\frac{z - z_1}{z - z_3},$$

а если $z_3 = \infty$ — то выражением

$$\frac{z - z_1}{z_2 - z_1}.$$

Если одна из точек z_1', z_2', z_3' является точкой ∞, то аналогичные замены нужно произвести в левой части соотношения (6).

Тем самым предложение 1 полностью доказано.

Чтобы ответить на вопрос об единственности построенного конформного преобразования, нам будет удобно рассмотреть предварительно *неподвижные точки* произвольного преобразования

$$z' = \frac{az + b}{cz + d} \tag{7}$$

из группы Conf^+, т. е. точки, оставляемые этим преобразованием на месте.

Эти точки определяются из уравнения

$$z = \frac{az + b}{cz + d},$$

т. е. из уравнения

$$cz^2 + (d - a)\, z - b = 0, \tag{8}$$

называемого *уравнением неподвижных точек.*

При $c \neq 0$ уравнение (8) имеет либо два различных корня, либо один двойной корень. При $c = 0$ и $d \neq a$ уравнение (8) имеет один корень $\frac{b}{d - a}$. Однако в этом случае преобразование (7) является преобразованием подобия и, по определению, оставляет на месте еще точку M_∞. Поэтому целесообразно считать, что при $c = 0$ уравнение (8) имеет также корень $z = \infty$. Наконец, при $c = 0$ и $d = a$ (т. е. в случае, когда преобразование (7) является параллельным переносом) уравнение (8) вообще не имеет конечных корней (конечно, если $b \neq 0$, т. е. если преобразование (7) не тождественно). В этом случае мы будем условно считать, что оно имеет двойной корень $z = \infty$.

Резюмируя, мы получаем, что

каждое конформное преобразование из группы Conf^+, *не являющееся тождественным преобразованием, имеет либо две неподвижные точки, либо одну двойную неподвижную точку.*

В частности,

если преобразование из группы Conf^+ *оставляет на месте три различные точки, то оно является тождественным преобразованием.*

Отсюда уже непосредственно вытекает ответ на вопрос об единственности конформного преобразования, переводящего три данные точки в три данные точки:

Предложение 2. *Группа* Conf^+ *содержит только одно преобразование, переводящее три различные точки* z_1, z_2, z_3 *в три различные точки* z_1', z_2', z_3'.

Доказательство. Если существует два таких преобразования Ω_1 и Ω_2, то преобразование $\Omega_1^{-1} \circ \Omega_2$ оставляет на месте три различные точки z_1, z_2, z_3 и, следовательно, $\Omega_1^{-1} \circ \Omega_2 = E$, что противоречит условию $\Omega_1 \neq \Omega_2$.

Аналогичные результаты имеют место и для преобразований из смежного класса Conf^-.

Предложение 3. *Для любых двух троек точек* z_1, z_2, z_3 *и* z_1', z_2', z_3' *пополненной плоскости, не содержащих одинаковых точек, существует в смежном классе* Conf^- *единственное преобразование, переводящее точки* z_1, z_2, z_3 *в точки* z_1', z_2', z_3'.

Доказательство. Ясно, что

через любые три различные точки пополненной плоскости проходит единственная окружность.

Действительно, в случае, когда данные точки конечны и не лежат на одной прямой, это утверждение известно из элементарной геометрии. Если же эти точки лежат на одной прямой, или если одна из них является бесконечно удаленной точкой, то искомой окружностью будет прямая, проходящая через эти точки.

В частности, через данные точки z_1, z_2, z_3 проходит некоторая окружность Σ. Пусть S — инверсия относительно этой окружности (симметрия, если окружность Σ является прямой) и пусть Ω^+ — преобразование из группы Conf^+, переводящее точки z_1, z_2, z_3 в точки z_1', z_2', z_3'. Рассмотрим преобразование

$$\Omega = \Omega^+ \circ S.$$

Являясь композицией преобразования S из смежного класса Conf^- и преобразования Ω^+ из группы Conf^+, это преобразование принадлежит смежному классу Conf^-. Кроме того, оно переводит точки z_1, z_2, z_3 в точки z_1', z_2', z_3', поскольку инверсия S оставляет точки z_1, z_2, z_3 на месте. Тем самым существование искомого преобразования доказано.

Пусть теперь существует два преобразования Ω_1 и Ω_2 из смежного класса Conf^-, переводящих точки z_1, z_2, z_3 в точки z_1', z_2', z_3'. Рассмотрим преобразование

$$\Omega_1^{-1} \circ \Omega_2.$$

Это преобразование принадлежит группе Conf^+ и оставляет на месте точки z_1, z_2, z_3. Поэтому оно является тождественным преобразованием, и, следовательно, $\Omega_1 = \Omega_2$.

Тем самым предложение 3 полностью доказано.

Подчеркнем, что в группе Conf существует, таким образом, два преобразования, переводящих точки z_1, z_2, z_3 в точки z_1', z_2', z_3': одно — из группы Conf^+, а другое — из смежного класса Conf^-.

Не нужно думать, что доказанное выше утверждение о неподвижных точках преобразований из группы Conf^+ остается справедливым для преобразований из смежного класса Conf^-. Напротив, легко видеть, что

преобразование

$$z' = \frac{a\bar{z} + b}{c\bar{z} + d}$$

из смежного класса Conf^- *может иметь бесконечно много неподвижных точек, заполняющих целую окружность.*

Действительно, уравнение неподвижных точек для такого преобразования имеет вид:

$$cz\bar{z} + dz - a\bar{z} - b = 0 \tag{9}$$

и потому при вещественных c и b и при $d = -\bar{a}$ является уравнением некоторой окружности.

Замечание 1. В общем случае уравнение (9) распадается на два уравнения, каждое из которых является уравнением окружности (или удовлетворяется тождественно). Действительно, положим

$$c = c_1 + ic_2, \quad b = -b_1 - ib_2, \quad d = \bar{a}_1 + i\bar{a}_2, \quad a = -a_1 - ia_2,$$

где c_1, c_2, b_1, b_2 — вещественные числа, а a_1 и a_2 — комплексные. Подставив эти выражения в уравнение (9) и отделив вещественную и мнимую части, мы, очевидно, получим уравнения

$$\begin{aligned} c_1 z\bar{z} + \bar{a}_1 z + a_1\bar{z} + b_1 = 0, \\ c_2 z\bar{z} + \bar{a}_2 z + a_2\bar{z} + b_2 = 0, \end{aligned} \tag{10}$$

каждое из которых является уравнением окружности или удовлетворяется тождественно.

Таким образом, каждое из уравнений (10) определяет либо пустое множество точек (окружность мнимого радиуса), либо точку (окружность нулевого радиуса), либо окружность (или прямую), либо всю плоскость. При этом в случае, когда оба уравнения (10) определяют окружности, эти окружности могут не пересекаться, касаться, пересекаться в двух точках и, наконец, совпадать. Кроме того, случай, когда оба уравнения (10) удовлетворяются тождественно, очевидно, невозможен (поскольку все коэффициенты a, b, c и d не могут одновременно

быть равны нулю). Следовательно, всего возможно 21 различных случаев. В одиннадцати из этих случаев неподвижных точек нет, в шести имеется одна неподвижная точка, в одном — две неподвижные точки и, наконец, в трех случаях неподвижные точки заполняют целую окружность. Тем самым доказано, что

преобразование из смежного класса Conf^- *либо не имеет неподвижных точек, либо имеет одну или две неподвижные точки, либо — целую окружность неподвижных точек.*

4. Конформные преобразования и ориентации

Пополненную плоскость мы будем считать *ориентированной*, если ориентирована исходная евклидова плоскость. Таким образом, в частности, пополненная плоскость, отнесенная к комплексной координате z, автоматически ориентирована.

Основная цель этого пункта — переформулировать для пополненной плоскости определение ориентации таким образом, чтобы было удобно исследовать поведение ориентаций при конформных преобразованиях. Обычное определение ориентации (см. п. 1 § 4 гл. 1) как, скажем, класса одноименных неколлинеарных троек точек для такого исследования плохо приспособлено, поскольку конформное преобразование неколлинеарную тройку точек может перевести в коллинеарную.

Мы сначала обсудим понятие ориентации пополненной плоскости на неформальном, наглядном, уровне, а затем дадим формальное определение.

Принципиальное отличие окружностей от прямых на евклидовой плоскости состоит в том (см. пп. 1 и 4 § 4 гл. 1), что ориентация окружности [1]) определяет ориентацию плоскости, а ориентация прямой — не определяет. Как мы знаем, это связано с тем, что две стороны прямой совершенно равноправны, а стороны окружности — нет. Однако на пополненной плоскости по отношению к конформным преобразованиям стороны окружности совершенно равноправны, ибо, например, инверсия относительно окружности переводит одну ее сторону в другую (подобно тому как симметрия относительно прямой переставляет ее стороны). Поэтому на пополненной плоскости мы для каждой окружности Σ должны специально указывать ее сторону.

Если теперь окружность Σ ориентирована (т. е. на ней указано некоторое направление движения), то мы сопоставим этой ориентации правую ориентацию плоскости (ориентацию «против часовой стрелки», см. п. 2 § 4 гл. 1), если при движении по окружности в данном направлении выбранная сторона остается слева, и левую ориентацию плоскости (ориентацию «по часовой стрелке»), если при этом движении выбранная сторона

[1]) То есть области, на которые окружность разбивает плоскость.

остается справа. Обратно, если на плоскости выбрана некоторая ориентация o, то мы можем сопоставить ей ориентацию окружности Σ, обладающую тем свойством, что при движении по окружности в положительном (относительно этой ориентации) направлении данная сторона остается слева, если ориентация o — правая, и справа, если ориентация o — левая. Наконец, задание ориентации o плоскости и ориентации окружности Σ определяет некоторую сторону окружности, а именно, сторону, которая при движении по окружности в положительном направлении остается слева, если ориентация o — правая, и остается справа, если ориентация o — левая.

Таким образом,

из трех объектов:

1) *ориентация плоскости*,

2) *ориентация окружности*,

3) *сторона окружности*,

любые два однозначно определяют третий.

Замечание 1. Все это справедливо как для «настоящих» окружностей, так и для прямых («окружностей, проходящих через точку M_∞»). При этом ясно, что для прямых мы получаем в точности результаты п. 4 § 3 гл. 1. Кроме того, если для окружности мы выберем сторону, содержащую ее центр, то получающееся соответствие между ориентациями плоскости и окружности будет совпадать с соответствием, описанным в п. 1 § 4 гл. 1 (поскольку при движении по окружности, скажем, против часовой стрелки ее центр остается слева).

Окружности и прямые отличаются еще и в том отношении, что для задания ориентации на прямой нам нужны не три точки A, B, C, а только две точки A, B, такие, что $A \prec B$ в данной ориентации. Однако фактически мы и на прямой имеем дело с тремя точками: третьей точкой является бесконечно удаленная точка M_∞, и ориентация, в которой $A \prec B$, может быть описана как направление на прямой, в котором точка B разделяет точки A и M_∞. Иными словами, задавая направление на прямой двумя точками A и B, мы, на самом деле, дополнительно требуем, чтобы при движении от точки A к точке B в этом направлении мы не проходили через точку M_∞. Если мы от этого последнего условия откажемся (подобно тому, как мы отказались выбирать сторону окружности обязательно там, где расположен ее центр), то для задания ориентации на прямой нам также понадобятся три точки A, B, C (одна из которых может быть точкой M_∞).

Подчеркнем, что если, например, точка C лежит между точками A и B (в элементарно-геометрическом смысле), то ориентация, определенная точками A, B, C, будет обладать тем свойством, что при движении в положительном направлении (по отношению к этой ориентации) от точки A к точке C мы должны будем пройти через точку B, т. е. это будет направление, в котором $C \prec A$.

Суть дела здесь в том, что, пополнив прямую точкой M_∞, мы превращаем ее в замкнутую линию, аналогичную окружности.

Согласно сказанному выше, чтобы задать ориентацию плоскости, достаточно задать ориентацию некоторой окружности (или прямой) и выбрать для этой окружности сторону. Чтобы задать ориентацию окружности, достаточно задать на ней три точки A, B, C, а чтобы задать сторону, достаточно задать произвольную точку D, расположенную на этой стороне. Следовательно,

ориентация плоскости однозначно определена заданием четырех точек A, B, C, D, не принадлежащих одной окружности.

Именно, чтобы определить по точкам A, B, C, D соответствующую ориентацию, нужно

1) провести через точки A, B, C окружность,

2) ориентировать эту окружность так, чтобы точка B разделяла точки A и C,

3) выбрать сторону окружности[1]) ABC, содержащую точку D.

Тогда ориентация плоскости, определенная окружностью ABC и выбранной ее стороной, и будет ориентацией, соответствующей четверке A, B, C, D.

Называя две четверки точек A, B, C, D и A', B', C', D' *одноименными*, если они определяют одну и ту же ориентацию плоскости, мы можем, следовательно, отождествить ориентации плоскости с классами одноименных четверок точек (не принадлежащих одной окружности).

Конечно, такое понимание ориентаций плоскости не будет тавтологией только тогда, когда мы сумеем описать одноименность четверок независимо от понятия ориентации.

Решать эту задачу нам будет удобно в следующей постановке: *определить по данной четверке точек A, B, C, D, не принадлежащих одной окружности, задает ли она данную (скажем, правую) ориентацию плоскости или нет.* При этом ответ должен быть получен в терминах, не связанных (по крайней мере явно) с понятием ориентации.

Предположим сначала, что точка B является бесконечно удаленной точкой M_∞. Тогда окружность ABC будет прямой AC, ориентированной так, что $C \prec A$. Рассмотрим наряду с этой прямой также и прямую AD (т. е. окружность ABD), ориентированную соответствующим образом (так, чтобы $D \prec A$). Поскольку прямые AC и AD пересекаются в точке A и обе ориен-

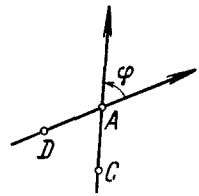
---

[1]) Окружность, проходящую через точки A, B, C и ориентированную так, чтобы точка B разделяла точки A и C, мы будем обозначать символом ABC.

тированы, а на плоскости задана ориентация, то определен угол φ от прямой AD к прямой AC. При этом

точки A, M_∞, C, D тогда и только тогда определяют данную ориентацию плоскости, когда угол φ положителен:

$$0 < \varphi < \pi.$$



Это утверждение очевидно из чертежа.

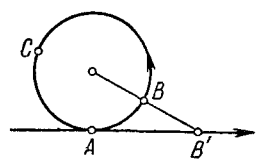
Задание. Дайте формальное доказательство.

Таким образом, вопрос о том, определяет ли четверка A, M_∞, C, D данную ориентацию плоскости, однозначно решается в зависимости от величины угла φ.

Чтобы получить аналогичный ответ в случае, когда точка B конечна, мы должны предварительно определить, что следует понимать под углом от одной ориентированной окружности к другой.

Пусть Σ — произвольная окружность, A — некоторая ее точка и α — касательная к окружности Σ в точке A. Заметим, что задание ориентации окружности Σ однозначно определяет некоторую ориентацию прямой α. Наглядно эту ориентацию касательной можно описать следующим образом: представим себе, что материальная точка, двигаясь по окружности в положительном направлении, освобождается от связей в точке A. Тогда она будет продолжать движение по касательной α. Направление этого движения и определяет ориентацию касательной α.

Более формально, ориентацию касательной, соответствующей данной ориентации окружности Σ, можно определить следующим способом. Рассмотрим на окружности точки A, B, C, определяющие заданную ориентацию окружности и выбранные так, чтобы точка B находилась от точки A на расстоянии, меньшем четверти окружности. Спроектировав точку B из центра окружности на прямую α, мы получим на прямой α некоторую точку B' Тогда ориентация прямой α задается направлением, в котором $A \prec B'$. Конечно, это определение нуждается в проверке корректности, т. е. в доказательстве независимости получающейся ориентации на прямой от выбора точки B.



Упражнение. Докажите корректность определения ориентации касательной.

Таким образом, мы видим, что

касательная к ориентированной окружности естественным образом определяется как ориентированная прямая.

Это утверждение остается справедливым и для случая, когда окружность Σ является прямой, если под касательной к прямой понимать саму эту прямую.

Пусть теперь Σ_1 и Σ_2 — две ориентированные окружности, имеющие общую точку A, и пусть α_1 и α_2 — касательные к этим окружностям в точке A. Угол от ориентированной прямой α_1 к ориентированной прямой α_2 мы и будем называть *углом* от окружности Σ_1 к окружности Σ_2 в их общей точке A.

Согласно сказанному выше, угол от ориентированной прямой к ориентированной прямой попадает под это общее определение.

В изложенном определении молчаливо предполагалось, что точка A конечна. Чтобы перенести его на случай, когда $A = M_\infty$, следует определить, что такое *угол* φ_∞ между двумя ориентированными прямыми *в точке* M_∞.

Если эти прямые параллельны и ориентации их одинаковы, мы, по определению, положим

$$\varphi_\infty = 0.$$

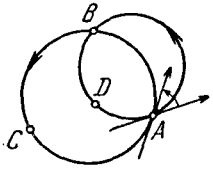
Аналогично, если прямые параллельны, но имеют противоположные ориентации, то мы положим

$$\varphi_\infty = \pi.$$

Наконец, если прямые не параллельны и образуют (в их конечной точке пересечения) угол φ, то мы, по определению, положим

$$\varphi_\infty = -\varphi.$$

Последнее определение оправдывается тем, что если две (настоящие) окружности пересекаются в точках A и B, то, как легко видеть, угол от первой окружности ко второй в точке A знаком отличается от угла в точке B.



Рассмотренный выше случай $B = M_\infty$ наводит на мысль о том, что

четверка точек A, B, C, D, не принадлежащих одной окружности, тогда и только тогда определяет данную ориентацию плоскости, когда угол от окружности ABD к окружности ABC в точке A положителен (в предположении, конечно, что этот угол вычисляется относительно данной ориентации плоскости и выбирается в пределах от $-\pi$ до π).

Это утверждение действительно верно, и его доказательство сводится к рассмотрению пяти возможных случаев: когда все точки A, B, C, D конечны, когда точка A является точкой M_∞, когда точка B является точкой M_∞ (случай, выше уже рассмотренный), когда точка C является точкой M_∞, и, наконец, когда точка D является точкой M_∞. Справедливость утверждения в

каждом из этих случаев непосредственно вытекает из чертежа. Формальные доказательства мы опустим.

Полученные результаты приобретут ценность только тогда, когда мы научимся вычислять угол φ (скажем, по координатам точек A, B, C, D). Выведем соответствующую формулу.

Предполагая все точки A, B, C, D конечными, и, следовательно, окружности ABD и ABC настоящими, рассмотрим прямые

$$CB, \quad CA, \quad DB, \quad DA \quad \text{и} \quad AB,$$

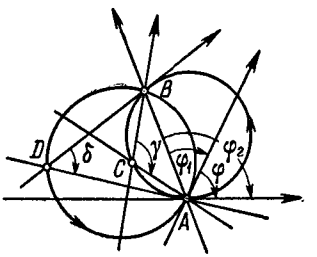
ориентированные в соответствии с указанным порядком точек (т. е., например, прямая CB ориентирована так, что $C \prec B$). Пусть α_1 — касательная в точке A к окружности ABD, а α_2 — касательная в точке A к окружности ABC. Кроме того, пусть

γ — угол от прямой CB к прямой CA,

δ — угол от прямой DB к прямой DA,

φ_1 — угол от прямой AB к касательной α_1,

φ_2 — угол от прямой AB к касательной α_2.



По определению,

$$\varphi = \varphi_2 - \varphi_1.$$

Кроме того, по известной теореме о вписанных углах (справедливой и для ориентированных углов):

$$\varphi_1 = \delta, \quad \varphi_2 = \gamma.$$

Следовательно,

$$\varphi = \gamma - \delta.$$

Предположим теперь, что на плоскости нам задана комплексная координата z, определяющая данную ориентацию o (т. е. такая, что соответствующие прямоугольные координаты x, y определяют ориентацию o).

Пусть

$$z_1, \quad z_2, \quad z_3, \quad z_4$$

— комплексные координаты точек A, B, C, D. Согласно известной геометрической интерпретации аргумента комплексного числа аргумент $\arg(z_2 - z_3)$ числа $z_2 - z_3$ равен углу от оси Ox к прямой CB, а аргумент $\arg(z_1 - z_3)$ числа $z_1 - z_3$ равен углу от оси Ox к прямой CA. Следовательно, угол γ равен разности этих аргументов, т. е. аргументу соответствующего частного:

$$\gamma = \arg \frac{z_1 - z_3}{z_2 - z_3}.$$

Аналогично,

$$\delta = \arg \frac{z_1 - z_4}{z_2 - z_4}.$$

Следовательно,

$$\varphi = \arg W, \tag{1}$$

где

$$W = \frac{z_1 - z_3}{z_2 - z_3} : \frac{z_1 - z_4}{z_2 - z_4}. \tag{2}$$

Пусть теперь одна (и только одна) из точек A, B, C, D является бесконечно удаленной точкой M_∞. Оказывается, что

формула (1) *остается справедливой, если считать, что*

$$\begin{aligned}
W &= \frac{1}{z_2 - z_3} : \frac{1}{z_2 - z_4} = \frac{z_2 - z_4}{z_2 - z_3} && \text{при } z_1 = \infty,\\
W &= (z_1 - z_3) : (z_1 - z_4) = \frac{z_1 - z_3}{z_1 - z_4} && \text{при } z_2 = \infty,\\
W &= 1 : \frac{z_1 - z_4}{z_2 - z_4} = \frac{z_2 - z_4}{z_1 - z_4} && \text{при } z_3 = \infty,\\
W &= \frac{z_1 - z_3}{z_2 - z_3} : 1 = \frac{z_1 - z_3}{z_2 - z_3} && \text{при } z_4 = \infty,
\end{aligned} \tag{3}$$

т. е. если определять число W формулой, получающейся из формулы (2) *соответствующим предельным переходом.*

Задание. Докажите это утверждение.

Если точки A, B, C, D принадлежат одной окружности, т. е. если окружности ABD и ABC совпадают, то естественно считать (по определению!), что $\varphi = 0$, если ориентации окружностей ABD и ABC совпадают, и $\varphi = \pi$, если эти ориентации различны. Оказывается, что

формула (1) *остается справедливой и в этом случае.*

Задание. Докажите это утверждение (не забудьте рассмотреть случай, когда окружности ABD и ABC проходят через точку M_∞).

Сопоставляя все сказанное мы немедленно получаем, что

точки A, B, C, D тогда и только тогда определяют ориентацию, задаваемую координатой z, когда аргумент числа W положителен.

Теперь мы уже можем перейти к формальным определениям.

Определение 1. Число W, определенное формулой (2) (или формулами (3)), называется *двойным отношением* точек A, B, C, D.

Из геометрической интерпретации модуля разности двух комплексных чисел непосредственно вытекает, что

модуль $|W|$ двойного отношения четырех точек не зависит от выбора координатной системы z.

Действительно, если, например, все точки A, B, C, D конечны, то

$$|W| = \frac{|AC|}{|BC|} : \frac{|AD|}{|BD|}.$$

Что же касается аргумента двойного отношения W, то, как показывает формула (1), его абсолютная величина имеет инвариантный геометрический смысл (и потому не зависит от выбора координатной системы z), а знак определяется ориентацией плоскости, задаваемой координатой z (и потому при переходе к другой координатной системе может измениться).

Таким образом,

с точностью до знака аргумента (т. е. с точностью до перехода к комплексно-сопряженному числу $\overline{W}$) двойное отношение не зависит от выбора координатной системы (т. е. определено корректно).

Замечание 2. Мы доказали корректность определения числа W, пользуясь геометрическими соображениями. Методологически более выдержанным было бы прямое алгебраическое доказательство, использующее формулы перехода от одной координатной системы к другой.

Упражнение. Дайте алгебраическое доказательство корректности определения двойного отношения четырех точек.

Из формулы (1) непосредственно вытекает следующее

Предложение 1. *Точки A, B, C, D тогда и только тогда принадлежат одной окружности, когда их двойное отношение W вещественно.*

Задание. Докажите предложение 1 прямым вычислением, не используя формулы (1) (не забудьте рассмотреть случай, когда окружность является прямой).

Определение 2. Две четверки точек пополненной плоскости, каждая из которых состоит из точек, не принадлежащих одной окружности, называются *одноименными*, если двойные отношения W этих четверок имеют аргументы $\arg W$ одного знака (конечно, при условии, что эти аргументы выбраны так, чтобы $-\pi < \arg W < \pi$).

Согласно предложению 1 это определение имеет смысл (аргументы двойных отношений W не могут быть равны ни нулю, ни π). Кроме того, ясно, что это определение корректно, т. е. не зависит от выбора координатной системы z.

Очевидно, что отношение одноименности четверок точек пополненной плоскости является отношением эквивалентности.

Определение 3. Соответствующие классы эквивалентности называются *ориентациями* пополненной плоскости. Пополненная плоскость, для которой выбрана ориентация, называется *ориентированной.*

Проведенные выше неформальные рассмотрения показывают, что это определение равносильно определению, принятому в начале этого пункта.

Теперь мы можем геометрически охарактеризовать различие между конформными преобразованиями из группы Conf^+ и из смежного класса Conf^-.

Определение 4. Мы будем говорить, что конформное преобразование Ω *сохраняет ориентации*, если для любой четверки точек A, B, C, D, не принадлежащих одной окружности, четверка

$$A' = \Omega(A), \quad B' = \Omega(B), \quad C' = \Omega(C), \quad D' = \Omega(D)$$

(также не принадлежащая одной окружности) одноименна четверке A, B, C, D (определяет ту же ориентацию). Если же для любой четверки A, B, C, D четверка A', B', C', D' определяет противоположную ориентацию, мы будем говорить, что конформное преобразование Ω *обращает ориентации.*

Теорема 1. *Конформное преобразование тогда и только тогда принадлежит группе* Conf^+ (*смежному классу* Conf^-), *когда оно сохраняет* (*обращает*) *ориентации.*

Замечание 3. Априори могли бы существовать конформные преобразования, не сохраняющие и не обращающие ориентаций, т. е. такие, что для одной четверки A, B, C, D четверка A', B', C', D' определяет ту же ориентацию, что и четверка A, B, C, D, а для другой четверки A, B, C, D — противоположную. Согласно теореме 1 таких конформных преобразований не существует.

Упражнение. Докажите этот факт непосредственно (не пользуясь теоремой 1).

Мы выведем теорему 1 из следующего предложения:

Предложение 2. *Для любого преобразования из группы* Conf^+ *и любой четверки различных точек A, B, C, D двойное отношение W' точек A', B', C', D' равно двойному отношению W точек A, B, C, D:*

$$W' = W.$$

Аналогично, для любого преобразования из смежного класса Conf^-:

$$W' = \overline{W}.$$

Доказательство. По определению, конформное преобразование из группы Conf^+ записывается в комплексной координате z формулой вида

$$z' = \frac{az + b}{cz + d}.$$

При этом, без ограничения общности, можно предполагать, что эта запись нормирована, т. е. что

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = 1.$$

Пусть z_1, z_2, z_3, z_4 — координаты точек A, B, C, D, а z_1', z_2', z_3', z_4' — координаты точек A', B', C', D'.

Тогда

$$z_1' - z_3' = \frac{az_1 + b}{cz_1 + d} - \frac{az_3 + b}{cz_3 + d} = \frac{(ad - bc)(z_1 - z_3)}{(cz_1 + d)(cz_3 + d)} = \frac{z_1 - z_3}{(cz_1 + d)(cz_3 + d)},$$

и аналогично,

$$z_2' - z_3' = \frac{z_2 - z_3}{(cz_2 + d)(cz_3 + d)},$$

$$z_1' - z_4' = \frac{z_1 - z_4}{(cz_1 + d)(cz_4 + d)},$$

$$z_2' - z_4' = \frac{z_2 - z_4}{(cz_2 + d)(cz_4 + d)}.$$

Поэтому

$$\frac{z_1' - z_3'}{z_2' - z_3'} = \frac{z_1 - z_3}{z_2 - z_3} \cdot \frac{cz_2 + d}{cz_1 + d},$$

$$\frac{z_1' - z_4'}{z_2' - z_4'} = \frac{z_1 - z_4}{z_2 - z_4} \cdot \frac{cz_2 + d}{cz_1 + d}$$

и, следовательно,

$$\frac{z_1' - z_3'}{z_2' - z_3'} : \frac{z_1' - z_4'}{z_2' - z_4'} = \frac{z_1 - z_3}{z_2 - z_3} : \frac{z_1 - z_4}{z_2 - z_4},$$

т. е. $W' = W$.

Равенство

$$W' = \overline{W}$$

для конформных преобразований из Conf^- доказывается аналогично.

Тем самым предложение 2 полностью доказано.

Доказательство теоремы 1. Если $W' = W$, то $\arg W' = \arg W$, а если $W' = \overline{W}$, то $\arg W' = -\arg W$. Поэтому в силу определения 1 четверки A, B, C, D и A', B', C', D' определяют в первом случае одну и ту же ориентацию, а во втором — противоположные ориентации.

5. Конформная геометрия

Определение 1. Свойство фигур на пополненной плоскости называется *конформно инвариантным*, если вместе с некоторой фигурой X этим свойством обладает и любая фигура X', получающаяся из фигуры X произвольным конформным преобразованием. Изучение конформно инвариантных свойств составляет предмет *конформной геометрии.*

Пополненная плоскость, на которой рассматривается конформная геометрия, называется *конформной плоскостью.*

Все точки конформной плоскости совершенно равноправны, так что, в частности, понятие конечных точек и бесконечно удаленной точки в конформной геометрии отсутствует.

Замечание 1. Можно, конечно, рассматривать лишь конформные преобразования, оставляющие точку M_∞ на месте. В соответствующей *конформно-евклидовой геометрии* понятие конечных точек смысла уже имеет.

Взаимоотношение между конформной и конформно-евклидовой геометриями полностью аналогично взаимоотношению между проективной и аффинно-проективной геометриями.

Поскольку конформные преобразования, оставляющие на месте точку M_∞, являются (см. п. 2) не чем иным, как преобразованиями подобия (возможно, не сохраняющими ориентаций); конформно-евклидова геометрия отличается от *геометрии подобия* (изучающей свойства, одинаковые у всех подобных фигур) лишь наличием дополнительной точки M_∞.

Таким образом, взаимоотношение между конформно-евклидовой геометрией и геометрией подобия полностью аналогично взаимоотношению между аффинно-проективной и аффинной геометриями.

Формально-аксиоматическое определение конформной плоскости без труда получается на основе общих идей § 3 гл. 2.

Введем в рассмотрение множество $\mathbb{C}^+$ всех комплексных чисел, пополненное символом ∞, и группу $\mathrm{Conf}(\mathbb{C}^+)$ его дробно-линейных и сопряженно дробно-линейных преобразований

$$z' = \frac{az+b}{cz+d}, \qquad \begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0.$$
$$z' = \frac{a\bar{z}+b}{c\bar{z}+d},$$

Определение 2. Множество $\mathfrak{P}$, для которого задано семейство $\mathrm{Coor}(\mathfrak{P})$ его биективных отображений

$$\alpha\colon \mathfrak{P} \to \mathbb{C}^+$$

на множество $\mathbb{C}^+$, называется *конформной плоскостью*, если выполнены следующие две аксиомы:

Аксиома 1. $\varphi \circ \alpha \in \mathrm{Coor}(\mathfrak{P})$ *для любых* $\alpha \in \mathrm{Coor}(\mathfrak{P})$ *и* $\varphi \in \mathrm{Conf}(\mathbb{C}^+)$.

Аксиома 2. $\alpha' \circ \alpha^{-1} \in \mathrm{Conf}(\mathbb{C}^+)$ *для любых* $\alpha, \alpha' \in \mathrm{Coor}(\mathfrak{P})$.

Согласно этому определению множество $\mathbb{C}^+$ можно превратить в конформную плоскость, полагая

$$\mathrm{Coor}(\mathbb{C}^+) = \mathrm{Conf}(\mathbb{C}^+).$$

Эта плоскость называется *стандартной конформной плоскостью.*

Конформная плоскость в смысле определения 1 является, конечно, конформной плоскостью в смысле определения 2.

Изоморфизмы конформных плоскостей определяются очевидным образом и для них имеют место известные нам общие теоремы. При этом, *автоморфизмами конформных плоскостей являются конформные преобразования и только они* (это лишь иная формулировка предложения 1 п. 2).

Важный вариант конформной геометрии мы получим, заменив группу Conf группой Conf^+ конформных преобразований, сохраняющих ориентации. Мы будем называть эту геометрию *специальной конформной геометрией.* Чтобы получить описывающие ее аксиомы, нужно в аксиомах 1 и 2 группу $\mathrm{Conf}(\mathbb{C}^+)$ заменить группой $\mathrm{Conf}^+(\mathbb{C}^+)$ дробно-линейных преобразований.

Поскольку

$$\mathrm{Conf}^+(\mathbb{C}^+) \subset \mathrm{Conf}(\mathbb{C}^+),$$

каждая специальная конформная плоскость автоматически определяется как конформная плоскость. Чтобы, обратно, от конформной плоскости перейти к специальной конформной плоскости, нужно зафиксировать некоторую координатную систему (ср., например, в п. 5 § 1 гл. 3 обсуждение аналогичной связи между аффинно-проективной и проективной геометриями). При этом две координатные системы α и α' тогда и только тогда приводят к одной и той же специальной конформной плоскости, когда

$$\alpha' \circ \alpha^{-1} \in \mathrm{Conf}^+(\mathbb{C}^+),$$

т. е. когда они одноименны. Другими словами, чтобы получить из конформной плоскости специальную конформную плоскость, нужно зафиксировать некоторую ориентацию. По этой причине специальная конформная геометрия называется также *конформной геометрией ориентированной плоскости.*

Замечание 2. Комплексная координата z, которую мы использовали выше, к сожалению, не вполне адекватна конформной геометрии, поскольку точка M_∞ имеет в ней особую координату $z = \infty$. Более отвечают сути дела две однородные координаты z_1

и z_2, связанные с координатой z соотношением)

$$z = \frac{z_1}{z_2}$$

(и принимающие значения не в $\mathbb{C}^+$, а в $\mathbb{C}$). В этих координатах дробно-линейные преобразования записываются формулами

$$\rho z_1' = az_1 + bz_2,$$
$$\rho z_2' = cz_1 + dz_2.$$

Но это — в точности формулы, описывающие проективные преобразования комплексной проективной прямой.

Следовательно,

конформные ориентированные плоскости — это не что иное, как комплексные проективные прямые, а конформные преобразования, сохраняющие ориентацию, — это проективные преобразования комплексных проективных прямых.

Другими словами,

конформная геометрия ориентированной плоскости — это в точности проективная геометрия комплексной прямой.

В конформной геометрии понятие прямой смысла не имеет. Оно заменяется понятием *окружности.* Отметим, что, тем не менее, понятия центра и радиуса окружности в конформной геометрии отсутствуют. Две области, на которые окружность разбивает конформную плоскость, совершенно равноправны. Они называются *полуплоскостями*, определенными данной окружностью.

Поскольку любые три точки конформной плоскости можно конформным преобразованием перевести в любые другие три точки, фигура, состоящая из трех точек, никаких конформных инвариантов не имеет. Тем более это справедливо для фигуры, состоящей из двух точек. Поэтому никакого аналога понятия расстояния в конформной геометрии не имеется.

Напротив, фигура, состоящая из четырех точек, обладает семиинвариантом W — двойным отношением этих точек. Этот семиинвариант не меняется при преобразованиях из подгруппы Conf^+ и переходит в комплексно-сопряженное число при преобразованиях из смежного класса Conf^-. Следовательно, в конформной геометрии ориентированной плоскости он является инвариантом.

Таким образом, в проективной геометрии комплексной прямой имеется инвариант четырех точек — их двойное отношение. Оказывается, что построение этого инварианта может быть перенесено в проективную геометрию плоскости или пространства (и притом над любым полем). В проективной геометрии этот инвариант играет роль, сравнимую с ролью инварианта двух точек — расстояния в евклидовой геометрии, и никакое более или менее глубокое развитие проективной геометрии без него невозможно. К сожалению, все эти вопросы далеко выходят за рамки нашего элементарного курса.

В отличие от понятия длины понятие угла (между окружностями) в конформной геометрии смысл имеет, поскольку согласно формуле (1) п. 4 угол между окружностями выражается через инвариант W.

При формально-аксиоматическом построении конформной геометрии формулу (1) п. 4 мы должны принять за определение. Это делается следующим образом.

Пусть Σ_1 и Σ_2 — две ориентированные окружности на конформной ориентированной плоскости, пересекающиеся в точках A и B. Выберем на окружности Σ_1 точку C, отличную от точек A, B и такую, что тройка (A, B, C) определяет данную ориентацию окружности Σ_1. Аналогично, на окружности Σ_2 выберем точку D, отличную от точек A, B и такую, что тройка (A, B, D) определяет данную ориентацию окружности Σ_2. Пусть W — двойное отношение точек A, B, C, D.

Определение 3. Аргумент $\arg W$ двойного отношения W называется *углом* от ориентированной окружности Σ_1 к ориентированной окружности Σ_2 в точке A на ориентированной конформной плоскости.

Конечно, в рамках формально-аксиоматического построения конформной геометрии это определение требует проверки корректности, т. е. доказательства независимости аргумента $\arg W$ от выбора точек C и D.

Упражнение. Докажите корректность определения 2.

Задание. Докажите, что угол от окружности Σ_1 к окружности Σ_2 в точке B равен $-\arg W$.

На неориентированной конформной плоскости *углом* между (неориентированными) окружностями Σ_1 и Σ_2 называется абсолютная величина $|\arg W|$ аргумента двойного отношения точек A, B, C, D, где A и B, как и выше, — общие точки окружностей Σ_1 и Σ_2, а C и D — их произвольные точки, отличные от точек A и B. Это определение также корректно.

Дальнейшее развитие конформной геометрии мы предоставим инициативе читателя.

§ 4. ГРУППЫ И ГЕОМЕТРИИ

1. Геометрии с данной группой автоморфизмов

Все разнообразные «геометрии», изученные в предыдущих главах, строились по некоей единой схеме. Рассмотрим эту схему в общем виде.

В первую очередь мы фиксируем некоторую *числовую модель* $\mathfrak{M}_0$, элементы («точки») которой тем или иным способом строятся из чисел (вообще говоря, принадлежащих произвольному полю $\mathbb{K}$), и некоторую *группу преобразований* G_0 этой модели. Например, для аффинной геометрии плоскости $\mathfrak{M}_0 = \mathbb{R}^2$ и $G_0 = \mathrm{Aff}(\mathbb{R}^2)$, для евклидовой геометрии $\mathfrak{M}_0 = \mathbb{R}^2$ и $G_0 = \mathrm{Ort}(\mathbb{R}^2)$,

для проективной геометрии $\mathfrak{M}_0 = \mathbb{RP}^2$ и $G_0 = \mathrm{Proj}(\mathbb{RP}^2)$, для конформной геометрии $\mathfrak{M}_0 = \mathbb{C}^+$ и $G_0 = \mathrm{Conf}(\mathbb{C}^+)$, и т. д.

Далее, мы рассматриваем произвольное множество $\mathfrak{M}$ и семейство $\mathrm{Coor}(\mathfrak{M})$ его биективных отображений («координатных систем») на множество $\mathfrak{M}_0$, подчиненное известным аксиомам 1 и 2. Это множество мы и объявляем «областью действия» (моделью) соответствующей геометрии. Тот факт, что геометрия (при данных $\mathfrak{M}_0$ и G_0) получается только «одна», находит строгое выражение в утверждении об изоморфности различных ее моделей (свойство полноты аксиом).

Автоморфизмы каждой модели $\mathfrak{M}$ совпадают с преобразованиями «по равенству координат» и составляют группу $G = \mathrm{Aut}(\mathfrak{M})$, изоморфную группе $G_0 = \mathrm{Aut}(\mathfrak{M}_0)$ (изоморфизм группы G на группу G_0 задается произвольной координатной системой α: $\mathfrak{M} \to \mathfrak{M}_0$ и определяется формулой $\Phi \mapsto \alpha \circ \Phi \circ \alpha^{-1}$).

После того как группа G построена, мы можем забыть о числовой модели $\mathfrak{M}_0$ и определить геометрию модели $\mathfrak{M}$ как науку, изучающую свойства подмножеств множества $\mathfrak{M}$, инвариантные относительно группы G.

Числовая модель $\mathfrak{M}_0$ и ее группа автоморфизмов G_0 нам нужна для того, чтобы в произвольной модели $\mathfrak{M}$ можно было говорить о координатах и тем самым применить к изучению модели $\mathfrak{M}$ метод аналитической геометрии.

Самым общим образом можно рассматривать произвольное множество $\mathfrak{M}$ и произвольную группу G его преобразований и называть «геометрией» этого множества науку, изучающую свойства подмножеств, инвариантные относительно группы G. В этой общности геометрия растворяется в общей теории групп преобразований и теряет свою специфику.

Изложенный общий взгляд на геометрию был выдвинут в конце XIX в. немецким математиком Ф. Клейном. Он позволяет обозреть с единой точки зрения всевозможные геометрии и расположить их в определенном иерархическом порядке.

Например, возможна ситуация, когда две геометрии имеют одну и ту же числовую модель $\mathfrak{M}_0$ и отличаются лишь соответствующими группами G_0 и H_0, причем группа H_0 является подгруппой группы G_0:

$$H_0 \subset G_0$$

(стандартный пример: евклидова геометрия с группой $\mathrm{Ort}(\mathbb{R}^2)$ и аффинная геометрия с группой $\mathrm{Aff}(\mathbb{R}^2)$).

В этом случае любая модель $\mathfrak{N}$ геометрии с группой H_0 естественным образом определяется как модель $\mathfrak{M}$ геометрии с группой G_0: достаточно принять за координатные системы модели $\mathfrak{M}$ всевозможные отображения вида $\varphi \circ \alpha$, где α — произвольная координатная система модели $\mathfrak{N}$, а φ — произвольное преобразование из группы G_0; см. п. 3 § 3 гл. 2, где все это подробно обсуждено для случая $H_0 = \mathrm{Ort}(\mathbb{R}^2)$ и $G_0 = \mathrm{Aff}(\mathbb{R}^2)$. В частности,

каждая координатная система модели $\mathfrak{N}$ будет координатной системой модели $\mathfrak{M}$:

$$\mathrm{Coor}\,(\mathfrak{N}) \subset \mathrm{Coor}\,(\mathfrak{M})$$

(«любые прямоугольные координаты являются аффинными координатами») и любой автоморфизм модели $\mathfrak{N}$ будет автоморфизмом модели $\mathfrak{M}$ («любое ортогональное преобразование является аффинным преобразованием»), так что группа H автоморфизмов модели $\mathfrak{N}$ будет подгруппой группы G автоморфизмов модели $\mathfrak{M}$:

$$H \subset G.$$

При этом для каждой координатной системы α модели $\mathfrak{N}$ (рассматриваемой как координатная система модели $\mathfrak{M}$) соответствующий изоморфизм

$$\Phi \longmapsto \alpha \circ \Phi \circ \alpha^{-1}$$

группы G на группу G_0 будет отображать подгруппу H на подгруппу H_0 и его ограничение на эту подгруппу будет совпадать с изоморфизмом $H \to H_0$, определенным системой α, рассматриваемой как координатная система модели $\mathfrak{N}$:

$$\begin{array}{ccc} H & \subset & G \\ \downarrow & & \downarrow \\ H_0 & \subset & G_0 \end{array}$$

Поскольку $H \subset G$, любое свойство, инвариантное относительно G, будет инвариантно и относительно H, так что любое утверждение геометрии с группой G будет справедливо и в геометрии с группой H. В этом смысле геометрия с «меньшей» группой H «богаче» геометрии с «большей» группой G, т. е. геометрия с группой G является «частью» геометрии с группой H («аффинная геометрия является частью евклидовой»).

Обратный переход от геометрии с группой G_0 к геометрии с группой H_0 осуществляется выбором для любой модели $\mathfrak{M}$ геометрии с группой G_0 произвольной координатной системы

$$\alpha_0\colon\ \mathfrak{M} \to \mathfrak{M}_0.$$

Когда эта система выбрана, мы можем построить модель $\mathfrak{M}_{\alpha_0}$ геометрии с группой H_0, считая, что как множество она совпадает с моделью $\mathfrak{M}$, и принимая за ее координатные системы всевозможные отображения вида $\varphi \circ \alpha_0$, где φ — произвольное преобразование из группы H_0. Модель $\mathfrak{M}_{\alpha_0}$ зависит, естественно, от системы α_0, причем

$$\mathfrak{M}_{\alpha_0} = \mathfrak{M}_{\alpha_1}$$

тогда и только тогда, когда

$$\alpha_1 \circ \alpha_0^{-1} \in H_0$$

(см., например, в п. 5 § 1 гл. 4 случай $H_0 = \mathrm{Aff}(\mathbb{R}P^2)$, $G_0 = \mathrm{Proj}(\mathbb{R}P^2)$).

Отметим две важные особенности этой конструкции.

Во-первых, модель $\mathfrak{M}_{\alpha_0}$ определена для любой координатной системы α_0. В частности, для случая $H_0 = \mathrm{Ort}(\mathbb{R}^2)$ и $G_0 = \mathrm{Aff}(\mathbb{R}^2)$ это означает, что в аффинную плоскость можно так ввести метрику (измерение углов и длин), чтобы любой, наперед заданный аффинный репер Oe_1e_2 оказался бы прямоугольным репером.

Во-вторых, модели геометрии с группой H_0, имеющие вид $\mathfrak{M}_{\alpha_0}$, обладают дополнительной структурой, которой не обладают произвольные модели. Эта дополнительная структура состоит в том, что семейство $\mathrm{Coor}(\mathfrak{M}_{\alpha_0})$ координатных систем модели $\mathfrak{M}_{\alpha_0}$ вполне определенным образом вложено в семейство $\mathrm{Coor}(\mathfrak{M})$ координатных систем модели $\mathfrak{M}$.

Например, для евклидовой геометрии утверждение, что данный аффинный репер является евклидовым (прямоугольным) репером, зависит, кроме всего прочего, от выбора эталона длины, т. е. отрезка, длина которого принята за единицу. Если мы возьмем другой эталон длины, то мы получим совсем другие евклидовы координатные системы.

Поэтому при изучении геометрии на моделях вида $\mathfrak{M}_{\alpha_0}$ нужно очень внимательно следить, чтобы доказываемые утверждения не зависели от вложения $\mathrm{Coor}(\mathfrak{M}_{\alpha_0})$ в $\mathrm{Coor}(\mathfrak{M})$ (конечно, никто не запрещает рассматривать утверждения, зависящие от этого вложения, — например, в евклидовой геометрии утверждения, зависящие от выбора эталона длины, — но это будут уже утверждения, не принадлежащие рассматриваемой геометрии).

Для дополнительной иллюстрации изложенных общих соображений мы рассмотрим в следующих пунктах несколько интересных примеров. При этом мы, как правило, ограничимся геометриями, числовой моделью которых является множество $\mathbb{R}^2$.

Геометрии с моделью $\mathbb{R}^2$ можно представлять себе действующими на обычной, неформально рассматриваемой, плоскости. Как мы знаем (см. п. 2 § 1), для того чтобы в геометрии на плоскости с группой автоморфизмов G имело смысл понятие прямой, необходимо (и достаточно), чтобы группа G была подгруппой аффинной группы $\mathrm{Aff}(2)$. Такого рода геометрии мы будем называть *геометриями аффинного типа*. За недостатком места мы в основном ими и ограничимся.

Аналогично, геометрии с моделью $\mathbb{R}P^2$ можно представлять себе действующими на расширенной плоскости. Для того чтобы в такой геометрии имело смысл понятие прямой, необходимо и достаточно, чтобы ее группа автоморфизмов G была подгруппой группы $\mathrm{Proj}(2)$. Некоторые из таких *геометрий проективного типа* чрезвычайно интересны, но, к сожалению, у нас нет возможности здесь их рассмотреть.

По определению, чтобы получить какую-нибудь геометрию аффинного типа, нужно выделить в группе $\mathrm{Aff}(2)$ определенную

подгруппу. Существует некий общий метод выделения подгрупп группы Aff(2) (и, вообще, любой группы преобразований G), которым мы и будем в дальнейшем пользоваться. Этот метод состоит в том, что фиксируется некоторый геометрический объект (точка, ориентация, бивектор, направление и т. п.) и рассматриваются все преобразования, оставляющие этот объект инвариантным. Ясно, что эти преобразования образуют подгруппу группы Aff(2) (конечно, эта подгруппа может оказаться неинтересной: например, совпадать со всей группой Aff(2), или, наоборот, содержать только тождественное преобразование E). Вопрос о том, можно ли таким образом задать произвольную подгруппу группы Aff(2), мы рассматривать не будем.

2. Простейшие геометрии аффинного типа

Наиболее близкой к аффинной геометрии геометрией аффинного типа является геометрия с группой $\mathrm{Aff}^+(2)$ всех аффинных преобразований, сохраняющих ориентации. Мы будем называть эту геометрию *специальной аффинной геометрией.*

Отметим, что в то время как имеет смысл говорить об аффинной геометрии над произвольным полем K (см. п. 4 § 3 гл. 2), специальная аффинная геометрия определена лишь над полем $\mathbb{R}$.

«Евклидовым» вариантом специальной аффинной геометрии является *специальная евклидова геометрия* с группой $\mathrm{Ort}^+(2)$ всех сохраняющих ориентации ортогональных преобразований. Эта геометрия отличается от евклидовой геометрии только тем, что симметричные фигуры в ней считаются различными.

Чтобы проиллюстрировать различие между евклидовой и специальной евклидовой геометрией, можно, например, указать, что понятие векторного произведения (п. 6 § 6 гл. 1) принадлежит специальной евклидовой геометрии (пространства); в собственно евклидовой геометрии это понятие отсутствует.

Плоскость, в которой действует специальная аффинная (или специальная евклидова) геометрия, естественно называть *специальной аффинной* (соответственно — *евклидовой*) *плоскостью.* Чтобы из аффинной плоскости $\mathfrak{M}$ получить специальную аффинную плоскость $\mathfrak{M}_{\text{спец}}$, нужно, следуя общей схеме, изложенной в п. 1, выбрать некоторую аффинную координатную систему α: $\mathfrak{M} \to \mathbb{R}^2$ и рассмотреть все аффинные координатные системы вида $\varphi \circ \alpha$, где $\varphi \in \mathrm{Aff}^+(\mathbb{R}^2)$. При этом две аффинные координатные системы α_1 и α_2 тогда и только тогда приводят к одной и той же специальной аффинной плоскости, когда $\alpha_1 \circ \alpha_2^{-1} \in \in \mathrm{Aff}^+(\mathbb{R}^2)$, т. е. когда эти системы одноименны (определяют одну и ту же ориентацию). Следовательно, переход от аффинной плоскости к специальной аффинной плоскости заключается в выборе на плоскости некоторой ориентации. В этом смысле можно сказать, что специальные аффинные плоскости являются не чем иным, как *ориентированными аффинными плоскостями.*

Однако выбор двух различных ориентаций приводит к различным (хотя, конечно, и изоморфным) специальным аффинным плоскостям. Поэтому, представляя себе специальную аффинную плоскость как ориентированную аффинную плоскость, надо постоянно следить за тем, чтобы результаты не зависели от выбора ориентации. Это является проявлением общего феномена зависимости от вложения $\mathrm{Coor}(\mathfrak{M}_{a_0})$ в $\mathrm{Coor}(\mathfrak{M})$, который мы обсуждали в п. 1.

Конечно, все это с равным правом справедливо и для специальных евклидовых плоскостей.

Геометрия с группой $\mathrm{Aff}_O(2)$ всех аффинных преобразований плоскости, оставляющих на месте некоторую точку O, называется *центроаффинной геометрией*, а соответствующая плоскость — *центрированной аффинной плоскостью*. Центроаффинная геометрия отличается от аффинной геометрии тем, что точка O играет в ней особую роль.

Координатными системами центроаффинной геометрии являются аффинные координатные системы с началом в точке O.

Центроаффинная геометрия определена, конечно, и для случая произвольного поля $\mathbb{K}$.

Точки центрированной аффинной плоскости находятся в естественном биективном соответствии с векторами из $\mathrm{Vect}(2)$ (точке M соответствует ее радиус-вектор $\overrightarrow{OM}$). Поэтому в центроаффинной геометрии можно, в частности, говорить о *сумме точек* и о *произведении точки на число* (понятия, в общей аффинной геометрии смысла не имеющие). Вообще, поскольку группа $\mathrm{Aff}_O(2)$ изоморфна группе обратимых линейных операторов $\mathrm{Op}(2)$, можно сказать, что *центроаффинная геометрия является не чем иным, как геометрией линеала* $\mathrm{Vect}(2)$ *относительно группы* $\mathrm{Op}(2)$.

Вариантом центроаффинной геометрии, «различающим симметричные фигуры», является *специальная центроаффинная геометрия* с группой $\mathrm{Aff}_O^+(2)$ всех сохраняющих ориентации аффинных преобразований, оставляющих точку O на месте. С учетом сказанного выше можно сказать, что специальная центроаффинная геометрия является не чем иным, как геометрией ориентированного линеала $\mathrm{Vect}(2)$.

Евклидовым вариантом центроаффинной геометрии является *центроевклидова геометрия* с группой $\mathrm{Ort}_O^+(2)$, состоящей из всех ортогональных преобразований, оставляющих точку O на месте. Эта геометрия совпадает с «метрической геометрией векторов» из $\mathrm{Vect}(2)$. В ней имеют смысл понятия *длины* точки и *угла* между двумя точками (впрочем, обычно длину точки, т. е. длину соответствующего радиус-вектора, называют *нормой* точки, а угол между двумя точками, т. е. угол между соответствующими радиус-векторами, — *угловым расстоянием* между этими точками).

Можно также рассматривать *специальную центроевклидову геометрию* с группой $\mathrm{Ort}_O^+(2) = \mathrm{Rot}_O(2)$ всех вращений вокруг точки O. Эта геометрия отличается от центроевклидовой геометрии только тем, что симметричные фигуры (с осью симметрии, проходящей через точку O) в ней считаются различными.

Пусть Φ — произвольное аффинное преобразование плоскости. В некоторой системе аффинных координат x, y с репером Oe_1e_2 это преобразование выражается, как мы знаем, формулами

$$\begin{aligned} x' &= a_1^1 x + a_2^1 y + b^1, \\ y' &= a_1^2 x + a_2^2 y + b^2, \end{aligned} \tag{1}$$

где

$$A = \begin{pmatrix} a_1^1 & a_2^1 \\ a_1^2 & a_2^2 \end{pmatrix}$$

— матрица индуцированного линейного оператора Φ в базисе e_1, e_2. В любом другом базисе e_1', e_2' матрицей оператора Φ является матрица $C^{-1}AC$, где C — матрица перехода от базиса e_1, e_2 к базису e_1', e_2' (см. п. 3 § 7 гл. 1). Так как

$$|C^{-1}AC| = |C|^{-1} \cdot |A| \cdot |C| = |A|,$$

то, следовательно,

определитель $|A|$ матрицы A не зависит от выбора репера Oe_1e_2.

Этот определитель мы будем называть *определителем аффинного преобразования* Φ и будем его обозначать символом $|\Phi|$ или $\det \Phi$.

Согласно только что доказанному, это определение корректно (не зависит от выбора репера Oe_1e_2).

Геометрический смысл определителя аффинного преобразования выясняется следующим предложением:

Предложение 1. *Пусть X — произвольная плоская фигура и пусть X' — фигура, получающаяся из неё аффинным преобразованием Φ. Тогда площадь фигуры X' равна площади фигуры X, умноженной на абсолютную величину определителя преобразования Φ:*

$$\text{площ. } X' = |\det \Phi| \cdot \text{площ. } X.$$

Доказательство. Выберем на плоскости некоторую прямоугольную систему координат Oxy. Тогда, как известно из анализа, площадь фигуры X будет равна абсолютной величине интеграла

$$\iint\limits_X dx\,dy.$$

С другой стороны, согласно формуле замены переменных в двойной интеграле при любом преобразовании плоскости, выражающемся в координатах x, y формулами

$$\begin{aligned} x' &= f(x, y), \\ y' &= g(x, y), \end{aligned} \tag{2}$$

имеет место равенство

$$\iint\limits_{X'} dx\,dy = \iint\limits_{X} \frac{\partial(f, g)}{\partial(x, y)}\,dx\,dy,$$

где X' — образ множества X при преобразовании (2), а

$$\frac{\partial(f, g)}{\partial(x, y)} = \begin{vmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \\ \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{vmatrix}$$

— якобиан этого преобразования.

Для завершения доказательства остается заменить, что в нашем случае преобразование (2) выражается формулами (1) и потому имеет постоянный якобиан, равный определителю преобразования Ф.

Для читателей, не знакомых с интегральным исчислением, дадим доказательство, не использующее интегралов.

Пусть снова Oxy — произвольная система прямоугольных координат на плоскости. Всевозможные прямые вида $x = n\varepsilon$ и $y = m\varepsilon$, где $\varepsilon > 0$, а n и m — произвольные целые числа, разбивают плоскость на квадраты K_ε со сторонами, равными ε (и потому имеющими площадь ε^2). Пусть N_ε — число всех таких квадратов, содержащихся в X. Тогда по определению площади

$$\text{площ. } X = \lim_{\varepsilon \to 0} (N_\varepsilon \varepsilon^2).$$

Предположим, что наше аффинное преобразование Ф является композицией двух сжатий: одного — к оси Ox с коэффициентом k_1 и другого — к оси Oy с коэффициентом k_2. При этом преобразовании квадраты K_ε перейдут в прямоугольники K'_ε со сторонами длины $k_1\varepsilon$ и $k_2\varepsilon$ (и потому имеющие площадь $k_1k_2\varepsilon^2$), а число всех таких прямоугольников, содержащихся в X', будет равно N_ε. Поскольку при вычислении площади фигуры X' мы с равным правом вместо квадратов можем использовать прямоугольники K'_ε, то

$$\text{площ. } X' = \lim_{\varepsilon \to 0} (N_\varepsilon \cdot k_1k_2\varepsilon^2) = k_1k_2 \cdot \lim_{\varepsilon \to 0} (N_\varepsilon \varepsilon^2) = k_1k_2 \cdot \text{площ. } X.$$

Поскольку определитель преобразования Ф равен, очевидно, k_1k_2, тем самым для рассматриваемых преобразований предложение 1 доказано.

Пусть теперь преобразование Ф ортогонально. При этом преобразовании квадраты K_ε перейдут в некоторые квадраты K'_ε той же площади ε^2, но только, возможно, повернутые (и сдвинутые), причем число квадратов K'_ε, содержащихся в X', будет по-прежнему равно числу N_ε квадратов K_ε, содержащихся в X. Следовательно,

$$\text{площ. } X' = \lim_{\varepsilon \to 0} (N_\varepsilon \varepsilon^2) = \text{площ. } X.$$

Таким образом, и для ортогональных преобразований (имеющих, как мы знаем, определители ± 1) предложение 1 справедливо.

Для того чтобы доказать это предложение для любых аффинных преобразований, остается вспомнить, что каждое такое преобразование является композицией некоторого ортогонального преобразования и двух сжатий к взаимно перпендикулярным прямым (которые мы можем принять за оси координат). Поскольку определитель композиции двух аффинных преобразований равен, очевидно, произведению определителей сомножителей, предложение 1 тем самым доказано полностью.

Следствие. *Площадь s эллипса с полуосями a и b (точнее, площадь области, ограниченной этим эллипсом) выражается формулой*

$$s = \pi ab.$$

Доказательство. Площадь круга радиуса a равна, как известно, πa^2. Но эллипс с полуосями a и b получается из этого круга сжатием плоскости к одному из его диаметров с коэффициентом $k = \frac{b}{a}$. Следовательно,

$$s = k \cdot (\pi a^2) = \pi ab.$$

Определение 1. Аффинное преобразование Φ называется *эквиаффинным*, если $|\Phi| = \pm 1$.

Согласно предложению 1,

аффинное преобразование тогда и только тогда эквиаффинно, когда оно сохраняет площади, т. е. когда площадь любой плоской фигуры X равна площади преобразованной фигуры X'.

К числу эквиаффинных преобразований принадлежат все ортогональные преобразования. Однако существуют и неортогональные эквиаффинные преобразования.

Задание. Приведите пример такого преобразования.

Так как

$$|\Psi \circ \Phi| = |\Psi| \cdot |\Phi|,$$

то

совокупность всех эквиаффинных преобразований является группой.

Геометрия с этой группой называется *эквиаффинной геометрией*, а соответствующие плоскости — *эквиаффинными плоскостями*. В отличие от аффинной геометрии, в эквиаффинной геометрии имеет смысл понятие площади. Она является естественной областью, в которой целесообразно строить теорию площадей.

Две аффинные координатные системы с реперами Oe_1e_2 и $O'e_1'e_2'$ тогда и только тогда определяют (согласно общей схеме из п. 1) одну и ту же эквиаффинную плоскость, когда бивекторы $e_1 \wedge e_2$ и $e_1' \wedge e_2'$ либо совпадают, либо отличаются знаком, т. е. когда параллелограммы, построенные на векторах e_1, e_2 и e_1', e_2',

имеют одну и ту же площадь. Это показывает, что *эквиаффинная плоскость является не чем иным, как аффинной плоскостью, на которой задан эталон площади* (область, площадь которой принята за единицу).

Конечно, и здесь при выборе различных эталонов площади мы получаем из одной аффинной плоскости различные эквиаффинные плоскости.

При формулировке утверждений эквиаффинной геометрии, относящихся к площадям, необходимо соблюдать определенную осторожность. Например, доказанное выше следствие о площади эллипса, как оно сформулировано, не принадлежит эквиаффинной геометрии, поскольку в нем фигурируют длины полуосей a и b, смысла в эквиаффинной геометрии не имеющие. Чтобы получить «эквиаффинную» формулировку этого следствия, достаточно, однако, заметить, что согласно теореме Аполлония (см. п. 6 § 2 гл. 5) площадь параллелограмма, построенного на произвольной паре сопряженных радиусов эллипса, равна ab. Поэтому мы можем сказать, что

отношение площади эллипса к площади параллелограмма, построенного на паре его сопряженных радиусов, равно π.

Эта формулировка в эквиаффинной геометрии уже вполне осмыслена.

Замечание 1. Ясно, что теорема Аполлония очевидным образом вытекает из возможности представления произвольного эллипса как образа некоторой окружности при аффинном преобразовании, поскольку для случая окружности она тривиальна (площадь квадрата, построенного на паре перпендикулярных радиусов окружности, равна квадрату радиуса окружности).

Замечание 2. Утверждение о площади эллипса может быть сформулировано также и следующим образом:

отношение площадей эллипса и параллелограмма, построенного на паре его сопряженных радиусов, равно π.

Обратим внимание на тонкое, но существенное, различие между двумя последними формулировками: в то время как в первой формулировке мы говорим об отношении площади одной фигуры (эллипса) к площади другой фигуры (параллелограмма), во второй формулировке речь идет об отношении площадей этих фигур. Поскольку согласно предложению 1 при любом аффинном преобразовании площади всех фигур умножаются на одно и то же число, не зависящее от фигуры, то для любых двух фигур X и Y отношение их площадей аффинно инвариантно. Другими словами, хотя в аффинной геометрии и нельзя говорить о площади одной отдельно взятой фигуры, но понятие *отношения площадей* двух фигур имеет полный смысл (подобно тому как имеет смысл понятие отношения длин двух параллельных отрезков). Таким образом, вторая из приведенных выше формулировок имеет смысл в аффинной геометрии, тогда как первая — только в эквиаффинной.

Вариант эквиаффинной геометрии, в котором фундаментальной группой считается группа эквиаффинных преобразований, сохраняющих ориентации, конечно, также возможен. В этой геометрии имеет смысл понятие *ориентированной площади.*

Определение 2. Аффинное преобразование Ф называется *преобразованием подобия*, если оно сохраняет углы между прямыми, т. е. если любые две (пересекающиеся) прямые α и β оно переводит в прямые α' и β', образующие тот же угол.

Ясно, что

все преобразования подобия образуют группу.

Соответствующая геометрия называется *геометрией подобия.* Фигуры, равные в этой геометрии, т. е. фигуры, переводящиеся друг в друга преобразованием подобия, называются *подобными.*

Покажем, что это понятие подобия совпадает с известным читателю из элементарного курса, т. е. что две фигуры тогда и только тогда подобны, когда после соответствующего перемещения они гомотетичны. Для этого, очевидно, достаточно показать, что любое преобразование подобия является композицией некоторого ортогонального преобразования (движения или движения плюс симметрия) и некоторой *гомотетии*, т. е. преобразования, выражающегося в соответствующим образом подобранной системе прямоугольных координат x, y формулами вида

$$x' = hx,$$
$$y' = hy,$$

где h — *коэффициент гомотетии.*

В силу предложения 1 п. 2 § 2 для этого в свою очередь достаточно показать, что преобразование вида

$$\begin{aligned} x' &= k_1 x, \\ y' &= k_2 y, \end{aligned} \qquad k_1 > 0, \quad k_2 > 0, \tag{3}$$

тогда и только тогда сохраняет углы между прямыми, когда $k_1 = k_2$, т. е. когда оно является гомотетией.

Но это ясно. Действительно, преобразование (3) переводит ось абсцисс $y = 0$ в себя, а прямую с уравнением

$$y = ax \tag{3}$$

— в прямую с уравнением

$$y = \frac{k_2}{k_1} ax.$$

Поэтому угол между осью абсцисс и прямой (4) тогда и только тогда сохраняется при преобразовании (3), когда

$$\frac{k_2}{k_1} a = a,$$

т. е. когда $k_1 = k_2$.

Мы не будем здесь рассматривать геометрию подобия, поскольку она достаточно подробно изучается в школе. (Заметим, кстати, что на самом деле школьный курс геометрии представляет собой смешение теорем из целого ряда геометрий — в основном, евклидовой геометрии и геометрии подобия.)

Чтобы перейти от геометрии подобия к евклидовой геометрии, достаточно выбрать эталон длины.

Ряд более или менее интересных вариантов геометрии подобия мы получим, накладывая на преобразования подобия те или иные дополнительные требования (сохранение ориентаций, сохранение данной точки O и т. п.). В геометрии, основывающейся на сохраняющих ориентации преобразованиях подобия (*специальной геометрии подобия*), имеет смысл понятие ориентированного угла.

Поскольку, как мы уже видели (см. п. 2 § 3), любое сохраняющее ориентации преобразование подобия записывается в комплексной координате $z = x + iy$ формулой

$$z' = az + b, \quad a \neq 0, \quad a,\ b \in \mathbb{C},$$

специальная геометрия подобия по существу совпадает с аффинной геометрией комплексной прямой.

Наложение условия эквиаффинности снова приводит к евклидовой геометрии, поскольку, как легко видеть,

преобразование подобия тогда и только тогда эквиаффинно, когда оно ортогонально.

3. Геометрии Галилея и Пуансо

Поскольку любое аффинное преобразование переводит любое направление на плоскости (пучок параллельных прямых) снова в некоторое направление, имеет смысл говорить об аффинных преобразованиях, оставляющих на месте данное направление ω.

Все такие преобразования образуют подгруппу $\mathrm{Aff}_\omega(2)$ группы $\mathrm{Aff}(2)$. Соответствующую геометрию мы будем называть *аффинной геометрией с особыми прямыми*, поскольку прямые направления ω играют в ней особую роль.

Координатными системами этой геометрии являются аффинные координатные системы Oxy, ось абсцисс $y = 0$ которых имеет направление ω (является особой прямой).

Пусть Oxy — такая координатная система. Поскольку аффинное преобразование.

$$x' = a_1^1 x + a_2^1 y + b^1,$$
$$y' = a_1^2 x + a_2^2 y + b^2$$

переводит ось абсцисс $y = 0$ в прямую с параметрическими уравнениями

$$x' = a_1^1 x + b^1,$$
$$y' = a_1^2 x + b^2$$

(здесь x', y' текущие координаты точек прямой, а x — параметр),

т. е. в прямую

$$a_1^2x' - a_1^1y' = a_1^2b^1 - a_1^1b^1,$$

то это преобразование тогда и только тогда принадлежит группе $\mathrm{Aff}_\omega(2)$, когда $a_1^2 = 0$, т. е. когда оно записывается формулами

$$\begin{aligned} x' &= a_1^1x + a_2^1y + b^1, \\ y' &= a_2^2y + b^2. \end{aligned} \tag{1}$$

Здесь числа b^1, b^2 и a_2^1 могут быть произвольными вещественными числами, а числа a_1^1 и a_2^2 должны быть отличны от нуля.

Если мы ограничимся лишь преобразованиями группы $\mathrm{Aff}_\omega(2)$, сохраняющими площади (эквиаффинными), то мы получим аналогичную геометрию (*эквиаффинную геометрию с особыми прямыми*), в которой дополнительно имеется понятие площади. В преобразованиях (1) для этой геометрии числа a_1^1 и a_2^2 должны удовлетворять соотношению

$$a_1^1a_2^2 = \pm 1.$$

Точно так же мы можем получить еще одну геометрию, ограничившись преобразованиями группы $\mathrm{Aff}_\omega(2)$, оставляющими на месте некоторую точку O. В этой геометрии, кроме особых прямых, имеется, таким образом, и *особая точка* O. В преобразованиях (1) для такой геометрии коэффициенты b^1 и b^2 должны быть равны нулю.

Накладывая на преобразования оба условия, мы получим геометрию с особыми прямыми и особой точкой, в которой имеет смысл понятие площади. Преобразования (1) для этой геометрии должны, естественно, удовлетворять обоим ограничениям:

$$a_1^1a_2^2 = \pm 1, \quad b^1 = b^2 = 0.$$

Конечно, для всех этих трех геометрий возможны варианты, возникающие, когда мы ограничиваемся лишь аффинными преобразованиями, сохраняющими ориентации.

Еще одна геометрия получается, если на преобразования (1) наложить условие

$$a_1^1 = \pm 1,$$

означающее, что индуцированные этими преобразованиями отображения особых прямых должны быть ортогональными отображениями.

Вариантом последней геометрии является геометрия, преобразования которой характеризуются условием

$$a_1^1 = 1$$

(индуцированные отображения особых прямых являются параллельными переносами).

И для этих двух геометрий возможны варианты, когда мы ограничиваемся либо преобразованиями, оставляющими на месте данную точку O, либо преобразованиями, сохраняющими ориентации, либо эквиаффинными преобразованиями, либо преобразованиями, обладающими любыми из этих свойств.

Таким образом, всего мы получаем четырнадцать различных геометрий. Все эти геометрии тесно связаны друг с другом и, изучив одну из них, мы без особого труда можем перейти к любой другой. Мы выберем для более детального изучения геометрию, преобразования которой характеризуются условиями $a_1^1 = a_2^2 = 1$, т. е. выражаются формулами

$$\begin{aligned} x' &= x + a_2^1 y + b^1, \\ y' &= y + b^2. \end{aligned} \tag{2}$$

Иными словами, мы будем рассматривать геометрию с особыми прямыми, автоморфизмы которой

1) сохраняют ориентации;
2) эквиаффинны;
3) для любой особой прямой индуцируют параллельный перенос этой прямой на соответствующую особую прямую.

Изменив обозначения, мы можем преобразования (2) записать в следующем виде:

$$\begin{aligned} x' &= x + vt + x_0, \\ t' &= t + t_0. \end{aligned}$$

В этом виде они совпадают с известными из механики *преобразованиями Галилея*, описывающими переход от одной инерциальной системы отсчета (x, t) на прямой к другой такой системе (x', t'), движущейся относительно первой с постоянной скоростью v. Согласно принципу относительности Галилея механический (точнее, кинематический) смысл имеют только те свойства движений, которые инвариантны относительно преобразований Галилея. Это означает, что интересующая нас геометрия по существу совпадает с кинематикой движений на прямой и отличается от нее лишь словесной оболочкой (терминологией). Каждое утверждение этой геометрии может быть переформулировано как утверждение кинематики, и наоборот.

На этом основании мы будем называть рассматриваемую геометрию *геометрией Галилея*, а преобразования (2) — *галилеевыми преобразованиями*.

Чтобы перейти от аффинной плоскости к галилеевой, нужно выбрать

1) особые прямые;
2) ориентацию особых прямых;
3) ориентацию плоскости;
4) эталон длины на особых прямых;
5) эталон площади.

Тогда каждая галилеева координатная система будет представлять собой аффинную координатную систему, репер Oe_1e_2 которой обладает следующими свойствами:

1) вектор e_1 параллелен особым прямым и положительно ориентирован (по отношению к данной ориентации этих прямых);

2) базис e_1, e_2 положительно ориентирован (по отношению к данной ориентации плоскости);

3) длина вектора e_1 равна единице (по отношению к данному эталону длин на особых прямых);

4) площадь параллелограмма, построенного на векторах e_1, e_2, равна единице (по отношению к данному эталону площадей).

В каждой из таких координатных систем Oxy галилеевы преобразования записываются формулами вида (2).

В дальнейшем мы будем считать фиксированной некоторую галилееву координатную систему Oxy.

На евклидовой плоскости расстояние между двумя точками $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$ может быть охарактеризовано как единственный (с точностью до постоянного множителя) ортогональный инвариант фигуры, состоящей из этих двух точек. Легко видеть, что единственным галилеевым инвариантом этой фигуры является разность $y_2 - y_1$. (Инвариантность этой разности очевидна; ее единственность вытекает из того, что любые пары точек с одной и той же разностью $y_2 - y_1$ можно некоторым галилеевым преобразованием перевести друг в друга.) По аналогии с евклидовой плоскостью абсолютную величину

$$d = |y_2 - y_1|$$

разности $y_2 - y_1$ мы будем называть *галилеевым расстоянием* между точками $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$.

Заметим, что в отличие от евклидова расстояния галилеево расстояние может быть равно нулю даже тогда, когда точки M_1 и M_2 не совпадают. Именно, ясно, что

галилеево расстояние d между точками M_1 и M_2 тогда и только тогда равно нулю, когда эти точки принадлежат одной особой прямой.

В этом случае для точек $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$ существует еще один галилеев инвариант, а именно, число $x_2 - x_1$. Действительно, согласно формулам (2)

$$x_2' - x_1' = x_2 - x_1 + a_2^1(y_2 - y_1)$$

и потому $x_2' - x_1' = x_2 - x_1$, если $y_2 - y_1 = 0$.

Таким образом, разность $x_2 - x_1$ инвариантна не всегда, а только при некотором специальном расположении точек $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$. На этом основании она называется *относительным инвариантом*.

Абсолютную величину

$$\delta = |x_2 - x_1|$$

относительного инварианта $x_2 - x_1$ мы будем называть *особым расстоянием* между точками $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$. Подчеркнем, что оно имеет смысл тогда и только тогда, когда галилеево расстояние между этими точками равно нулю.

Ясно, что

точки M_1 и M_2 тогда и только тогда совпадают, когда оба расстояния d и δ между ними равны нулю.

По аналогии с евклидовой геометрией *окружностью в геометрии Галилея* называется геометрическое место точек, галилеево расстояние которых от данной точки $M_0(x_0, y_0)$ (центра окружности) равно данному числу R (*радиусу* окружности). Аналитически окружность задается уравнением

$$|y - y_0| = R$$

и потому представляет собой пару особых прямых

$$y = y_0 \pm R$$

(сливающихся в одну при $R = 0$). Заметим, что в отличие от евклидовой геометрии галилеева окружность имеет бесконечно много центров (но однозначно определенный радиус).

Продолжая аналогию с евклидовой геометрией, мы определим *галилеев угол* φ (точнее «галилееву величину угла») между двумя неособыми прямыми α_1 и α_2, пересекающимися в точке $M_0(x_0, y_0)$, как особое расстояние между точками пересечения N_1 и N_2 прямых α_1 и α_2 с окружностью радиуса $R = 1$ с центром в точке $M_0(x_0, y_0)$ (точнее, с одной из особых прямых, составляющих эту окружность; скажем, для определенности, с прямой $y = y_0 + 1$). Для особых прямых угол не определяется.

Пусть прямые α_1 и α_2 заданы уравнениями с угловым коэффициентом (относительно оси ординат):

$$x = k_1 y + b_1, \quad x = k_2 y + b_2.$$

Тогда их точка пересечения $M_0(x_0, y_0)$ имеет ординату

$$y_0 = -\frac{b_2 - b_1}{k_2 - k_1}$$

и потому точки N_1 и N_2 (являющиеся точками пересечения прямых α_1 и α_2 с прямой $y = y_0 + 1$) имеют абсциссы

$$x_1 = k_1 + \frac{b_1 k_2 - b_2 k_1}{k_2 - k_1},$$

$$x_2 = k_2 + \frac{b_1 k_2 - b_2 k_1}{k_2 - k_1}.$$

Следовательно, особое расстояние между ними равно

$$|x_2 - x_1| = |k_2 - k_1|.$$

Таким образом,

галилеев угол φ *между неособыми прямыми с угловыми коэффициентами* k_1 *и* k_2 *выражается формулой*

$$\varphi = |k_2 - k_1|.$$

Мы могли бы принять эту формулу за определение угла φ. Но тогда мы, конечно, должны были бы доказывать галилееву инвариантность величины $|k_2 - k_1|$.

Если прямые параллельны ($k_2 = k_1$), то для них определена величина

$$b = |b_2 - b_1|,$$

равная длине произвольного заключенного между ними отрезка особой прямой. Эту величину мы будем называть *особым углом* между параллельными прямыми.

Имея в своем распоряжении основные понятия расстояния и угла, мы можем теперь без особого труда развивать геометрию Галилея сколь угодно далеко. Мы этого здесь делать не будем, отсылая читателя к специальной литературе [1]). Конечно, теоремы этой геометрии совсем не похожи на теоремы геометрии Евклида. Например, в любом треугольнике бо́льшая сторона равна в геометрии Галилея сумме двух других сторон и, аналогично, больший угол равен сумме двух других углов.

Мы скажем только несколько слов о *геометрии прямых* в геометрии Галилея.

Каждая неособая прямая

$$x = ky + b$$

в геометрии Галилея определяется двумя «координатами» b и k. При галилеевом преобразовании (2) такая прямая переходит в прямую с уравнением

$$x' - a_2^1(y' - b^2) - b^1 = k(y' - b^2) + b,$$

т. е. с уравнением

$$x' = k'y' + b',$$

где

$$b' = b - b^2k + (b^1 - a_2^1b^2),$$

$$k' = k + a_2^1.$$

Это означает, что сопоставив прямой $x = ky + b$ точку с координатами $x = b$ и $y = k$, мы преобразованной прямой должны будем сопоставить точку с координатами

$$\begin{aligned} x' &= x + Ay + B_1, \\ y' &= y + B_2, \end{aligned} \tag{3}$$

[1]) См., например, И. М. Яглом, Принцип относительности Галилея и неевклидова геометрия, «Наука», 1969.

где

$$A = -b^2, \quad B_1 = b^1 - a_2^1 b^2, \quad B_2 = a_2^1.$$

Поскольку преобразования (3) — также галилеевы и каждое галилеево преобразование можно представить в виде (3) (при соответствующем выборе чисел b^1, b^2 и a_2^1), мы можем, следовательно, сказать, что

геометрия неособых прямых на галилеевой плоскости также является галилеевой геометрией.

Из этого утверждения легко следует, что на галилеевой плоскости справедлив следующий общий принцип:

Принцип двойственности. *Любая теорема геометрии Галилея (относящаяся к точкам и прямым) переходит в верную теорему при одновременных заменах*

«точка»	$\Rightarrow$ *«неособая прямая»*,
«расстояние»	$\Rightarrow$ *«угол»*,
«особое расстояние»	$\Rightarrow$ *«особый угол»*,
«лежат на»	$\Rightarrow$ *«проходят через»*,
«лежат на особой прямой»	$\Rightarrow$ *«параллельны»*

и т. д.

Задание. Докажите принцип двойственности.

В заключение мы приведем «словарь» для перевода основных понятий геометрии Галилея на язык кинематики (см. выше). Читателю настоятельно рекомендуется как следует его осознать и по возможности продолжить.

Геометрические понятия	Кинематические понятия
Точка	Мгновенное положение движущейся материальной точки (событие), характеризующееся местом (координата x) и временем (координата t)
Неособая прямая	Мировая линия (траектория событий) равномерно движущейся точки
Особая прямая	Совокупность всех событий, происходящих в данный момент времени
Галилеево расстояние	Временной интервал между событиями
Особое расстояние	Пространственный интервал между одновременными событиями
Галилеев угол	Относительная скорость
Особый угол	Расстояние между покоящимися друг относительно друга точками
Окружность радиуса R	Совокупность всех событий, опережающих данное событие на время R и отстающих от него на то же время

Упражнение. Дайте кинематическое истолкование принципа двойственности.

Для аналогичного геометрического описания кинематики движений в плоскости служит *пространственная геометрия Галилея*, преобразования которой являются композициями параллельных переносов, вращений вокруг оси времени Ot и преобразований вида

$$x' = x + v_1 t,$$
$$y' = y + v_2 t,$$
$$t' = t,$$

где v_1 и v_2 — компоненты скорости инерциальной системы отсчета (x, y, t) относительно системы (x', y', t').

Обозначая координату t более привычным (в геометрии) символом z, мы можем наиболее общее такое преобразование записать формулами

$$x' = x\cos\varphi - y\sin\varphi + zv_1 + x_0,$$
$$y' = x\sin\varphi + y\cos\varphi + zv_2 + y_0,$$
$$z' = z + z_0.$$

Естественно, что эта геометрия значительно сложнее плоской геометрии Галилея, и мы ею заниматься здесь не будем. Заметим только, что в этой геометрии на каждой плоскости $z = \text{const}$ имеет место обычная евклидова геометрия.

Для механики наиболее важен, конечно, случай движений в пространстве. Геометрически это соответствует рассмотрению *геометрии Галилея четырехмерного пространства — времени*. Необходимых для описания такой геометрии средств мы пока еще не имеем.

Интересным вариантом пространственной геометрии Галилея является *центрогалилеева геометрия*, преобразования которой имеют вид

$$x' = x\cos\varphi - y\sin\varphi + zv_1,$$
$$y' = x\sin\varphi + y\cos\varphi + zv_2, \qquad (4)$$
$$z' = z.$$

Рассмотрим в этой геометрии произвольную плоскость, не проходящую через начало координат O (являющееся особой точкой центрогалилеевой геометрии). При соответствующей нормировке коэффициентов уравнение такой плоскости может быть единственным образом записано в виде

$$Ax + By + Cz = 1.$$

Без труда проверяется, что преобразование (4) переводит эту плоскость в плоскость

$$A'x' + B'y' + C'z' = 1,$$

для которой

$$A' = A\cos\varphi - B\sin\varphi,$$
$$B' = A\sin\varphi + B\cos\varphi,$$
$$C' = -Av_1 - Bv_2 + C.$$

Это означает, что

геометрия плоскостей центральногалилеевой геометрии (не проходящих через особую точку O) является геометрией, преобразования которой задаются формулами вида

$$x' = x\cos\varphi - y\sin\varphi,$$
$$y' = x\sin\varphi + y\cos\varphi, \qquad (5)$$
$$z' = ax + by + z.$$

Напомним теперь, что любая система сил на плоскости полностью характеризуется (при выбранной точке O) некоторым вектором (*главным вектором* системы) и некоторым числом (*главным моментом* системы). Обозначая координаты (в некоторой прямоугольной системе координат) главного вектора символами x, y, а главный момент системы — символом z, мы можем, таким образом, сказать, что в любой прямоугольной системе координат каждая система сил на плоскости характеризуется тремя «координатами» x, y, z. При повороте координатных осей координаты x, y преобразуются, естественно, как координаты вектора, т. е. по формулам

$$x' = x\cos\varphi - y\sin\varphi,$$
$$y' = x\sin\varphi + y\cos\varphi,$$

а координата z остается неизменной. При переносе же начала координат O в некоторую точку $O'(x_0, y_0)$ координаты x, y, являясь координатами вектора, не меняются, а координата z (главный момент системы) переходит, как показывается в механике, в координату

$$z' = y_0x - x_0y + z.$$

Это означает, что статика систем сил на плоскости изучает свойства «точек» (x, y, z), инвариантные относительно преобразований вида

$$x' = x\cos\varphi - y\sin\varphi,$$
$$y' = x\sin\varphi + y\cos\varphi,$$
$$z' = y_0x - x_0y + z.$$

Поскольку эти преобразования только обозначениями отличаются от преобразований (5), мы видим, что

«геометрия плоской статики» совпадает с геометрией плоскостей центрогалилеевой геометрии, не проходящих через особую точку.

Общая теория систем сил была впервые разработана французским механиком Пуансо. В честь него геометрия с группой преобразований (5) называется *геометрией Пуансо*. Любое утверждение этой геометрии может быть переформулировано как некоторое утверждение о системах сил на плоскости, и наоборот.

Расстоянием между точками $M_1(x_1, y_1, z_1)$ и $M_2(x_2, y_2, z_2)$ в геометрии Пуансо естественно называть величину

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

инвариантную относительно всех преобразований (5).

Когда это расстояние равно нулю, для точек $M_1(x_1, y_1 z_1)$ и $M_2(x_2, y_2, z_2)$ инвариантна также разность

$$z_2 - z_1.$$

Абсолютная величина

$$\delta = |z_2 - z_1|$$

этой разности называется *особым расстоянием* между точками $M_1(x_1, y_1, z_1)$ и $M_2(x_2, y_2, z_2)$.

Дальнейшее построение геометрии Пуансо (определение угла, окружностей, сфер, построение теории треугольников и т. п.) мы оставим инициативе читателя.

Упражнение. Опишите геометрию Пуансо на плоскости.

4. Геометрия Минковского

Целый «куст» геометрий получается, когда мы выбираем на плоскости не одно, а два (различных) направления ω_1 и ω_2, и рассматриваем аффинные преобразования, переводящие прямые каждого из этих направлений снова в прямые того же направления и удовлетворяющие, возможно, некоторым дополнительным условиям (эквиаффинность, сохранение ориентаций и т. п.).

Мы здесь ограничимся геометрией, преобразования которой:

1) переводят прямые каждого из двух направлений ω_1 и ω_2 в прямые того же направления;

2) сохраняют ориентации прямых этих направлений;

3) эквиаффинны (сохраняют площади).

Такие преобразования называются *преобразованиями Лоренца*, а соответствующая геометрия — *геометрией Минковского* (или *псевдоевклидовой геометрией*). В этой геометрии прямые направлений ω_1 и ω_2 играют особую роль. Их принято называть *изотропными прямыми* геометрии Минковского (а их направления ω_1 и ω_2 — *изотропными направлениями*).

Чтобы получить из аффинной плоскости плоскость Минковского, следует выбрать

1) направления ω_1 и ω_2;

2) ориентацию прямых направления ω_1;
3) ориентацию прямых направления ω_2;
4) эталон площади.

Тогда *координатными системами Минковского* будут аффинные координатные системы, реперы Oe_1e_2 которых обладают следующими свойствами:

1) вектор e_1 параллелен прямым направления ω_1 и положительно ориентирован (по отношению к данной ориентации этих прямых);

2) вектор e_2 параллелен прямым направления ω_2 и положительно ориентирован (по отношению к данной ориентации прямых этого направления);

3) площадь параллелограмма, построенного на векторах e_1, e_2, равна единице (по отношению к данному эталону площади).

В дальнейшем мы будем считать фиксированной некоторую координатную систему Минковского Oxy.

В такой координатной системе преобразования Лоренца имеют вид

$$\begin{aligned} x' &= \lambda x + x_0, \\ y' &= \frac{1}{\lambda}\, y + y_0, \end{aligned} \tag{1}$$

где λ, x_0, y_0 — произвольные числа, причем $\lambda > 0$.

Запишем преобразование Лоренца (1) в координатах X, Y, связанных с координатами x, y формулами

$$x = \frac{X - Y}{\sqrt{2}},$$
$$y = \frac{X + Y}{\sqrt{2}}.$$

Преобразование (1) в этих координатах имеет вид

$$\frac{X' - Y'}{\sqrt{2}} = \lambda\left(\frac{X - Y}{\sqrt{2}}\right) + x_0,$$
$$\frac{X' + Y'}{\sqrt{2}} = \frac{1}{\lambda}\left(\frac{X + Y}{\sqrt{2}}\right) + y_0,$$

т. е. вид

$$\begin{aligned} X' &= \frac{1}{2}\left(\lambda + \frac{1}{\lambda}\right)X - \frac{1}{2}\left(\lambda - \frac{1}{\lambda}\right)Y + X_0, \\ Y' &= -\frac{1}{2}\left(\lambda - \frac{1}{\lambda}\right)X + \frac{1}{2}\left(\lambda + \frac{1}{\lambda}\right)Y + Y_0. \end{aligned} \tag{2}$$

Положив $\beta = \frac{\lambda^2 - 1}{\lambda^2 + 1}$, мы получим, что

$$\frac{1}{2}\left(\lambda + \frac{1}{\lambda}\right) = \frac{1}{\sqrt{1 - \beta^2}},$$
$$\frac{1}{2}\left(\lambda - \frac{1}{\lambda}\right) = \frac{\beta}{\sqrt{1 - \beta^2}}.$$

Обозначая координату X прежним символом x, а координату Y — символом t, мы получаем, следовательно, для преобразования (1) формулы вида

$$x' = \frac{x - \beta t}{\sqrt{1 - \beta^2}} + x_0,$$

$$t' = \frac{-\beta x + t}{\sqrt{1 - \beta^2}} + t_0.$$

В специальной теории относительности Эйнштейна эти формулы описывают переход от одной инерциальной системы отсчета (x, t) к другой такой системе (x', t'), движущейся относительно первой с постоянной скоростью $v = \beta c$ (где c — скорость света). Таким образом, подобно тому как геометрия Галилея является переложением на геометрический язык классической (галилеевой) кинематики, так и *геометрия Минковского является переложением кинематики теории относительности Эйнштейна.* Именно для наглядного изображения и лучшего прояснения релятивистской кинематики ее впервые Минковский и ввел.

Систему координат X, Y мы будем называть *физической системой координат*, соответствующей данной системе координат x, y.

Каждое неизотропное направление на плоскости Минковского характеризуется его угловым коэффициентом k (относительно, как обычно, оси абсцисс Ox). Произвольная прямая

$$y = kx + b$$

этого направления переходит при преобразовании (1) в прямую

$$\lambda (y' - y_0) = \frac{k}{\lambda} (x' - x_0) + b,$$

т. е. в прямую

$$y = \frac{k}{\lambda^2} x + \left(y_0 - \frac{k}{\lambda^2} x_0 + \frac{b}{\lambda} \right),$$

направление которой характеризуется угловым коэффициентом k/λ^2.

Мы видим, таким образом, что знак коэффициента k — один и тот же для всех прямых, переводимых друг в друга преобразованиями Лоренца, и потому этот знак должен иметь в геометрии Минковского определенный геометрический смысл.

Прямые с $k > 0$ называются *пространственноподобными*, а прямые с $k < 0$ — *времяподобными*. Эти названия распространяются также и на отрезки этих прямых.

Эта терминология объясняется тем, что, как можно легко показать, для любой времяподобной прямой существует физическая система координат, осью ординат которой (т. е. с физической точки зрения — осью времени) является данная прямая. Это означает, что для любого времяподобного отрезка $\overline{M_1M_2}$ существует инерциальная система отсчета, в которой события, изображаемые точками M_1 и M_2, происходят в одном и том же месте (но в разное время).

Аналогично, для любой пространственноподобной прямой существует физическая система координат, осью абсцисс (т. е. осью «мест») которой является эта прямая, так что для любого пространственноподобного отрезка $\overline{M_1M_2}$ существует инерциальная система отсчета, в которой события M_1 и M_2 происходят в одно и то же время (но в разных местах).

Для любой точки M_0 совокупность всех точек M, обладающих тем свойством, что отрезок $\overline{M_0M}$ пространственноподобен, называется *областью одновременности* точки M_0, а совокупность всех точек M, обладающих тем свойством, что отрезок $\overline{M_0M}$ времяподобен, — *областью одноместности* точки M_0.

Для начала координат O область одновременности состоит из первого и третьего квадрантов (т. е. характеризуется неравенством $xy > 0$), а область одноместности — из второго и четвертого квадрантов (характеризуется неравенством $xy < 0$).

Вообще, для любой точки M_0 область одновременности состоит из одной пары вертикальных углов, образованных изотропными прямыми, проходящими через точку M_0, а область одноместности — из другой пары.

Ясно, что для любых двух точек $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$ величина

$$(x_2 - x_1)(y_2 - y_1)$$

инвариантна относительно всех преобразований Лоренца (1). Корень квадратный

$$d = \sqrt{2(x_2 - x_1)(y_2 - y_1)}$$

из удвоенной этой величины называется *расстоянием* между точками M_1 и M_2 (или *длиной* отрезка $\overline{M_1M_2}$) в геометрии Минковского. Ясно, что

отрезок M_1M_2 тогда и только тогда изотропен (принадлежит изотропной прямой) когда его длина равна нулю.

Таким образом, в геометрии Минковского расстояние между точками может быть равно нулю даже тогда, когда эти точки не совпадают.

Предположим, что отрезок $\overline{M_1M_2}$ неизотропен. Тогда его угловой коэффициент выражается формулой

$$k = \frac{y_2 - y_1}{x_2 - x_1}$$

и потому для длины d отрезка $\overline{M_1M_2}$ мы получаем формулу

$$d = |x_2 - x_1| \sqrt{2k},$$

показывающую, что

отрезок $\overline{M_1M_2}$ тогда и только тогда пространственноподобен (времяподобен), когда его длина вещественна (соответственно, чисто мнима).

Замечание 1. До сих пор мы не фиксировали знака квадратного корня в определении длины. Теперь мы можем это сделать. Именно, мы потребуем, чтобы для пространственноподобных отрезков длина была неотрицательным числом, а для времяподобных — имела неотрицательную мнимую часть.

Простое вычисление показывает, что для пространственноподобного отрезка $\overline{M_1M_2}$ расстояние d равно обычному расстоянию между событиями, изображаемыми точками M_1 и M_2, вычисленному в инерциальной системе отсчета, *в которой эти события одновременны*. Аналогично, для времяподобного отрезка $\overline{M_1M_2}$ расстояние d равно умноженному на ic времени, протекшему между событиями M_1 и M_2 в инерциальной системе отсчета, *в которой эти события одноместны* (где c — скорость света).

Окружностью в геометрии Минковского называется, по аналогии с евклидовой геометрией, геометрическое место точек, находящихся на одном и том же расстоянии R (*радиусе* окружности) от данной точки M_0 (*центра* окружности). При этом радиус окружности может быть как вещественным (неотрицательным) числом, так и чисто мнимым (с положительной мнимой частью). Аналитически окружность радиуса R с центром в точке $M_0(x_0, y_0)$ задается уравнением

$$(x - x_0)(y - y_0) = \frac{R^2}{2} \tag{3}$$

и потому на евклидовой плоскости является (при $R \neq 0$) гиперболой с центром в точке $M_0(x_0, y_0)$. При $R = 0$ эта окружность представляет собой пару изотропных прямых, проходящих через точку M_0. При R вещественном окружность с центром M_0 расположена в области одновременности для точки M_0, а при R чисто мнимом — в области одноместности. Окружности, радиусы которых имеют один и тот же модуль $|R|$, сопряжены (как гиперболы, см. п. 6 § 2 гл. 5).

Имея понятие окружности, мы могли бы теперь по аналогии с геометриями Евклида и Галилея определить угол между двумя пересекающимися в точке M_0 прямыми как длину дуги, высекаемой этими прямыми на окружности единичного радиуса с центром в точке M_0. Однако это определение требует (в достаточной мере сложного) выяснения понятия «длина дуги окружности» в геометрии Минковского. Поэтому мы пойдем другим путем. При этом для простоты мы будем считать, что точка M_0 совпадает с точкой O (что, очевидно, не ограничивает общности).

В физических координатах X, Y, соответствующих координатам x, y, окружность радиуса R с центром в точке O имеет уравнение

$$\frac{X - Y}{\sqrt{2}} \cdot \frac{X + Y}{\sqrt{2}} = \frac{R^2}{2},$$

т. е. уравнение

$$\frac{X^2}{R^2} - \frac{Y^2}{R^2} = 1.$$

Следовательно (см. п. 3 § 1 гл. 5), положив (при $R^2 > 0$)

$$\begin{aligned} X &= R \operatorname{ch} \theta, \\ Y &= R \operatorname{sh} \theta, \end{aligned} \tag{4}$$

мы получим параметрические уравнения правой ветви этой окружности.

Эти уравнения вполне аналогичны параметрическим уравнениям

$$X = R\cos\varphi,$$
$$Y = R\sin\varphi$$

евклидовой окружности (в прямоугольных координатах X, Y), в которых параметр φ, соответствующий произвольной точке M окружности, является (ориентированным) углом, образованным прямой OM с осью абсцисс Ox.

Следуя этой аналогии, мы для любой точки M ветви (4) примем, по определению, за угол от оси Ox к прямой OM в геометрии Минковского значение параметра θ, отвечающее в силу уравнений (4) этой точке.

Далее, для любых двух точек M_1 и M_2 ветви (4) мы определим *угол* θ *от прямой* OM_1 *к прямой* OM_2 формулой

$$\theta = \theta_2 - \theta_1,$$

где θ_2 и θ_1 — значения параметра, отвечающие точкам M_2 и M_1.

За угол *между прямыми* OM_1 и OM_2 мы примем абсолютную величину угла θ.

Заметим, что, в отличие от евклидовой геометрии, угол θ от оси Ox к прямой OM принимает любые вещественные значения: когда прямая OM, вращаясь, приближается к асимптоте $X = Y$ гиперболы (4) (являющейся изотропной прямой $x = 0$), угол θ стремится к $+\infty$, а когда эта прямая приближается к асимптоте $X = -Y$ (т. е. к изотропной прямой $y = 0$), угол θ стремится к $-\infty$.

Мы ввели угол между прямыми в геометрии Минковского, руководствуясь довольно формальной аналогией с евклидовой геометрией. Более содержательное оправдание нашего определения можно получить, заметив, что в евклидовой геометрии угол φ между двумя прямыми, пересекающимися в точке M_0, равен удвоенной площади кругового сектора, высекаемого этими прямыми из круга единичного радиуса с центром в точке M_0. Аналогом этого сектора в геометрии Минковского является гиперболический сектор $M_0M_1M_2$, ограниченный отрезками $\overline{M_0M_1}$, $\overline{M_0M_2}$ и дугой $\overset{\frown}{M_1M_2}$ гиперболы (4). Пользуясь, скажем, известными из интегрального исчисления формулами для площадей плоских фигур, можно без труда показать, что площадь этого сектора равна $\frac{\theta_2 - \theta_1}{2}$. С другой стороны, поскольку преобразования Лоренца эквиаффинны, понятие евклидовой площади сохраняется и в геометрии Минковского. Поэтому так же, как и в евклидовой геометрии,

угол между двумя прямыми в геометрии Минковского равен удвоенной площади сектора, высекаемого этими прямыми из «единичного круга».

Упражнение. Проверьте, что это утверждение верно и в геометрии Галилея.

Точка M, имеющая в физической координатной системе координаты $X = R\,\mathrm{ch}\,\theta$ и $Y = R\,\mathrm{sh}\,\theta$, имеет в исходной

координатной системе Oxy координаты

$$x = \frac{R}{\sqrt{2}}(\operatorname{ch}\theta - \operatorname{sh}\theta),$$

$$y = \frac{R}{\sqrt{2}}(\operatorname{ch}\theta + \operatorname{sh}\theta).$$

Поэтому угловой коэффициент k прямой OM (в координатной системе Oxy) выражается формулой

$$k = \frac{\operatorname{ch}\theta + \operatorname{sh}\theta}{\operatorname{ch}\theta - \operatorname{sh}\theta} = \frac{1 + \operatorname{th}\theta}{1 - \operatorname{th}\theta}.$$

Следовательно,

$$\operatorname{th}\theta = \frac{k-1}{k+1}.$$

Пусть теперь нам даны две точки M_1 и M_2 и пусть k_1 — угловой коэффициент прямой OM_1, а k_2 — угловой коэффициент прямой OM_2. Тогда для угла θ между прямыми OM_1 и OM_2 мы получаем, что

$$\operatorname{th}\theta = \operatorname{th}(\theta_2 - \theta_1) = \frac{\operatorname{th}\theta_2 - \operatorname{th}\theta_1}{1 - \operatorname{th}\theta_1 \operatorname{th}\theta_2} = \frac{\dfrac{k_2-1}{k_2+1} - \dfrac{k_1-1}{k_1+1}}{1 - \dfrac{k_1-1}{k_1+1}\cdot\dfrac{k_2-1}{k_2+1}} = \frac{k_2 - k_1}{k_2 + k_1}.$$

Тем самым доказано, что
угол θ в геометрии Минковского между прямыми

$$y = k_1 x + b_1, \quad y = k_2 x + b_2$$

вычисляется по формуле

$$\operatorname{th}\theta = \frac{k_2 - k_1}{k_2 + k_1}. \tag{5}$$

Замечание 2. Последнюю формулу мы могли бы принять за о п р е д е л е н и е угла θ. Этот путь был бы, конечно, самым простым. Однако геометрический смысл угла был бы при этом совершенно потерян.

Внимательный читатель, без сомнения, уже заметил, что наше определение угла пригодно не для любых прямых. Действительно, кривая (4) является ветвью окружности, квадрат R^2 радиуса которой положителен (т. е. является ветвью окружности вещественного радиуса R). Поэтому для любой ее точки M прямая OM пространственноподобна. Следовательно, наше определение угла пригодно только для пространственноподобных прямых.

Но для окружности мнимого радиуса iR параметрические уравнения некоторой ее ветви имеют вид

$$X = R\operatorname{sh}\theta,$$

$$Y = R\operatorname{ch}\theta,$$

и для любой точки M этой ветви прямая OM времяподобна. Поэтому, поступая, как и выше, мы примем параметр θ за угол от прямой OM к оси ординат Oy, а затем определим *угол от прямой OM_1 к прямой OM_2* формулой

$$\theta = \theta_1 - \theta_2$$

(обратим внимание на перемену ролей точек M_1 и M_2). Тем самым угол будет определен и для времяподобных прямых. Ясно, что он также будет равен удвоенной площади соответствующего гиперболического сектора. Формула (5) для него также останется справедливой.

Для прямых, одна из которых времяподобна, а другая пространственноподобна, угол не определяется (формула (5) дает в этом случае для θ комплексное значение).

Имея в своем распоряжении понятия длины и угла, мы можем теперь развивать геометрию Минковского сколь угодно далеко. Мы этого делать не будем. В следующем пункте мы осветим аналогию между геометриями Евклида и Минковского с другой, совершенно неожиданной стороны.

5. Комплексная евклидова геометрия

Как уже отмечалось (замечание 3 п. 4 § 3 гл. 2), построение евклидовой геометрии возможно и над полем комплексных чисел $\mathbb{C}$. В этой геометрии имеет смысл понятие *расстояния* d между любыми точками $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$, определяющегося обычной формулой

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Однако получающаяся геометрия малоудовлетворительна, поскольку это расстояние может быть любым комплексным числом (и к тому же определенным только с точностью до знака) и может быть равно нулю даже в том случае, если M_1 и M_2 различны. По этой причине *угол* между двумя прямыми (также определяемый по обычной формуле) оказывается не всегда существующим (и, вообще говоря, представляет собой комплексное число). Правда, с аналогичными неприятностями мы уже встречались в геометрии Минковского, но ввиду возможности физической интерпретации этой геометрии нам приходилось с ними мириться. В то же время комплексная евклидова геометрия появляется лишь в порядке не слишком глубокого обобщения и никаких интересных приложений и интерпретаций даже внутри математики (не говоря уже о физике) не имеет. Поэтому разбираться в ее сложностях в принципе не очень интересно. В крайнем случае это было бы оправдано, если бы теоремы этой геометрии оказались в достаточной мере красивы и изящны. Однако на самом деле все обстоит как раз

наоборот (это общее правило: неудачная математическая теория, не имеющая приложений, никогда не отличается изяществом и красотой). По всем этим причинам комплексную евклидову геометрию мы выше и не рассматривали.

Иначе дело обстоит, конечно, для *евклидовой вещественно-комплексной геометрии* (п. 5 § 3 гл. 2), на множестве вещественных точек которой имеет место обычная евклидова геометрия. Однако эта геометрия имеет лишь вспомогательное значение: она служит для удобного описания феноменов вещественной геометрии, которые трудно описать, не выходя в «мнимую область». Ее группа автоморфизмов по существу совпадает с группой автоморфизмов (ортогональных преобразований) вещественной евклидовой геометрии, и отличие между этими геометриями состоит только в том, что первая имеет «больше» точек.

Заметим, что «вещественно-комплексный» вариант возможен для любой вещественной геометрии. Можно говорить, например, о *вещественно-комплексной геометрии Галилея*, *вещественно-комплексной геометрии Минковского* и даже о *вещественно-комплексной конформной геометрии*.

Переход от евклидовой комплексной геометрии к евклидовой вещественно-комплексной геометрии можно описать также следующим образом (ср. п. 5 § 1 гл. 4). Мы выбираем в комплексной плоскости $\mathfrak{M}$ вещественную плоскость $\mathfrak{M}^{\text{вещ}}$, характеризующуюся требованием вещественности координат всех ее точек (по отношению к данной координатной системе), и ограничиваемся рассмотрением только тех преобразований, которые эту вещественную плоскость переводят в себя.

Чтобы получить затем вещественную геометрию, остается откинуть мнимые точки.

Таким образом, вещественная евклидова геометрия получается из комплексной в два этапа: на первом этапе мы, оставляя неизменным множество точек, сокращаем группу до группы всех вещественных преобразований, а на втором — сокращаем множество точек до множества, состоящего из точек, все координаты которых вещественны.

Поскольку каждая линия в вещественной плоскости $\mathfrak{M}^{\text{вещ}}$ получается сечением соответствующей линии в комплексной плоскости $\mathfrak{M}$ плоскостью $\mathfrak{M}^{\text{вещ}}$, иногда говорят, что вещественная евклидова геометрия является *сечением* комплексной евклидовой геометрии, получающимся наложением условий

«x вещественно», «y вещественно».

Другое сечение комплексной евклидовой геометрии мы получим, наложив на координаты x, y условия

«x вещественно», «y чисто мнимо».

Множество всех точек (x, y), координаты которых удовлетворяют этим условиям, мы обозначим через $\mathfrak{M}^{\text{м}}$. Каждая точка

этого множества характеризуется двумя вещественными координатами X, Y, связанными с координатами x, y соотношениями

$$x = X,$$
$$y = iY.$$

Таким образом, множество $\mathfrak{M}^{\text{м}}$ мы также можем рассматривать как вещественную плоскость (с координатами X, Y).

Преобразование

$$\begin{aligned} x' &= a_1^1 x + a_2^1 y + b^1, \\ y' &= a_1^2 x + a_2^2 y + b^2 \end{aligned} \tag{1}$$

тогда и только тогда переводит каждую точку из $\mathfrak{M}^{\text{м}}$ снова в точку из $\mathfrak{M}^{\text{м}}$, когда для любых вещественных чисел X и Y числа

$$a_1^1 X + i a_2^1 Y + b^1 \quad \text{и} \quad \frac{1}{i}(a_1^2 X + i a_2^2 Y + b^2)$$

также вещественны. Очевидно, что это имеет место тогда и только тогда, когда числа a_1^1, a_2^2 и b^1 вещественны, а числа a_2^1, a_1^2 и b^1 чисто мнимы. Мы положим

$$\begin{aligned} a_1^1 &= A_1^1, \quad a_2^1 = iA_2^1, \quad b^1 = B^1, \\ a_1^2 &= iA_1^2, \quad a_2^2 = A_2^2, \quad b^2 = iB^2. \end{aligned}$$

Тогда условия

$$(a_1^1)^2 + (a_2^1)^2 = 1, \quad a_1^1 a_1^2 + a_2^1 a_2^2 = 0, \quad (a_1^2)^2 + (a_2^2)^2 = 1$$

ортогональности преобразования (1) перепишутся в следующем виде:

$$(A_1^1)^2 - (A_2^1)^2 = 1, \quad A_1^1 A_1^2 + A_2^1 A_2^2 = 0, \quad -(A_1^2)^2 + (A_2^2)^2 = 1. \tag{2}$$

Мы найдем общее решение этих уравнений при дополнительном предположении

$$A_1^1 > 0, \quad A_2^2 > 0. \tag{3}$$

Таким образом, мы несколько сократим интересующую нас группу. Это сокращение совершенно несущественно, и от геометрии с сокращенной группой мы всегда без особого труда можем вернуться к геометрии с полной группой.

Мы будем искать коэффициент A_1^1 в виде

$$A_1^1 = \frac{1}{2}\left(\mu + \frac{1}{\mu}\right),$$

где μ — некоторый положительный параметр. Подставляя это выражение в первое из уравнений (2), мы немедленно получим, что

$$(A_2^1)^2 = \frac{1}{4}\left(\mu + \frac{1}{\mu}\right)^2 - 1 = \frac{1}{4}\left(\mu - \frac{1}{\mu}\right)^2.$$

Заменяя в случае необходимости μ на $1/\mu$, мы можем общее решение этого уравнения записать в следующем виде:

$$A_2^1 = \frac{1}{2}\left(\mu - \frac{1}{\mu}\right).$$

Аналогично, общее решение третьего из уравнений (2) имеет вид

$$A_1^2 = \frac{1}{2}\left(\lambda - \frac{1}{\lambda}\right), \quad A_2^2 = \frac{1}{2}\left(\lambda + \frac{1}{\lambda}\right),$$

где λ — некоторый положительный параметр.

Подставляя эти выражения во второе уравнение (2), мы после тривиальных преобразований получим, что

$$\lambda = \frac{1}{\mu}.$$

Заметив, что преобразование, индуцированное на плоскости $\mathfrak{M}^{\text{м}}$ преобразованием (1), имеет в координатах X, Y вид

$$X' = A_1^1 X - A_2^1 Y + B^1,$$
$$Y' = A_1^2 X + A_2^2 Y + B^2,$$

мы окончательно получаем (полагая $B^1 = X_0$ и $B^2 = Y_0$), что

$$X' = \frac{1}{2}\left(\lambda + \frac{1}{\lambda}\right) X + \frac{1}{2}\left(\lambda - \frac{1}{\lambda}\right) Y + X_0,$$
$$Y' = \frac{1}{2}\left(\lambda - \frac{1}{\lambda}\right) X + \frac{1}{2}\left(\lambda + \frac{1}{\lambda}\right) Y + Y_0.$$

Поскольку эти формулы совпадают с формулами, выражающими преобразование Лоренца в физических координатах (см. п. 4), мы видим, следовательно, что

сечением евклидовой комплексной геометрии, получающимся наложением условий

«*x вещественно*», «*y чисто мнимо*»,

(*а также условий* (3)) *является геометрия Минковского.*

Таким образом,

как евклидова геометрия, так и геометрия Минковского (*псевдоевклидова геометрия*) *являются сечениями комплексной евклидовой геометрии.*

В частности, отсюда вытекает, что

подстановка $y \mapsto iy$ *переводит формулы евклидовой геометрии в формулы геометрии Минковского.*

Последнее обстоятельство читатель, возможно, уже давно заметил. Теперь же он знает и объяснение этому.

Обратим внимание на то, что установленная связь между евклидовой геометрией и геометрией Минковского позволяет нам почти автоматически получать теоремы геометрии Минковского из соответствующих евклидовых теорем.

6. Унитарная геометрия

Вернемся теперь снова к задаче «разумного» обобщения евклидовой геометрии на комплексный случай. Как мы видели, «прямолинейное» обобщение, приводящее к комплексной евклидовой геометрии, по существу не достигает цели. Оказывается, что более удовлетворительная геометрия получается, если мы наложим на коэффициенты преобразования (1) п. 5 следующие условия, сводящиеся в вещественном случае к тем же соотношениям ортогональности:

$$a_1^1\bar{a}_1^1 + a_2^1\bar{a}_2^1 = 1, \quad a_1^1\bar{a}_1^2 + a_2^1\bar{a}_2^2 = 0, \quad a_1^2\bar{a}_1^2 + a_2^2\bar{a}_2^2 = 1. \quad (1)$$

Здесь, конечно, необходимо в первую очередь доказать, что преобразования, коэффициенты которых удовлетворяют этим условиям, образуют группу. Мы докажем сначала, что группу составляют всевозможные матрицы

$$A = \begin{pmatrix} a_1^1 & a_2^1 \\ a_1^2 & a_2^2 \end{pmatrix}, \quad (2)$$

элементы которых удовлетворяют условиям (1).

Определение 1. Матрицы (2), элементы которых удовлетворяют условиям (1), называются *унитарными матрицами.*

Полагая

$$\bar{A} = \begin{pmatrix} \bar{a}_1^1 & \bar{a}_2^1 \\ \bar{a}_1^2 & \bar{a}_2^2 \end{pmatrix},$$

мы можем условия (1) переписать в следующем виде:

$$A\bar{A}^{\top} = E. \quad (3)$$

Перейдя в этом соотношении к определителям, мы немедленно получим равенство

$$|A| \cdot \overline{|A|} = 1,$$

означающее, что

определитель унитарной матрицы является комплексным числом, по модулю равным единице (и потому отличным от нуля).

Умножив соотношение (3) на матрицу A^{-1}, получаем

$$A^{-1} = \bar{A}^{\top}.$$

Таким образом,

матрицей, обратной к унитарной матрице A, является матрица $\bar{A}^{\top}$.

Поэтому

для любой унитарной матрицы A имеет место соотношение

$$\bar{A}^{\top}A = E.$$

Поскольку $A=\overline{(\overline{A}^{\mathrm{T}})}^{\mathrm{T}}$, мы можем это соотношение переписать в следующем виде:

$$\overline{A}^{\mathrm{T}}\,\overline{(\overline{A}^{\mathrm{T}})}^{\mathrm{T}}=E,$$

т. е. в виде

$$A^{-1}\overline{A^{-1}}^{\mathrm{T}}=E,$$

показывающем, что

матрица A^{-1}, обратная к унитарной матрице A, унитарна.

Наконец, если

$$A\overline{A}^{\mathrm{T}}=E,\quad B\overline{B}^{\mathrm{T}}=E,$$

то

$$(AB)\,\overline{(AB)}^{\mathrm{T}}=AB\overline{B}^{\mathrm{T}}\overline{A}^{\mathrm{T}}=A\overline{A}^{\mathrm{T}}=E.$$

Следовательно,

произведение AB двух унитарных матриц A и B также является унитарной матрицей.

Заметим, что все сказанное справедливо не только для матриц второго порядка (с которых мы начинали), но и для матриц любого порядка n. Совокупность всех унитарных матриц порядка n принято обозначать символом $U(n)$. Последние два доказанные выше утверждения означают, что

множество $U(n)$ всех унитарных матриц порядка n является группой.

Вернемся теперь к преобразованиям (1) п. 5, коэффициенты которых удовлетворяют соотношениям (1). В матричном виде эти преобразования записываются формулой

$$\begin{pmatrix}x'\\y'\end{pmatrix}=A\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}b^1\\b^2\end{pmatrix},\tag{4}$$

где A — произвольная унитарная матрица.

Из того, что унитарные матрицы образуют группу, немедленно вытекает, что

преобразования (6) *составляют группу.*

Соответствующая геометрия называется *унитарной геометрией.*

Без труда проверяется, что в унитарной геометрии фигура, состоящая из двух точек $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$, имеет инвариант, выражающийся формулой

$$(x_2-x_1)(\bar{x}_2-\bar{x}_1)+(y_2-y_1)(\bar{y}_2-\bar{y}_1)=|x_2-x_1|^2+|y_2-y_1|^2.$$

Этот инвариант является неотрицательным вещественным числом и арифметический квадратный корень из него

$$d=\sqrt{|x_2-x_1|^2+|y_2-y_1|^2}$$

называется (*унитарным*) *расстоянием* между точками $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$.

Это расстояние равно нулю тогда и только тогда, когда точки M_1 и M_2 совпадают. Поэтому можно ожидать, что унитарная геометрия и является искомым «разумным» обобщением евклидовой геометрии на комплексный случай. Оказывается, что это действительно так: во всяком случае, многие утверждения евклидовой геометрии естественным образом переносятся в унитарную геометрию, и получающаяся теория не только красива и изящна, но и имеет многочисленные применения. К сожалению, у нас здесь нет ни места, ни времени для более детального ознакомления с этой замечательной геометрией.

Отметим только, что унитарная геометрия тоже «не без греха»: окружности в ней выражаются уравнениями вида

$$|x - x_0|^2 + |y - y_0|^2 = R^2$$

и потому не являются алгебраическими линиями. Это обстоятельство, кажущееся на первый взгляд малосущественным, на самом деле приводит часто к серьезным осложнениям (или, если хотите,— к интересной специфике унитарной геометрии, резко отличающей ее от евклидовой).

Накладывая на унитарную геометрию ограничения

«x вещественно», «y вещественно»,

мы получаем, очевидно, евклидову геометрию. Интересно отметить, что ограничения

«x вещественно», «y чисто мнимо»

также приводят к евклидовой геометрии.

7. Геометрия Пуанкаре — Лобачевского

В соответствии с общими принципами п. 1 каждая подгруппа группы Conf определяет на конформной плоскости некоторую геометрию, в которой имеет смысл понятие окружности.

Примером такой геометрии является специальная конформная геометрия с группой Conf^+.

Другой пример мы получим, зафиксировав на конформной плоскости некоторую окружность и рассмотрев подгруппу группы Conf (или группы Conf^+), состоящую из всех конформных преобразований, переводящих эту окружность в себя.

Ясно, что для любых двух окружностей получающиеся таким образом геометрии изоморфны. Поэтому достаточно ограничиться некоторой одной окружностью. В качестве такой окружности мы выберем вещественную ось $y = 0$ на стандартной конформной плоскости $\mathbb{C}^+$ (напомним, что $z = x + iy$). Кроме того, мы ограничимся лишь преобразованиями из группы Conf^+. Соответствующая геометрия называется *геометрией Пуанкаре.*

По определению, группой автоморфизмов геометрии Пуанкаре является подгруппа Conf_0^+ группы Conf^+, состоящая из

всех дробно-линейных преобразований

$$z' = \frac{az + b}{cz + d},$$

переводящих прямую $y = 0$ в себя, т. е. обладающих тем свойством, что для любого вещественного z число z' вещественно.

Полагая $z = 0$, мы видим, что число b/d должно быть вещественным, а полагая $z = \pm 1$,— что вещественными должны быть и числа $(\pm a + b)/(\pm c + d)$. Отсюда непосредственно вытекает, что коэффициенты a, b, c, d должны быть пропорциональны вещественным числам. Ясно, что это необходимое условие также и достаточно. Сокращая коэффициенты на общий множитель, мы получаем, таким образом, что

любое преобразование из группы Conf_0^+ *имеет вид*

$$z' = \frac{az + b}{cz + d}, \tag{1}$$

где a, b, c, d *— такие вещественные числа, что* $\begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0$.

Нормируя коэффициенты a, b, c, d, мы можем добиться того, чтобы последний определитель был равен ± 1. Ясно, что

совокупность Conf_0^{++} *всех преобразований* (1), *для которых*

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = 1,$$

является подгруппой группы Conf_0^+.

Геометрически преобразования из группы Conf_0^{++} характеризуются тем, что они сохраняют ориентацию вещественной оси.

Оказывается целесообразным вместо геометрии с группой Conf_0^+ изучать геометрию с группой Conf_0^{++}, устроенную несколько проще. Эти две геометрии настолько тесно связаны, что обратный переход от геометрии с группой Conf_0^{++} к геометрии с группой Conf_0^+ никаких затруднений не вызывает.

Каждое преобразование из группы Conf_0^{++} переводит точку $z = x + iy$ в точку $z' = x' + iy'$, для которой

$$x' = \frac{(ax + b)(cx + d) + acy^2}{(cx + d)^2 + (cy)^2}, \qquad y' = \frac{y}{(cx + d)^2 + (cy)^2}.$$

Следовательно, если $y > 0$, то $y' > 0$, а если $y < 0$, то $y' < 0$.

Это показывает, что геометрия с группой Conf_0^{++} в определенном смысле распадается на две геометрии, «областями действия» которых являются «верхняя полуплоскость» $y > 0$ и «нижняя полуплоскость» $y < 0$. Эти две геометрии очевидным образом изоморфны, и потому достаточно ограничиться изучением лишь геометрии точек верхней полуплоскости. Эта геометрия называется *геометрией Пуанкаре на верхней полуплоскости.*

В связи с этим верхнюю полуплоскость $y > 0$ часто называют *плоскостью Пуанкаре*. Она обозначается символом $\mathbb{H}^+$.

От каждой окружности из $\mathbb{C}^+$ в верхней полуплоскости $\mathbb{H}^+$ остается, вообще говоря, одна дуга. Точнее, возможны следующие четыре случая:

1) остается вся окружность Σ;

2) остается вся окружность, за исключением одной точки (случай, когда окружность Σ касается вещественной оси);

3) остается полуокружность (случай, когда окружность Σ ортогональна вещественной оси);

4) остается дуга окружности, не являющаяся полуокружностью (случай, когда окружность Σ образует с вещественной осью угол $\neq \pi/2$).

Полуокружности, ортогональные вещественной оси (случай 3), называются *прямыми* геометрии Пуанкаре. К ним принадлежат, в частности, евклидовы полупрямые, ортогональные вещественной оси (т. е. «вертикальные» полупрямые, параллельные мнимой оси $x = 0$). Ясно, что

через любые две различные точки полуплоскости $\mathbb{H}^+$ *проходит точно одна прямая.*

Это свойство и оправдывает термин «прямая» в применении к этим полуокружностям.

Поскольку прямая в геометрии Пуанкаре является незамкнутой линией, направление на ней однозначно характеризуется двумя точками. Прямую, проходящую через точки z_1, z_2 и ориентированную в направлении от точки z_1 к точке z_2, мы будем обозначать символом $[z_1, z_2]$.

Прямая $[z_1, z_2]$ пересекает вещественную ось $y = 0$ в двух точках x_1 и x_2 (одной из этих точек может быть точка ∞). Предполагая, что эти точки занумерованы так, что точка x_1 предшествует точке z_1 на ориентированной прямой $[z_1, z_2]$, рассмотрим двойное отношение точек z_1, z_2, x_1, x_2:

$$W(z_1, z_2) = \frac{z_1 - x_1}{z_2 - x_1} : \frac{z_1 - x_2}{z_2 - x_2}.$$

При произвольном преобразовании из группы Conf_0^{++} точки z_1, z_2 переходят в некоторые точки z_1', z_2', прямая $[z_1, z_2]$ переходит в прямую $[z_1', z_2']$, а точки x_1, x_2 — в точки пересечения x_1', x_2' прямой $[z_1', z_2']$ с вещественной осью. Поэтому, в силу инвариантности двойного отношения,

$$W(z_1', z_2') = W(z_1, z_2).$$

Таким образом,

в геометрии Пуанкаре две точки обладают инвариантом $W(z_1, z_2)$.

Это наводит на мысль, что в геометрии Пуанкаре должно иметь смысл какое-то понятие расстояния между точками (являющееся, вообще говоря, некоторой функцией числа W). Чтобы

выяснить этот вопрос, изучим более внимательно инвариант $W(z_1, z_2)$.

Во-первых, ясно, что

число $W(z_1, z_2)$ вещественно.

Действительно, точки z_1, z_2, x_1, x_2 принадлежат, по построению, одной окружности.

Заметим теперь, что в группе Conf_0^{++} существует (и даже не одно) преобразование, переводящее точку x_1 в точку 0, а точку x_2 — в точку ∞. Например, если обе точки x_1 и x_2 конечны, такое преобразование можно задать формулой

$$z' = \frac{1}{x_1 - x_2} \cdot \frac{z - x_1}{z - x_2}.$$

Прямую $[z_1, z_2]$ это преобразование переводит в прямую, проходящую через точки 0 и ∞, т. е. в мнимую ось $x = 0$. Пусть iy_1 и iy_2 — образы точек z_1 и z_2 при этом преобразовании. Поскольку дробно-линейное преобразование сохраняет порядок точек на окружности, точка 0 предшествует точке iy_1 на мнимой оси, точка iy_1 предшествует точке iy_2, а точка iy_2 предшествует точке ∞. Другими словами,

$$0 < y_1 < y_2 < \infty.$$

Поскольку, в силу инвариантности,

$$W(z_1, z_2) = W(iy_1, iy_2) = \frac{iy_1 - 0}{iy_2 - 0} = \frac{y_1}{y_2},$$

отсюда немедленно вытекает, что

вещественное число $W(z_1, z_2)$ положительно и не превосходит единицы:

$$0 < W(z_1, z_2) < 1.$$

Наконец, легко видеть, что

для любых трех точек z_1, z_2, z_3, принадлежащих одной прямой (геометрии Пуанкаре) и расположенных так, что точка z_2 находится между точками z_1 и z_3, имеет место равенство

$$W(z_1, z_3) = W(z_1, z_2)\, W(z_2, z_3).$$

Действительно, условие на расположение точек z_1, z_2, z_3 означает, что ориентированные прямые $[z_1, z_2]$, $[z_2, z_3]$ и $[z_1, z_3]$ совпадают. Поэтому

$$W(z_1, z_2) = \frac{z_1 - x_1}{z_2 - x_1} : \frac{z_1 - x_2}{z_2 - x_2},$$

$$W(z_2, z_3) = \frac{z_2 - x_1}{z_3 - x_1} : \frac{z_2 - x_2}{z_3 - x_2},$$

$$W(z_1, z_3) = \frac{z_1 - x_1}{z_3 - x_1} : \frac{z_1 - x_2}{z_3 - x_2},$$

где x_1 и x_2 — одни и те же во всех трех формулах. Требуемое соотношение вытекает отсюда непосредственно.

Мы видим, таким образом, что число

$$|z_1, z_2| = -\log W(z_1, z_2)$$

для любых (различных) точек z_1, z_2 верхней полуплоскости положительно (если основание логарифмов больше единицы) и обладает тем свойством, что

$$|z_1, z_3| = |z_1, z_2| + |z_2, z_3|$$

каждый раз, когда точка z_2 принадлежит прямой $[z_1, z_3]$ и расположена между точками z_1 и z_3.

Это показывает, что число $|z_1, z_2|$ обладает одним из важнейших свойств евклидова расстояния. Оно называется *расстоянием Пуанкаре* между точками z_1 и z_2.

Для случая $z_1 = z_2$, мы, естественно, полагаем

$$|z_1, z_1| = 0.$$

Тогда расстояние $|z_1, z_2|$ будет определено для любой пары z_1, z_2 точек полуплоскости $\mathbb{H}^+$ и будет равно нулю тогда и только тогда, когда эти точки совпадают.

Расстояние $|z_1, z_2|$, очевидно, симметрично:

$$|z_1, z_2| = |z_2, z_1|,$$

поскольку при перестановке точек z_1, z_2 переставляются также и точки x_1, x_2.

Кроме того, можно показать, что оно удовлетворяет неравенству треугольника, т. е.

$$|z_1, z_3| \leqslant |z_1, z_2| + |z_2, z_3|$$

для любых трех точек z_1, z_2, z_3.

Задание. Докажите неравенство треугольника для расстояния Пуанкаре.

Заметим, что расстояние Пуанкаре определено только с точностью до множителя (зависящего от основания логарифмов). Чтобы его фиксировать, нужно, как и в евклидовой геометрии, выбрать эталон длины, т. е. отрезок, длина которого принимается за единицу.

При приближении точки z_2 по прямой $[z_1, z_2]$ к точке x_2 расстояние $|z_1, z_2|$ стремится, очевидно, к бесконечности. Аналогично, оно стремится к бесконечности и при $z_1 \to x_1$. Таким образом, точки x_1 и x_2 мы можем рассматривать как *бесконечно удаленные точки* прямой $[z_1, z_2]$.

Обратим внимание на то, что, в отличие от евклидовой геометрии, прямые в геометрии Пуанкаре имеют, таким образом, две различные бесконечно удаленные точки, которые можно условно называть «левой» и «правой». Для вертикальных полупрямых одной из этих точек является конечная точка вещественной оси, а другой — точка ∞.

В евклидовой геометрии параллельные прямые можно рассматривать как прямые, имеющие общую бесконечно удаленную

(несобственную) точку. По аналогии мы будем две прямые плоскости Пуанкаре называть *параллельными*, если они имеют общую бесконечно удаленную точку. Поскольку прямая имеет в геометрии Пуанкаре две бесконечно удаленные точки, через каждую точку, не принадлежащую прямой, проходят две прямые, параллельные данной: «левая» и «правая». Любая прямая, лежащая «между» этими параллельными прямыми, не пересекает данной прямой.

Читатель, знакомый с геометрией Лобачевского, немедленно обнаружит в этом аксиому параллельности Лобачевского. Более того, без труда проверяется, что все аксиомы геометрии Лобачевского выполнены в геометрии Пуанкаре (об этих аксиомах см. дополнение к § 3 гл. 2). Это означает, что

геометрия Пуанкаре является одной из интерпретаций геометрии Лобачевского.

Упражнение. Проверьте аксиомы геометрии Лобачевского в геометрии Пуанкаре.

Поэтому обычно термин «геометрия Пуанкаре» не употребляется: предпочитают говорить об *интерпретации Пуанкаре геометрии Лобачевского.*

Таким образом, в интерпретации Пуанкаре

точками плоскости Лобачевского считаются, по определению, точки верхней полуплоскости $\mathbb{H}^+$,

прямыми — полуокружности, ортогональные оси $y = 0$,

движениями — преобразования из группы Conf_0^{++}.

Замечательным свойством интерпретации Пуанкаре является тот факт, что *углы между прямыми в этой интерпретации совпадают с обычными евклидовыми углами* (между окружностями).

Как уже отмечалось, интерпретацию Пуанкаре необязательно строить в верхней полуплоскости: можно взять любую область в $\mathbb{C}^+$, ограниченную окружностью, например, *единичный круг*

$$|z| < 1,$$

ограниченный единичной окружностью $|z| = 1$. Соответствующая интерпретация называется *интерпретацией Пуанкаре в единичном круге.* В этой интерпретации прямыми являются полуокружности, ортогональные единичной окружности.

Эту интерпретацию мы по существу уже рассматривали в дополнении к § 3 гл. 4. Теперь мы, во-первых, можем указать, как измерять длины отрезков в этой интерпретации (длина отрезка с концами в точках z_1 и z_2 равна логарифму двойного отношения этих точек и точек пересечения с единичной окружностью окружности, проходящей через точки z_1 и z_2 и ортогональной единичной окружности), а во-вторых, можем описать все движения: ими будут сохраняющие ориентации конформные преобразования, переводящие единичный круг в себя.

Найдем явные формулы для этих движений.

Пусть сохраняющее ориентации конформное преобразование переводит в себя единичный круг. Рассмотрим точку z_0 единичного круга, переходящую при этом преобразовании в центр 0 круга. Ясно, что точка $1/\bar{z}_0$, симметричная точке z_0 относительно единичной окружности, переходит при этом преобразовании в точку ∞, симметричную точке 0. Действительно, симметричная точка характеризуется (см. п. 1 § 3) тем, что любая окружность, проходящая через точку z_0 и ортогональная единичной окружности, проходит и через точку $1/\bar{z}_0$. Поэтому любое конформное преобразование, переводящее в себя единичную окружность, переводит симметричные точки в симметричные.

Но ясно, что любое сохраняющее ориентации конформное преобразование, переводящее точку z_0 в точку 0, а точку $1/\bar{z}$ — в точку ∞, выражается формулой

$$z' = a\frac{z - z_0}{z - \frac{1}{\bar{z}}},$$

т. е. формулой

$$z' = (-a\bar{z}_0)\frac{z - z_0}{1 - \bar{z}_0 z}, \qquad (2)$$

где a — произвольное число.

Условие, что конформное преобразование (2) переводит единичную окружность в единичную, т. е. условие, что $|z'| = 1$ при $|z| = 1$, накладывает некоторые ограничения на число a. Действительно, при $|z| = 1$

$$|1 - \bar{z}_0 z| = \left|\frac{1}{z} - \bar{z}_0\right| = |\bar{z} - \bar{z}_0| = |z - z_0|.$$

Поэтому $|z'| = 1$ при $|z| = 1$ тогда и только тогда, когда $|a\bar{z}_0| = 1$. Полагая $-a\bar{z}_0 = e^{i\theta}$, мы получаем, что

любое сохраняющее ориентации конформное преобразование, переводящее единичный круг в себя, имеет вид

$$z' = e^{i\theta}\frac{z - z_0}{1 - \bar{z}_0 z},$$

где $|z_0| < 1$, *а* θ *произвольно.*

У нас здесь нет ни времени, ни места подробно изучать геометрию Пуанкаре — Лобачевского. Мы ограничимся лишь тем, что охарактеризуем линии, изображающиеся евклидовыми окружностями (или, точнее, их дугами).

Как выше было отмечено, эти линии принадлежат к четырем различным типам 1)—4). Линии типа 3) мы уже рассмотрели: это прямые геометрии Пуанкаре. Займемся поэтому линиями оставшихся типов 1), 2) и 4).

Рассмотрим сначала линии типа 1) (окружности, целиком лежащие в верхней полуплоскости). Окружности, ортогонально пересекающие такую окружность Σ и ортогональные вещественной оси $y=0$, составляют, очевидно, пучок (являющийся пересечением гиперболической связки окружностей, ортогональных окружности Σ, и гиперболической связки окружностей, ортогональных вещественной оси). Этому пучку принадлежит, в частности, вертикальная прямая, проходящая через евклидов центр окружности Σ. Поскольку эта прямая пересекает, как нетрудно видеть, любую другую окружность пучка, этот пучок является эллиптическим пучком, и потому все его окружности проходят через две фиксированные точки. Поскольку линией центров этого пучка является, очевидно, вещественная ось, эти точки симметричны относительно этой оси, и потому одна (и только одна) из них лежит в верхней полуплоскости. Следовательно, переходя от окружностей к соответствующим полуокружностям (прямым плоскости Пуанкаре), мы получаем, что

все прямые геометрии Пуанкаре, ортогональные окружности Σ, *проходят через одну точку* z_0 *верхней полуплоскости.*

Эта точка называется *центром Пуанкаре* окружности Σ. Она расположена на той же вертикальной прямой, что и евклидов центр этой окружности, но ниже его (ближе к вещественной оси).

В евклидовой геометрии окружность с центром в точке A может быть определена как линия, ортогонально секущая все прямые, проходящие через точку A. Принимая то же определение окружности и в геометрии Пуанкаре, мы получаем, следовательно, что

окружностями геометрии Пуанкаре являются евклидовы окружности, целиком расположенные в верхней полуплоскости.

Однако при этом

центром окружности в геометрии Пуанкаре является ее центр Пуанкаре (а не евклидов центр).

Теперь естественно возникает вопрос, является ли окружность геометрии Пуанкаре геометрическим местом точек, равноудаленных от ее центра? Ответ оказывается утвердительным, т. е.

линия на плоскости Пуанкаре тогда и только тогда является окружностью с центром z_0, *когда любая ее точка удовлетворяет уравнению*

$$|z_0, z| = r,$$

где r *— некоторое число (называемое радиусом Пуанкаре этой окружности).*

Это утверждение удобнее доказывать в интерпретации не на верхней полуплоскости, а в единичном круге. При этом, поскольку соответствующим неевклидовым движением (конформным преобразованием) любую точку круга можно перевести в

его центр 0, мы без ограничения общности можем предполагать, что $z_0 = 0$.

Таким образом, для доказательства нашего утверждения достаточно доказать, что

геометрическое место точек z единичного круга, удовлетворяющих уравнению

$$|0, z| = r,$$

представляет собой окружность с центром в точке 0 *(некоторого евклидова радиуса $R < 1$).*

С этой целью мы заметим, что для любой точки $z \neq 0$ единичного круга прямой Пуанкаре, проходящей через точки 0 и z, является, очевидно, обыкновенная евклидова прямая. Поэтому эта прямая пересекает единичную окружность в точках $\frac{z}{|z|}$ и $-\frac{z}{|z|}$ и, следовательно, по определению,

$$|0, z| = -\log\left(\frac{0-\frac{z}{|z|}}{z-\frac{z}{|z|}} : \frac{0+\frac{z}{|z|}}{z+\frac{z}{|z|}}\right) = \log\frac{1-|z|}{1+|z|}.$$

Таким образом, соотношение $|0, z| = r$ равносильно соотношению $|z| = R$, где $R = (k^r - 1)/(k^r + 1)$ (здесь k — основание логарифмов). Это полностью доказывает наше утверждение.

Окружности типа 2) (касающиеся вещественной оси) называются в геометрии Пуанкаре *орициклами.* Заметим, что орициклом является, в частности, любая прямая $y = y_0$, параллельная вещественной оси (она касается оси в точке ∞).

Пучок прямых геометрии Пуанкаре, ортогональных орициклу, является, очевидно, параболическим пучком с центром в точке, в которой орицикл касается вещественной оси. Это означает, что

орицикл, касающийся вещественной оси в точке x_0, можно рассматривать как окружность с центром Пуанкаре в точке x_0.

В этом отношении орициклы аналогичны прямым на евклидовой плоскости (которые можно рассматривать как окружности с несобственными центрами).

Пусть Σ — окружность, проходящая через точки z_0, z_1, z_2. Оставляя точки z_1, z_2 неподвижными, будем придвигать точку z_0 к вещественной оси (например, двигая ее вертикально вниз). В пределе получится, очевидно, орицикл. Таким образом, можно сказать, что орициклы являются предельными положениями окружностей. На этом основании они называются также *предельными линиями.*

Аналогичными предельными линиями на евклидовой плоскости являются прямые.

Орициклы обладают многими замечательными свойствами, но у нас нет возможности их здесь изучать.

Дуги окружностей типа 4) (пересекающие вещественную ось под углом $\neq\pi/2$) называются *эквидистантами*. Это название объясняется тем, что

геометрическое место точек плоскости Пуанкаре, равноудаленных от некоторой прямой, состоит из двух эквидистант, проходящих через бесконечно удаленные точки этой прямой.

В этой формулировке под *расстоянием от точки до прямой* понимается, как и в евклидовой геометрии, длина перпендикуляра, опущенного из данной точки на прямую.

Поскольку неевклидовым движением любую прямую мы можем перевести в любую другую прямую, нам достаточно доказать это утверждение лишь для случая, когда данная прямая является мнимой осью

$$x=0.$$

Перпендикулярами к этой прямой будут, очевидно, всевозможные полуокружности с центром в точке 0. Такая полуокружность, проходящая через точку z, имеет радиус $|z|$ и потому пересекает мнимую ось в точке $i|z|$, а вещественную ось — в точках $\pm|z|$. Следовательно, по определению, расстояние от точки z до прямой $x=0$ равно

$$\left|\log\left(\frac{z-|z|}{i|z|-|z|}:\frac{z+|z|}{i|z|+|z|}\right)\right|=\left|\log\left(\frac{z-|z|}{z+|z|}\,i\right)\right|.$$

Поэтому условие, что точка z находится от прямой $x=0$ на данном расстоянии r, выражается формулой

$$\left|\log\left(\frac{z-|z|}{z+|z|}\,i\right)\right|=r,$$

т. е. формулой

$$\frac{z-|z|}{z+|z|}=-ik^{\pm r},$$

где k — основание логарифмов. Полагая

$$e^{i\theta}=\frac{1+ik^r}{1-ik^r},$$

мы можем это условие записать в следующем виде:

$$z=e^{\pm i\theta}|z|,$$

показывающем, что точки z, удаленные от прямой $x=0$ на расстояние r, расположены на двух лучах, исходящих из точки 0 и образующих с вещественной осью углы $\pm\theta$.

Для завершения доказательства остается заметить, что эти лучи и являются дугами окружностей, проходящих через бесконечно удаленные точки 0 и ∞ прямой $x=0$.

Прямая, от которой равноудалены точки данной эквидистанты, называется ее *базой*.

В соответствии с тем, что существуют три типа пучков окружностей — эллиптические, параболические и гиперболические,— на плоскости Пуанкаре можно различать три типа пучков прямых. Пучки первого (эллиптического) типа состоят из всех прямых геометрии Пуанкаре, проходящих через данную точку z_0 (*центр* пучка). Пучки второго (параболического) типа состоят из всех прямых, проходящих через данную бесконечно удаленную точку x_0, т. е. из всех прямых, параллельных некоторой данной прямой. Наконец, пучки третьего (гиперболического) типа состоят из всех прямых, ортогональных (т. е. перпендикулярных) некоторой данной прямой (*базе* пучка). (На евклидовой плоскости последние два типа пучков совпадают, на плоскости Пуанкаре они различны.)

Выше мы доказали, что любая окружность ортогонально сечет все прямые пучка, центр которого совпадает с центром окружности. Аналогично можно показать, что

любой орицикл является линией, ортогонально секущей прямые некоторого пучка паралелльных прямых (бесконечно удаленной точкой которых является центр орицикла), а любая эквидистанта является линией, ортогонально секущей прямые некоторого пучка перпендикуляров (базой которого является база эквидистанты).

Задание. Докажите это утверждение.

Пусть $\mathfrak{M}$ — произвольное множество и G — некоторая его группа преобразований. Для любой точки $A \in \mathfrak{M}$ множество всех точек, которые можно получить из этой точки преобразованиями группы G, называется *орбитой* точки A (по отношению к группе G).

Например, на евклидовой плоскости орбитой точки A по отношению к группе Rot_O вращений вокруг некоторой точки $O \neq A$ является окружность с центром в точке O, проходящая через точку A. Если мы на евклидовой плоскости выберем некоторое направление ω и рассмотрим группу Trans_ω всех параллельных переносов на векторы, параллельные прямым направления ω, то орбитой произвольной точки A по отношению к этой группе будет прямая направления ω, проходящая через точку A.

В геометрии Пуанкаре аналогом группы Rot_O является подгруппа группы Conf_0^{++}, состоящая из всех преобразований, оставляющих на месте данную точку z_0 (такие преобразования естественно называть *неевклидовыми вращениями* вокруг точки z_0). Поскольку все преобразования из этой группы не меняют расстояний Пуанкаре, то орбита произвольной точки z относительно этой группы содержится, очевидно, в окружности с центром Пуанкаре z_0 и радиусом, равным $|z_0, z|$. Но легко видеть, что и обратно, эта орбита содержит всю указанную окружность, поскольку для любых двух точек z_1 и z_2, находящихся от точки z_0 на одном и том же расстоянии, существует неевклидово вра-

щение вокруг точки z_0, переводящее точку z_1 в точку z_2. Действительно, этим вращением будет, как нетрудно видеть, сохраняющее ориентации конформное преобразование, оставляющее на месте точки z_0 и $\bar{z}_0$ и переводящее точку z_1 в точку z_2.

Тем самым доказано, что аналогично евклидовым окружностям

окружности плоскости Пуанкаре с центром в точке z_0 являются орбитами группы неевклидовых вращений вокруг этой точки.

Параллельные переносы на евклидовой плоскости можно определить как движения, относительно которых инвариантно некоторое направление, т. е. некоторая несобственная точка. По аналогии, *переносом* (прилагательное «параллельный» здесь неуместно) *плоскости Пуанкаре* мы будем называть ее движение (преобразование из группы Conf_0^{++}), оставляющее на месте некоторую точку x_0 вещественной оси. Все переносы, оставляющие на месте данную точку x_0, образуют, очевидно, группу (являющуюся аналогом группы Trans_ω). Оказывается, что

орбитами этой группы являются орициклы, имеющие точку x_0 своим центром Пуанкаре.

Задание. Докажите это утверждение.

Параллельные переносы на евклидовой плоскости можно также определить как движения, переводящие в себя некоторую прямую. По аналогии, *сдвигами плоскости Пуанкаре* вдоль некоторой прямой мы будем называть ее движения, переводящие в себя эту прямую. Вообще говоря, сдвиг не является переносом, а перенос — сдвигом. Все сдвиги плоскости Пуанкаре вдоль данной прямой образуют группу (также являющуюся аналогом группы Trans_ω) и

орбитами этой группы являются эквидистанты, базой которых служит данная прямая, а также сама эта прямая.

Задание. Докажите это утверждение.

Резюмируя, мы можем, таким образом, сказать, что

окружности, орициклы, эквидистанты и прямые являются орбитами некоторых групп движений плоскости Пуанкаре.

Отсюда, в частности, следует, что все эти линии обладают *свойством свободной подвижности*, т. е. любой отрезок такой линии можно как угодно двигать по этой линии.

СОДЕРЖАНИЕ

Предисловие 7

Глава 1. **Векторное исчисление** 11

§ 1*. Понятие вектора 11
1. Предварительное определение вектора 11
2. Отношения эквивалентности 13
3. Окончательное определение вектора 15
4. Векторы на прямой, на плоскости и в пространстве 19

§ 2*. Векторы на прямой 21
1. Ориентации прямой 21
2. Длина и величина вектора на прямой 25
3. Отношение векторов на прямой 26
4. Сложение векторов на прямой. Лемма Шаля 29
5. Алгебраические свойства линейных операций 30
6. Теорема об изоморфизме 32

§ 3*. Линейные операции над векторами на плоскости и в пространстве 34
1. Определение линейных операций 34
2. Алгебраические свойства линейных операций 37
3. Линейная зависимость 38
4. Геометрический смысл линейной зависимости 42
5. Базисы и координаты 46
6. Проекции и координаты 50
7. Преобразование координат при замене базиса 54
Дополнение. Теорема о ранге матрицы 58

§ 4*. Ориентации прямой, плоскости и пространства 63
1. Понятие ориентации 63
2. Правые и левые ориентации 67
3. Произведения ориентаций 68
4. Стороны прямой на плоскости и плоскости в пространстве . . . 70
5. Деформации базисов и ориентации 73
6. Резюме 79
Дополнение. О понятии угла 80

§ 5*. Метрическая теория векторов 82
1. Длина вектора и угол между векторами 82
2. Скалярное произведение векторов 86
3. Применение скалярного умножения к доказательству геометрических теорем 88
4. Выражение скалярного произведения в координатах 90
5. Ортонормированные базисы 94
6. Ортогональные матрицы 98

§ 6. Поливекторы 105
1. Бивекторы 105
2. Линейные операции над бивекторами 110
3. Линейная теория бивекторов 116
4. Метрическая теория бивекторов 121

5. Тривекторы 126
6. Векторное и смешанное произведения 133
§ 7. Линейные операторы 135
1. Отображения и преобразования 135
2. Кольцо линейных операторов 138
3. Описание линейных операторов 140
4. Обратимые линейные операторы 145
5. Операторы, действующие по равенству координат 149
6. Обратимые линейные операторы и ориентации 151
7. Изометричные операторы 155
8. Свойства изометричных операторов 157

Глава 2. **Метод координат** 164

§ 1*. Координаты на прямой, в плоскости и в пространстве 164
1. Аффинные координаты 164
2. Замена аффинных координат 169
3. Деление отрезка в данном отношении 172
4. Прямоугольные координаты 173
5. Полярные, сферические и цилиндрические координаты 175
6. Однородные координаты 177
§ 2. Уравнения линий и поверхностей 184
1. Задание линий и поверхностей уравнениями 184
2. Алгебраические линии 187
3. Параметрические уравнения линий и поверхностей 191
§ 3. Координатно-аксиоматическое построение геометрии 195
1. Основные положения аксиоматического метода 195
2. Аксиоматика евклидовой геометрии 198
3. Аксиоматика аффинной геометрии 208
4. Афинная геометрия над полем комплексных чисел 217
5. Вещественно-комплексная геометрия 221
Дополнение. Аксиоматика Гильберта 226

Глава 3. **Линии и поверхности первого порядка** 239

§ 1*. Прямая на плоскости 239
1. Прямая как линия первого порядка 239
2. Параметрические и канонические уравнения прямой 242
3. Взаимное расположение прямых на плоскости 247
4. Полуплоскости, на которые прямая разбивает плоскость . . . 251
5. Прямая на евклидовой плоскости 253
§ 2*. Плоскость в пространстве 258
1. Плоскость как поверхность первого порядка 258
2. Параметрические уравнения плоскости 260
3. Взаимное расположение плоскостей в пространстве 263
4. Полупространства, на которые плоскость разбивает пространство 266
5. Плоскость в евклидовом пространстве 268
§ 3*. Прямая в пространстве 271
1. Прямая в аффинном пространстве 271
2. Взаимное расположение прямых и плоскостей 275
3. Прямая в евклидовом пространстве 278
4. Расстояние между двумя прямыми в пространстве 282

Глава 4. **Геометрии прямых, плоскостей и окружностей** 290

§ 1. Геометрия прямых на плоскости 290
1*. Пучки прямых 290
2. Расширенная плоскость 294
3. Полнота и непротиворечивость аксиом геометрии расширенной плоскости 298
4. Координаты на расширенной плоскости 302
5. Проективная плоскость 307
6. Интерпретации проективной геометрии и их применения . . . 316

7. Конфигурационная геометрия 322
Дополнение. Трилинейные координаты 326
§ 2. Геометрия плоскостей в пространстве 331
1*. Пучки плоскостей 332
2*. Связки плоскостей 333
3. Расширенное пространство 337
4. Проективное пространство 341
Дополнение. О геометрии прямых в пространстве 346
§ 3. Геометрия окружностей на плоскости 349
1. Степень точки относительно окружности 349
2. Связки окружностей 353
3. Пучки окружностей 362
4. Пучки как пересечения связок 370
5. Прямые как окружности 377
6. Окружности на вещественно-комплексной плоскости 386
Дополнение. Геометрии параболической и гиперболической связок 396

Глава 5. **Элементарная теория линий и поверхностей второго порядка** . 398
§ 1*. Линии второго порядка 398
1. Параболы 398
2. Эллипсы 401
3. Гиперболы 411
4. Уравнения эллипса, параболы и гиперболы в полярных координатах 423
§ 2. Некоторые дополнительные свойства линий второго порядка . . . 426
1. Эллипс, парабола и гипербола как конические сечения 426
2. Взаимное расположение конических сечений и прямых 430
3. Прямые, касающиеся конических сечений 436
4. Семейства софокусных эллипсов и гипербол 443
5. Диаметры конических сечений 447
6. Теоремы Аполлония 455
§ 3*. Поверхности второго порядка 460
1. Эллипсоиды 460
2. Двуполостные гиперболоиды 463
3. Однополостные гиперболоиды 466
4. Прямолинейные образующие однополостного гиперболоида . . . 469
5. Гиперболические параболоиды 480
6. Эллиптические параболоиды 488
7. Конусы второго порядка 490
8. Цилиндры второго порядка 496

Глава 6. **Общая теория линий второго порядка** 499
§ 1*. Классификация линий второго порядка 499
1. Линии второго порядка на евклидовой плоскости 499
2. Инварианты уравнений линий второго порядка 509
3. Определение вида линии второго порядка по инвариантам ее уравнения 513
4. Линии второго порядка на аффинной плоскости. Теорема единственности 519
5. Центры линий второго порядка 527
6. Асимптоты и диаметры линий второго порядка 533
7. Приведение уравнений линий второго порядка к простейшему виду 540
8. Главные направления и диаметры линий второго порядка . . . 544
Дополнение. Классификация поверхностей второго порядка . 548
§ 2. Проективная теория линий второго порядка 554
1. Линии второго порядка на аффинно-проективной и проективной плоскостях 554
2. Пересечение прямой и линии второго порядка 558

3. Поляры и полюсы 563
Дополнение. Поляры в пространстве 568
4. Теорема Безу 569
Дополнение. Дифференциально-геометрическое истолкование кратности точки пересечения 577
§ 3. Геометрия линий второго порядка 579
1. Пучки линий второго порядка 579
2. Описание пучков линий второго порядка 583
Дополнение. Еще раз о пучках окружностей 593
3. Линии второго порядка, проходящие через пять точек 594
Дополнение. Поверхности второго порядка, проходящие через девять точек 596
4. Теорема Штурма 598
5. Теорема Паскаля 601
6. Квадратичные пучки прямых 603
7. Фокусы линий второго порядка 612

Глава 7. **Геометрические преобразования** 617
§ 1. Аффинные, проективные и ортогональные преобразования 617
1. Аффинные преобразования 617
2. Линейный оператор, индуцированный аффинным преобразованием 621
3. Общий вид афинных преобразований 628
4. Проективные преобразования 633
5. Ортогональные преобразования 637
§ 2. Разложение аффинных и ортогональных преобразований в композицию более простых 644
1. Аффинные и ортогональные преобразования прямой 644
2. Разложение аффинных преобразований плоскости и пространства 647
3. Разложение ортогональных преобразований плоскости 652
4. Разложение ортогональных преобразований пространства . . . 658
5. Представление движений пространства с помощью кватернионов 665
§ 3. Конформные преобразования 677
1. Инверсия относительно окружности 677
2. Пополненная плоскость 679
3. Свойства конформных преобразований 686
4. Конформные преобразования и ориентации 694
5. Конформная геометрия 704
§ 4. Группы и геометрии 707
1. Геометрии с данной группой автоморфизмов 707
2. Простейшие геометрии аффинного типа 711
3. Геометрии Галилея и Пуансо 718
4. Геометрия Минковского 727
5. Комплексная евклидова геометрия 734
6. Унитарная геометрия 738
7. Геометрия Пуанкаре — Лобачевского 740