МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. ЛОМОНОСОВА

А. В. Овчинников

АЛГЕБРА И ГЕОМЕТРИЯ ДЛЯ СТУДЕНТОВ-ФИЗИКОВ

Лекционный курс

Семестр 1

МОСКВА
ФИЗИЧЕСКИЙ ФАКУЛЬТЕТ
МГУ им. М. В. ЛОМОНОСОВА
2016

УДК 51(07), 512.6, 514.1
ББК 22.1я73, 22.144, 22.151.5
О-35

Рецензенты:

кафедра общей топологии и геометрии
механико-математического факультета
МГУ им. М. В. Ломоносова
(заведующий кафедрой профессор *Ю. В. Садовничий*);

академик РАН *Р. В. Гамкрелидзе*

Редакторы *М. О. Корпусов, В. Л. Попов*

Рекомендовано к изданию методической комиссией отделения
прикладной математики физического факультета МГУ

Печатается по решению Учёного совета
физического факультета МГУ

Овчинников Алексей Витальевич

О-35 Алгебра и геометрия для студентов-физиков. Лекционный курс. Семестр 1. — М.: Физический факультет МГУ, 2016. — 360 с.

ISBN 978-5-8279-0138-9

Учебное пособие создано на базе курса лекций, читаемых автором на физическом факультете МГУ им. М. В. Ломоносова, и содержит изложение начал алгебры и аналитической геометрии. В книге обсуждаются традиционные для курса вопросы: алгебра векторов, линейные многообразия на плоскости и в пространстве, линии и поверхности второго порядка. Алгебраическая часть курса содержит теорию матриц, определителей, систем линейных уравнений. Для студентов физико-математических специальностей университетов, а также других специальностей с повышенной математической подготовкой.

УДК 51(07), 512.6, 514.1
ББК 22.1я73, 22.144, 22.151.5

ISBN 978-5-8279-0138-9

Оглавление

Предисловие

Учебное пособие, предлагаемое вниманию читателей, представляет собой расширенную запись курса лекций по дисциплине «аналитическая геометрия», которые автор в течение многих лет читает на физическом факультете Московского государственного университета им. М. В. Ломоносова.

В настоящее время аналитическая геометрия представляет собой не столько раздел математической науки, сколько учебный курс, который знакомит студентов с методом координат и является источником наглядных представлений и примеров для других учебных курсов, в первую очередь анализа и механики. Основные цели преподавания аналитической геометрии — продемонстрировать студентам общие принципы приложения алгебры и анализа к геометрии, приучить начинающих к оперированию математическими объектами, которые не встречались в школьном курсе (речь идёт о векторах, матрицах, определителях).

Современное изложение аналитической геометрии существенно использует алгебраический аппарат; по этой причине в предлагаемый курс для студентов физических специальностей, учебный план которых не содержит алгебры как отдельной дисциплины, включён довольно значительный объём алгебраического материала: это теория матриц, определителей, систем линейных уравнений. (Этим и объясняется присутствие слов «алгебра и геометрия» в названии книги.) Кроме того, для расширения математического кругозора студентов и ознакомления их с современной научной терминологией книга снабжена приложениями, содержащими краткое обсуждение важнейших общематематических понятий: множеств, отображений, простейших алгебраических систем и т. д. Собственно геометрическая часть курса вполне традиционна и содержит теорию векторов, линейных многообразий (прямых и плоскостей) и квадрик (линий и поверхностей второго порядка) в аффинных и евклидовых пространствах.

В отличие от большинства учебников по аналитической геометрии для физиков, изложение в предлагаемом курсе построено на системе аксиом Вейля. Это делает изложение более строгим и почти не зависящим от школьного курса геометрии, основанного на одном из вариантов системы аксиом Гильберта, роль которой в современной математике весьма незначительна. Аксиоматика Вейля позволяет изложить геометрию быстро и просто, чётко отделяя аффинные аспекты от метрических, что весьма важно для формирования правильного понимания будущим учёным-физиком природы изучаемых пространств.

При попытке записать лекции, предназначенные для студентов-первокурсников, приходится выбирать между логической последовательностью изложения материала и последовательностью, диктуемой соображениями педагогического характера. В настоящем пособии предпочтение отдано второму способу организации материала; каждое новое понятие вводится лишь после мотивации его необходимости и разъяснения на примерах.

Автору бывает трудно избежать искушения включить в конспект лекций некоторые темы, не входящие в программу курса, но доступные пониманию добросовестных студентов и полезные для развития их математической эрудиции. Некоторые из этих сюжетов, в разные годы либо обсуждавшиеся на лекциях или консультациях, либо предлагавшиеся заинтересованным студентам для самостоятельной проработки, напечатаны петитом; при первом чтении учебника их можно пропустить без ущерба для понимания остального текста.

Автор выражает глубокую признательность за помощь и поддержку при работе над книгой А. В. Бадьину, А. Н. Боголюбову, Е. Е. Букжалёву, В. Ф. Бутузову, В. Т. Волкову, П. В. Голубцову, А. В. Ильяшенко, С. Л. Ичиро Делир Магрифу, С. Б. Кадомцеву, В. В. Колыбасовой, Н. Ч. Крутицкой, Н. Т. Левашовой, Д. В. Лукьяненко, Г. Н. Медведеву, И. Е. Могилевскому, Н. Н. Нефёдову, С. А. Овчинниковой, А. А. Панину, Л. В. Перовой, В. Ю. Попову, Д. Д. Соколову, Т. Танигучи, М. Чино, М. А. Чино, Н. Е. Шапкиной, А. А. Широнину, А. А. Шишкину и всем коллегам по кафедре математики физического факультета МГУ, участвовавшим в обсуждении курса. Особую важность имели замечания и ценные советы профессоров М. О. Корпусова (МГУ им. М. В. Ломоносова) и В. Л. Попова (Математический институт им. В. А. Стеклова РАН), которые тщательно отредактировали рукопись, и академика РАН Р. В. Гамкрелидзе, который оказал неоценимое влияние на представления автора о том, как нужно преподавать математику.

Автор

Список обозначений

$\mathbb{N}$ множество натуральных чисел: $\mathbb{N} = \{1, 2, \dots\}$;
$\mathbb{N}_0$ множество натуральных чисел с нулём: $\mathbb{N}_0 = \{0, 1, 2, \dots\}$;
$\mathbb{Z}$ множество целых чисел;
$\mathbb{Q}$ множество рациональных чисел;
$\mathbb{R}$ множество вещественных (действительных) чисел;
$\mathbb{C}$ множество комплексных чисел;
$\mathbb{K}$ произвольное числовое поле, в частности, $\mathbb{Q}$, $\mathbb{R}$ или $\mathbb{C}$;
$\mathbb{K}[x]$ множество всех многочленов от переменной x с коэффициентами из числового поля $\mathbb{K}$;
$\mathbb{K}^n$ множество столбцов высоты n с элементами из числового поля $\mathbb{K}$;
$\mathbb{K}^{m\times n}$ множество матриц размера $m\times n$ (т.е. состоящих из m строк и n столбцов) с элементами из числового поля $\mathbb{K}$;
$\stackrel{\text{def}}{=\!=}$ знак «равенства по определению»: $A \stackrel{\text{def}}{=\!=} B$ читается «A по определению равно B»;
$\wedge$ логическая конъюнкция (союз «и»);
$\vee$ логическая дизъюнкция (союз «или»);
$\Rightarrow$ логическая импликация: $A \Rightarrow B$ читается «из A следует B»;
$\Leftrightarrow$ логическая эквивалентность: $A \Leftrightarrow B$ читается «A эквивалентно (равносильно) B»;
$\forall$ квантор общности ($\forall x$ читается «для всех x»);
$\exists$ квантор существования ($\exists x$ читается «существует такой x, что...»);

ГЛАВА 1

Введение

АНАЛИТИЧЕСКАЯ геометрия — это раздел геометрии, в котором геометрические фигуры и их свойства исследуются средствами алгебры, что становится возможным благодаря методу координат. При этом точки описываются наборами чисел, более сложные геометрические фигуры — уравнениями и неравенствами, и это позволяет превратить геометрическую задачу в алгебраическую.

1. Аксиоматический метод в геометрии

Преподавание геометрии, следуя традиции, восходящей к Евклиду[1], основывается на аксиоматическом методе.

1.1. В настоящее время в истории математики общепринятой является точка зрения, что создателями геометрии как науки были древнегреческие учёные, превратившие египетское землемерное ремесло в строгую научную дисциплину. Перейдя от коллекционирования геометрических рецептов к поиску общих закономерностей, античные философы составили первые систематические трактаты по геометрии, центральное место среди которых занимают «Начала» Евклида. Более двух тысячелетий этот труд считался образцовым изложением в духе аксиоматического метода.

1.2. *Аксиоматический метод* основан на том, что из небольшого числа очевидных принципов, называемых *аксиомами*, с помощью дедуктивных рассуждений, т.е. рассуждений «от общего к частному», выводятся разнообразные заключения. Природа дедуктивного вывода такова, что она гарантирует истинность заключения, если только истинны исходные аксиомы. Сущность аксиоматического метода может быть кратко описана следующим образом:

(i) перечисляются *основные понятия* теории;

(ii) устанавливаются *основные отношения* между основными понятиями;

[1] Евклид — древнегреческий математик (около 350–270 до н.э.).

(iii) формулируются свойства основных отношений, принимаемые без доказательства — *аксиомы*;
(iv) логическим путём выводятся следствия аксиом — *теоремы*, в совокупности образующие математическую теорию.

1.3. *Основные понятия* невозможно определить при помощи более простых понятий; можно лишь разъяснить их смысл на примерах. Действительно, если некоторое понятие D определяется через более простое понятие C, которое, в свою очередь, нуждается в определении посредством ещё более простого понятия B и так далее, то в конце концов мы должны будем прийти к такому понятию A, которое уже не удаётся определить с помощью более простых понятий. Понятие A и является основным.

1.4. Основные понятия теории возникают в результате *абстрагирования* от свойств реальных объектов; так, в «Началах» Евклида точка определяется как «то, что не имеет частей», линия — как «длина без ширины», прямая — как «линия, которая одинаково расположена по отношению ко всем своим точкам». Формулировки подобных «определений» содержат понятия, которые сами требуют определения, например, «длина», «одинаковое расположение»; по сути дела, они представляют собой просто наглядные описания простейших геометрических образов и бесполезны при доказательстве теорем. В современных аксиоматических схемах построения геометрии наглядные описания геометрических образов полностью отсутствуют.

1.5. *Аксиомы* составляются с учётом накопленного эмпирического материала таким образом, чтобы весь этот эмпирический материал мог быть выведен из аксиом логически. Но объекты и отношения, о которых говорится в аксиомах, не обязаны обладать какой-либо спецификой или иметь определённый внешний вид. Геометрия оперирует понятиями, возникающими из опыта в результате абстракции объектов реального мира, при которой принимаются во внимание лишь *некоторые*, существенные именно для геометрии свойства реальных объектов; в логических рассуждениях приходится иметь дело именно с этими свойствами, и лишь они должны быть отмечены в определениях и аксиомах. Все остальные свойства, которые мы привыкли воображать, когда слышим слова «точка», «прямая», «плоскость», в логическом построении геометрии никакой роли не играют и не должны упоминаться среди основных положений.

1.6. Конечно, геометрическая интуиция и соображения наглядности весьма важны и при выводе новых фактов, и при интерпретации доказанных результатов, однако строгие доказательства должны быть основаны исключительно на аксиомах и не апеллировать к каким-либо наглядным представлениям или интуитивным концепциям.

1.7. В естественных науках истинность всех утверждений проверяется при помощи эксперимента или наблюдения. В математике функцию эксперимента выполняет *доказательство.*

1.8. Идея доказательства, как и аксиоматического метода в целом, восходит к античным учёным. Аксиомы и только они принимаются как абсолютно верные утверждения без доказательств; все остальные утверждения выводятся логически из аксиом. При таком подходе зачастую подлежат доказательству интуитивно очевидные факты, которые в число аксиом не входят. С формально-логической точки зрения доказывать эти факты необходимо; именно это и называется *математической строгостью.* Однако довольно часто в целях экономии времени подобные «почти очевидные» доказательства опускают, заменяя их интуитивными соображениями, аналогиями и эвристическими аргументами, которые, впрочем, могут быть доведены до математически строгих доказательств посредством рутинной работы.

1.9. Строгий логический анализ основ геометрии, предпринятый Д. Гильбертом[1], показал недостаточность аксиоматики, содержащейся в «Началах» Евклида. Гильберт предложил систему аксиом, которая — с незначительными изменениями — лежит в основе курса геометрии, преподаваемого в средней школе и поныне. В этой системе основными понятиями являются *точка*, *прямая* и *плоскость*. Основных отношений в системе Гильберта пять:

(1) *инцидентность* («точка лежит на прямой», «прямая проходит через точку» и т. п.); свойства отношения инцидентности описываются восемью аксиомами, четыре из которых планиметрические и четыре — стереометрические;
(2) *порядок* (понятие «лежать между») — четыре аксиомы;
(3) *конгруэнтность* (равенство геометрических фигур) — пять аксиом;
(4) *параллельность* — одна аксиома;
(5) *непрерывность* — две аксиомы.

1.10. Система аксиом Гильберта, будучи безупречной с логической точки зрения, обладает рядом недостатков, важнейшие среди которых таковы:

(a) рассуждения, основанные на гильбертовой аксиоматике, довольно сложны и, как правило, требуют хорошего пространственного воображения и определённой изобретательности;

[1] Д. Гильберт (D. Hilbert) — немецкий математик (1862–1943).

(б) гильбертова система аксиом с трудом поддаётся обобщению на многомерный случай: при попытке такого обобщения происходит добавление новых основных понятий и аксиом;
(в) нигде, кроме элементарной геометрии, эта система аксиом не используется, и потому практическая ценность её в современной математике крайне низка.

1.11. Существуют иные системы аксиом, в основе которых могут лежать как другие отношения между основными понятиями (например, в системе аксиом Шура вместо конгруэнтности используется движение, а в системе Кагана — расстояние), так и другие основные понятия. Отметим, что все системы аксиом логически эквивалентны, т.е. выводятся одна из другой; при этом аксиомы одной системы являются теоремами другой и наоборот.

1.12. Одна из важнейших систем аксиом геометрии — это система, предложенная Г. Вейлем[1] в 1918 г. в его книге «Пространство, время, материя», которая явилась одним из первых в истории науки систематических изложений общей теории относительности А. Эйнштейна[2]. В системе аксиом Вейля основными понятиями являются *точка* и *вектор*, а прямые и плоскости определяются как множества точек. Эта векторно-точечная аксиоматика по сравнению с системой Гильберта намного удобнее и проще. Удобство её заключается в универсальности: различные части аксиоматики Вейля используются практически во всех разделах математики. Упрощение же теории обусловлено тем, что векторная аксиоматика явно основана на теории вещественных чисел, в то время как любая система аксиом «гильбертова» типа должна содержать аксиомы, воспроизводящие некоторые требуемые для геометрии свойства чисел.

2. Координаты и уравнения

1.13. Идея координат и уравнения линии появилась ещё в древнегреческой математике; например, Архимед[3] и Аполлоний[4] в своих сочинениях приводили так называемые *симптомы* конических сечений (эллипса, гиперболы и параболы), которые очень похожи на современные уравнения этих линий. Однако из-за невысокого уровня развития алгебры в античности и отсутствия интереса к линиям, отличным от прямой и окружности, метод координат развития не получил.

---

[1] Г. Вейль (H. Weyl) — немецкий математик и физик-теоретик (1885–1955).

[2] А. Эйнштейн (A. Einstein) — выдающийся физик-теоретик, общественный деятель-гуманист, лауреат Нобелевской премии по физике (1879–1955).

[3] Архимед из Сиракуз — древнегреческий математик, физик и инженер (287–212 до н.э.).

[4] Аполлоний Пергский — древнегреческий математик (262–190 до н.э.).

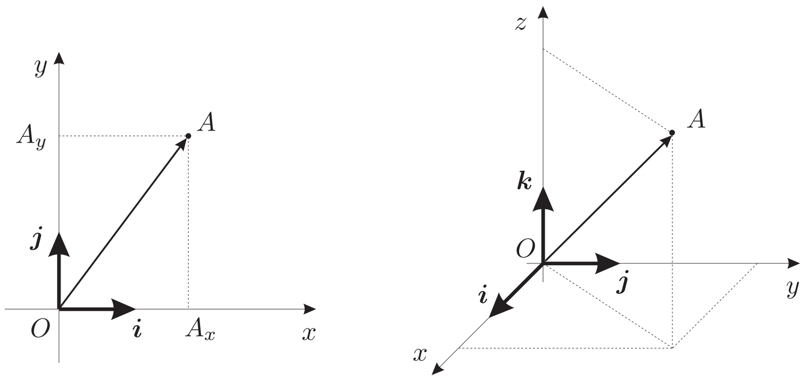


Рис. 1.1. Декартова прямоугольная система координат на плоскости и в пространстве

1.14. Первые шаги в развитии метода координат связаны с именами Виета, Ферма, Декарта (конец эпохи Возрождения, XVI–XVII вв.). Эффективность метода быстро сделала его весьма популярным, и исследования в области геометрии и анализа в Новое время (XVII–XVIII вв.) уже немыслимы без его применения. В XVIII столетии аналитическая геометрия, используя мощные методы математического анализа, дала начало новому разделу математики — дифференциальной геометрии.

1.15. Аналитическая и дифференциальная геометрия тесно связаны с механикой. Ещё Декарт начал рассматривать в геометрии кривые, отличные от прямых и конических сечений, в том числе «механические» кривые, являющиеся траекториями движения точек, и сформулировал тезис о том, что у каждой кривой имеется определяющее её уравнение. Линии и поверхности, задаваемые уравнениями, более сложными, нежели линейные и квадратичные уравнения с несколькими переменными, являются объектом изучения в дифференциальной геометрии, поскольку их исследование требует не только алгебраических средств, но и развитого дифференциального и интегрального исчисления.

A. Системы координат.

1.16. ***Декартовы прямоугольные координаты*** хорошо знакомы читателю из школьного курса математики. Система декартовых прямоугольных координат па плоскости состоит из двух взаимно перпендикулярных прямых, на каждой из которых выбрано одно из двух возможных направлений и фиксирована масштабная единица (одинаковая для обеих осей); эти прямые называются *осями координат*, а точка O их пересечения — *началом координат.*

Единичные направленные отрезки вдоль координатных осей называются *ортами* и обозначаются $\boldsymbol{i}, \boldsymbol{j}$. Пусть A — произвольная точка плоскости. Направленный отрезок $\overrightarrow{OA}$ называется *радиус-вектором* точки A. Ортогональные проекции произвольной точки плоскости A на оси координат обозначим A_x и A_y. Величины x и y направленных отрезков OA_x и OA_y (т.е. их длины с учётом знака: если направленный отрезок сонаправлен с соответствующей осью — берётся знак «+», в противном случае — знак «−») называются (*декартовыми*) *координатами* (*абсциссой* и *ординатой*) точки A (а также направленного отрезка $\overrightarrow{OA}$); обозначения $A(x, y)$, $\overrightarrow{OA}(x, y)$ и $\overrightarrow{OA} = (x, y)$. Аналогично определяются декартовы прямоугольные координаты в пространстве; третья координата z называется *аппликатой*, а орт оси Oz обозначается $\boldsymbol{k}$.

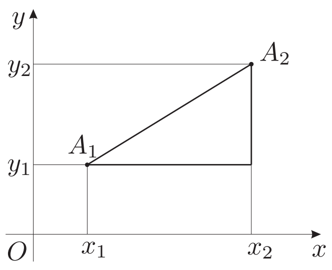


Рис. 1.2.

1.17. Очевидно, расстояние $|A_1A_2|$ между двумя точками $A_1(x_1, y_1)$ и $A_2(x_2, y_2)$ на плоскости может быть вычислено по формуле

$$|A_1A_2| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

являющейся простым следствием теоремы Пифагора (см. рис. 1.2). Аналогичная формула имеет место и в пространственном случае: расстояние между точками $A_1(x_1, y_1, z_1)$ и $A_2(x_2, y_2, z_2)$ равно

$$|A_1A_2| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}. \qquad (1.1)$$

1.18. В некоторых задачах (например, в задачах кристаллографии) удобнее использовать *косоугольные декартовы координаты*; от прямоугольных они отличаются тем, что координатные оси не перпендикулярны (вследствие чего вместо ортогональных проекций используются параллельные), а масштабные единицы вдоль разных осей могут быть различными.

1.19. Кроме декартовых координат в математике используются и другие типы координатных систем — так называемые *криволинейные координаты*, простейшие из которых описаны ниже.

1.20. ***Полярные координаты*** на плоскости определяются следующим образом. На плоскости фиксируется луч OP, начало которого называется *полюсом* (а сам луч — *полярной осью*). Положение любой точки A плоскости описывается упорядоченной парой

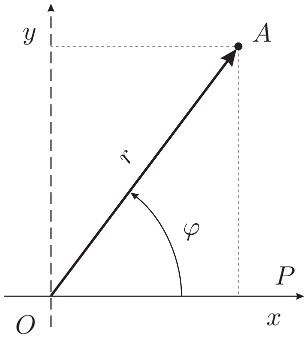


Рис. 1.3. Полярные координаты на плоскости

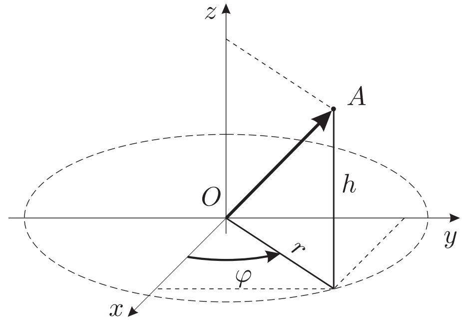


Рис. 1.4. Цилиндрические координаты в пространстве

чисел (r, φ), где r — расстояние от точки A до полюса, а φ — ориентированный угол между полярной осью и направленным отрезком (радиус-вектором) $\overrightarrow{OA}$, отсчитываемый против часовой стрелки. Значение угла φ для самой точки O не определено: она вполне задаётся значением $r = 0$. Очевидно,

$$0 \leqslant r < +\infty, \quad 0 \leqslant \varphi < 2\pi.$$

Во многих задачах удобно считать, что угол φ определён с точностью до слагаемого вида $2\pi n$, $n \in \mathbb{Z}$; это записывается следующим образом:

$$0 \leqslant \varphi < 2\pi \pmod{2\pi}.$$

1.21. Если совместить полярную систему координат с прямоугольной декартовой системой координат так, чтобы полюс совпал с началом декартовых координат, а полярная ось — с положительной полуосью Ox (см. рис. 1.3), то легко получить формулы, выражающие декартовы координаты через полярные:

$$x = r\cos\varphi, \quad y = r\sin\varphi. \tag{1.2}$$

Формулы обратного перехода имеют вид

$$r = \sqrt{x^2 + y^2}, \quad \cos\varphi = \frac{x}{\sqrt{x^2 + y^2}}, \quad \sin\varphi = \frac{y}{\sqrt{x^2 + y^2}}. \tag{1.3}$$

1.22. В некоторых задачах полярному радиусу разрешают принимать отрицательные значения, условившись, что при $r < 0$ точка откладывается по другую сторону от полюса, т.е. вместо точки (r, φ), где $r < 0$, изображается точка $(-r, \varphi + \pi)$.

1.23. ***Цилиндрические координаты*** в пространстве являются аналогом полярных координат на плоскости; их удобно определить, предварительно введя прямоугольную декартову систему координат $Oxyz$. Тогда цилиндрические координаты точки A в пространстве — это тройка чисел (r, φ, h), где r и φ — полярные координаты ортогональной проекции точки A на координатную плоскость Oxy, а h — ортогональная проекция радиус-вектора $\overrightarrow{OA}$ на ось Oz.

Формулы, связывающие цилиндрические координаты точки с её прямоугольными декартовыми координатами, имеют вид

$$x = r\cos\varphi, \quad y = r\sin\varphi, \quad z = h. \tag{1.4}$$

1.24. ***Сферические координаты*** точки A определим, предварительно введя прямоугольную декартову систему координат $Oxyz$. Сферическими координатами точки A называются три числа (r, θ, φ), определяемые следующим образом: r — расстояние от начала координат O до точки A; θ — ориентированный угол между осью Oz и радиус-вектором $\overrightarrow{OA}$, называемый *зенитным углом*; φ — полярный угол проекции A' точки A на плоскость Oxy в этой плоскости, называемый *азимутальным углом* (см. рис. 1.5). Точка O вполне задаётся значением одной лишь радиальной координаты $r = 0$; угловые координаты θ и φ для неё не определены. Для точек оси Oz (т.е. при $\theta = 0$ или $\theta = \pi$) не определено значение угла φ. Диапазоны изменения значений сферических координат:

$$0 \leqslant r < +\infty, \quad 0 \leqslant \theta \leqslant \pi, \quad 0 \leqslant \varphi < 2\pi \pmod{2\pi}.$$

Формулы, связывающие сферические координаты (r, θ, φ) точки с её прямоугольными декартовыми координатами, имеют вид

$$x = r\cos\varphi\sin\theta, \quad y = r\sin\varphi\sin\theta, \quad z = r\cos\theta. \tag{1.5}$$

1.25. Часто используется другой вариант сферических координат, в котором вместо зенитного угла θ используется угол ϑ, отсчитываемый не от оси Oz, а от ортогональной проекции $\overrightarrow{OA'}$ радиус-вектора $\overrightarrow{OA}$ на плоскость Oxy, и называемый (географической) *широтой*. В этом случае азимутальный угол (обозначим его ϕ) называют (географической) *долготой*. Такие координаты (их называют иногда *географическими координатами*) имеют следующие диапазоны изменения:

$$0 \leqslant r < +\infty, \quad -\frac{\pi}{2} \leqslant \vartheta \leqslant \frac{\pi}{2}, \quad -\pi < \phi \leqslant \pi \pmod{2\pi}.$$

Формулы, связывающие географические координаты (r, ϑ, ϕ) точки с её прямоугольными декартовыми координатами, весьма похожи на соответствующие формулы для сферических координат:

$$x = r\cos\phi\cos\vartheta, \quad y = r\sin\phi\cos\vartheta, \quad z = r\sin\vartheta. \tag{1.6}$$

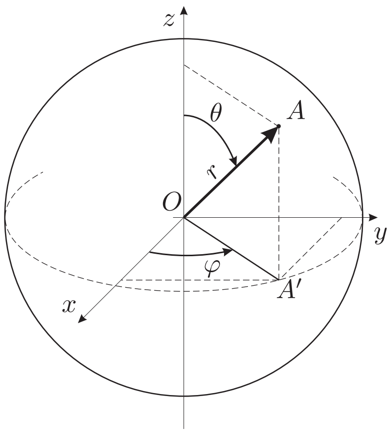


Рис. 1.5. Сферические координаты

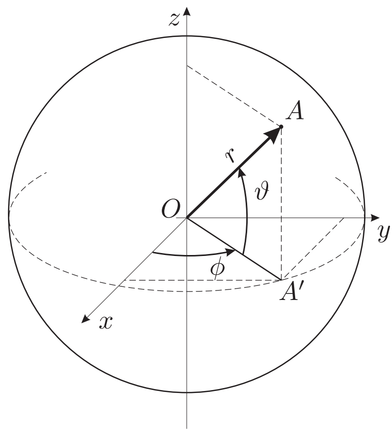


Рис. 1.6. Географические координаты

Б. Простейшие задачи. Предполагается, что из школьного курса читателю известно понятие вектора — направленного отрезка, а также определение равных направленных отрезков (см. также п. 5.2, с. 90).

1.26. ***Координаты направленного отрезка*** $\overrightarrow{BC}$, начало которого не совпадает с началом координат, определяются как координаты равного ему направленного отрезка $\overrightarrow{OA}$, приложенного в начале координат (см. п. 1.16).

Пусть $\overrightarrow{BC}$ — направленный отрезок с концами $B(x_B, y_B, z_B)$, $C(x_C, y_C, z_C)$ (см. рис. 1.7); координаты точек B и C заданы относительно некоторой декартовой (прямоугольной или косоугольной) системы координат. Поскольку равные направленные отрезки имеют равные проекции, получаем

$$x_A = x_C - x_B, \quad y_A = y_C - y_B, \quad z_A = z_C - z_B,$$

так что

$$\overrightarrow{BC} = \left(x_C - x_B,\ y_C - y_B,\ z_C - z_B\right). \tag{1.7}$$

1.27. ***Деление отрезка в заданном отношении.*** Пусть точки A и B заданы своими координатами относительно произвольной (прямоугольной или косоугольной) декартовой системы координат: $A(x_1, y_1, z_1)$, $B(x_2, y_2, z_2)$. Получим формулы, выражающие координаты точки $M(x, y, z)$, которая делит заданный отрезок AB в отношении $\alpha : \beta$, считая от точки A.

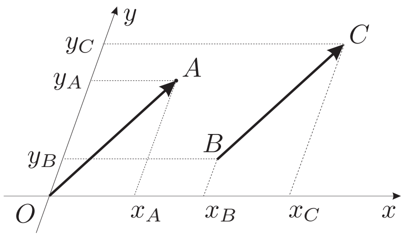


Рис. 1.7. Координаты направленного отрезка

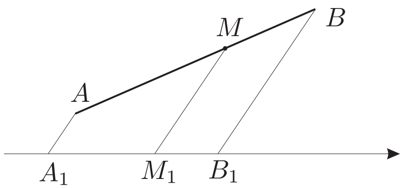


Рис. 1.8. Деление отрезка в заданном отношении

Спроектировав точки A, B и M на любую из координатных осей, например, на ось Ox (см. рис. 1.8), получим, что проекция M_1 точки M делит отрезок A_1B_1, где A_1 и B_1 — проекции точек A и B соответственно, в том же отношении $\alpha : \beta$, считая от точки A_1. Решая уравнение

$$\frac{\alpha}{\beta} = \frac{|A_1M_1|}{|M_1B_1|} = \frac{x - x_1}{x_2 - x}$$

относительно x и аналогичные уравнения для y и z, получаем:

$$x = \frac{\beta x_1 + \alpha x_2}{\alpha + \beta}, \quad y = \frac{\beta y_1 + \alpha y_2}{\alpha + \beta}, \quad z = \frac{\beta z_1 + \alpha z_2}{\alpha + \beta}. \tag{1.8}$$

1.28. ***Центр масс системы материальных точек.*** Пусть в точках $A_1(x_1, y_1, z_1)$ и $A_2(x_2, y_2, z_2)$ помещены точечные частицы, массы которых m_1 и m_2. Центр масс этой системы — это точка B, которая делит отрезок A_1A_2 на части, обратно пропорциональные массам, т.е. в отношении $m_2 : m_1$. По формулам (1.8) находим координаты центра масс:

$$x_B = \frac{m_1x_1 + m_2x_2}{m_1 + m_2}, \quad y_B = \frac{m_1y_1 + m_2y_2}{m_1 + m_2}, \quad z_B = \frac{m_1z_1 + m_2z_2}{m_1 + m_2}. \tag{1.9}$$

Найдём центр масс системы, состоящей из трёх частиц, массы которых m_1, m_2, m_3, помещённых в точках $A_1(x_1, y_1, z_1)$, $A_2(x_2, y_2, z_2)$, $A_3(x_3, y_3, z_3)$. Пусть $B(x, y, z)$ — центр масс системы частиц m_1 и m_2, координаты которой найдены выше. Центр масс системы трёх частиц расположен в точке C, которую можно рассматривать как центр масс системы двух частиц: массой $m_1 + m_2$ в точке B и массой m_3 в точке A_3. По формулам (1.9) получаем

$$x_C = \frac{(m_1 + m_2)x_B + m_3x_3}{(m_1 + m_2) + m_3} = \frac{m_1x_1 + m_2x_2 + m_3x_3}{m_1 + m_2 + m_3}$$

и аналогично для y_C и z_C.

Для системы k частиц, имеющих массы $m_1, \ldots, m_k$ и расположенных в точках $A_1(x_1, y_1, z_1), \ldots, A_k(x_k, y_k, z_k)$, с помощью метода математической индукции[1] можно доказать следующие формулы для координат центра масс:

$$x_C = \frac{m_1x_1 + \cdots + m_kx_k}{m_1 + \cdots + m_k}$$

и аналогично для координат y_C и z_C.

[1] См. теорему III.12, с. 318.

1.29. ***Направляющие косинусы направленного отрезка.*** Рассмотрим в *прямоугольной* декартовой системе координат $Oxyz$ направленный отрезок $\overrightarrow{OA}$, где A — точка с координатами (x, y, z). Обозначив через α, β, γ углы, которые $\overrightarrow{OA}$ образует с осями координат, можем записать ортогональные проекции направленного отрезка $\overrightarrow{OA}$ на координатные оси (т.е. его координаты):

$$x = \Pr_{Ox} \overrightarrow{AB} = \left|\overrightarrow{AB}\right| \cdot \cos\alpha,$$
$$y = \Pr_{Oy} \overrightarrow{AB} = \left|\overrightarrow{AB}\right| \cdot \cos\beta,$$
$$z = \Pr_{Oz} \overrightarrow{AB} = \left|\overrightarrow{AB}\right| \cdot \cos\gamma.$$

Поскольку длина направленного отрезка $\overrightarrow{AB}$ равна

$$\left|\overrightarrow{AB}\right| = \sqrt{x^2 + y^2 + z^2},$$

отсюда получаем

$$\cos\alpha = \frac{x}{\sqrt{x^2 + y^2 + z^2}},$$
$$\cos\beta = \frac{y}{\sqrt{x^2 + y^2 + z^2}},$$
$$\cos\gamma = \frac{z}{\sqrt{x^2 + y^2 + z^2}}.$$

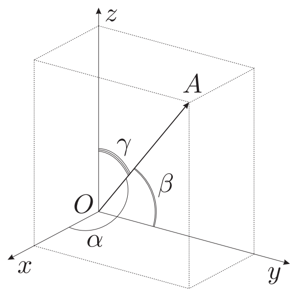


Рис. 1.9.

Числа $\cos\alpha$, $\cos\beta$ и $\cos\gamma$ называются *направляющими косинусами* направленного отрезка $\overrightarrow{OA}$. Складывая квадраты этих равенств, находим зависимость между направляющими косинусами:

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1. \tag{1.10}$$

Направляющими косинусами направленного отрезка $\overrightarrow{BC}$, начало которого не совпадает с началом координат, называют направляющие косинусы направленного отрезка $\overrightarrow{OA}$, приложенного в начале координат и равного $\overrightarrow{BC}$.

B. Уравнения линий и поверхностей. Понятия линии и поверхности являются одними из самых трудных математических понятий. Мы не будем останавливаться на их строгом определении; для наших целей достаточно наглядного интуитивного представления о линии и поверхности.

Будем считать, что на плоскости (в пространстве) задана некоторая прямоугольная декартова система координат.

1.30. *Уравнение линии* на плоскости — это уравнение вида $F(x, y) = 0$, каждое решение которого представляет собой прямоугольные декартовы координаты некоторой точки линии, причём каждая точка линии представляется некоторым решением данного уравнения.

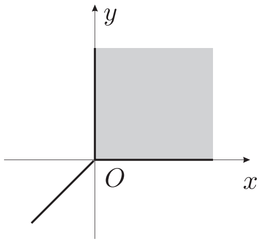


Рис. 1.10.

1.31. Для того чтобы множество решений уравнения $F(x, y) = 0$ соответствовало интуитивному представлению о «линии», функция $F(x, y)$ должна обладать достаточно «хорошими» свойствами, изучение которых выходит за рамки курса аналитической геометрии. Однако нетрудно привести пример, в котором уравнение «достаточно простой структуры» описывает геометрическую фигуру, которую трудно назвать «линией» (см. рис. 1.10): $x - |x| - y + |y| = 0$.

1.32. Некоторые геометрические свойства линии могут быть установлены непосредственно по её уравнению. Например, предположим, что функция $F(x, y)$ обладает свойством

$$F(-x, y) = F(x, y).$$

Если точка $A(x_0, y_0)$ удовлетворяет уравнению $F(x, y) = 0$, то точка $A'(-x_0, y_0)$, симметричная точке A относительно оси Oy, также удовлетворяет этому уравнению. Поэтому линия, определяемая уравнением $F(x, y) = 0$, будет симметричной относительно оси Oy. Аналогично, если

$$F(x, -y) = F(x, y),$$

то линия, определяемая уравнением $F(x, y) = 0$, симметрична относительно оси Ox.

1.33. *Уравнение поверхности* в пространстве определяется аналогично, но содержит три переменные:

$$G(x, y, z) = 0.$$

1.34. Ясно, что соответствие между линиями (поверхностями) и их уравнениями не является однозначным: если умножить обе части уравнения на одно и то же ненулевое число, то получившееся — другое! — уравнение будет описывать ту же самую линию (поверхность).

1.35. Вместо уравнений можно рассматривать неравенства; это позволяет описывать области на плоскости и в пространстве. Такие задачи в нашем курсе рассматриваться не будут.

Г. Параметрические уравнения.

1.36. Линия на плоскости может быть задана как множество точек, координаты которых вычисляются по формулам

$$x = \varphi(t), \quad y = \psi(t), \quad \alpha \leqslant t \leqslant \beta,$$

где t — вспомогательная величина, называемая *параметром.* При изменении параметра в промежутке $[\alpha, \beta]$ точка плоскости с координатами $(\varphi(t), \psi(t))$ описывает некоторую линию.

1.37. Этот способ пригоден и для задания линий в пространстве:

$$x = \varphi(t), \quad y = \psi(t), \quad z = \chi(t), \quad \alpha \leqslant t \leqslant \beta.$$

С точки зрения механики параметрические уравнения линии — это закон движения материальной точки, параметр t — время.

1.38. Например, окружность радиуса R с центром в начале координат можно задать параметрическими уравнениями

$$x = R\cos t, \quad y = R\sin t,$$

$$0 \leqslant t < 2\pi \pmod{2\pi}.$$

Параметр t представляет собой угол между осью Ox и радиус-вектором точки окружности. Записанные уравнения описывают равномерное движение точки по окружности (вращение) с фазой t (угловая скорость точки равна 1).

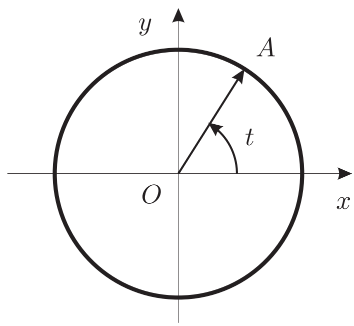


Рис. 1.11. Параметризация окружности

1.39. Поверхности также можно задавать параметрическими уравнениями

$$x = \varphi(u, v), \quad y = \psi(u, v), \quad z = \chi(u, v),$$

где параметры u, v изменяются независимо в некоторой области D вспомогательной координатной плоскости Ouv. Такой способ задания поверхностей особенно удобен в дифференциальной геометрии.

* * *

В настоящем курсе изложение построено на системе аксиом Вейля, в которой основными понятиями являются векторы и точки. Поэтому в дальнейшем фрагменты текста, содержащие слова «прямая», «плоскость» и другие подобные термины элементарной геометрии, встретившиеся *до* их формального определения в рамках выбранной системы аксиом, должны рассматриваться лишь как нестрогие интуитивно-наглядные описания обсуждаемых объектов.

ГЛАВА 2

Комплексные числа. Многочлены

Из школьного курса алгебры хорошо известно, что квадратное уравнение с отрицательным дискриминантом не имеет корней. Эту ситуацию возможно исправить, введя в рассмотрение «числа» вида $a + b\sqrt{-1}$ (их называют комплексными), после чего квадратные уравнения с отрицательным дискриминантом становятся разрешимыми. Более того, «корни», полученные таким образом, обладают всеми свойствами, присущими вещественным корням, например, удовлетворяют теореме Виета. Разумеется, возникает вопрос о *существовании* комплексных чисел и их свойствах. Этим вопросам посвящена данная глава.

1. Поле комплексных чисел $\mathbb{C}$

A. Алгебраическая замкнутость поля.

2.1. Под *числовым полем* $\mathbb{K}$ будем подразумевать множество рациональных чисел $\mathbb{Q}$ или множество вещественных чисел $\mathbb{R}$ (в дальнейшем к этому списку добавится множество комплексных чисел $\mathbb{C}$). Каждое из этих числовых множеств замкнуто относительно четырёх арифметических операций: сложения, вычитания, умножения и деления на ненулевое число; это означает, что результатом операции над произвольными элементами множества $\mathbb{K}$ также является элемент множества $\mathbb{K}$. Напомним свойства арифметических операций:

(1) коммутативность сложения:
$$\forall a, b \in \mathbb{K} \quad a + b = b + a;$$

(2) ассоциативность сложения:
$$\forall a, b, c \in \mathbb{K} \quad (a + b) + c = a + (b + c);$$

(3) коммутативность умножения:
$$\forall a, b \in \mathbb{K} \quad ab = ba;$$

(4) ассоциативность умножения:
$$\forall a, b, c \in \mathbb{K} \quad (ab)c = a(bc);$$

(5) дистрибутивность умножения относительно сложения:

$$\forall a, b, c \in \mathbb{K} \quad (a + b)c = ac + bc.$$

2.2. В алгебре под *полем* понимают любое множество, снабжённое двумя операциями, которые называются сложением и умножением и обладают перечисленными выше свойствами, причём в этом множестве однозначно разрешимы уравнения $a + x = b$ и $kx = b$ (при любых a, b, k, причём $k \neq 0$).

2.3. Строгое определение поля см. в. приложении IV (определение IV.35, с. 341; для его понимания нужно прочитать приложение IV с самого начала).

2.4. Множества натуральных чисел $\mathbb{N}$ и целых чисел $\mathbb{Z}$ не являются полями, поскольку в них не выполняется операция деления (а в $\mathbb{N}$ ещё и операция вычитания).

2.5. Подмножество $\mathbb{L} \subset \mathbb{K}$ поля $\mathbb{K}$, само являющееся полем, называется *подполем* в $\mathbb{K}$. Поле $\mathbb{K}$ при этом называется *расширением* поля $\mathbb{L}$. Например, поле вещественных чисел $\mathbb{R}$ является расширением поля рациональных чисел $\mathbb{Q}$.

2.6. Пусть $\mathbb{K}$ — некоторое поле[1]. *Многочленом* степени n (от одной переменной x) над полем $\mathbb{K}$ называется функция $P : \mathbb{K} \to \mathbb{K}$ вида

$$P(x) = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-2} x^2 + a_{n-1} x + a_n, \quad (2.1)$$

где $a_j \in \mathbb{K}$, $j = 0, 1, \dots, n$, причём $a_0 \neq 0$. Если $a_0 = 1$, то многочлен называется *нормированным*. Множество всех многочленов от переменной x над полем $\mathbb{K}$ обозначается $\mathbb{K}[x]$.

2.7. Каждое слагаемое в (2.1) называется *одночленом*, а показатель степени переменной x в одночлене — *степенью одночлена*. Степень нулевого одночлена $0x^k \equiv 0$ считается равной $-\infty$. *Степень* многочлена $\deg P(x)$ — это наибольшая из степеней входящих в его состав одночленов. В дальнейшем при необходимости будем указывать степень многочлена при помощи нижнего индекса: запись $P_n(x)$ обозначает многочлен степени n.

2.8. *Корнем* многочлена $P(x) \in \mathbb{K}[x]$ называется такое число $c \in \mathbb{K}$, что $P(c) = 0$.

2.9. Отметим, что один и тот же многочлен можно рассматривать как элемент разных множеств $\mathbb{K}[x]$; при этом свойства одного и того же многочлена в различных множествах $\mathbb{K}[x]$ различны. Например, многочлен $x^2 - 2 \in \mathbb{Q}[x]$ не имеет корней, тогда как $x^2 - 2 \in \mathbb{R}[x]$ имеет два корня $\pm\sqrt{2}$.

[1] Можно для простоты считать, что это одно из числовых полей $\mathbb{Q}$ или $\mathbb{R}$.

2.10. Поле $\mathbb{K}$ называется *алгебраически замкнутым*, если любой многочлен из $\mathbb{K}[x]$ имеет корень $c \in \mathbb{K}$.

2.11. Поле рациональных чисел $\mathbb{Q}$ не является алгебраически замкнутым, так как, например, многочлен $x^2 - 2 \in \mathbb{Q}[x]$ не имеет корней (разумеется, речь идёт лишь о рациональных корнях). Если же рассматривать этот многочлен не как элемент множества $\mathbb{Q}[x]$, а как элемент множества $\mathbb{R}[x]$, то он имеет два (вещественных) корня $\pm\sqrt{2}$. Таким образом, переход от поля $\mathbb{Q}$ к его расширению $\mathbb{R}$ позволяет расширить класс многочленов, имеющих корни. Построение указанного расширения осуществляется введением иррациональных чисел.

Однако поле вещественных чисел $\mathbb{R}$ всё ещё не является алгебраически замкнутым: многочлен $x^2 + 1$ не имеет вещественных корней. Поэтому возникает задача построения такого расширения поля $\mathbb{R}$ вещественных чисел, которое было бы алгебраически замкнутым.

Б. Существование поля $\mathbb{C}$. Алгебраическая форма записи комплексных чисел.

2.12. Определение. *Полем комплексных чисел* $\mathbb{C}$ называется поле, обладающее следующими свойствами:

$\mathbb{C}$1: поле $\mathbb{C}$ содержит поле вещественных чисел $\mathbb{R}$ в качестве подполя (т.е. $\mathbb{C}$ является расширением $\mathbb{R}$);

$\mathbb{C}$2: поле $\mathbb{C}$ содержит такой элемент i, что $i^2 = -1$;

$\mathbb{C}$3: среди полей с указанными свойствами поле $\mathbb{C}$ минимально, т.е. если $K \subset \mathbb{C}$ — какое-либо подполе, содержащее $\mathbb{R}$ и i, то $K = \mathbb{C}$.

2.13. Докажем, что *если поле $\mathbb{C}$ существует, то каждый его элемент можно представить в виде $x + iy$, где i — какой-либо элемент поля $\mathbb{C}$, обладающий свойством $i^2 = -1$.*

Рассмотрим следующее подмножество поля $\mathbb{C}$:

$$K = \left\{ x + iy \;\middle|\; x, y \in \mathbb{R} \right\} \subset \mathbb{C}.$$

Требуется доказать, что $K = \mathbb{C}$.

2.14. Сначала докажем, что[1]

$$x + iy = a + ib \;\Leftrightarrow\; (x = a) \;\wedge\; (y = b).$$

Запишем равенство $x + iy = a + ib$ в виде $x - a = i(b - y)$ и предположим, что $b - y \neq 0$; тогда $i = \dfrac{x - a}{b - y}$, т.е. $i \in \mathbb{R}$, что невозможно. Итак, $y = b$, но тогда и $x = a$.

[1] Знак $\wedge$ означает логическую связку «*и*» (см. определение I.14, с. 294).

2.15. Из свойств арифметических операций в поле и соотношения $i^2 = -1$ следует, что для любых элементов $a+ib$ и $c+id$ этого подмножества выполняются следующие соотношения:

$$(a+ib)+(c+id) = (a+c)+i(b+d), \tag{2.2}$$

$$(a+ib)\cdot(c+id) = (ac-bd)+i(ad+bc); \tag{2.3}$$

таким образом, *множество K замкнуто относительно операций сложения и умножения.*

2.16. Будем искать число $-(a+ib)$, *противоположное* $a+ib$, в виде $x+iy$; для этого требуется решить уравнение $(a+ib)+(x+iy)=0$:

$$(a+ib)+(x+iy)=0 \;\Leftrightarrow\; (a+x)+i(b+y)=0 \;\Leftrightarrow$$
$$\Leftrightarrow \begin{cases} x=-a \\ y=-b \end{cases} \Leftrightarrow\; -(a+ib)=(-a)+i(-b).$$

Итак, *множество K замкнуто относительно операции взятия противоположного элемента и, следовательно, относительно операции вычитания*:

$$(a+ib)-(c+id)=(a-c)+i(b-d).$$

2.17. Будем искать число $(a+ib)^{-1}$, *обратное* к числу $a+ib\neq 0$, в виде $x+iy$; для этого нужно решить уравнение $(a+ib)\cdot(x+iy)=1$:

$$(a+ib)\cdot(x+iy)=1 \;\Leftrightarrow\; (ax-by)+i(bx+ay)=1 \;\Leftrightarrow\; \begin{cases} ax-by=1, \\ bx+ay=0. \end{cases}$$

Чтобы решить эту систему, умножим первое уравнение на a, второе на b и сложим полученные уравнения; получим

$$a^2x+b^2x=a \quad\Rightarrow\quad x=\frac{a}{a^2+b^2} \quad\Rightarrow\quad y=\frac{-b}{a^2+b^2},$$

так что

$$(a+ib)^{-1}=\frac{a}{a^2+b^2}+i\frac{-b}{a^2+b^2} \overset{\text{def}}{=\!=} \frac{a-ib}{a^2+b^2}, \tag{2.4}$$

т.е. *множество K замкнуто относительно операции взятия обратного элемента и, следовательно, относительно операции деления*:

$$\frac{a+ib}{c+id}=\frac{ac+bd}{c^2+d^2}+i\frac{bc-ad}{c^2+d^2}$$

(проверьте самостоятельно!).

2.18. На практике деление комплексных чисел осуществляют, домножая числитель $a+ib$ и знаменатель $c+id$ дроби $\dfrac{a+ib}{c+id}$ на число

$c - id$; в результате в знаменателе образуется вещественное число $(c+id)(c-id) = c^2 + d^2$, после чего деление не представляет труда:

$$\frac{a+ib}{c+id} = \frac{(a+ib)(c-id)}{(c+id)(c-id)} = \frac{(ac+bd)+i(bc-ad)}{c^2+d^2}.$$

Число $c - id$ называется *сопряжённым* к числу $c + id$.

2.19. Поскольку множество K замкнуто относительно операций сложения, вычитания, умножения и деления, оно является *полем*, причём это поле содержит все вещественные числа и такой элемент i, что $i^2 = -1$. Таким образом, $K = \mathbb{C}$.

2.20. Теперь докажем *существование* поля комплексных чисел; для этого нужно построить какую-либо его *модель* из уже известных математических объектов. Предыдущее исследование подсказывает, как это сделать. (Другая модель поля $\mathbb{C}$ будет построена в гл. 3; см. с. 61.)

2.21. Рассмотрим множество[1] $\mathbb{R}\times\mathbb{R}$ всех упорядоченных пар вещественных чисел и введём на нём операции сложения и умножения формулами, подсказанными соотношениями (2.2) и (2.3):

$$(a,b)+(c,d) = (a+c,\ b+d), \tag{2.5}$$

$$(a,b)\cdot(c,d) = (ac-bd,\ ad+bc). \tag{2.6}$$

2.22. Коммутативность и ассоциативность этих операций, а также дистрибутивность умножения относительно сложения проверяются непосредственно, например,

$$\begin{aligned}(a,b)\cdot\Big[(c_1,d_1)+(c_2,d_2)\Big] &= (a,b)\cdot(c_1+c_2, d_1+d_2) = \\ &= \Big(a(c_1+c_2)-b(d_1+d_2),\ a(d_1+d_2)+b(c_1+c_2)\Big) = \\ &= \Big(ac_1-bd_1,\ ad_1+bc_1\Big)+\Big(ac_2-bd_2,\ ad_2+bc_2\Big) = \\ &= (a,b)\cdot(c_1,d_1)+(a,b)\cdot(c_2,d_2).\end{aligned}$$

2.23. Из очевидных соотношений

$$(a,b)+(0,0) = (a,b), \quad (a,b)\cdot(1,0) = (a,b)$$

следует, что пары $(0,0)$ и $(1,0)$ играют роль нуля и единицы. Формула (2.4) подсказывает, как должен выглядеть элемент, обратный

[1] Символом $A\times B$ обозначается *декартово произведение* множеств A и B, состоящее из упорядоченных пар (a,b), где $a\in A$, $b\in B$ (см. определение II.25, с. 306).

к $(a, b) \neq (0; 0)$:

$$(a, b)^{-1} = \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2} \right); \tag{2.7}$$

непосредственная проверка (умножение (a, b) на $(a, b)^{-1}$ при помощи формулы (2.6)) доказывает справедливость этой формулы. Таким образом, множество $\mathbb{R} \times \mathbb{R}$, операции в котором определяются формулами (2.5) и (2.6), является *полем*, которое и является требуемой моделью поля $\mathbb{C}$.

2.24. Далее, согласно формулам (2.5) и (2.6)

$$(a, 0) + (c, 0) = (a + c, 0), \quad (a, 0) \cdot (c, 0) = (ac, 0),$$

т.е. операции над па́рами с нулевыми вторыми компонентами сводятся к операциям над первыми компонентами, что позволяет отождествить пары такого вида с вещественными числами, $(a, 0) \equiv a$, и считать *поле* $\mathbb{R}$ *подполем поля* $\mathbb{C}$.

2.25. Положим $i \overset{\text{def}}{=\!=} (0, 1)$; тогда

$$i^2 = -1, \ (0, b) = (0, 1) \cdot (b, 0) = ib, \ (a, b) = (a, 0) + (0, b) = a + ib.$$

Элемент i называется *мнимой единицей.*

2.26. Итак, *каждый элемент* (a, b) *поля* $\mathbb{C}$ *может быть записан в виде* $a + ib$; эта форма записи называется *алгебраической*, а числа a и b — *вещественной* и *мнимой* частями рассматриваемого комплексного числа; для них приняты обозначения

$$a = \mathrm{Re}(a, b) = \mathrm{Re}(a + ib), \quad b = \mathrm{Im}(a, b) = \mathrm{Im}(a + ib).$$

Обратите внимание, что $\mathrm{Im}\, z$ — это вещественное число (коэффициент при мнимой единице); так, $\mathrm{Im}(2 + 3i) = 3$ (а не $3i$).

Итак, мы доказали существование поля $\mathbb{C}$, построив его модель, состоящую из упорядоченных пар вещественных чисел, операции над которыми определяются формулами (2.5) и (2.6).

2.27. Комплексное число $a - ib$ называется *сопряжённым* к числу $z = a + ib$ и обозначается $\bar{z}$ (используется также обозначение z^*); очевидно,

$$\mathrm{Re}\, z = \frac{z + \bar{z}}{2}, \quad \mathrm{Im}\, z = \frac{z - \bar{z}}{2i}.$$

2.28. Ясно, что $z = \bar{z}$ тогда и только тогда, когда $z \in \mathbb{R}$.

Следующее утверждение проверяется прямыми вычислениями.

2.29. Предложение. *Операция сопряжения* $z \mapsto \bar{z}$ *обладает следующими свойствами:*

(1) $\overline{z_1 \pm z_2} = \bar{z}_1 \pm \bar{z}_2$;
(2) $\overline{z_1 \cdot z_2} = \bar{z}_1 \cdot \bar{z}_2$;
(3) $\overline{z_1 / z_2} = \bar{z}_1 / \bar{z}_2$;
(4) $\bar{\bar{z}} = z$ (*инволютивность*).

2.30. В отличие от поля $\mathbb{R}$, в поле комплексных чисел $\mathbb{C}$ нет отношения порядка, т.е. комплексные числа нельзя сравнивать.

B. Тригонометрическая форма записи.

2.31. Комплексное число $z = x+iy$ можно изобразить точкой (x, y) координатной плоскости либо *радиус-вектором* этой точки (т.е. направленным отрезком с началом в $(0,0)$ и концом в (x, y)). При сложении чисел соответствующие векторы складываются (по правилу параллелограмма, известному читателю из школьного курса геометрии); разность чисел $z_2 - z_1$ изображается вектором с началом в точке z_1 и концом в точке z_2.

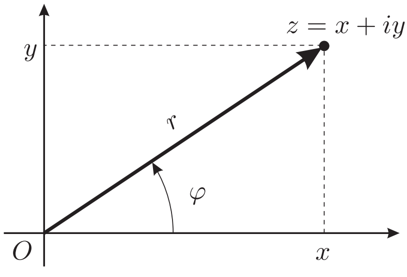


Рис. 2.1.

2.32. Точки на плоскости можно описывать не только декартовыми, но и полярными координатами. В этом случае точка, изображающая комплексное число z, описывается парой своих полярных координат, в качестве которых выступают длина r радиус-вектора этой точки и (ориентированный) угол φ между положительным направлением оси Ox и этим радиус-вектором (см. рис. 2.1).

2.33. Полярные координаты точки $z = x + iy = (x, y)$ называются соответственно *модулем* и *аргументом* комплексного числа z и обозначаются $|z|$ и $\arg z$. Очевидно, для комплексного числа $z = x + iy$

$$|z|^2 = x^2 + y^2 = z \cdot \bar{z}.$$

Поскольку аргумент определён не однозначно, а лишь с точностью до слагаемого, кратного 2π, рассматривают также *множество всех значений аргумента* данного комплексного числа z; оно обозначается $\operatorname{Arg} z$. В таком случае $\arg z$ называют *главным значением аргумента* и накладывают на него ограничение, обычно $\arg z \in [0, 2\pi)$ или $\arg z \in (-\pi, \pi]$. Итак,

$$\operatorname{Arg} z = \Big\{ \arg z + 2\pi n,\ n \in \mathbb{Z} \Big\},\ \arg z \in [0, 2\pi)\ (\text{или } \arg z \in (-\pi, \pi]).$$

Аргумент числа 0 не определён.

2.34. Используя формулы связи декартовых координат и полярных координат

$$x = r\cos\varphi, \quad y = r\sin\varphi,$$

получаем для комплексного числа $z = x + iy$ так называемую *тригонометрическую форму записи*:

$$z = x + iy = r\cos\varphi + ir\sin\varphi = r\big(\cos\varphi + i\sin\varphi\big).$$

Г. Формула Муавра.

2.35. Перемножим два комплексных числа, записанных в тригонометрической форме:

$$\begin{aligned} &r_1(\cos\varphi_1 + i\sin\varphi_1)\cdot r_2(\cos\varphi_2 + i\sin\varphi_2) = \\ &= r_1r_2\Big(\cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2 + i\cos\varphi_1\sin\varphi_2 + i\sin\varphi_1\cos\varphi_2\Big) = \\ &= r_1r_2\Big(\cos(\varphi_1+\varphi_2) + i\sin(\varphi_1+\varphi_2)\Big). \qquad (2.8) \end{aligned}$$

2.36. Таким образом,

$$|z_1\cdot z_2| = |z_1|\cdot|z_2|, \qquad \operatorname{Arg}(z_1\cdot z_2) = \operatorname{Arg} z_1 + \operatorname{Arg} z_2.$$

Вторая из этих формул понимается в том смысле, что для любого $\varphi \in \operatorname{Arg}(z_1\cdot z_2)$ существуют такие $\varphi_1 \in \operatorname{Arg} z_1$ и $\varphi_2 \in \operatorname{Arg} z_2$, что $\varphi = \varphi_1 + \varphi_2$.

2.37. Отсюда очевидным образом получается так называемая *формула Муавра*[1]:

$$\Big(\cos\varphi + i\sin\varphi\Big)^n = \cos n\varphi + i\sin n\varphi, \qquad (2.9)$$

верная для любых натуральных показателей степеней n. Эта формула позволяет легко выполнять рутинные тригонометрические преобразования функций кратного аргумента.

2.38. Например, запишем формулу Муавра (2.9) для $n = 3$:

$$(\cos\varphi + i\sin\varphi)^3 = \cos 3\varphi + i\sin 3\varphi.$$

Развернём левую часть по формуле бинома Ньютона (см. (III.11), с. 326):

$$\cos^3\varphi + 3\cos^2\varphi\cdot i\sin\varphi + 3\cos\varphi\cdot i^2\sin^2\varphi + i^3\sin^3\varphi = \cos 3\varphi + i\sin 3\varphi.$$

Приравнивая вещественные и мнимые части левой и правой частей этого равенства, получаем

$$\cos 3\varphi = \cos^3\varphi - 3\cos\varphi\sin^2\varphi, \quad \sin 3\varphi = 3\cos^2\varphi\sin\varphi - \sin^3\varphi.$$

[1] А. де Муавр (A. de Moivre) — английский математик французского происхождения (1667–1754).

2.39. При помощи формулы (2.4) легко проверить соотношение

$$\Bigl(\cos\varphi + i\sin\varphi\Bigr)^{-1} = \cos\varphi - i\sin\varphi;$$

отсюда немедленно вытекает, что формула (2.9) верна для всех $n \in \mathbb{Z}$.

Д. Формула Эйлера. Показательная форма записи.

2.40. Рассмотрим функцию $f(\varphi) = \cos\varphi + i\sin\varphi$. Из (2.8) следует, что она обладает свойством показательной функции

$$f(\varphi_1)\cdot f(\varphi_2) = f(\varphi_1+\varphi_2).$$

Положим по определению

$$e^{i\varphi} \overset{\text{def}}{=\!=} \cos\varphi + i\sin\varphi.$$

Эта формула называется *формулой Эйлера*[1]. Она является *определением* функции $e^{i\varphi}$ и потому не нуждается в доказательстве, хотя математический анализ доставляет множество доводов в пользу целесообразности такого определения.

2.41. Для любого $z = x + iy \in \mathbb{C}$ положим по определению

$$e^{z} \overset{\text{def}}{=\!=} e^{x}\bigl(\cos y + i\sin y\bigr).$$

2.42. Нетрудно доказать, что для любых $z_1, z_2 \in \mathbb{C}$

$$e^{z_1}\cdot e^{z_2} = e^{z_1+z_2}.$$

Действительно, если $z_1 = x_1 + iy_1$ и $z_2 = x_2 + iy_2$, то

$$e^{z_1}\cdot e^{z_2} = e^{x_1}\Bigl(\cos y_1 + i\sin y_1\Bigr)\cdot e^{x_2}\Bigl(\cos y_2 + i\sin y_2\Bigr) = \\ = e^{x_1+x_2}\Bigl(\cos(y_1+y_2) + i\sin(y_1+y_2)\Bigr) = e^{z_1+z_2}.$$

2.43. Итак, комплексные числа могут быть записаны в *показательной форме*

$$z = r\,e^{i\varphi}, \quad \text{где } r = |z|,\ \varphi \in \operatorname{Arg} z,$$

которая, в сущности, является удобным сокращением для тригонометрической формы. Отметим, что показательная функция обладает свойствами

$$e^{i\pi} = -1, \quad e^{2i\pi} = 1;$$

следовательно, она периодична с периодом $2i\pi$:

$$e^{z+2i\pi} = e^{z}.$$

---

[1] Л. Эйлер (L. Euler) — швейцарский, немецкий и российский математик и механик (1707–1783).

2.44. Тригонометрическая и показательная формы записи удобны при возведении комплексных чисел в степень:

$$\Big[r(\cos\varphi + i\sin\varphi)\Big]^n = r^n(\cos n\varphi + i\sin n\varphi), \qquad \Big[r\,e^{i\varphi}\Big]^n = r^n \cdot e^{in\varphi}.$$

2. Извлечение корней из комплексных чисел

А. Формула извлечения корня.

2.45. Число w называется *корнем n-й степени из числа z*, если $w^n = z$:

$$w = \sqrt[n]{z} \quad \Leftrightarrow \quad w^n = z.$$

2.46. Выведем формулу для извлечения корня из комплексного числа. Представим числа w, z в показательной форме:

$$w = Re^{i\Phi}, \quad z = re^{i\varphi}.$$

Наша задача — по данным r, φ найти R, Φ. Учитывая периодичность показательной функции (см. п. 2.43), имеем:

$$\Big(Re^{i\Phi}\Big)^n = re^{i\varphi} \quad \Leftrightarrow \quad R^n e^{in\Phi} = re^{i\varphi} \quad \Leftrightarrow$$

$$\begin{cases} R^n = r, \\ n\Phi = \varphi + 2\pi k, \quad k \in \mathbb{Z}, \end{cases} \quad \Leftrightarrow \quad \begin{cases} R = r^{1/n}, \\ \Phi = \dfrac{\varphi}{n} + \dfrac{2\pi k}{n}, \quad k \in \mathbb{Z}. \end{cases}$$

Здесь $r^{1/n}$ означает арифметическое значение корня n-й степени из положительного вещественного числа.

2.47. Теорема. *Существует ровно n различных значений корня n-й степени из ненулевого комплексного числа $z = r\,e^{i\varphi}$, которые выражаются формулой*

$$\sqrt[n]{z} = r^{1/n} \exp\left(i\frac{\varphi + 2\pi k}{n}\right) \equiv r^{1/n}\left(\cos\frac{\varphi + 2\pi k}{n} + i\sin\frac{\varphi + 2\pi k}{n}\right),$$

где индекс k принимает значения[1] $0, 1, \ldots, n-1$. *Геометрически эти корни изображаются вершинами правильного n-угольника, вписанного в окружность радиуса $r^{1/n}$.*

Доказательство. Требуется доказать, что два значения корня n-й степени из числа $r\,e^{i\varphi}$

$$w_k = r^{1/n} \exp\left(i\frac{\varphi + 2\pi k}{n}\right), \quad w_l = r^{1/n} \exp\left(i\frac{\varphi + 2\pi l}{n}\right)$$

[1] Точнее, индекс k пробегает *полную систему вычетов* по модулю n; см. определение IV.45, с. 343.

равны в том и только том случае, когда целое число $k-l$ делится на n. Напомним, что модули равных комплексных чисел равны, а аргументы могут отличаться на слагаемое, кратное 2π. Поэтому

$$w_k = w_l \ \Leftrightarrow \ \frac{\varphi + 2\pi k}{n} = \frac{\varphi + 2\pi l}{n} + 2\pi t,\ t \in \mathbb{Z},\ \Leftrightarrow \ \frac{k-l}{n} = t \in \mathbb{Z}.$$

Следовательно, различные значения $\sqrt[n]{z}$ будут получаться при $k = 0, 1, \dots, n-1$. □

2.48. Пример. Найдём все значения $\sqrt[3]{-i}$.

Решение. Поскольку $|-i| = 1$, $\arg(-i) = 3\pi/2$, имеем:

$$\sqrt[3]{-1} = \left(1 \cdot e^{i\frac{3\pi}{2}}\right)^{1/3} = \exp i\frac{\frac{3\pi}{2} + 2\pi k}{3}, \quad k = 0, 1, 2.$$

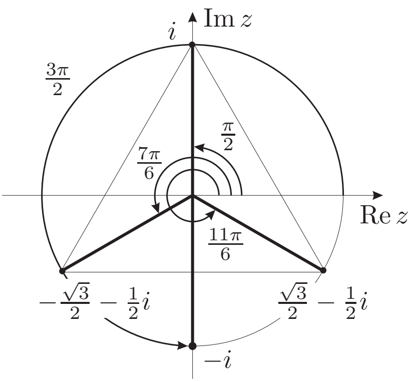


Рис. 2.2. К примеру 2.48

При $k = 0$ получаем

$$\left(\sqrt[3]{-i}\right)_0 = \exp\frac{i\pi}{2} = i;$$

при $k = 1$

$$\left(\sqrt[3]{-i}\right)_1 = \exp\frac{7i\pi}{6} = -\frac{\sqrt{3}}{2} - \frac{1}{2}i;$$

при $k = 2$

$$\left(\sqrt[3]{-i}\right)_1 = \exp\frac{11i\pi}{6} = \frac{\sqrt{3}}{2} - \frac{1}{2}i.$$

Найденные значения корня изображены на рис. 2.2; видно, что соответствующие точки являются вершинами правильного треугольника.

2.49. Отметим, что среди n значений корня n-й степени из комплексного числа нет оснований предпочитать какое-либо одно значение остальным, поэтому понятие «арифметического значения корня» не вводится.

Б. Извлечение квадратного корня.

2.50. Извлечение квадратного корня можно осуществить, не обращаясь к показательной форме. Пусть $\sqrt{a+ib} = x + iy$, где $b \neq 0$. Тогда

$$a + ib = (x+iy)^2 = x^2 - y^2 + 2ixy \ \Leftrightarrow \ \begin{cases} x^2 - y^2 = a, \\ 2xy = b, \end{cases}$$

причём нас интересуют только вещественные решения этой системы. Нам уже известно, что задача имеет ровно два решения. Возводя в квадрат уравнения системы и складывая полученные равенства, находим

$$\begin{cases} \left(x^2 - y^2\right)^2 = a^2 \\ 4x^2y^2 = b^2 \end{cases} \Rightarrow \left(x^2 + y^2\right)^2 = a^2 + b^2 \Rightarrow x^2 + y^2 = \sqrt{a^2 + b^2},$$

где берётся арифметическое (положительное) значение корня из положительного вещественного числа. Учитывая первое уравнение $x^2 - y^2 = a$ исходной системы, получаем

$$2x^2 = \sqrt{a^2 + b^2} + a > 0, \quad 2y^2 = \sqrt{a^2 + b^2} - a > 0,$$

откуда

$$x = \sigma_1 \sqrt{\frac{\sqrt{a^2 + b^2} + a}{2}}, \quad y = \sigma_2 \sqrt{\frac{\sqrt{a^2 + b^2} - a}{2}},$$

где σ_1, σ_2 принимают значения ± 1, а символ $\sqrt{}$ обозначает *арифметический* квадратный корень. Найденные числа x и y удовлетворяют первому уравнению системы $x^2 - y^2 = a$. Проверим выполнение второго уравнения $2xy = b$:

$$2\sigma_1\sigma_2 \sqrt{\frac{\sqrt{a^2 + b^2} + a}{2}} \cdot \sqrt{\frac{\sqrt{a^2 + b^2} - a}{2}} = b \Leftrightarrow \sigma_1\sigma_2\sqrt{b^2} = b,$$

откуда

$$\sigma_1\sigma_2 = \begin{cases} -1, & \text{если } b < 0, \\ 1, & \text{если } b > 0. \end{cases}$$

так что $\sigma_2 = \sigma_1 \cdot \operatorname{sign} b$, где символ sign обозначает функцию «знак вещественного числа»[1]. Итак, приходим к окончательной формуле

$$\sqrt{a + ib} = \pm \left(\sqrt{\frac{\sqrt{a^2 + b^2} + a}{2}} + i \operatorname{sign} b \sqrt{\frac{\sqrt{a^2 + b^2} - a}{2}} \right). \quad (2.10)$$

[1]Эта функция определяется следующим образом:

$$\operatorname{sign} x = \begin{cases} -1, & x < 0, \\ 0, & x = 0, \\ 1, & x > 0. \end{cases}$$

3. Элементарные функции комплексного аргумента

А. Тригонометрические функции.

2.51. Из формулы Эйлера

$$e^{i\varphi} = \cos\varphi + i\sin\varphi$$

получаем, заменяя φ на $-\varphi$ и учитывая, что косинус — чётная функция, а синус — нечётная:

$$e^{-i\varphi} = \cos\varphi - i\sin\varphi.$$

Складывая или вычитая полученные равенства, находим

$$\cos\varphi = \frac{e^{i\varphi} + e^{-i\varphi}}{2}, \quad \sin\varphi = \frac{e^{i\varphi} - e^{-i\varphi}}{2i}.$$

Эти формулы позволяют ввести определения синуса и косинуса комплексного аргумента:

$$\cos z \stackrel{\text{def}}{=\!=} \frac{e^{iz} + e^{-iz}}{2}, \quad \sin z \stackrel{\text{def}}{=\!=} \frac{e^{iz} - e^{-iz}}{2i}. \tag{2.11}$$

2.52. Основное тригонометрическое тождество

$$\cos^2 z + \sin^2 z = 1$$

легко проверяется. Все формулы для тригонометрических функций, справедливые для вещественных значений аргумента, остаются верными и для комплексных.

Б. Гиперболические функции.

2.53. *Гиперболические функции* определяются равенствами, похожими на (2.11):

$$\operatorname{ch} z \stackrel{\text{def}}{=\!=} \frac{e^{z} + e^{-z}}{2}, \quad \operatorname{sh} z \stackrel{\text{def}}{=\!=} \frac{e^{z} - e^{-z}}{2}. \tag{2.12}$$

Эти функции называются соответственно гиперболическим косинусом и гиперболическим синусом; в зарубежной литературе их принято обозначать $\cosh z$ и $\sinh z$.

2.54. Сравнивая формулы (2.11) и (2.12), находим связь между гиперболическими и тригонометрическими функциями:

$$\operatorname{ch} iz = \cos z, \qquad \cos iz = \operatorname{ch} z,$$
$$\operatorname{sh} iz = i\sin z, \qquad \sin iz = i\operatorname{sh} z.$$

Эти соотношения позволяют получать соотношения для гиперболических функций из соответствующих соотношений для тригонометрических функций, например,

$$\operatorname{ch}^2 x - \operatorname{sh}^2 x = \cos^2 ix - \left(\frac{1}{i}\sin ix\right)^2 = \cos^2 ix + \sin^2 ix = 1,$$

$$\begin{aligned}\operatorname{ch} 2x = \cos 2ix &= \cos^2 ix - \sin^2 ix = \\ &= \operatorname{ch}^2 x - (i\sin x)^2 = \operatorname{ch}^2 x + \operatorname{sh}^2 x,\\ \cos(x+iy) &= \cos x \cos iy - \sin x \sin iy = \\ &= \cos x \operatorname{ch} y - i \sin x \operatorname{sh} y.\end{aligned}$$

4. Многочлены

A. Алгебраические операции над многочленами. Пусть $\mathbb{K}$ — одно из числовых полей $\mathbb{Q}$, $\mathbb{R}$ или $\mathbb{C}$. Напомним, что символом $\mathbb{K}[x]$ обозначается множество всех многочленов от переменной x с коэффициентами из поля $\mathbb{K}$.

2.55. Множество многочленов $\mathbb{K}[x]$ замкнуто относительно операций сложения, вычитания и умножения, т.е. результатом выполнения указанных операций над многочленами снова является многочлен. При этом для *степени* суммы и произведения многочленов $P(x)$ и $Q(x)$ имеют место соотношения

$$\deg\big(P(x)+Q(x)\big) \leqslant \max\big\{\deg P(x), \deg Q(x)\big\},$$
$$\deg\big(P(x)Q(x)\big) = \deg P(x) + \deg Q(x).$$

Для того чтобы эти соотношения были справедливы и в том случае, когда среди многочленов $P(x)$, $Q(x)$, $P(x)+Q(x)$, $P(x)Q(x)$ имеются нулевые (степени которых, согласно определению 2.7, равны $-\infty$), нужно считать, что $-\infty \leqslant n$ и $-\infty + n = -\infty$ для любого n.

2.56. Очевидно, произведение ненулевых многочленов

$$P(x) = a_0 + \dots + a_n x^n, \quad Q(x) = b_0 + \dots + b_m x^m$$

также является ненулевым многочленом. Действительно, произведение равно

$$P(x)Q(x) = a_0 b_0 + \dots + a_n b_m x^{m+n}.$$

Так как $a_n b_m \neq 0$, то $P(x)Q(x)$ — ненулевой многочлен.

2.57. Таким образом, множество многочленов $\mathbb{K}[x]$ является коммутативным ассоциативным кольцом с единицей, не имеющим делителей нуля (см. определение IV.31, с. 340).

2.58. Будем говорить, что многочлен $P(x)$ *делится* на многочлен $D(x)$, если существует такой многочлен $Q(x)$, называемый *частным* от деления, что $P(x) = D(x)Q(x)$; при этом будем писать

$$Q(x) = \frac{P(x)}{D(x)}.$$

Ясно, что

$$\deg Q(x) = \deg P(x) - \deg D(x).$$

Деление многочленов осуществляется при помощи алгоритма «деления уголком», известного из школьного курса.

2.59. Один многочлен не всегда делится на другой, однако всегда возможно деление с остатком, похожее на деление с остатком в множестве целых чисел.

2.60. Предложение (теорема о делении с остатком). *Для любых многочленов $P(x), D(x) \in \mathbb{K}[x]$, где $D(x) \neq 0$, существуют такие многочлены $Q(x)$ (неполное частное) и $R(x)$ (остаток), что*

$$P(x) = D(x)Q(x) + R(x), \quad \deg R(x) < \deg D(x).$$

Многочлены $Q(x)$ и $R(x)$ определены этими условиями однозначно.

Доказательство. Существование. Если $\deg P(x) < \deg D(x)$, то положим $Q(x) = 0$ и $R(x) = P(x)$. Если $\deg P(x) \geqslant \deg D(x)$, то $Q(x)$ и $R(x)$ можно найти обычной процедурой «деления уголком». Именно, пусть

$$P(x) = a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n, \quad D(x) = b_0x^m + b_1x^{m-1} + \dots + b_{m-1}x + b_m,$$

где $a_0 \neq 0$, $b_0 \neq 0$, $n \geqslant m$. Рассмотрим многочлен

$$P_1(x) = P(x) - \frac{a_0}{b_0}x^{n-m}D(x).$$

Ясно, что $\deg P_1(x) < \deg P(x)$. Если $\deg P_1(x) < \deg D(x)$, то можно взять

$$Q(x) = \frac{a_0}{b_0}x^{n-m}, \quad R(x) = P_1(x).$$

В противном случае поступаем с многочленом $P_1(x)$ так же, как с $P(x)$. В конце концов получим такой многочлен

$$Q(x) = c_0x^{n-m} + c_1x^{n-m-1} + \dots + c_{n-m},$$

что $\deg\big(P(x) - Q(x)D(x)\big) < \deg D(x)$. Он и будет неполным частным от деления $P(x)$ на $D(x)$, а многочлен $R(x) = P(x) - Q(x)D(x)$ будет остатком.

Единственность. Теперь докажем, что неполное частное $Q(x)$ и остаток $R(x)$ определены однозначно. Предположим обратное, т.е. пусть

$$P(x) = Q_1(x)D(x) + R_1(x) = Q_2(x)D(x) + R_2(x), \quad Q_1(x) \neq Q_2(x),$$
$$\deg R_1(x) < \deg D(x), \quad \deg R_2(x) < \deg D(x).$$

Тогда

$$R_1(x) - R_2(x) = \big(Q_2(x) - Q_1(x)\big)D(x) \quad \Rightarrow$$
$$\Rightarrow \quad \deg\big(R_1(x) - R_2(x)\big) = \deg\big(Q_2(x) - Q_1(x)\big) + \deg D(x) \geqslant \deg D(x),$$

противоречие. Следовательно, $Q_1(x) = Q_2(x)$ и потому $R_1(x) = R_2(x)$. $\square$

Б. Теорема Безу.

2.61. Особое значение имеет деление с остатком на многочлен первой степени $x - c$ (линейный двучлен). В этом случае остаток имеет степень <1, т.е. является элементом поля $\mathbb{K}$ (числом). Таким образом, результат деления с остатком многочлена $P(x)$ на двучлен $x - c$ имеет вид

$$P(x) = (x - c)Q(x) + R, \quad R \in \mathbb{K}. \tag{2.13}$$

Полагая здесь $x = c$, видим, что *остаток R при делении многочлена $P(x)$ на линейный двучлен $x - c$ равен значению многочлена $P(x)$ в точке c*:

$$R = P(c).$$

2.62. Предложение (теорема Безу[1]). *Многочлен $P(x) \in \mathbb{K}[x]$ делится без остатка на $x - c$ тогда и только тогда, когда число c является корнем этого многочлена: $P(c) = 0$.*

Доказательство. Необходимость. Если $P(x)$ делится без остатка на $x - c$, т.е. $P(x) = (x - c)Q(x)$, то $P(c) = (c - c)Q(c) = 0$.

Достаточность. Пусть $P(c) = 0$. Тогда, положив $x = c$ в равенстве (2.13), получим $R = P(c) = 0$. □

В. Схема Горнера. Деление многочлена $P(x)$ на линейный двучлен $x - c$ удобно производить при помощи следующего алгоритма, называемого *схемой Горнера*[2]. Запишем равенство $P(x) = (x - c)Q(x) + R$ в развёрнутом виде:

$$\begin{aligned} &a_0x^n + a_1x^{n-1} + a_2x^{n-2} \cdots + a_{n-1}x + a_n = \\ &\quad = (x - c)\Big(b_0x^{n-1} + b_1x^{n-2} + b_2x^{n-3} \cdots + b_{n-2}x + b_{n-1}\Big) + R = \\ &\quad = b_0x^n + (b_1 - cb_0)x^{n-1} + \cdots + (b_{n-1} - cb_{n-2})x + (R - cb_{n-1}). \end{aligned}$$

Приравнивая коэффициенты при одинаковых степенях переменной x, получаем

$$\begin{gathered} b_0 = a_0, \quad b_1 = cb_0 + a_1, \quad b_2 = cb_1 + a_2, \quad \ldots, \\ b_{n-1} = cb_{n-2} + a_{n-1}, \quad R = cb_{n-1} + a_n. \end{gathered} \tag{2.14}$$

Эти формулы позволяют последовательно находить b_0, b_1, b_2, $\ldots$, b_{n-1} и R. Результаты расчётов удобно записывать в виде таблицы

	a_0	a_1	$\ldots$	a_{k-1}	a_k	$\ldots$	a_{n-1}	a_n
c	b_0	b_1	$\ldots$	b_{k-1}	$b_k = cb_{k-1} + a_k$	$\ldots$	b_{n-1}	R

[1] Э. Безу (É. Bésout) — французский математик (1730–1783).

[2] У. Дж. Горнер (W. G. Horner) — британский математик (1786–1837).

Элементы нижней строки этой таблицы вычисляются последовательно по формулам (2.14): $b_0 = a_0$, а каждый последующий элемент равен сумме предыдущего элемента, умноженного на c, и элемента, находящегося над ним.

Так как $P(c) = R$, схемой Горнера удобно пользоваться и для вычисления значения многочлена в точке c.

2.63. Пример. Разделим с остатком $2x^5 - 3x^3 + 4x^2 + 5$ на $x - 3$. Схема Горнера имеет следующий вид:

	2	0	-3	4	0	5
3	2	$3 \cdot 2 + 0 = 6$	$3 \cdot 6 - 3 = 15$	$3 \cdot 15 + 4 = 49$	$3 \cdot 49 + 0 = 147$	$3 \cdot 147 + 5 = 446$

Получаем неполное частное $Q(x) = 2x^4 + 6x^3 + 15x^2 + 49x + 147$ и остаток $R = 446$.

2.64. Пример. Значение многочлена $(1+i)x^3 + (2-i)x^2 + 1 - 2i$ в точке $c = 2 + i$ вычисляется следующим образом:

	$1+i$	$2-i$	0	$1-2i$
$2+i$	$1+i$	$3+2i$	$4+7i$	$2+16i$

Итак, $f(2+i) = 2 + 16i$.

Г. Корни многочлена.

2.65. Предложение. *Количество корней ненулевого многочлена* $P(x) \in \mathbb{K}[x]$ *в поле* $\mathbb{K}$ *не превосходит его степени* $\deg P(x)$.

Доказательство. Применим индукцию по степени многочлена[1]. База индукции ($n = 0$): многочлен $a_0 \neq 0$ нулевой степени корней не имеет. Допустим, что теорема доказана для многочленов степени $n - 1$, и докажем её для многочлена $P(x)$ степени n.

Если у многочлена $P(x)$ нет корней в $\mathbb{K}$, то утверждение теоремы верно. Если корни имеются и c_1 — один из корней, то по теореме Безу

$$P(x) = (x - c_1)Q(x), \quad \deg Q(x) = \deg P(x) - 1.$$

Если c_2 — какой-либо корень многочлена $P(x)$, отличный от c_1, то

$$0 = P(c_2) = \underbrace{(c_2 - c_1)}_{\neq 0} Q(c_2) \quad \Rightarrow \quad Q(c_2) = 0,$$

т.е. любой корень многочлена $P(x)$, кроме c_1, является также и корнем многочлена $Q(x)$ степени $n - 1$, который в силу предположения индукции имеет не более $n - 1$ корней. Следовательно, $P(x)$ имеет не более n корней. $\square$

[1] См. теорему III.12, с. 318.

2.66. Пусть число $c \in \mathbb{K}$ является корнем многочлена $P(x) \in \mathbb{K}[x]$, т.е. $P(c) = 0$. Согласно теореме Безу, многочлен $P(x)$ делится на $x - c$:

$$P(x) = (x - c)Q(x).$$

Может оказаться, что элемент c является также корнем частного $Q(x)$, т.е. $Q(c) = 0$ и потому

$$P(x) = (x - c)^2 Q_1(x).$$

2.67. *Кратностью* корня c многочлена $P(x)$ называется такое натуральное число k, что многочлен $P(x)$ делится на $(x - c)^k$, но не делится на $(x - c)^{k+1}$. Иными словами,

$$P(x) = (x - c)^k S(x), \quad S(c) \neq 0.$$

Корень c называется в этом случае k-*кратным*; при $k = 1$ используется также термин *простой* корень.

2.68. Удобно считать, что это определение применимо и для $k = 0$: в этом случае корень кратности 0 — это элемент поля $\mathbb{K}$, не являющийся корнем многочлена $P(x)$.

Следующее утверждение уточняет предложение 2.65.

2.69. Предложение. *Число корней многочлена с учётом их кратностей (т.е. если каждый корень считается столько раз, какова его кратность), не превосходит степени многочлена, причём равенство имеет место тогда и только тогда, когда многочлен разлагается на линейные множители.*

Д. Основная теорема алгебры.

2.70. Теорема (основная теорема алгебры)**.** *Каждый многочлен*[1] $P(z) \in \mathbb{C}[z]$ *имеет корень* $c \in \mathbb{C}$. *Иными словами, поле комплексных чисел алгебраически замкнуто.*

Доказательство этой теоремы мы откладываем до курса теории функций комплексной переменной, в котором оно окажется весьма коротким и изящным.

2.71. Таким образом, *каждый многочлен степени n в поле $\mathbb{C}$ имеет ровно n корней, если каждый корень считать столько раз, какова его кратность.* Отсюда вытекает, что любой многочлен $P(z) \in \mathbb{C}[z]$ разлагается в произведение линейных сомножителей:

$$P(z) = a_0(x - c_1)^{m_1} \dots (x - c_p)^{m_p},$$

где m_k — кратность корня c_k; при этом $\sum m_k = \deg P(z)$.

[1] Переменная, принимающая комплексные значения, традиционно обозначается буквой z.

Е. Многочлены с вещественными коэффициентами.

2.72. Каждый многочлен $P(x) \in \mathbb{R}[x]$ степени n с вещественными коэффициентами можно рассматривать и как многочлен $P(z) \in \mathbb{C}[z]$. В этом случае он имеет ровно n *комплексных* корней, если считать каждый корень столько раз, какова его кратность; вещественных же корней этот многочлен может не иметь вовсе. Однако комплексные корни у многочленов с вещественными коэффициентами могут появляться только сопряжёнными парами.

2.73. Предложение. *Пусть $P(z) \in \mathbb{C}[z]$ — многочлен с вещественными коэффициентами. Тогда для любого $z \in \mathbb{C}$ имеем*

$$P(\bar{z}) = \overline{P(z)}.$$

Доказательство. Пусть $P(z) = a_0 z^n + a_1 z^{n-1} + \dots + a_{n-1} z + a_n$. Поскольку коэффициенты вещественны,

$$\bar{a}_0 = a_0, \quad \bar{a}_1 = a_1, \quad \dots, \quad \bar{a}_n = a_n.$$

Поэтому

$$\begin{aligned}\overline{P(z)} &= \overline{a_0 z^n + a_1 z^{n-1} + \dots + a_{n-1} z + a_n} = \\ &= \bar{a}_0 \bar{z}^n + \bar{a}_1 \bar{z}^{n-1} + \dots + \bar{a}_{n-1} \bar{z} + \bar{a}_n = \\ &= a_0 \bar{z}^n + a_1 \bar{z}^{n-1} + \dots + a_{n-1} \bar{z} + a_n = P(\bar{z}). \qquad \square\end{aligned}$$

2.74. Предложение. *Если $P(z)$ — многочлен с вещественными коэффициентами и комплексное число c является его корнем, то сопряжённое число $\bar{c}$ также является корнем.*

Доказательство. В силу предложения 2.73 доказательство тривиально: $P(\bar{c}) = \overline{P(c)} = \bar{0} = 0$. □

2.75. Пусть c, $\bar{c}$ — пара сопряжённых комплексных корней многочлена $P(z)$ с вещественными коэффициентами. В разложение многочлена на множители входит произведение

$$(z - c)(z - \bar{c}) = z^2 - (c + \bar{c})z + c\bar{c} = z^2 - 2(\operatorname{Re} c)z + |c|^2,$$

являющееся квадратным трёхчленом; отметим, что дискриминант этого трёхчлена отрицателен:

$$D = 4(\operatorname{Re} c)^2 - 4|c|^2 < 0, \text{ так как } |c|^2 = (\operatorname{Re} c)^2 + (\operatorname{Im} c)^2 > (\operatorname{Re} c)^2,$$

так что трёхчлен не имеет вещественных корней.

2.76. Итак, мы убедились, что наличие у многочлена корней тесно связано с разложением этого многочлена на множители. Многочлен $P(x) \in \mathbb{K}[x]$, который можно разложить на множители ненулевых степеней, также являющиеся элементами множества $\mathbb{K}[x]$, называется *приводимым*; в противном случае, т.е. когда такого разложения не существует, многочлен называется *неприводимым*. Неприводимые многочлены в множестве $\mathbb{K}[x]$ являются аналогами простых чисел в множестве $\mathbb{Z}$ целых чисел.

2.77. Таким образом, *каждый многочлен с вещественными коэффициентами может быть разложен в произведение линейных множителей и неприводимых квадратных трёхчленов*:

$$P(x) = a(x - c_1)^{\alpha_1} \cdots (x - c_s)^{\alpha_s}(x + p_1x + q_1)^{\beta_1} \cdots (x + p_rx + q_r)^{\beta_r}.$$

Итак, в множестве $\mathbb{R}[x]$ многочленов с вещественными коэффициентами неприводимыми являются лишь многочлены первой степени и квадратные трёхчлены с отрицательным дискриминантом.

Ж. Многочлены с рациональными коэффициентами.

2.78. Множество $\mathbb{Q}[x]$ многочленов с рациональными коэффициентами устроено ещё сложнее: в нём имеются неприводимые многочлены любых степеней (приведите примеры самостоятельно).

2.79. Рассмотрим вопрос о нахождении *рациональных* корней многочленов с рациональными коэффициентами. Прежде всего отметим, что любой многочлен с рациональными коэффициентами можно представить в виде $\dfrac{a}{b}P(x)$, где $a, b \in \mathbb{Z}$, а $P(x)$ — многочлен с *целыми* коэффициентами, поэтому мы ограничимся рассмотрением именно таких многочленов.

2.80. Предложение. *Если рациональное число $x_0 = p/q$, где p и q — взаимно простые целые числа, является корнем многочлена*

$$P(x) = a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n$$

с целыми коэффициентами, то q является делителем старшего коэффициента a_0, а p — делителем свободного члена a_n.

Доказательство. Если $x_0 = p/q$ — корень многочлена $P(x)$, то

$$P(x_0) = a_0\frac{p^n}{q^n} + a_1\frac{p^{n-1}}{q^{n-1}} + \cdots + a_{n-1}\frac{p}{q} + a_n = 0.$$

Умножив обе части последнего равенства на q^n, запишем его в виде

$$a_0p^n = -a_1p^{n-1}q - \cdots - a_{n-1}pq^{n-1} - a_nq^n.$$

Так как все слагаемые правой части этого равенства делятся на q, то и левая часть должна делиться на q, а поскольку p и q взаимно просты, то делиться на q должен коэффициент a_0. Аналогично доказывается, что a_n делится на p. $\square$

2.81. Пример. Найдём указанным способом рациональные корни многочлена $3x^3 + x^2 + x - 2$. Делителями старшего коэффициента являются числа ± 1, ± 3, а делителями свободного члена — числа ± 1, ± 2. Поэтому рациональными корнями этого многочлена могут быть только числа ± 1, ± 2, $\pm 1/3$, $\pm 2/3$. Вычисляя значения многочлена в этих точках, убеждаемся, что лишь число $2/3$ является корнем. Других рациональных корней у данного многочлена нет.

2.82. В случае *нормированного* многочлена ($a_0 = 1$) с целыми коэффициентами рациональное число $x_0 = p/q$ может быть его корнем только при $q = \pm 1$. Итак, *все рациональные корни нормированного многочлена с целыми коэффициентами суть целые числа, являющиеся делителями свободного члена.*

2.83. Пример. Рассмотрим многочлен $x^3 + 4x^2 + 7x + 6$. Делителями свободного члена являются числа ± 1, ± 2, ± 3, ± 6. Вычисляя значения многочлена в этих точках, обнаруживаем, что единственным рациональным корнем является число -2.

5. Формула Кардано

Важную роль в укоренении комплексных чисел в математическом обиходе сыграла формула решения кубических уравнений, открытая в XVI в. и известная в настоящее время под названием формулы Кардано[1].

А. Вывод формулы Кардано.

2.84. Кубическое уравнение

$$y^3 + ay^2 + by + c = 0$$

может быть преобразовано с помощью замены переменной к виду, не содержащему квадрата неизвестной: положив $y = x - a/3$, получим

$$\left(x - \frac{a}{3}\right)^3 + a\left(x - \frac{a}{3}\right)^2 + b\left(x - \frac{a}{3}\right) + c = 0$$

или

$$x^3 + \left(b - \frac{1}{3}a^2\right)x + \left(\frac{2}{27}a^3 - \frac{1}{3}ab + c\right) = 0.$$

Таким образом, не ограничивая общности, можем считать, что кубическое уравнение имеет так называемый приведённый вид

$$x^3 + px + q = 0. \tag{2.15}$$

2.85. Будем искать корни этого уравнения в виде суммы двух слагаемых $x = u + v$. Подставляя это соотношение в уравнение, последовательно получим

$$(u + v)^3 + p(u + v) + q = 0,$$

[1] Дж. Кардано (G. Cardano) — итальянский математик, инженер, философ, медик и астролог (1501–1576).

$$u^3 + 3u^2v + 3uv^2 + v^3 + p(u+v) + q = 0,$$
$$u^3 + 3uv(u+v) + v^3 + p(u+v) + q = 0,$$
$$u^3 + v^3 + (3uv + p)(u+v) + q = 0.$$

Положим $3uv + p = 0$; тогда $u^3 + v^3 + q = 0$. Ясно, что если выполняются два последние уравнения, то число $x = u + v$ является решением уравнения (2.15). Итак, требуется решить систему, состоящую из двух уравнений

$$u^3 + v^3 = -q, \quad 3uv = -p.$$

Возведя второе уравнение в куб, получим уравнение

$$u^3v^3 = -\frac{p^3}{27}.$$

Таким образом, для чисел u^3 и v^3 известны сумма и произведение, поэтому согласно теореме Виета они являются корнями квадратного уравнения

$$z^2 + qz - \frac{p^3}{27} = 0.$$

Имеем

$$u^3 = -\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}, \quad v^3 = -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}},$$

откуда

$$u = \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}, \quad v = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

Теперь для решения кубического уравнения (2.15) получаем окончательную формулу:

$$x = \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}. \tag{2.16}$$

2.86. Рассмотрим многочлен $(x-1)(x-2)(x+3) \equiv x^3 - 7x + 6$ и попытаемся решить кубическое уравнение $x^3 - 7x + 6 = 0$. Здесь $p = -7$, $q = 6$, так что

$$x = \sqrt[3]{-3 - \sqrt{-\frac{100}{27}}} + \sqrt[3]{-3 + \sqrt{-\frac{100}{27}}}.$$

Мы обнаруживаем, что решение уравнения, имеющего три вещественных корня ($x_1 = 1$, $x_2 = 2$, $x_3 = -3$), требует извлечения квадратного корня из отрицательного вещественного числа — операции, невыполнимой в поле вещественных чисел.

Б. Анализ формулы Кардано.

2.87. При решении кубического уравнения $x^3 + px + q = 0$ по формуле Кардано (2.16) приходится извлекать два кубических корня, для каждого из которых в поле комплексных чисел имеется три значения, т.е. всего имеется девять комбинаций корней. Однако нужно брать только те комбинации, в которых произведение корней равно $uv = -p/3$. Обозначим через ω_1 и ω_2 кубические корни из единицы:

$$\omega_1 = -\frac{1}{2} + i\frac{\sqrt{3}}{2}, \quad \omega_2 = -\frac{1}{2} - i\frac{\sqrt{3}}{2}.$$

Если u_1, v_1 — одна подходящая пара значений u и v, то остальные значения для u будут $u\omega_1$ и $u\omega_2$, а соответствующие значения для v будут $v\omega_2$ и $v\omega_1$. Поэтому по формуле Кардано получаются три корня уравнения:

$$x_1 = u_1 + v_1, \quad x_2 = u_1\omega_1 + v_1\omega_2, \quad x_3 = u_1\omega_2 + v_1\omega_1.$$

2.88. Проведём анализ формулы Кардано (2.16) в предположении, что коэффициенты уравнения p и q являются вещественными числами. Из вида формулы ясно, что знак выражения $\dfrac{q^2}{4} + \dfrac{p^3}{27}$ существенно влияет на характер корней уравнения.

2.89. ***Случай 1:*** $\dfrac{q^2}{4} + \dfrac{p^3}{27} > 0$. В этом случае числа

$$-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}, \quad -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}$$

вещественны и различны. Если значение u_1 первого кубического корня взято вещественным, то и для второго корня нужно взять вещественное значение, и корни в этом случае будут

$$\begin{aligned} x_1 &= u_1 + v_1 \in \mathbb{R}, \\ x_2 &= u_1\omega_1 + v_1\omega_2 = -\frac{u_1 + v_1}{2} + i\frac{u_1 - v_1}{2}\sqrt{3}, \\ x_3 &= u_1\omega_2 + v_1\omega_1 = -\frac{u_1 + v_1}{2} - i\frac{u_1 - v_1}{2}\sqrt{3}, \end{aligned}$$

т.е. в этом случае имеется один вещественный и два комплексно сопряжённых корня.

2.90. ***Случай 2:*** $\dfrac{q^2}{4} + \dfrac{p^3}{27} = 0$. В этом случае числа

$$-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}, \quad -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}$$

вещественны и равны. Вещественному значению u_1 опять соответствует вещественное значение v_1, но теперь $u_1 = v_1$. Комплексные значения кубических корней подбираются так, чтобы $uv = -p/3$. В рассматриваемом случае уравнение имеет три вещественных корня, но среди них имеются два равных:

$$x_1 = u_1 + v_1 = 2u_1 \in \mathbb{R}, \; x_2 = u_1\omega_1 + v_1\omega_2 = -u_1, \; x_3 = u_1\omega_2 + v_1\omega_1 = -u_1.$$

2.91. ***Случай 3:*** $\dfrac{q^2}{4} + \dfrac{p^3}{27} < 0$; это возможно только при $p < 0$. Введём обозначение $p' = -p > 0$. Под знаками кубических корней оказываются сопряжённые комплексные числа

$$-\frac{q}{2} - i\sqrt{\frac{p'^3}{27} - \frac{q^2}{4}}, \quad -\frac{q}{2} + i\sqrt{\frac{p'^3}{27} - \frac{q^2}{4}}.$$

Для извлечения кубического корня найдём модули и аргументы этих чисел; отметим, что их модули r равны, а главные значения аргумента φ, выбираемые в промежутке $(-\pi, \pi]$, отличаются знаком. Имеем

$$r^2 = \left(-\frac{q}{2}\right)^2 + \left(\sqrt{\frac{p'^3}{27} - \frac{q^2}{4}}\right)^2 = \frac{p'^3}{27}, \quad \cos\varphi = -\frac{q}{2r},$$

откуда $r = \sqrt{p'^3/27}$ и

$$u_1 = \sqrt[3]{-\frac{q}{2} + i\sqrt{\frac{p'^3}{27} - \frac{q^2}{4}}} = r^{1/3}\left(\cos\frac{\varphi + 2\pi k}{3} + i\sin\frac{\varphi + 2\pi k}{3}\right) =$$
$$= \sqrt{\frac{p'}{3}}\left(\cos\frac{\varphi + 2\pi k}{3} + i\sin\frac{\varphi + 2\pi k}{3}\right),$$

где $k = 0, 1, 2$. Поскольку $uv = p'/3$, находим

$$v = \frac{p'/3}{u} = \sqrt{\frac{p'}{3}}\left(\cos\frac{\varphi + 2\pi k}{3} - i\sin\frac{\varphi + 2\pi k}{3}\right),$$

т.е. $v = \bar{u}$. Таким образом,

$$x = u + v = u + \bar{u} = 2\operatorname{Re} u = 2\sqrt{\frac{p'}{3}}\cos\frac{\varphi + 2\pi k}{3}, \quad k = 0, 1, 2,$$

т.е. уравнение имеет три попарно различных вещественных корня. Нахождение этих корней по формуле Кардано без привлечения комплексных чисел оказывается невозможным.

ГЛАВА 3

Матрицы

МАТРИЦА — одно из основных понятий современной математики, возникновение которого относится к довольно отдалённым временам. Наиболее ранние упоминания о матрицах содержатся в древнекитайских математических трактатах при изложении метода решения систем линейных уравнений. Развитие современной теории матриц относится к середине XIX в.

1. Матрицы и линейные операции над ними

A. Основные определения.

3.1. Определение. *Матрицей* размера $m \times n$ (или $(m \times n)$-матрицей) над множеством $\mathcal{X}$ называется упорядоченный набор mn элементов этого множества, записанных в виде прямоугольной таблицы, состоящей из m строк и n столбцов. Подчеркнём, что при указании размера матрицы сначала указывается количество её строк, а затем — количество столбцов. Элементы матрицы обозначаются буквами, снабжёнными двумя индексами, которые указывают положение элемента внутри матрицы; при этом используются две системы обозначений:

(1) *верхний* индекс указывает номер строки, *нижний* — номер столбца, не пересечении которых стоит элемент:

(2) *первый* индекс указывает номер строки, а *второй* — номер столбца:

$$A = \begin{pmatrix} a_1^1 & a_2^1 & \cdots & a_n^1 \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^m & a_2^m & \cdots & a_n^m \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}.$$

Используются также следующие сокращённые обозначения, смысл которых вполне очевиден:

$$A = (a_k^j)_n^m, \quad B = (b_{jk})_{mn}.$$

Иногда для явного указания будем записывать размер матрицы под буквой, обозначающей саму матрицу: $\underset{m\times n}{A}$.

3.2. В дальнейшем полезным окажется также следующее обозначение элементов матрицы: если $A = (a_k^j)_n^m$ или $B = (b_{jk})_{mn}$, то $\{A\}_k^j \overset{\text{def}}{=\!=} a_k^j$ и $\{B\}_{jk} \overset{\text{def}}{=\!=} b_{jk}$; фигурные скобки здесь обозначают «операцию извлечения элемента из матрицы».

3.3. Множество всех матриц размера $m \times n$, элементы которых принадлежат множеству $\mathcal{X}$, обозначается $\mathcal{X}^{m\times n}$. Мы будем работать в основном с матрицами, элементами которых являются числа из некоторого числового поля $\mathbb{K}$; множество таких матриц обозначается $\mathbb{K}^{m\times n}$. Наиболее важным для нас будет множество $\mathbb{R}^{m\times n}$ матриц с вещественными элементами.

3.4. Часто используются матрицы, состоящие из одной строки или из одного столбца; они называются *вектор-строками* и *вектор-столбцами* соответственно.

Множество всех вектор-строк длины m (т.е. состоящих из m элементов) обозначается

$$\mathbb{K}^{*m} = \bigl\{A = (a_1, a_2, \ldots, a_m) \mid a_1, a_2, \ldots, a_m \in \mathbb{K}\bigr\}$$

и называется *арифметическим пространством строк.*

3.5. Множество всех вектор-столбцов высоты n (т.е. состоящих из n элементов) обозначается

$$\mathbb{K}^n = \left\{ X = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix} \;\middle|\; x^1, x^2, \ldots, x^n \in \mathbb{K} \right\}$$

и называется *арифметическим пространством* $\mathbb{K}^n$ (или кратко *пространством* $\mathbb{K}^n$). Элементы столбцов нумеруются верхними индексами в соответствии с принятым выше соглашением; нижние индексы использовать неудобно, поскольку может возникнуть путаница со строками.

3.6. Во многих случаях удобно считать, что структурными элементами матрицы $A \in \mathcal{X}^{m\times n}$ являются не отдельные её элементы, а целые строки или столбцы. В этом случае вместо круглых скобок будем писать ограничители $\|\cdot\|$ и использовать нумерацию по системе «верхний-нижний индекс»; столбцы матрицы A при этом будем обозначать A_k, $k = 1, \ldots, n$, а строки — A^j, $j = 1, \ldots, m$:

$$A = \begin{pmatrix} a_1^1 & a_2^1 & \cdots & a_n^1 \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^m & a_2^m & \cdots & a_n^m \end{pmatrix} = \left\| A_1 \quad A_2 \quad \ldots \quad A_n \right\|,$$

где

$$A_1 = \begin{pmatrix} a_1^1 \\ a_1^2 \\ \vdots \\ a_1^m \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_2^1 \\ a_2^2 \\ \vdots \\ a_2^m \end{pmatrix}, \quad, \ldots, \quad A_n = \begin{pmatrix} a_n^1 \\ a_n^2 \\ \vdots \\ a_n^m \end{pmatrix},$$

либо

$$A = \begin{pmatrix} a_1^1 & a_2^1 & \cdots & a_n^1 \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^m & a_2^m & \cdots & a_n^m \end{pmatrix} = \begin{Vmatrix} A^1 \\ A^2 \\ \vdots \\ A^m \end{Vmatrix},$$

где

$$A^1 = \begin{pmatrix} a_1^1 & a_2^1 & \cdots & a_n^1 \end{pmatrix}, \quad \ldots, \quad A^m = \begin{pmatrix} a_1^m & a_2^m & \cdots & a_n^m \end{pmatrix}.$$

3.7. Более общим является понятие *блочной матрицы*, т.е. матрицы, элементы которой сами являются матрицами, например

$$A = \left(\begin{array}{cc:c} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ \hdashline a_{31} & a_{32} & a_{33} \end{array}\right) = \left\| \begin{array}{c:c} A_{11} & A_{12} \\ \hdashline A_{21} & A_{22} \end{array} \right\|;$$

здесь матрицу A размера 3×3 можно рассматривать как блочную матрицу размера 2×2 с блоками

$$A_{11} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad A_{12} = \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix},$$

$$A_{21} = \begin{pmatrix} a_{31} & a_{32} \end{pmatrix}, \qquad A_{22} = \begin{pmatrix} a_{33} \end{pmatrix}.$$

Для обозначения блочных матриц также будем использовать ограничители $\| \cdot \|$.

3.8. Выделяют следующие специальные виды матриц:

(i) *нулевая матрица* O: все её элементы равны нулю; в частности, рассматриваются нулевые столбцы и нулевые строки;

(ii) *квадратная матрица*: количество её строк равно количеству столбцов,

$$\begin{pmatrix} a_1^1 & a_2^1 & \cdots & a_n^1 \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^n & a_2^n & \cdots & a_n^n \end{pmatrix}.$$

Порядком квадратной матрицы называют количество её строк или столбцов. *Главная диагональ* квадратной матрицы состоит из элементов a_1^1, a_2^2, …, a_n^n, а *побочная диагональ* — из элементов a_n^1, a_{n-1}^2, …, a_2^{n-1}, a_1^n. Сумма элементов главной диагонали называется *следом* (англ. *trace*) квадратной матрицы и обозначается $\operatorname{tr} A$:

$$\operatorname{tr} A = a_1^1 + a_2^2 + \cdots + a_n^n = \sum_{j=1}^{n} a_j^j;$$

(iii) *диагональная матрица*: квадратная матрица, у которой все элементы вне главной диагонали равны нулю,

$$\begin{pmatrix} a_1^1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & a_2^2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & a_n^n \end{pmatrix} = \operatorname{diag}\left(a_1^1, a_2^2, \ldots, a_{n-1}^{n-1}, a_n^n\right);$$

(iv) *верхнетреугольная* (или *правая треугольная*) и *нижнетреугольная* (или *левая треугольная*) матрица: у первой из них $a_k^j = 0$ для всех $j > k$, а у второй $a_k^j = 0$ для всех $j < k$:

$$\begin{pmatrix} * & * & * & \cdots & * & * \\ 0 & * & * & \cdots & * & * \\ 0 & 0 & * & \cdots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & * & * \\ 0 & 0 & 0 & \cdots & 0 & * \end{pmatrix}, \qquad \begin{pmatrix} * & 0 & 0 & \cdots & 0 & 0 \\ * & * & 0 & \cdots & 0 & 0 \\ * & * & * & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ * & * & * & \cdots & * & 0 \\ * & * & * & \cdots & * & * \end{pmatrix}.$$

Б. Линейные операции над матрицами.

3.9. Определение. Матрицы A и B называются *равными* (обозначение $A = B$), если их размеры совпадают, а элементы, стоящие на одинаковых местах, равны между собой:

$$\forall j = 1, \ldots, m,\ \forall k = 1, \ldots, n \quad a_k^j = b_k^j.$$

3.10. Определение. *Сумма матриц* $A = (a_k^j)_n^m$ *и* $B = (b_k^j)_n^m$ одинакового размера $m \times n$ и *произведение матрицы* $A = (a_k^j)_n^m$ *на число* α определяются следующим образом:

$$\forall j = 1, \ldots, m,\ \forall k = 1, \ldots, n$$

$$\{A + B\}_k^j = \{A\}_k^j + \{B\}_k^j, \quad \{\alpha A\}_k^j = \alpha \{A\}_k^j.$$

Операции сложения матриц и умножения матрицы на число называются *линейными операциями*. Говорят, что линейные операции выполняются *поэлементно*; это означает, что каждый элемент результирующей матрицы зависит только от элементов исходных матриц, расположенных на том же месте.

3.11. Определение умножения матрицы на число согласовано с определением сложения матриц в том смысле, что для любого натурального n справедливы равенства

$$nA = \underbrace{A + \cdots + A}_{n \text{ слагаемых}}.$$

3.12. Предложение. *Линейные операции над матрицами обладают следующими свойствами:*

(1) *коммутативность сложения:*

$$\forall A, B \in \mathbb{K}^{m \times n} \quad A + B = B + A;$$

(2) *ассоциативность сложения:*

$$\forall A, B, C \in \mathbb{K}^{m \times n} \quad (A + B) + C = A + (B + C);$$

(3) *свойство нулевой матрицы:*

$$\forall A \in \mathbb{K}^{m \times n} \quad A + O = A,$$

где $O \in \mathbb{K}^{m \times n}$ *— нулевая матрица размера* $m \times n$;

(4) *существование противоположной матрицы:*

$$\forall A \in \mathbb{K}^{m \times n} \quad \exists A' \in \mathbb{K}^{m \times n} \quad A + A' = O;$$

очевидно, $A' = (-1) \cdot A$;

(5) *свойство единицы:*

$$\forall A \in \mathbb{K}^{m \times n} \quad 1 \cdot A = A;$$

(6) *ассоциативность умножения на число*:

$$\forall \alpha, \beta \in \mathbb{K},\ \forall A \in \mathbb{K}^{m\times n} \quad \alpha(\beta A) = (\alpha\beta)A;$$

(7) *дистрибутивность относительно сложения матриц*:

$$\forall \alpha \in \mathbb{K},\ \forall A, B \in \mathbb{K}^{m\times n} \quad \alpha(A + B) = \alpha A + \alpha B;$$

(8) *дистрибутивность относительно сложения чисел*:

$$\forall \alpha, \beta \in \mathbb{K},\ \forall A \in \mathbb{K}^{m\times n} \quad (\alpha + \beta)A = \alpha A + \beta A.$$

Доказательство этих утверждений очевидным образом вытекает из определения линейных операций над матрицами.

3.13. Если имеется несколько матриц $A, B, C, \ldots \in \mathbb{K}^{m\times n}$ и такое же количество числовых коэффициентов $\alpha, \beta, \gamma, \ldots \in \mathbb{K}$, то выражение $\alpha A + \beta B + \gamma C + \ldots$ называется *линейной комбинацией* матриц $A, B, C, \ldots$ с коэффициентами $\alpha, \beta, \gamma, \ldots$.

3.14. Определение 3.10 вводит сумму лишь двух матриц. Однако предыдущее определение линейной комбинации корректно благодаря ассоциативности операции сложения матриц (см. п. IV.10 и предложение IV.11, с. 336).

3.15. Линейные операции над блочными матрицами производятся по тем же правилам, что и в случае обычных числовых матриц, однако сложение блочных матриц возможно только в случае, когда блочная структура слагаемых одинакова, т.е. матрицы-слагаемые разбиты на одинаковое число блоков попарно совпадающих размеров.

2. Произведение матриц

A. Определение и примеры.

3.16. Определение. *Произведением* матриц $A = (a_l^j)_p^m \in \mathbb{K}^{m\times p}$ и $B = (b_k^l)_n^p \in \mathbb{K}^{p\times n}$ называется матрица $AB \in \mathbb{K}^{m\times n}$, элементы которой вычисляются по формуле

$$\{AB\}_k^j = \sum_{l=1}^{p} a_l^j\, b_k^l.$$

Таким образом, элемент $\{AB\}_k^j$, расположенный в j-й строке и k-м столбце матрицы AB, равен сумме попарных произведений элементов j-й строки матрицы A и элементов k-го столбца матрицы B. Ясно, что перемножить матрицы можно лишь в том случае, когда *количество столбцов первой из перемножаемых матриц равно количеству строк второй*. Схема на рис. 3.1 иллюстрирует правило умножения матриц.

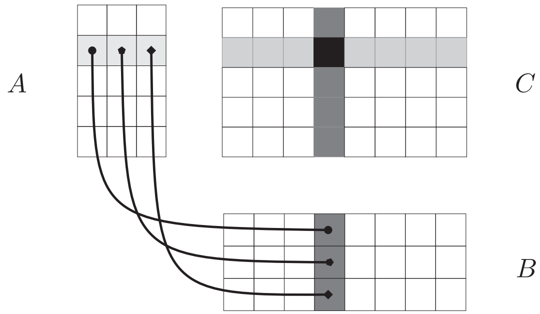


Рис. 3.1. Произведение матриц

3.17. Рассмотрим примеры:

$$\begin{pmatrix}1 & 2\\3 & 4\end{pmatrix}\cdot\begin{pmatrix}5\\6\end{pmatrix}=\begin{pmatrix}1\cdot 5+2\cdot 6\\3\cdot 5+4\cdot 6\end{pmatrix}=\begin{pmatrix}17\\39\end{pmatrix};$$

произведение

$$\begin{pmatrix}5\\6\end{pmatrix}\cdot\begin{pmatrix}1 & 2\\3 & 4\end{pmatrix}$$

не существует. Далее,

$$\begin{pmatrix}1 & 2 & 3\\4 & 5 & 6\end{pmatrix}\cdot\begin{pmatrix}1 & 2\\3 & 4\\5 & 6\end{pmatrix}=\begin{pmatrix}22 & 28\\49 & 64\end{pmatrix},$$

$$\begin{pmatrix}1 & 2\\3 & 4\\5 & 6\end{pmatrix}\cdot\begin{pmatrix}1 & 2 & 3\\4 & 5 & 6\end{pmatrix}=\begin{pmatrix}9 & 12 & 15\\19 & 26 & 33\\29 & 40 & 51\end{pmatrix};$$

этот пример показывает, что при перестановке местами матриц-сомножителей произведения получаются разными; более того, они имеют разные размеры. В примерах

$$\begin{pmatrix}1 & 2\\3 & 4\end{pmatrix}\cdot\begin{pmatrix}5 & 6\\7 & 8\end{pmatrix}=\begin{pmatrix}19 & 22\\43 & 50\end{pmatrix},\quad\begin{pmatrix}5 & 6\\7 & 8\end{pmatrix}\cdot\begin{pmatrix}1 & 2\\3 & 4\end{pmatrix}=\begin{pmatrix}23 & 34\\31 & 46\end{pmatrix}.$$

размеры матриц-произведений AB и BA равны, но сами произведения различны; таким образом, *умножение матриц не обладает свойством коммутативности.* Однако существуют матрицы, произведение которых не зависит от порядка сомножителей:

$$\begin{pmatrix}1 & 2\\0 & 1\end{pmatrix}\cdot\begin{pmatrix}2 & 3\\0 & 2\end{pmatrix}=\begin{pmatrix}2 & 7\\0 & 2\end{pmatrix};\quad\begin{pmatrix}2 & 3\\0 & 2\end{pmatrix}\cdot\begin{pmatrix}1 & 2\\0 & 1\end{pmatrix}=\begin{pmatrix}2 & 7\\0 & 2\end{pmatrix}.$$

Такие матрицы называются *коммутирующими.*

3.18. В отличие от произведения чисел, которое отлично от нуля при ненулевых сомножителях, произведение двух ненулевых матриц может оказаться нулевой матрицей:

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

3.19. Определение. *Единичной матрицей* $\mathbb{1}$ называется диагональная матрица, все диагональные элементы которой равны 1. Элементы единичной матрицы обозначаются

$$\delta_k^j = \begin{cases} 1, & \text{если } j = k, \\ 0, & \text{если } j \neq k, \end{cases} \tag{3.1}$$

и называются *символами Кронекера*[1].

3.20. Умножение блочных матриц также производится согласно определению 3.16: каждый блок матрицы-произведения равен сумме попарных произведений блоков матриц-сомножителей, однако требуется, чтобы все «горизонтальные размеры» блоков в первом сомножителе совпадали с соответствующими «вертикальными размерами» во втором. Например,

$$\left\| \begin{matrix} \underset{m_1\times n_1}{A_{11}} & \underset{m_1\times n_2}{A_{12}} \\ \\ \underset{m_2\times n_1}{A_{21}} & \underset{m_2\times n_2}{A_{22}} \end{matrix} \right\| \cdot \left\| \begin{matrix} \underset{n_1\times p}{B_1} \\ \\ \underset{n_2\times p}{B_2} \end{matrix} \right\| = \left\| \begin{matrix} \underbrace{A_{11}B_1 + A_{12}B_2}_{m_1\times p} \\ \underbrace{A_{21}B_1 + A_{22}B_2}_{m_2\times p} \end{matrix} \right\|.$$

3.21. Предложение. *Операция умножения матриц обладает следующими свойствами:*

(1) *ассоциативность:*

$$\forall A \in \mathbb{K}^{m\times s},\ \forall B \in \mathbb{K}^{s\times p},\ \forall C \in \mathbb{K}^{p\times n} \quad A(BC) = (AB)C;$$

(2) *дистрибутивность слева:*

$$\forall A \in \mathbb{K}^{m\times s},\ \forall B, C \in \mathbb{K}^{s\times n} \quad A(B + C) = AB + AC;$$

(3) *дистрибутивность справа:*

$$\forall A, B \in \mathbb{K}^{m\times s},\ \forall C \in \mathbb{K}^{s\times n} \quad (A + B)C = AC + BC;$$

(4) *свойство единичной матрицы:*

$$\forall A \in \mathbb{K}^{m\times n} \quad \mathbb{1}_m A = A\mathbb{1}_n = A,$$

где $\mathbb{1}_m$ и $\mathbb{1}_n$ — единичные матрицы порядков m и n соответственно;

(5) *свойство следа: если $A \in \mathbb{K}^{m\times n}$ и $B \in \mathbb{K}^{n\times m}$, то*

$$\operatorname{tr}(AB) = \operatorname{tr}(BA).$$

[1] Л. Кро́некер (L. Kronecker) — немецкий математик (1823–1891).

Доказательство. Докажем лишь некоторые из перечисленных свойств. Проверим свойство ассоциативности (1). Пусть

$$A = (a^i_j)^m_s, \quad B = (b^j_k)^s_p, \quad C = (c^k_l)^p_n.$$

Убедимся в совпадении размеров матриц: AB имеет размер $m \times p$, $(AB)C$ — размер $m \times n$, BC — размер $s \times n$, $A(BC)$ — размер $m \times n$. Таким образом, размеры матриц $A(BC)$ и $(AB)C$ одинаковы.

Вычислим соответствующие элементы матриц:

$$
\begin{aligned}
\{A(BC)\}^i_l &= \sum_{j=1}^{s} a^i_j \, \{BC\}^j_l && \text{(определение произведения } A(BC)) \\
&= \sum_{j=1}^{s} a^i_j \left(\sum_{k=1}^{p} b^j_k \, c^k_l \right) && \text{(определение произведения } BC) \\
&= \sum_{j=1}^{s} \left(\sum_{k=1}^{p} a^i_j \, b^j_k \, c^k_l \right) && \text{(внесение множителя в скобку)} \\
&= \sum_{k=1}^{p} \left(\sum_{j=1}^{s} a^i_j \, b^j_k \, c^k_l \right) && \text{(перемена порядка суммирования)} \\
&= \sum_{k=1}^{p} \left(\sum_{j=1}^{s} a^i_j \, b^j_k \right) c^k_l && \text{(вынесение общего множителя)} \\
&= \sum_{k=1}^{p} \left(\sum_{j=1}^{s} \{AB\}^i_k \right) c^k_l = \{(AB)C\}^i_l.
\end{aligned}
$$

Докажем утверждение (5). Оба произведения AB и BA являются квадратными матрицами, а потому можно говорить о следах AB и BA: $AB \in \mathbb{K}^{m\times m}$, $BA \in \mathbb{K}^{n\times n}$. Имеем:

$$
\begin{aligned}
\{AB\}^i_j &= \sum_{k=1}^{n} a^i_k b^k_j, \quad \operatorname{tr}(AB) = \sum_{i=1}^{m} \{AB\}^i_i = \sum_{i=1}^{m} \sum_{k=1}^{n} a^i_k b^k_i, \\
\{BA\}^i_j &= \sum_{l=1}^{m} b^i_l a^l_j, \quad \operatorname{tr}(BA) = \sum_{i=1}^{n} \{BA\}^i_i = \sum_{i=1}^{n} \sum_{l=1}^{m} b^i_l a^l_i. \qquad \square
\end{aligned}
$$

3.22. Из установленных выше свойств операций над матрицами вытекает, что множество всех квадратных матриц $\mathbb{K}^{n\times n}$ является некоммутативным ассоциативным кольцом с единицей, обладающим делителями нуля (см. п. 3.18 и п. IV.2, с. 338).

3.23. Если A — квадратная матрица, то её натуральная *степень* A^p, $p = 0, 1, \ldots$, определяется следующим образом:

$$A^0 = \mathbb{1}, \quad A^p = \underbrace{A \cdot \ldots \cdot A}_{p \text{ раз}}.$$

Обозначение p-й степени матрицы совпадает с обозначением её p-й строки; на практике путаницы обычно не возникает, ибо из контекста всегда ясно, о чём идёт речь.

3.24. Пусть $P(x) = a_0 x^p + a_1 x^{p-1} + \cdots + a_{p-1} x + a_p$ — многочлен степени p от переменной x с коэффициентами из $\mathbb{K}$. Тогда под *многочленом от матрицы* понимается выражение

$$P(A) = a_0 A^p + a_1 A^{p-1} + \cdots + a_{p-1} A + a_p \mathbb{1}.$$

Б. Структура произведения матриц.

3.25. Проанализируем строение столбцов матрицы $C = AB$, где

$$A = (a_l^j)_p^m \in \mathbb{K}^{m \times p}, \quad B = (b_k^l)_n^p \in \mathbb{K}^{p \times n},$$

Представим матрицу A в виде совокупности столбцов

$$A = \left\| A_1,\ A_2,\ \ldots,\ A_p \right\|;$$

здесь

$$A_1 = \begin{pmatrix} a_1^1 \\ a_1^2 \\ \vdots \\ a_1^m \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_2^1 \\ a_2^2 \\ \vdots \\ a_2^m \end{pmatrix}, \quad \ldots, \quad A_p = \begin{pmatrix} a_p^1 \\ a_p^2 \\ \vdots \\ a_p^m \end{pmatrix}.$$

Элементы матрицы $\underset{m\times n}{C} = \underset{m\times p}{A}\ \underset{p\times n}{B}$ вычисляются по формуле

$$c_k^j = \sum_{l=1}^{p} a_l^j\, b_k^l, \qquad \begin{array}{l} j = 1, \ldots, m, \\ k = 1, \ldots, n. \end{array}$$

Представим матрицу C в виде совокупности столбцов:

$$C = \left\| C_1,\ C_2,\ \ldots,\ C_n \right\|$$

и проанализируем структуру k-го столбца C_k матрицы C:

$$C_k = \begin{pmatrix} c_k^1 \\ \vdots \\ c_k^m \end{pmatrix} = \begin{pmatrix} \sum\limits_{l=1}^{p} a_l^1\, b_k^l \\ \vdots \\ \sum\limits_{l=1}^{p} a_l^m\, b_k^l \end{pmatrix} = \sum_{l=1}^{p} \begin{pmatrix} a_l^1 \\ \vdots \\ a_l^m \end{pmatrix} b_k^l = \sum_{l=1}^{p} A_l\, b_k^l = A \cdot B_k.$$

Таким образом, получаем следующее предложение (утверждения (3) и (4), касающиеся строк матрицы $C = AB$, докажите самостоятельно).

3.26. Предложение. *Рассмотрим матрицы* $A = (a_l^j)_p^m \in \mathbb{K}^{m\times p}$ *и* $B = (b_k^l)_n^p \in \mathbb{K}^{p\times n}$.

(1) *k-й столбец матрицы $C = AB$ равен линейной комбинации столбцов матрицы A с коэффициентами, равными элементам k-го столбца матрицы B:*

$$C_k = \sum_{l=1}^{p} A_l \ b_k^l.$$

(2) *k-й столбец матрицы $C = AB$ равен произведению матрицы A на k-й столбец матрицы B:*

$$C_k = A \cdot B_k.$$

(3) *s-я строка матрицы AB равна линейной комбинации строк матрицы B с коэффициентами, равными элементам s-й строки матрицы A:*

$$C^s = \sum_{k=1}^{p} a_k^s \ B^k.$$

(4) *s-я строка матрицы AB равна произведению s-й строки матрицы A на матрицу B.*

$$C^s = A^s \cdot B.$$

3. Обратная матрица

A. Определение и свойства.

3.27. Определение. Квадратная матрица A^{-1} называется *обратной* для квадратной матрицы A, если выполняются соотношения

$$AA^{-1} = A^{-1}A = \mathbb{1}.$$

Матрица A называется *обратимой*, если она имеет обратную.

3.28. Обратите внимание, что понятие обратной матрицы вводится только для квадратных матриц.

3.29. Не все квадратные матрицы обратимы; очевидно, для нулевой матрицы обратная не существует. Вопрос о существовании обратной матрицы решён в гл. 6 (см. теорему 6.57, с. 143), а алгоритм вычисления разработан в гл. 4 (см. п. 4.48, с. 85).

3.30. Предложение. *Обратимые матрицы обладают следующими свойствами:*

(1) *Единичная матрица обратима и* $\mathbb{1}^{-1} = \mathbb{1}$.

(2) *Если матрица* A *обратима, то обратная матрица* A^{-1} *единственна.*

(3) *Если матрицы* A *и* B *обратимы, то их произведение также обратимо и* $(AB)^{-1} = B^{-1}A^{-1}$.

(4) *Если матрица* A *обратима, то её обратная матрица* A^{-1} *также обратима и* $(A^{-1})^{-1} = A$.

Доказательство. Утверждение (1) является очевидным следствием определения 3.27.

(2) Пусть B и C — две матрицы, обратных к A, т.е.

$$AB = BA = \mathbb{1}, \quad AC = CA = \mathbb{1}.$$

В силу ассоциативности операции умножения матриц имеем

$$B = B\mathbb{1} = B(AC) = (BA)C = \mathbb{1}C = C.$$

(3) Снова в силу ассоциативности имеем

$$(AB)\cdot(B^{-1}A^{-1}) = \big((AB)B^{-1}\big)A^{-1} = \\ = \big(A(BB^{-1})\big)A^{-1} = (A\mathbb{1})A^{-1} = AA^{-1} = \mathbb{1}.$$

Аналогично доказывается, что $(B^{-1}A^{-1})\cdot(AB) = \mathbb{1}$.

(4) Пользуясь предыдущим фактом, получаем

$$(A^{-1})^{-1}\cdot A^{-1} = (AA^{-1})^{-1} = \mathbb{1}^{-1} = \mathbb{1}. \qquad \square$$

Б. Обращение (2×2)**-матриц.** В случае матриц порядка 2 имеется простая формула для нахождения обратной матрицы.

3.31. Определение. *Определителем* (2×2)-матрицы

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{3.2}$$

называется число $ad - bc$; определитель обозначается $\det A$ или $|A|$ (см. гл. 6). Матрица A называется *невырожденной*, если $\det A \neq 0$, и *вырожденной* в противном случае.

3.32. Предложение. *Матрица* (3.2) *обратима тогда и только тогда, когда она невырождена, т.е.* $\det A \neq 0$. *Обратная матрица вычисляется по формуле*

$$A^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{3.3}$$

Доказательство. В случае $ad - bc \neq 0$ существование обратной матрицы для A доказывается непосредственной проверкой:

$$\frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Единственность была доказана в предложении 3.30.

Если же $ad - bc = 0$, то столбцы матрицы A линейно зависимы:

$$A = \begin{pmatrix} a & \lambda a \\ c & \lambda c \end{pmatrix};$$

поэтому для любой матрицы B в силу предложения 3.26(1) столбцы произведения AB также будут пропорциональны, т.е. равенство $AB = \mathbb{1}$ невозможно ни для какой матрицы B. $\square$

3.33. Формула (3.3) может быть получена следующим образом. Будем искать обратную матрицу в виде

$$\begin{pmatrix} x & y \\ u & v \end{pmatrix}.$$

Умножим эту матрицу на A и приравняем результат $\mathbb{1}$:

$$\begin{pmatrix} x & y \\ u & v \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ax + cy & bx + dy \\ au + cv & bu + dv \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

В результате получим две системы уравнений:

$$\begin{cases} ax + cy = 1, \\ bx + dy = 0, \end{cases} \qquad \begin{cases} au + cv = 0, \\ bu + dv = 1. \end{cases}$$

Решим первую из этих систем. Умножим первое уравнение на d, второе — на $(-c)$ и сложим полученные уравнения:

$$+\begin{cases} adx + cdy = d, \\ -bcx - dcy = 0 \end{cases} \quad \Rightarrow \quad (ad - bc)x = d.$$

Далее, умножая первое уравнение на $(-b)$, второе — на a и складывая полученные уравнения, получим

$$+\begin{cases} -abx - bcy = -b, \\ abx + ady = 0 \end{cases} \quad \Rightarrow \quad (ad - bc)y = -b.$$

Если $ad - bc \neq 0$, то

$$x = \frac{d}{ad - bc}, \quad y = \frac{-b}{ad - bc}.$$

Решая аналогичным образом вторую систему, получим

$$\begin{cases} au + cv = 0, \\ bu + dv = 1 \end{cases} \quad \Rightarrow \quad u = \frac{-c}{ad - bc}, \quad v = \frac{a}{ad - bc}.$$

4. Транспонирование

А. Определение и свойства.

3.34. Матрицей, *транспонированной* по отношению к матрице $A = (a_k^j)_n^m \in \mathbb{K}^{m\times n}$, называется матрица $A^T \in \mathbb{K}^{n\times m}$ с элементами

$$\{A^T\}_j^k \overset{\text{def}}{=\!=} \{A\}_k^j.$$

Очевидно, количество строк (столбцов) транспонированной матрицы равно количеству столбцов (строк) исходной матрицы.

Например,

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

3.35. Предложение. *Операция транспонирования обладает следующими свойствами:*

(1) *линейность:*

$$\forall \alpha, \beta \in \mathbb{K},\ \forall A, B \in \mathbb{K}^{m\times n} \quad (\alpha A + \beta B)^T = \alpha A^T + \beta B^T;$$

(2) *инволютивность:*

$$\forall A \in \mathbb{K}^{m\times n} \quad (A^T)^T = A;$$

(3) *правило транспонирования произведения:*

$$\forall A \in \mathbb{K}^{m\times p},\ \forall B \in \mathbb{K}^{p\times n} \quad (AB)^T = B^T A^T;$$

(4) *транспонирование обратной матрицы: если матрица A обратима, то*

$$(A^T)^{-1} = (A^{-1})^T.$$

Доказательство. Докажем правило транспонирования произведения. Сначала проверим совпадение размеров: AB имеет размер $m \times n$, $(AB)^T$ — размер $n \times m$, A^T — размер $p \times m$, B^T — размер $n \times p$, $B^T A^T$ — размер $n \times m$, как и $(AB)^T$ (обратите внимание, что произведение $A^T B^T$ не существует, если $m \neq n$). Проверим равенство элементов:

$$\begin{aligned}
\{(AB)^T\}_i^j &= \{AB\}_j^i && \text{(определение транспонирования)}\\
&= \sum_{k=1}^{p} a_k^i b_j^k && \text{(определение произведения матриц)}\\
&= \sum_{k=1}^{p} \{A^T\}_i^k\ \{B^T\}_k^j && \text{(определение транспонирования)}
\end{aligned}$$

$$= \sum_{k=1}^{p} \{B^T\}_k^j \; \{A^T\}_i^k \quad \text{(перестановка числовых сомножителей)}$$

$$= \{B^T A^T\}_i^j \qquad \text{(определение произведения матриц).}$$

Докажем, что матрица $(A^{-1})^T$ является обратной для A^T:

$$(A^{-1})^T A^T = (AA^{-1})^T = \mathbb{1}^T = \mathbb{1}. \qquad \square$$

Использование операции транспонирования позволяет экономить место при записи столбцов, например, четырёхэлементный столбец с элементами $1, 2, 3, 4$ можно записать в виде $(1, 2, 3, 4)^T$.

Б. Симметричные и кососимметричные матрицы.

3.36. Квадратная матрица A называется *симметричной*, если $A^T = A$, и *кососимметричной* (антисимметричной), если $A^T = -A$. Очевидно, все элементы главной диагонали кососимметричной матрицы равны нулю. Примеры:

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ 3 & 7 & 4 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 & -2 \\ -1 & 0 & 3 \\ 2 & -3 & 0 \end{pmatrix}.$$

Множества всех симметричных и кососимметричных матриц порядка n обозначаются соответственно $S\mathbb{K}^{n\times n}$ и $A\mathbb{K}^{n\times n}$.

3.37. Легко видеть, что любая линейная комбинация симметричных (кососимметричных) матриц также является симметричной (кососимметричной) матрицей. Для произведения матриц аналогичное свойство не имеет места:

$$\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 4 & 5 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 14 & 17 \\ 23 & 28 \end{pmatrix}, \; \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}.$$

3.38. Предложение. *Любая квадратная матрица A может быть единственным образом представлена в виде суммы симметричной и кососимметричной матриц.*

Доказательство. Представим матрицу A в виде

$$A = \underbrace{\frac{1}{2}(A + A^T)}_{=A_1} + \underbrace{\frac{1}{2}(A - A^T)}_{=A_2}.$$

Легко проверяется, что матрица A_1 симметрична, а A_2 кососимметрична:

$$A_1^T = \left(\frac{1}{2}(A + A^T)\right)^T = \frac{1}{2}\big(A^T + (A^T)^T\big) = \frac{1}{2}\big(A^T + A\big) = A_1,$$

$$A_2^T = \left(\frac{1}{2}(A - A^T)\right)^T = \frac{1}{2}\bigl(A^T - (A^T)^T\bigr) = \frac{1}{2}\bigl(A^T - A\bigr) = -A_2;$$

таким образом, искомое представление найдено. Докажем его единственность. Предположим, что разложение матрицы A в сумму симметричной и кососимметричной матриц не единственно, т.е. существуют два различных разложения $A = A_1 + A_2 = A_1' + A_2'$, где матрицы A_1, A_1' симметричны, а матрицы A_2, A_2' кососимметричны. Тогда

$$\underbrace{A_1 - A_1'}_{\text{симм.}} = \underbrace{A_2' - A_2}_{\text{кососимм.}}.$$

Однако единственная матрица, являющаяся одновременно симметричной и кососимметричной — это нулевая: если $C = C^T = -C^T$, то $2C^T = O$, $C^T = O$ и $C = O$. Поэтому $A_1 - A_1' = A_2' - A_2 = O$, т.е. $A_1 = A_1'$ и $A_2 = A_2'$. $\square$

5. Матричная модель поля комплексных чисел

3.39. Рассмотрим (2×2)-матрицу

$$\mathbb{J} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Легко проверить, что $\mathbb{J}^2 = -\mathbb{1}$; это соотношение похоже на свойство мнимой единицы $i^2 = -1$. Убедимся, что множество вещественных (2×2)-матриц вида

$$Z = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = a\mathbb{1} + b\mathbb{J} \tag{3.4}$$

является моделью поля комплексных чисел, т.е. каждому комплексному числу $z = a + ib$ можно взаимно однозначно поставить в соответствие матрицу $Z = a\mathbb{1} + b\mathbb{J}$, причём арифметическим операциям над комплексными числами будут соответствовать операции над матрицами. Факт соответствия комплексного числа $z = a + ib$ и матрицы $Z = a\mathbb{1} + b\mathbb{J}$ будем записывать в виде $z \leftrightarrow Z$.

3.40. Для матриц $a\mathbb{1} + b\mathbb{J}$ и $c\mathbb{1} + d\mathbb{J}$ имеем

$$(a\mathbb{1} + b\mathbb{J}) \pm (c\mathbb{1} + d\mathbb{J}) = (a \pm c)\mathbb{1} + (b \pm d)\mathbb{J},$$
$$(a\mathbb{1} + b\mathbb{J}) \cdot (c\mathbb{1} + d\mathbb{J}) = (ac - bd)\mathbb{1} + (ad + bc)\mathbb{J},$$

что по форме совпадает с правилами (2.5) и (2.6) (см. с. 26) сложения и умножения комплексных чисел. Таким образом, если $z_1 \leftrightarrow Z_1$ и $z_2 \leftrightarrow Z_2$, то $z_1 \pm z_2 \leftrightarrow Z_1 \pm Z_2$ и $z_1 z_2 \leftrightarrow Z_1 Z_2$. Кроме того, поскольку единичная матрица коммутирует с любой матрицей того же порядка, ясно, что умножение матриц вида (3.4) коммутативно.

3.41. Поскольку матрица $\mathbb{J}$ кососимметрична, т.е. $\mathbb{J}^T = -\mathbb{J}$, транспонирование матрицы (3.4) приводит к сопряжённому комплексному числу:

$$(a\mathbb{1} + b\mathbb{J})^T = a\mathbb{1}^T + b\mathbb{J}^T = a\mathbb{1} - b\mathbb{J}.$$

Таким образом, если $z \leftrightarrow Z$, то $\bar{z} \leftrightarrow Z^T$.

3.42. Модулю комплексного числа $z = a + ib$ соответствует определитель матрицы $Z = a\mathbb{1} + b\mathbb{J}$; этот факт мы запишем в виде $|z| \leftrightarrow \det Z$.

3.43. Матрица, обратная к (3.4), равна

$$Z^{-1} = \frac{1}{a^2 + b^2} \begin{pmatrix} a & b \\ -b & a \end{pmatrix} = \frac{a}{a^2 + b^2}\mathbb{1} + \frac{-b}{a^2 + b^2}\mathbb{J},$$

что также совпадает по форме с (2.7), т.е. если $z \leftrightarrow Z$, то $z^{-1} \leftrightarrow Z^{-1}$.

3.44. Тригонометрическая форма записи комплексного числа приводит к следующему матричному представлению:

$$z = r(\cos\varphi + i\sin\varphi) \;\leftrightarrow\; Z = r\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}.$$

В гл. 10 выяснится геометрический смысл полученной матрицы.

3.45. Таким образом, мы построили реализацию комплексных чисел в виде вещественных (2×2)-матриц. Построенное взаимно однозначное соответствие между полем $\mathbb{C}$ и множеством матриц вида (3.4) является изоморфизмом (см. определение IV.60, с. 346).

ГЛАВА 4

Системы линейных уравнений

Системы линейных уравнений встречаются в математике и её приложениях весьма часто: они не только предоставляют аппарат теоретического анализа многих важных задач, но и являются мощным средством в вычислительной математике.

В этой главе мы изучим свойства линейных систем и их решений и разработаем алгоритм нахождения решений для произвольного числового поля $\mathbb{K}$.

1. Системы линейных уравнений

A. Основные определения.

4.1. Рассмотрим систему, состоящую из m линейных уравнений с n неизвестными:

$$\begin{cases} a_1^1x^1 + a_2^1x^2 + \dots + a_n^1x^n = b^1, \\ a_1^2x^1 + a_2^2x^2 + \dots + a_n^2x^n = b^2, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_1^mx^1 + a_2^mx^2 + \dots + a_n^mx^n = b^m. \end{cases}$$

Введём матрицы A и $\widetilde{A}$ — *основную* и *расширенную* матрицы системы, а также столбец правых частей B и столбец неизвестных X:

$$A = \begin{pmatrix} a_1^1 & a_2^1 & \dots & a_n^1 \\ a_1^2 & a_2^2 & \dots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^m & a_2^m & \dots & a_n^m \end{pmatrix} = \left\| A_1, A_2, \dots, A_n \right\|, \quad B = \begin{pmatrix} b^1 \\ b^2 \\ \vdots \\ b^m \end{pmatrix},$$

$$\widetilde{A} = \left(\begin{array}{cccc|c} a_1^1 & a_2^1 & \dots & a_n^1 & b^1 \\ a_1^2 & a_2^2 & \dots & a_n^2 & b^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1^m & a_2^m & \dots & a_n^m & b^m \end{array}\right) = \left\| A \mid B \right\|, \qquad X = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix};$$

тогда систему можно записать в матричном виде

$$AX = B,$$

а также в виде

$$A_1x^1 + A_2x^2 + \cdots + A_nx^n = B.$$

Будем называть линейную систему *квадратной*, если в ней количество уравнение равно количеству неизвестных, или, что то же самое, если основная матрицы системы является квадратной.

4.2. Упорядоченный набор $(x_1, \ldots, x_n)$ элементов произвольного множества будем называть *семейством* элементов.

4.3. Строгое определение семейства (упорядоченного набора) сформулировать не очень просто; см. по этому поводу приложение II (пп. II.22–II.24, с. 305, и определение II.63, с. 314).

4.4. Упорядоченный набор чисел $(x_0^1, x_0^2, \ldots, x_0^n)$ (или, в матричной записи, столбец $X_0 = (x_0^1, x_0^2, \ldots, x_0^n)^T$) называется *решением* системы, если при подстановке этих чисел в каждое из уравнений системы получается верное числовое равенство (при подстановке столбца в матричное уравнение получается верное матричное равенство).

Система называется *совместной*, если она имеет хотя бы одно решение, и *несовместной*, если она не имеет решений.

Две совместные системы называются *эквивалентными*, если множества их решений совпадают. Ясно, что эквивалентные системы содержат одинаковое число неизвестных; число же уравнений в них может быть различным.

Б. Однородные системы.

4.5. Система линейных уравнений называется *однородной*, если правые части всех уравнений равны нулю:

$$\begin{cases} a_1^1x^1 + a_2^1x^2 + \cdots + a_n^1x^n = 0, \\ a_1^2x^1 + a_2^2x^2 + \cdots + a_n^2x^n = 0, \\ \dotfill \\ a_1^mx^1 + a_2^mx^2 + \cdots + a_n^mx^n = 0; \end{cases}$$

в матричной записи

$$AX = O \quad \text{или} \quad A_1x^1 + A_2x^2 + \cdots + A_nx^n = O.$$

4.6. Однородная система всегда совместна: набор чисел $x^1 = 0$, $x^2 = 0$, …, $x^n = 0$ (или, что то же же, нулевой столбец) образуют её решение, называемое *тривиальным*. Однако однородная система может иметь и нетривиальные (т.е. ненулевые) решения. Например, одним из решений однородной системы $x^1 + x^2 = 0$ с двумя неизвестными, состоящей из одного уравнения, является пара чисел $x^1 = c$, $x^2 = -c$, где c — произвольное число.

4.7. Теорема (принцип линейной суперпозиции для однородных систем). *Если X_1, X_2 — два решения однородной системы $AX = 0$, то любая их линейная комбинация также является решением этой системы.*

Доказательство. Пусть столбцы X_1, X_2 являются решениями системы, т.е.

$$AX_1 = O, \quad AX_2 = O,$$

и пусть c^1, c^2 — произвольные числа[1]. Для линейной комбинации $c^1X_1 + c^2X_2$ имеем

$$A(c^1X_1 + c^2X_2) = c^1AX_1 + c^2AX_2 = O,$$

т.е. линейная комбинация $c^1X_1 + c^2X_2$ также является решением данной системы. □

4.8. Это означает, множество всех решений однородной системы образует подпространство в пространстве столбцов $\mathbb{K}^n$ (см. определение 5.29, с. 100).

4.9. Определение. Семейство столбцов $A_1, A_2, \ldots, A_p$ называется *линейно независимым*, если линейная комбинация этих столбцов

$$\alpha^1A_1 + \alpha^2A_2 + \cdots + \alpha^pA_p$$

обращается в нулевой столбец лишь при нулевых коэффициентах.

4.10. Детальное изучение линейной независимости столбцов будет проведено в гл. 5 (см. п. 4, с. 102). Сейчас лишь отметим, что ни один из столбцов линейно независимого семейства нельзя выразить как линейную комбинацию остальных. Действительно, предположим, что, например, $A_p = \beta^1A_1 + \cdots + \beta^{p-1}A_{p-1}$. Тогда линейная комбинация $\beta^1A_1 + \cdots + \beta^{p-1}A_{p-1} - A_p$, в которой по крайней мере один коэффициент (-1 при A_p) ненулевой, равна нулевому столбцу — противоречие.

4.11. Определение. *Фундаментальное семейство решений* (сокращённо ФСР) однородной системы — это такое семейство линейно независимых решений $X_1, X_2, \ldots, X_s$, что любое решение данной системы можно представить в виде их линейной комбинации:

$$X = c^1X_1 + c^2X_2 + \cdots + c^sX_s, \tag{4.1}$$

[1] Здесь верхние индексы являются номерами, а не показателями степени. Преимущества нумерации коэффициентов линейных комбинаций верхними индексами выяснятся впоследствии (см. с. 112).

где $c^1, c^2, \dots, c^s$ — произвольные числа. Выражение (4.1), позволяющее с помощью надлежащего выбора чисел $c^1, c^2, \dots, c^s$ получить любое решение однородной системы, называется *общим решением* данной системы.

4.12. Согласно определению 5.52 (см. с. 110) фундаментальное семейство решений однородной системы линейных уравнений является *базисом* в подпространстве пространства $\mathbb{K}^n$, состоящем из всех решений этой системы.

4.13. Матрица Φ, столбцами которой являются столбцы ФСР, называется *фундаментальной матрицей* данной системы:

$$\Phi = \left\| X_1,\ X_2,\ \dots,\ X_s \right\|.$$

Общее решение системы выражается через фундаментальную матрицу по формуле

$$X = \Phi \cdot \left(c^1, c^2, \dots, c^s \right)^T, \tag{4.2}$$

где $c^1, c^2, \dots, c^s$ — произвольные постоянные.

4.14. Формула (4.2) определяет отображение $\hat{\Phi} : \mathbb{K}^s \to \mathbb{K}^n$, которое каждому элементу пространства $\mathbb{K}^s$ ставит в соответствие некоторое решение однородной системы $AX = O$:

$$\hat{\Phi} : \mathbb{K}^s \to \mathbb{K}^n, \quad C \mapsto X = \Phi C.$$

B. Неоднородные системы.

4.15. Система $AX = B$ называется *неоднородной*, если $B \neq O$. Часто неоднородную систему $AX = B$ рассматривают вместе с однородной системой $AX = O$, которую называют *сопутствующей*.

4.16. Теорема. *Если X_1, X_2 — два решения неоднородной системы $AX = B$, то их разность $X_1 - X_2$ является решением сопутствующей однородной системы $AX = O$.*

Доказательство. Пусть $AX_1 = B$, $AX_2 = B$. Тогда

$$A(X_1 - X_2) = AX_1 - AX_2 = B - B = O,$$

что и требовалось. □

4.17. Таким образом, любое решение неоднородной системы (т.е. выражение для общего решения, ОРНС) можно представить в виде суммы некоторого частного решения этой неоднородной системы (ЧРНС) и подходящего решения однородной системы (т.е. выражения для её общего решения, ОРОС):

$$\text{ОРНС} = \text{ЧРНС} + \text{ОРОС}.$$

Это утверждение называется *принципом линейной суперпозиции.*

4.18. Итак, общее решение неоднородной системы имеет вид

$$X = X_0 + c^1 X_1 + \cdots + c^s X_s, \tag{4.3}$$

где X_0 — какое-либо частное решение этой системы, а $X_1, \ldots, X_s$ — ФСР сопутствующей однородной системы.

4.19. Это означает, множество всех решений неоднородной системы образует s-мерную плоскость в аффинном пространстве столбцов $\mathbb{K}^n_{\text{афф}}$ (см. п. 7.37, с. 166).

Г. Системы упрощённого вида.

4.20. Неизвестная x^k называется *базисной*, если она входит только в одно уравнение системы.

4.21. Система называется *системой упрощённого вида*, если в каждом уравнении имеется базисная неизвестная; в случае, если некоторое уравнение содержит более одной базисной неизвестной, выбирается одна из них[1]. Итак, *в системе упрощённого вида количество базисных неизвестных равно количеству уравнений в системе.*

4.22. Неизвестные, не являющиеся базисными, называются *свободными*. Происхождение терминов «базисная неизвестная» и «свободная неизвестная» станет ясным из дальнейшего изложения.

4.23. Рассмотрим однородную систему упрощённого вида, в которой для удобства рассуждений будем считать, что базисными неизвестными являются $x^1, \ldots, x^r$, а коэффициенты при этих неизвестных равны 1 (этого легко добиться умножением уравнения на подходящий коэффициент):

$$\begin{cases} \boxed{x^1} \qquad\qquad + a^1_{r+1} x^{r+1} + \cdots + a^1_n x^n = 0, \\ \qquad \boxed{x^2} \qquad + a^2_{r+1} x^{r+1} + \cdots + a^2_n x^n = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \qquad\qquad \boxed{x^r} + a^r_{r+1} x^{r+1} + \cdots + a^r_n x^n = 0. \end{cases}$$

[1] Например, в системе

$$\begin{cases} x^1 + x^2 \qquad + x^4 = 0 \\ \qquad\qquad x^3 + x^4 = 0 \end{cases}$$

неизвестные x^1 и x^2 входят только в первое уравнение, а x^3 — только во второе; в этом случае лишь одну из неизвестных x^1 или x^2 будем относить к числу базисных. Таким образом, в этой системе базисными неизвестными можно считать либо x^1 и x^3, либо x^2 и x^3.

Перенеся все слагаемые, содержащие свободные неизвестные, в правые части уравнений, перепишем систему в виде

$$\begin{cases} x^1 &= -a^1_{r+1}x^{r+1} - \dots - a^1_n x^n, \\ x^2 &= -a^2_{r+1}x^{r+1} - \dots - a^2_n x^n, \\ \dots & \dots\dots\dots\dots\dots\dots\dots\dots \\ x^r &= -a^r_{r+1}x^{r+1} - \dots - a^r_n x^n. \end{cases}$$

Если теперь вместо свободных неизвестных $x^{r+1}, \dots, x^n$ подставлять произвольные числа и вычислять значения базисных неизвестных $x^1, \dots, x^r$ по указанным формулам, то полученный набор чисел $x^1, \dots, x^n$ будет представлять собой некоторое решение данной однородной системы. Таким образом, полученные формулы представляют общее решение исходной системы.

4.24. Обозначив произвольные значения свободных неизвестных $x^{r+1}, \dots, x^n$ через $c^1, \dots, c^{n-r}$, представим общее решение системы в виде

$$X = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^r \\ x^{r+1} \\ x^{r+2} \\ \vdots \\ x^n \end{pmatrix} = \begin{pmatrix} -a^1_{r+1}c^1 - \dots - a^1_n c^{n-r} \\ -a^2_{r+1}c^1 - \dots - a^2_n c^{n-r} \\ \vdots \\ -a^r_{r+1}c^1 - \dots - a^r_n c^{n-r} \\ c^1 \\ c^2 \\ \vdots \\ c^{n-r} \end{pmatrix} =$$

$$= c^1 \begin{pmatrix} -a^1_{r+1} \\ -a^2_{r+1} \\ \vdots \\ -a^r_{r+1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \dots + c^{n-r} \begin{pmatrix} -a^1_n \\ -a^2_n \\ \vdots \\ -a^r_n \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Столбцы, фигурирующие в последнем выражении, образуют фундаментальное семейство решений данной однородной системы, которое называется *нормальным ФСР (НФСР), соответствующим базисным неизвестным* $x^1, \dots, x^r$ *и свободным неизвестным* $x^{r+1}, \dots, x^n$. Каждый столбец, входящий в НФСР, получается, если значение одной из свободных неизвестных взять равным 1, а остальные положить равными нулю (соответствующие элементы столбцов выделены). Линейная независимость этих столбцов очевидна.

4.25. Матрица, составленная из столбцов НФСР, называется нормальной фундаментальной матрицей рассматриваемой однородной системы:

$$\Phi = \begin{pmatrix} -a^1_{r+1} & -a^1_{r+2} & \dots & -a^1_n \\ -a^2_{r+1} & -a^2_{r+2} & \dots & -a^2_n \\ \vdots & \vdots & \ddots & \vdots \\ -a^r_{r+1} & -a^r_{r+2} & \dots & -a^r_n \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix};$$

выделенные элементы этой матрицы образуют единичную матрицу порядка $n - r$.

4.26. Ситуация в случае неоднородной системы аналогична. Перепишем упрощённую неоднородную систему

$$\begin{cases} x^1 \qquad\qquad + a^1_{r+1}x^{r+1} + \dots + a^1_n x^n = b^1, \\ \quad x^2 \qquad\quad + a^2_{r+1}x^{r+1} + \dots + a^2_n x^n = b^2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \qquad\quad x^r + a^r_{r+1}x^{r+1} + \dots + a^r_n x^n = b^r \end{cases}$$

в виде

$$\begin{cases} x^1 = b^1 - a^1_{r+1}x^{r+1} - \dots - a^1_n x^n, \\ x^2 = b^2 - a^2_{r+1}x^{r+1} - \dots - a^2_n x^n, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ x^r = b^r - a^r_{r+1}x^{r+1} - \dots - a^r_n x^n. \end{cases}$$

Положив значения всех свободных неизвестных равными нулю, получим частное решение неоднородной системы $(b^1, \ldots, b^r, 0, \ldots, 0)^T$, которое называется *базисным решением, отвечающим базисным неизвестным* $x^1, \ldots, x^r$ *и свободным неизвестным* $x^{r+1}, \ldots, x^n$. Согласно принципу линейной суперпозиции для неоднородных систем (см. п. 4.17) общее решение можно записать в виде

$$X = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^r \\ x^{r+1} \\ x^{r+2} \\ \vdots \\ x^n \end{pmatrix} = \underbrace{\begin{pmatrix} b^1 \\ b^2 \\ \vdots \\ b^r \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\text{ЧРНС}} \underbrace{+ c^1 \begin{pmatrix} -a^1_{r+1} \\ -a^2_{r+1} \\ \vdots \\ -a^r_{r+1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + c^{n-r} \begin{pmatrix} -a^1_n \\ -a^2_n \\ \vdots \\ -a^r_n \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}}_{\text{ОРОС}}.$$

4.27. Пример. Рассмотрим однородную упрощённую систему

$$\begin{cases} \boxed{x^1} \phantom{{}+x^2} + 3x^3 \phantom{{}+x^4} + x^5 = 0, \\ \boxed{x^2} + 4x^3 \phantom{{}+x^4} + 2x^5 = 0, \\ \boxed{x^4} + 2x^5 = 0 \end{cases} \Longrightarrow \begin{cases} x^1 = -3x^3 - x^5, \\ x^2 = -4x^3 - 2x^5, \\ x^4 = -x^5. \end{cases}$$

Базисными неизвестными являются x^1, x^2, x^4 (они выделены), свободными — x^3, x^5. Если вместо свободных неизвестных x^3 и x^5 подставлять произвольные числа, $x^3 = c^1$, $x^5 = c^2$, и вычислять x^1, x^2, x^4 по указанным формулам, то получим общее решение системы:

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \\ x^4 \\ x^5 \end{pmatrix} = \begin{pmatrix} -3c^1 - c^2 \\ -4c^1 - 2c^2 \\ c^1 \\ -c^2 \\ c^2 \end{pmatrix} = \underbrace{c^1 \overbrace{\begin{pmatrix} -3 \\ -4 \\ 1 \\ 0 \\ 0 \end{pmatrix}}^{X_1} + c^2 \overbrace{\begin{pmatrix} -1 \\ -2 \\ 0 \\ -1 \\ 1 \end{pmatrix}}^{X_2}}_{\text{ОРОС}}.$$

Нормальное фундаментальное семейство решений (X_1, X_2) и нормальная фундаментальная матрица Φ данной системы имеют вид

$$X_1 = \begin{pmatrix} -3 \\ -4 \\ \boxed{1} \\ 0 \\ \boxed{0} \end{pmatrix}, \quad X_2 = \begin{pmatrix} -1 \\ -2 \\ \boxed{0} \\ -1 \\ \boxed{1} \end{pmatrix}, \qquad \Phi = \begin{pmatrix} -3 & -1 \\ -4 & -2 \\ \boxed{1} & \boxed{0} \\ 0 & -1 \\ \boxed{0} & \boxed{1} \end{pmatrix}.$$

Здесь выделены элементы матрицы, в которых находятся значения свободных неизвестных; матрица, составленная из выделенных элементов, является единичной матрицей порядка 2. Общее решение системы представляется через нормальное фундаментальное семейство решений и через нормальную фундаментальную матрицу по формулам

$$X = c^1 X_1 + c^2 X_2 = c^1 \begin{pmatrix} -3 \\ -4 \\ 1 \\ 0 \\ 0 \end{pmatrix} + c^2 \begin{pmatrix} -1 \\ -2 \\ 0 \\ -1 \\ 1 \end{pmatrix}, \quad X = \begin{pmatrix} -3 & -1 \\ -4 & -2 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c^1 \\ c^2 \end{pmatrix}.$$

4.28. Пример. Теперь рассмотрим неоднородную систему упрощенного вида:

$$\left\{\begin{array}{lllll} \boxed{x^1} & & + \; 3x^3 & & + \;\; x^5 = 1, \\ & \boxed{x^2} & + \; 4x^3 & & + \; 2x^5 = 2, \\ & & & \boxed{x^4} & + \; 2x^5 = 3 \end{array}\right. \quad \Longrightarrow \quad \left\{\begin{array}{l} x^1 = 1 - 3x^3 - x^5, \\ x^2 = 2 - 4x^3 - 2x^5, \\ x^4 = 3 - x^5; \end{array}\right.$$

базисные неизвестные x^1, x^2, x^4 выделены. Общее решение системы имеет вид

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \\ x^4 \\ x^5 \end{pmatrix} = \begin{pmatrix} 1 - 3c^1 - c^2 \\ 2 - 4c^1 - 2c^2 \\ c^1 \\ 3 - c^2 \\ c^2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 \\ 2 \\ \boxed{0} \\ 3 \\ \boxed{0} \end{pmatrix}}_{\text{ЧРНС}} \underbrace{+c^1 \overbrace{\begin{pmatrix} -3 \\ -4 \\ \boxed{1} \\ 0 \\ \boxed{0} \end{pmatrix}}^{X_1} + c^2 \overbrace{\begin{pmatrix} -1 \\ -2 \\ \boxed{0} \\ -1 \\ \boxed{1} \end{pmatrix}}^{X_2}}_{\text{ОРОС}}$$

(здесь выделены значения свободных неизвестных). Базисное решение неоднородной системы представляет собой её частное решение, отвечающее нулевым значениям свободных переменных.

2. Алгоритм Гаусса—Жордана

A. Элементарные преобразования.

4.29. Итак, если система линейных уравнений (однородная или неоднородная) имеет упрощённый вид, то её общее решение легко выписывается. Чтобы решить систему произвольного вида, нужно преобразовать её к эквивалентной системе упрощённого вида. Опишем алгоритм такого преобразования.

4.30. Очевидно, следующие преобразования системы уравнений переводят её в эквивалентную систему:

(1) перестановка двух уравнений местами;
(2) умножение любого уравнения системы на ненулевое число;
(3) прибавление к любому уравнению системы другого уравнения, умноженного на произвольное число.

Эти преобразования систем линейных уравнений называются *элементарными преобразованиями*.

4.31. Иногда к элементарным преобразованиям причисляют

(4) удаление из системы уравнения вида

$$0 \cdot x^1 + 0 \cdot x^2 + \cdots + 0 \cdot x^n = 0.$$

4.32. Вместо преобразований самой системы удобно выполнять преобразования расширенной матрицы этой системы; ясно, что при этом элементарным преобразованиям системы соответствуют следующие элементарные преобразования строк расширенной матрицы:

(1) перестановка двух строк местами;
(2) умножение любой строки на ненулевое число;
(3) прибавление к любой строке другой строки, умноженной на произвольное число;
(4) [удаление из матрицы нулевых строк].

Любую матрицу можно привести к упрощённому виду с помощью указанных преобразований. Соответствующий алгоритм называется *алгоритмом Гаусса*[1]*—Жордана*[2]; он изложен ниже в п. B.

[1]К. Ф. Гаусс (C. F. Gauß) — немецкий математик (1777–1855).

[2]В. Йордан (W. Jordan) — немецкий геодезист (1842–1899). В русском языке распространена ошибочная транслитерация фамилии («Жордан»). М. Жордан (M. Jordan) — французский математик (1838–1922); к методу Гаусса–Жордана не имеет отношения.

Б. Матрицы упрощённого вида.

4.33. Говорят, что матрица имеет *упрощённый вид*, если она является расширенной матрицей линейной системы упрощённого вида[1]. Матрица упрощённого вида имеет следующую структуру:

(1) все нулевые строки располагаются ниже всех ненулевых;
(2) *ведущий элемент* каждой строки, т.е. первый ненулевой элемент строки (при отсчёте слева направо), располагается строго правее ведущего элемента строки, лежащей выше;
(3) ведущий элемент каждой ненулевой строки равен единице и является единственным ненулевым элементом в своём столбце; соответствующие строки и столбцы называются *базисными*.

Таким образом, матрица упрощённого вида выглядит следующим образом:

$$\begin{pmatrix} \boxed{1} & * & 0 & * & 0 & * & \cdots & 0 & * & * \\ 0 & 0 & \boxed{1} & * & 0 & * & \cdots & 0 & * & * \\ 0 & 0 & 0 & 0 & \boxed{1} & * & \cdots & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & * & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \boxed{1} & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix};$$

базисные столбцы и ведущие элементы выделены.

4.34. Очевидны следующие свойства матриц упрощённого вида:

(1) некоторые столбцы матрицы являются последовательными столбцами единичной матрицы; эти столбцы, являющиеся не чем иным, как базисными столбцами, отвечают базисным неизвестным системы линейных уравнений;
(2) любой столбец, не являющийся базисным, представляет собой линейную комбинацию *предшествующих* базисных столбцов.

[1] В англоязычной литературе такой вид матрицы называется «reduced row echelon form». В отечественной литературе устоявшегося названия нет: используются термины «упрощённый вид», «ступенчатый вид», «приведённый ступенчатый вид по строкам» и т. п.

В. Алгоритм Гаусса—Жордана.

4.35. Опишем один шаг алгоритма Гаусса—Жордана; то этот алгоритм позволяет привести к упрощённому виду матрицу, содержащую m строк, не более чем за m шагов.

4.36. ***Шаг № k.***

1. Среди строк матрицы с номерами $k, \ldots, m$ выберем одну из строк с наименьшим количеством нулей, считая от начала строки. Эту строку назовём *ведущей строкой*, а её первый ненулевой элемент — *ведущим элементом*; *ведущим столбцом* назовём столбец, в котором располагается ведущий элемент.
2. Разделим ведущую строку на ведущий элемент; в полученной новой строке ведущий элемент будет равен 1.
3. Переставим ведущую строку на k-е место.
4. Вычтем из каждой строки матрицы ведущую строку, умноженную на коэффициенты, подобранные таким образом, чтобы все элементы ведущего столбца (кроме самого ведущего элемента) обратились в нуль. После этого ведущий столбец будет представлять собой k-й столбец единичной матрицы.

Процесс завершается, когда каждая строка матрицы уже побывала в роли ведущей строки или когда ведущую строку выбрать невозможно.

4.37. Пример. Приведём к упрощённому виду матрицу

$$A = \begin{array}{c} \begin{matrix} A_1 & A_2 & A_3 & A_4 & A_5 & A_6 \end{matrix} \\ \begin{pmatrix} 0 & -2 & 6 & 2 & 8 & -2 \\ 2 & 1 & 1 & 0 & 0 & 1 \\ 0 & -1 & 3 & 1 & 4 & -1 \\ 3 & 1 & 3 & 0 & 1 & 0 \end{pmatrix} \end{array}.$$

Шаг 1. В качестве ведущей строки можно выбрать вторую или четвёртую строку; выберем вторую. Ведущий элемент равен 2; делим ведущую строку на 2 и переставляем на первое место:

$$\begin{pmatrix} 0 & -2 & 6 & 2 & 8 & -2 \\ \mathbf{2} & 1 & 1 & 0 & 0 & 1 \\ 0 & -1 & 3 & 1 & 4 & -1 \\ 3 & 1 & 3 & 0 & 1 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} \mathbf{1} & \frac{1}{2} & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & -2 & 6 & 2 & 8 & -2 \\ 0 & -1 & 3 & 1 & 4 & -1 \\ \mathbf{3} & 1 & 3 & 0 & 1 & 0 \end{pmatrix}.$$

Ведущей строкой теперь является первая строка, а ведущим столбцом — первый столбец.

Обратим в нуль все элементы ведущего (первого) столбца, кроме ведущего элемента; единственный такой элемент — это 3 в четвёртой строке (этот элемент выделен жирным шрифтом). Выполним следующее элементарное преобразование: к четвёртой строке прибавим ведущую (первую), умноженную на (-3):

$$\begin{pmatrix} \boxed{1} & \frac{1}{2} & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & -2 & 6 & 2 & 8 & -2 \\ 0 & -1 & 3 & 1 & 4 & -1 \\ \mathbf{3} & 1 & 3 & 0 & 1 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} \boxed{1} & \frac{1}{2} & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & -2 & 6 & 2 & 8 & -2 \\ 0 & -1 & 3 & 1 & 4 & -1 \\ 0 & -\frac{1}{2} & \frac{3}{2} & 0 & 1 & -\frac{3}{2} \end{pmatrix}.$$

Первый столбец полученной матрицы теперь является первым столбцом единичной матрицы порядка 4.

Шаг 2. На втором шаге в качестве ведущей строки можно выбрать вторую, третью или четвёртую строку. Выберем третью, разделим её на ведущий элемент (-1) (он выделен) и переставим на второе место:

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & -2 & 6 & 2 & 8 & -2 \\ 0 & \boxed{-1} & 3 & 1 & 4 & -1 \\ 0 & -\frac{1}{2} & \frac{3}{2} & 0 & 1 & -\frac{3}{2} \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & \boldsymbol{\frac{1}{2}} & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \boxed{1} & -3 & -1 & -4 & 1 \\ 0 & \mathbf{-2} & 6 & 2 & 8 & -2 \\ 0 & \boldsymbol{-\frac{1}{2}} & \frac{3}{2} & 0 & 1 & -\frac{3}{2} \end{pmatrix}.$$

Ведущей строкой теперь является вторая строка, а ведущим столбцом — второй столбец.

Обратим в нуль все элементы ведущего (второго) столбца, кроме ведущего элемента (они выделены жирным шрифтом). Выполним следующие элементарные преобразования:

(а) к первой строке прибавляем ведущую (вторую) строку, умноженную на $(-1/2)$;

(б) к третьей строке прибавляем ведущую (вторую) строку, умноженную на 2;

(в) к четвёртой строке прибавляем ведущую (вторую) строку, умноженную на $1/2$.

Результат выполнения:

$$\begin{pmatrix} 1 & \boldsymbol{\frac{1}{2}} & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \boxed{1} & -3 & -1 & -4 & 1 \\ 0 & \mathbf{-2} & 6 & 2 & 8 & -2 \\ 0 & \boldsymbol{-\frac{1}{2}} & \frac{3}{2} & 0 & 1 & -\frac{3}{2} \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 2 & \frac{1}{2} & 2 & 0 \\ 0 & \boxed{1} & -3 & -1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & -1 & -1 \end{pmatrix}.$$

Второй столбец полученной матрицы теперь является вторым столбцом единичной матрицы порядка 4.

Шаг 3. В качестве ведущей строки можно выбрать только четвёртую строку. Делим её на ведущий элемент $-1/2$ (он выделен) и переставляем на третье место:

$$\begin{pmatrix} 1 & 0 & 2 & \frac{1}{2} & 2 & 0 \\ 0 & 1 & -3 & -1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{-\frac{1}{2}} & -1 & -1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 2 & \boldsymbol{\frac{1}{2}} & 2 & 0 \\ 0 & 1 & -3 & \mathbf{-1} & -4 & 1 \\ 0 & 0 & 0 & \mathbf{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Ведущей строкой теперь является третья строка, а ведущим столбцом — четвёртый столбец.

Обратим в нуль элементы ведущего (четвёртого) столбца, кроме ведущего элемента (они выделены жирным шрифтом), для чего выполним следующие элементарные преобразования:

(а) к первой строке прибавим ведущую (третью) строку, умноженную на $-1/2$;

(б) к первой строке прибавляем ведущую (третью) строку.

Результат выполнения:

$$\begin{pmatrix} 1 & 0 & 2 & \boldsymbol{\frac{1}{2}} & 2 & 0 \\ 0 & 1 & -3 & \mathbf{-1} & -4 & 1 \\ 0 & 0 & 0 & \mathbf{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 2 & 0 & 1 & -1 \\ 0 & 1 & -3 & 0 & -2 & 3 \\ 0 & 0 & 0 & \mathbf{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Четвёртый столбец полученной матрицы теперь является третьим столбцом единичной матрицы порядка 4.

Ещё один шаг выполнить невозможно, так как четвертую строку нельзя выбрать в качестве ведущей: в ней нет ненулевых элементов. Процедура закончена, а полученная матрица имеет упрощённый вид:

$$A' = \begin{pmatrix} \mathbf{1} & 0 & 2 & 0 & 1 & -1 \\ 0 & \mathbf{1} & -3 & 0 & -2 & 3 \\ 0 & 0 & 0 & \mathbf{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Её первый, второй и четвёртый столбцы (A'_1, A'_2, A'_4) являются базисными (они представляют соответственно собой первый, второй и третий столбцы единичной матрицы размера порядка 4), а остальные столбцы являются линейными комбинациями базисных:

$$A'_3 = 2A'_1 - 3A'_2,$$
$$A'_5 = A'_1 - 2A'_2 + 2A'_4,$$
$$A'_6 = -A'_1 + 3A'_2 + 2A'_4.$$

Коэффициенты этих линейных комбинаций — это элементы соответствующих столбцов преобразованной матрицы.

Отметим, что указанные зависимости имеют место не только для столбцов преобразованной матрицы, но и для столбцов исходной:

$$A_3 = \begin{pmatrix} 6 \\ 1 \\ 3 \\ 3 \end{pmatrix} = 2 \begin{pmatrix} 0 \\ 2 \\ 0 \\ 3 \end{pmatrix} - 3 \begin{pmatrix} -2 \\ 1 \\ -1 \\ 1 \end{pmatrix} = 2A_1 - 3A_2,$$

$$A_5 = \begin{pmatrix} 8 \\ 0 \\ 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 3 \end{pmatrix} - 2 \begin{pmatrix} -2 \\ 1 \\ -1 \\ 1 \end{pmatrix} + 2 \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} = A_1 - 2A_2 + A_4,$$

$$A_6 = \begin{pmatrix} -2 \\ 1 \\ -1 \\ 0 \end{pmatrix} = - \begin{pmatrix} 0 \\ 2 \\ 0 \\ 3 \end{pmatrix} + 3 \begin{pmatrix} -2 \\ 1 \\ -1 \\ 1 \end{pmatrix} + 2 \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} = -A_1 + 3A_2 + 2A_4.$$

Разумеется, этот факт не случаен; он будет объяснён позже (см. теорему 5.46.2, с. 107).

Таким образом, решение системы линейных уравнений можно получить, приведя её расширенную матрицу к упрощённому виду при помощи алгоритма Гаусса—Жордана.

Отметим следующее интересное свойство однородных систем.

4.38. Предложение (достаточное условие нетривиальной разрешимости однородной системы)**.** *Если в однородной системе число неизвестных больше числа уравнений, то она имеет нетривиальное решение.*

Доказательство. После приведения системы к упрощённому виду в ней обязательно образуется хотя бы одна свободная неизвестная, так как количество всех неизвестных по условию больше количества уравнений, т.е. больше количества базисных неизвестных. Придав этой свободной неизвестной произвольное ненулевое значение, получим нетривиальное решение системы. □

Г. «Правило прямоугольников».

4.39. В рассмотренном выше примере 4.37 вычисления с дробями доставляли определённые неудобства, которых можно избежать, если применить следующую модификацию алгоритма[1], которая позволяет обратить в нуль все элементы ведущего столбца (кроме ведущего элемента). Разъясним эту модификацию на следующем примере. Рассмотрим матрицу

$$\begin{pmatrix} p & q & r & s \\ a & b & \boldsymbol{c} & d \\ x & y & z & u \end{pmatrix}.$$

Проведём один шаг алгоритма Гаусса—Жордана с ведущим элементом c; задачей является обращение в нуль всех элементов ведущего столбца (разумеется, кроме ведущего элемента). Для уничтожения элемента r (при условии $r \neq 0$) выполним предварительно следующие вспомогательные элементарные преобразования:

(а) первую строку матрицы умножим на ведущий элемент c;
(б) вторую (ведущую) строку умножим на r;

в результате получим матрицу

$$\begin{pmatrix} cp & cq & \boldsymbol{cr} & cs \\ ra & rb & \boldsymbol{rc} & rd \\ x & y & z & u \end{pmatrix}.$$

Вычитая теперь из первой строки вторую (ведущую) и после этого возвращая ведущей строке её первоначальный вид (т.е. деля её на r), получим

$$\begin{pmatrix} cp{-}ra & cq{-}rb & 0 & cs{-}rd \\ ra & rb & \boldsymbol{rc} & rd \\ x & y & z & u \end{pmatrix} \Rightarrow \begin{pmatrix} cp{-}ra & cq{-}rb & 0 & cs{-}rd \\ a & b & \boldsymbol{c} & d \\ x & y & z & u \end{pmatrix}.$$

Аналогично уничтожаем элемент z (если $z \neq 0$), для чего сначала выполняем вспомогательные преобразования:

(а) третью строку матрицы умножим на ведущий элемент c;
(б) вторую (ведущую) строку умножим на z;

в результате получаем матрицу

$$\begin{pmatrix} cp - ra & cq - rb & 0 & cs - rd \\ za & zb & \boldsymbol{zc} & zd \\ cx & cy & \boldsymbol{cz} & cu \end{pmatrix}.$$

[1]В англоязычной литературе она называется «fraction-free Gaussian elimination».

Вычитая из третьей строки вторую (ведущую) и после этого возвращая ведущей строке её первоначальный вид, получим

$$\begin{pmatrix} cp-ra & cq-rb & 0 & cs-rd \\ za & zb & \boldsymbol{zc} & zd \\ cx-za & cy-zb & 0 & cu-zd \end{pmatrix} \Rightarrow \begin{pmatrix} cp-ra & cq-rb & 0 & cs-rd \\ a & b & \boldsymbol{c} & d \\ cx-za & cy-zb & 0 & cu-zd \end{pmatrix}.$$

Цель достигнута: все элементы ведущего столбца, кроме ведущего элемента, равны нулю. Если все элементы исходной матрицы были целыми числами, то и в процессе вычислений нам не придётся работать с дробями.

4.40. Сформулируем теперь правило, по которому вычисляются элементы преобразованной матрицы:

$$\begin{pmatrix} p & q & r & s \\ a & b & \boldsymbol{c} & d \\ x & y & z & u \end{pmatrix} \Rightarrow \begin{pmatrix} cp-ra & cq-rb & 0 & cs-rd \\ a & b & \boldsymbol{c} & d \\ cx-za & cy-zb & 0 & cu-zd \end{pmatrix}.$$

Например, элемент p превращается в элемент $cp - ra$; отметим участвующие в этом выражении элементы исходной матрицы:

$$\begin{pmatrix} p & q & r & s \\ a & b & \boldsymbol{c} & d \\ x & y & z & u \end{pmatrix}.$$

Выражение $cp - ra$ представляет собой разность произведения преобразуемого элемента p на ведущий элемент c, расположенных в противоположных углах воображаемого прямоугольника, и произведения пары элементов матрицы, r и a, расположенных в двух других противоположных углах того же прямоугольника. Для остальных элементов ситуация аналогична:

$$q \mapsto cq - rb : \begin{pmatrix} p & q & r & s \\ a & b & \boldsymbol{c} & d \\ x & y & z & u \end{pmatrix}, \quad s \mapsto cs - rd : \begin{pmatrix} p & q & r & s \\ a & b & \boldsymbol{c} & d \\ x & y & z & u \end{pmatrix},$$

Полученное правило пересчёта элементов будем называть *правилом прямоугольников*; оно может быть сформулировано следующим образом:

(а) в матрице выбирается ведущий элемент a^s_t;
(б) ведущая строка (строка A^s) переписывается без изменений;
(в) все элементы ведущего столбца (столбца A_t), кроме ведущего элемента, заменяются нулями;
(г) каждый элемент a^i_j матрицы, не принадлежащий ведущей строке или ведущему столбцу, заменяется на число $a^i_j a^s_t - a^i_t a^s_j$.

Схематическая иллюстрация:

$$\begin{pmatrix} & \vdots & & \vdots & \\ \cdots & a_j^i & \cdots & a_t^i & \cdots \\ & \vdots & & \vdots & \\ \cdots & a_j^s & \cdots & \boldsymbol{a_t^s} & \cdots \\ & \vdots & & \vdots & \end{pmatrix} \Rightarrow \begin{pmatrix} & \vdots & & \vdots & \\ \cdots & a_j^i a_t^s - a_t^i a_j^s & \cdots & 0 & \cdots \\ & \vdots & & \vdots & \\ \cdots & a_j^s & \cdots & \boldsymbol{a_t^s} & \cdots \\ & \vdots & & \vdots & \end{pmatrix}.$$

4.41. Пример. Приведём к упрощённому виду матрицу из примера 4.37,

$$\begin{pmatrix} 0 & -2 & 6 & 2 & 8 & -2 \\ 2 & 1 & 1 & 0 & 0 & 1 \\ 0 & -1 & 3 & 1 & 4 & -1 \\ 3 & 1 & 3 & 0 & 1 & 0 \end{pmatrix},$$

используя при необходимости правило прямоугольников.

Шаг 1. В качестве ведущей выберем вторую строку и переставим её на первое место:

$$\begin{pmatrix} 0 & -2 & 6 & 2 & 8 & -2 \\ \mathbf{2} & 1 & 1 & 0 & 0 & 1 \\ 0 & -1 & 3 & 1 & 4 & -1 \\ 3 & 1 & 3 & 0 & 1 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} \mathbf{2} & 1 & 1 & 0 & 0 & 1 \\ 0 & -2 & 6 & 2 & 8 & -2 \\ 0 & -1 & 3 & 1 & 4 & -1 \\ 3 & 1 & 3 & 0 & 1 & 0 \end{pmatrix}.$$

На первом шаге требуется все элементы ведущего (первого) столбца, кроме самого ведущего элемента, сделать равными нулю. Два из них (во второй и третьей строках) и так равны нулю, поэтому соответствующие строки преобразовывать не требуется (впрочем, можно разделить их на (-2) и (-1) соответственно), а элементы третьей строки преобразуем по правилу прямоугольников (показано вычисление элемента последней строки второго столбца):

$$\begin{pmatrix} \mathbf{2} & 1 & 1 & 0 & 0 & 1 \\ 0 & -2 & 6 & 2 & 8 & -2 \\ 0 & -1 & 3 & 1 & 4 & -1 \\ 3 & 1 & 3 & 0 & 1 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 2 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & -3 & -1 & -4 & 1 \\ 0 & 1 & -3 & -1 & -4 & 1 \\ 0 & -1 & \mathbf{3 \cdot 2 - 3 \cdot 1 = 3} & 0 & 2 & -3 \end{pmatrix}.$$

Шаг 2. В качестве ведущей выберем вторую строку. Вычтя из третьей строки матрицы вторую, получим нулевую строку, которую сразу переставим на последнее место:

$$\begin{pmatrix} 2 & 1 & 1 & 0 & 0 & 1 \\ 0 & \mathbf{1} & -3 & -1 & -4 & 1 \\ 0 & 1 & -3 & -1 & -4 & 1 \\ 0 & -1 & 3 & 0 & 2 & -3 \end{pmatrix} \Rightarrow \begin{pmatrix} 2 & 1 & 1 & 0 & 0 & 1 \\ 0 & \mathbf{1} & -3 & -1 & -4 & 1 \\ 0 & -1 & 3 & 0 & 2 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

На втором шаге нужно обратить в нуль все элементы ведущего (второго) столбца (кроме самого ведущего элемента). В данной ситуации, по-видимому, вместо правила прямоугольников удобнее просто ведущую строку вычесть из первой и прибавить к третьей:

$$\begin{pmatrix} 2 & 1 & 1 & 0 & 0 & 1 \\ 0 & \mathbf{1} & -3 & -1 & -4 & 1 \\ 0 & -1 & 3 & 0 & 2 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 2 & 0 & 4 & 1 & 4 & 0 \\ 0 & \mathbf{1} & -3 & -1 & -4 & 1 \\ 0 & 0 & 0 & -1 & -2 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Шаг 3. В качестве ведущей строки выберем третью; после преобразований получаем

$$\begin{pmatrix} 2 & 0 & 4 & 1 & 4 & 0 \\ 0 & 1 & -3 & -1 & -4 & 1 \\ 0 & 0 & 0 & \mathbf{-1} & -2 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 2 & 0 & 4 & 1 & 4 & 0 \\ 0 & 1 & -3 & -1 & -4 & 1 \\ 0 & 0 & 0 & \mathbf{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow$$

$$\Rightarrow \begin{pmatrix} 2 & 0 & 4 & 0 & 2 & -2 \\ 0 & 1 & -3 & 0 & -2 & 3 \\ 0 & 0 & 0 & \mathbf{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} \mathbf{1} & 0 & 2 & 0 & 1 & -1 \\ 0 & \mathbf{1} & -3 & 0 & -2 & 3 \\ 0 & 0 & 0 & \mathbf{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

в последней матрице выделены базисные столбцы.

Д. Алгоритм Гаусса.

4.42. Алгоритм Гаусса отличается от алгоритма Гаусса—Жордана тем, что при выполнении элементарных преобразований на каждом шаге обращают в нуль не все элементы ведущего столбца (за исключением, разумеется, ведущего элемента), а только элементы, лежащие *ниже* ведущего элемента. В результате матрица приводится к так называемому *ступенчатому виду по строкам*[1]:

[1] В англоязычной литературе он называется «row echelon form».

$$\begin{pmatrix}
\boxed{*} & * & * & * & * & * & \cdots & * & * & * \\
0 & 0 & \boxed{*} & * & * & * & \cdots & * & * & * \\
0 & 0 & 0 & 0 & \boxed{*} & * & \cdots & * & * & * \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & * & * & * \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & \boxed{*} & * & * \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0
\end{pmatrix}.$$

Матрица ступенчатого вида имеет следующую структуру:

(a) все нулевые строки располагаются ниже всех ненулевых;

(b) *ведущий элемент* каждой строки, т.е. первый ненулевой элемент строки, располагается строго правее ведущего элемента строки, лежащей выше.

В отличие от определения матрицы упрощённого вида (ср. с. 73) здесь отсутствует требование, чтобы ведущий элемент каждой ненулевой строки был равен единице и являлся единственным ненулевым элементом в своём столбце. Очевидно, объём вычислений при использовании алгоритма Гаусса несколько меньше, чем при использовании алгоритма Гаусса—Жордана, однако он не позволяет непосредственно выписывать решение системы линейных уравнений и указывать линейные зависимости между столбцами матрицы.

3. Применения алгоритма Гаусса—Жордана

A. Связь элементарных преобразований строк и операции умножения матриц.

4.43. Предложение. *Пусть A — некоторая матрица, $R(A)$ — матрица, полученная из A элементарными преобразованиями строк R. Тогда*

$$R(A) = R(\mathbb{1}) \cdot A,$$

где $\mathbb{1}$ — единичная матрица, $R(\mathbb{1})$ — матрица, полученная из единичной с помощью тех же самых элементарных преобразований строк R. Матрица $R(\mathbb{1})$, называемая матрицей элементарных преобразований, *обратима.*

Доказательство. При умножении матриц каждый столбец матрицы-произведения равен произведению матрицы-первого сомножителя на соответствующий столбец матрицы-второго сомножителя (см. теорему 3.26(2), с. 56), поэтому достаточно проверить утверждение теоремы для матрицы A, состоящей из одного столбца.

4.44. Пусть R — перестановка строк с номерами s и t. Применяя это преобразование к единичной матрице, получаем матрицу элементарного преобразования

$$R(\mathbb{1}) = R\left(\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}\right) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

Умножая эту матрицу на столбец A, находим

$$R(\mathbb{1}) \cdot A = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \cdot \begin{pmatrix} a^1 \\ \vdots \\ a^s \\ \vdots \\ a^t \\ \vdots \\ a^n \end{pmatrix} = \begin{pmatrix} a^1 \\ \vdots \\ a^t \\ \vdots \\ a^s \\ \vdots \\ a^n \end{pmatrix} = R(A).$$

Обратимость матрицы $R(\mathbb{1})$ элементарного преобразования очевидна: обратная матрица, осуществляющая обратное элементарное преобразование, совпадает с исходной,

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

4.45. Пусть R — элементарное преобразование, состоящее в умножении строки с номером s на число $\lambda \neq 0$. Применяя это преобразование к единичной матрице, получим матрицу элементарного преобразования

$$R(\mathbb{1}) = R\left(\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}\right) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & \lambda & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

Умножая эту матрицу на столбец A, находим:

$$R(\mathbb{1}) \cdot A = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & \lambda & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \cdot \begin{pmatrix} a^1 \\ \vdots \\ a^s \\ \vdots \\ a^n \end{pmatrix} = \begin{pmatrix} a^1 \\ \vdots \\ \lambda a^s \\ \vdots \\ a^n \end{pmatrix} = R(A).$$

Обратимость матрицы $R(\mathbb{1})$ элементарного преобразования здесь также очевидна:

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & \lambda & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & \dfrac{1}{\lambda} & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

4.46. Наконец, рассмотрим элементарное преобразование, состоящее в прибавлении строки с номером s, умноженной на число λ, к строке с номером t (для определённости рассмотрим случай $s < t$):

$$R(\mathbb{1}) = R\left(\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}\right) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \lambda & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix},$$

$$R(\mathbb{1})\cdot A=\begin{pmatrix}1&\cdots&0&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&1&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&\lambda&\cdots&1&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&0&\cdots&0&\cdots&1\end{pmatrix}\begin{pmatrix}a^1\\ \vdots\\ a^s\\ \vdots\\ a^t\\ \vdots\\ a^n\end{pmatrix}=\begin{pmatrix}a^1\\ \vdots\\ a^s\\ \vdots\\ \lambda a^s+a^t\\ \vdots\\ a^n\end{pmatrix}=R(A).$$

Обратное преобразование состоит в вычитании умноженной на λ строки с номером s из строки с номером t; этому преобразованию соответствует матрица

$$\begin{pmatrix}1&\cdots&0&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&1&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&-\lambda&\cdots&1&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&0&\cdots&0&\cdots&1\end{pmatrix}.$$

4.47. Если несколько элементарных преобразований выполняются последовательно, то утверждение теоремы также остаётся верным. Действительно, пусть R_1 и R_2 — два элементарных преобразования, $R_2\circ R_1$ — их композиция (сначала выполняется R_1, затем R_2). Тогда

$$\begin{aligned}(R_2\circ R_1)(A)&=R_2\big(R_1(A)\big)=R_2(\mathbb{1})\cdot R_1(A)=\\ &=R_2(\mathbb{1})\cdot\big(R_1(\mathbb{1})\cdot A\big)=\big(R_2(\mathbb{1})\cdot R_1(\mathbb{1})\big)\cdot A=\\ &=R_2\big(R_1(\mathbb{1})\big)\cdot A=\big(R_2\circ R_1\big)(\mathbb{1})\cdot A.\end{aligned}$$

Итак, любая композиция элементарных преобразований строк матрицы эквивалентна умножению этой матрицы слева на некоторую обратимую квадратную матрицу. Теорема доказана. □

Б. Вычисление обратной матрицы.

4.48. Алгоритм Гаусса—Жордана доставляет удобный метод вычисления обратной матрицы. Действительно, если при помощи элементарных преобразований строк привести данную квадратную матрицу A к единичной, то, в силу предложения 4.43 матрицей преобразования будет A^{-1}. На практике удобно поступить следующим образом: составить блочную матрицу $\|A \mid \mathbb{1}\|$ и элементарными преобразованиями строк привести её левый блок к единичной матрице; тогда в правом блоке образуется матрица A^{-1}:

$$\|A \mid \mathbb{1}\| \xrightarrow[\text{строк}]{\text{эл. преобр.}} \|\mathbb{1} \mid A^{-1}\|.$$

Может оказаться, что приведение левого блока к единичной матрице невозможно; это означает, что матрица A не имеет обратной. Это явление будет объяснено в теореме 6.57 (см. с. 143).

4.49. Пример. Найдём матрицу, обратную к матрице

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 6 \end{pmatrix}.$$

Проводя вычисления согласно описанному рецепту, получаем:

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 3 & 4 & 0 & 1 & 0 \\ 3 & 4 & 6 & 0 & 0 & 1 \end{array}\right) \Rightarrow \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & -2 & -3 & -3 & 0 & 1 \end{array}\right) \Rightarrow$$

$$\Rightarrow \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & 2 & -1 & 0 \\ 0 & -2 & -3 & -3 & 0 & 1 \end{array}\right) \Rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & -1 & -3 & 2 & 0 \\ 0 & 1 & 2 & 2 & -1 & 0 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right) \Rightarrow$$

$$\Rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -2 & 0 & 1 \\ 0 & 1 & 0 & 0 & 3 & -2 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right) \Rightarrow A^{-1} = \begin{pmatrix} -2 & 0 & 1 \\ 0 & 3 & -2 \\ 1 & -2 & 1 \end{pmatrix}.$$

4.50. Аналогичную методику можно использовать для вычисления матрицы $A^{-1}B$:

$$\| A \mid B \| \xrightarrow[\text{строк}]{\text{эл. преобр.}} \| \mathbb{1} \mid A^{-1}B \|.$$

4.51. Если требуется найти матрицу AB^{-1}, можно подобным образом использовать элементарные преобразования столбцов, однако на практике поступают следующим образом:

$$\left\| B^T \;\middle|\; A^T \right\| \xrightarrow[\text{строк}]{\text{эл. преобр.}} \left\| \mathbb{1} \;\middle|\; (B^T)^{-1}A^T \right\| \longrightarrow$$

$$\xrightarrow[\text{рование}]{\text{транспони-}} \Big((B^T)^{-1}A^T\Big)^T = AB^{-1}.$$

В. Составление однородной системы по заданному фундаментальному семейству решений.

4.52. Пусть имеется семейство линейно независимых столбцов $X_1, \ldots, X_p \in \mathbb{K}^n$, $p < n$. Требуется составить однородную систему $AX = O$, для которой данный набор столбцов образует фундаментальное семейство решений. (Случай $p = n$ не представляет интереса, поскольку тогда столбцы $X_1, \ldots, X_n$ образуют базис в $\mathbb{K}^n$, фундаментальная матрица $\Phi = \|X_1, \ldots, X_n\| \in \mathbb{K}^{n\times n}$ имеет обратную (см. ниже теорему 6.57, с. 143) и из равенства $A\Phi = O$ вытекает $A = O$.)

4.53. Число уравнений в искомой системе не может быть менее $n - p$. Действительно, число p столбцов, составляющих ФСР, равно числу *свободных* неизвестных в системе, которое, в свою очередь, равно разности числа n *всех* неизвестных и числа *базисных* неизвестных, равного числу линейно независимых уравнений. Будем искать именно такие *минимальные* системы, т.е. системы, состоящие из $n - p$ линейно независимых уравнений.

4.54. Рассмотрим систему линейно независимых уравнений

$$\underset{m\times n}{A}\ \underset{n\times 1}{X} = \underset{m\times 1}{O};$$

иными словами, строки матрицы A линейно независимы. Пусть $\Phi = \|X_1, \ldots, X_p\|$ — фундаментальная матрица этой системы: столбцы $X_1, \ldots, X_p$ линейно независимы и $AX_j = O$ для любого $j = 1, \ldots, p$. Последние уравнения можно объединить в матричное уравнение

$$\underset{m\times n}{A}\ \underset{n\times p}{\Phi} = \underset{m\times p}{O}.$$

Транспонируем обе части это уравнения:

$$\underset{p\times n}{\Phi^T}\ \underset{n\times m}{A^T} = \underset{p\times m}{O}.$$

Таким образом, столбцы матрицы A^T (т.е. транспонированные строки матрицы A) являются решениями однородной системы

$$\underset{p\times n}{\Phi^T}\ \underset{n\times 1}{Y} = \underset{p\times 1}{O}.$$

Фундаментальное семейство решений этой системы состоит из $n - p \overset{\text{def}}{=\!=} m$ столбцов, которые после транспонирования становятся строками основной матрицы A искомой системы $AX = O$.

4.55. Пример. Найдём систему однородных линейных уравнений, для которой столбцы $X_1 = (1, 2, 3, 4)^T$ и $X_2 = (5, 6, 7, 8)^T$ образуют фундаментальное семейство решений.

Решим методом Гаусса—Жордана систему $\Phi^T Y = O$ (вычисления опущены):

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \end{pmatrix}.$$

НФСР этой системы имеет вид

$$Y_1 = \begin{pmatrix} 1, & -2, & 1, & 0 \end{pmatrix}^T, \quad Y_1 = \begin{pmatrix} 2, & -3, & 0, & 1 \end{pmatrix}^T,$$

так что основная матрица искомой системы

$$A = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 2 & -3 & 0 & 1 \end{pmatrix}.$$

ГЛАВА 5

Векторные пространства

Вектор в современной математике — также одно из фундаментальных понятий, но, в отличие от понятия матрицы, возникло оно относительно недавно. В элементарной геометрии вектор — это направленный отрезок. Интуитивно вектор понимается как объект, имеющий величину и направление. Естественным образом векторам ставятся в соответствие параллельные переносы, что объясняет сам термин (лат. vector — несущий). Интерпретация вектора как параллельного переноса позволяет ввести операции сложения векторов и умножения векторов на числа, что открывает возможность применять алгебраические методы при решении геометрических задач. Векторы широко используются и в физике: они представляют величины, имеющие помимо численного значения ещё и направление (силы, скорости и т. п.). Если в пространстве задана система координат, то вектор однозначно задаётся набором своих координат. Поэтому упорядоченные наборы чисел тоже называют векторами.

Современное изложение аналитической геометрии (а также и большинства физических курсов) базируется на теории векторов. Первые понятия исчисления векторов возникли в рамках геометрической модели комплексных чисел К. Ф. Гаусса. Позже У. Р. Гамильтон[1] и Г. Грассман[2] разработали алгебраические операции над векторами и некоторые операции векторного анализа. Впоследствии этот формализм использовал Дж. К. Максвелл[3] в своих трудах по электромагнетизму, тем самым обратив внимание учёных на новое исчисление. Современный вид придал векторному анализу О. Хевисайд[4].

---

[1] У. Р. Га́мильтон (W. R. Hamilton) — ирландский математик и физик (1805–1865).

[2] Г. Грассман (H. Grassmann) — немецкий физик, математик и филолог (1809–1877).

[3] Дж. К. Ма́ксвелл (J. C. Maxwell) — британский (шотландец по происхождению) физик и математик (1831–1879).

[4] О. Хевисайд (O. Heaviside) — английский учёный-самоучка, инженер, математик и физик (1850–1925).

Широкое применение теории векторов объясняется тем, что векторные представления адекватно передают суть многих геометрических и физических понятий и закономерностей. Кроме того, векторные формулы не зависят от выбора той или иной системы координат, т.е. имеют инвариантный характер и отражают сущность явления в чистом виде. Наконец, в векторных выкладках соединяются наиболее привлекательные черты алгебраического и геометрического методов исследования, благодаря чему векторные формулы отличаются краткостью и наглядностью.

Если принять вектор за основное понятие и зафиксировать в аксиомах свойства операций над векторами, доказанные в рамках элементарной геометрии как теоремы, то получится фундаментальное понятие векторного пространства. В рамках этого абстрактного определения векторами оказываются очень многие математические объекты: функции, матрицы и т. п. Изучением общих алгебраических свойств таких объектов занимается специальный раздел математики — *линейная алгебра*. Некоторые вопросы линейной алгебры, нужные для разработки геометрии, будут рассмотрены в этой и следующих главах; систематический курс излагается во втором семестре.

1. Векторы в элементарной геометрии

Напомним известные из курса средней школы сведения о векторах, понимаемых в элементарно-геометрическом смысле.

A. Направленные отрезки.

5.1. В элементарной геометрии *направленным отрезком* $\overrightarrow{AB}$ называется упорядоченная пара точек A и B, первая A из которых называется *началом*, а вторая B — *концом* направленного отрезка $\overrightarrow{AB}$. (Можно также считать, что в состав направленного отрезка $\overrightarrow{AB}$ помимо его начала A и конца B входят все точки, лежащие на прямой AB между точками A и B.) На чертеже направленный отрезок изображается стрелкой, начинающейся в точке A и заканчивающейся в точке B.

Если точки A и B совпадают, то направленный отрезок $\overrightarrow{AB}$ (точнее, $\overrightarrow{AA}$) называется *вырожденным* или *нулевым* и обозначается $\mathbf{0}$. Его длина по определению считается равной нулю, а направления он не имеет.

5.2. Направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{CD}$ называются *равными*, если середины отрезков AD и BC совпадают; обозначение $\overrightarrow{AB} = \overrightarrow{CD}$. На рис. 5.1 изображены все возможные случаи взаимного расположения равных направленных отрезков. Равные направленные отрезки имеют одинаковые длины и одинаковые направления, т.е. расположены на коллинеарных[1] прямых и направлены в одну сторону.

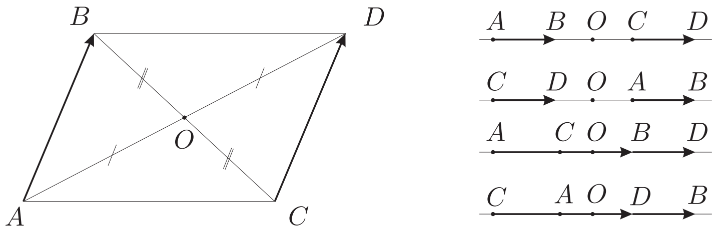


Рис. 5.1. Равные направленные отрезки

Б. Понятие вектора.

5.3. Отношение равенства направленных отрезков является отношением эквивалентности в смысле определения II.33 (см. с. 308). Поэтому множество всех направленных отрезков разбивается на непересекающиеся классы эквивалентности (см. теорему II.36, с. 308). Каждый такой класс эквивалентности называется *вектором* (или *свободным вектором*). Векторы обычно обозначаются строчными полужирными буквами: $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ и т. д. Ясно, что равенство векторов означает совпадение соответствующих классов эквивалентности.

5.4. Отложить вектор $\boldsymbol{a}$ от точки A означает построить направленный отрезок $\overrightarrow{AB}$ с началом в точке A, входящий в класс эквивалентности, описываемый вектором $\boldsymbol{a}$. Направленный отрезок $\overrightarrow{AB}$ называется *представителем* или *реализацией* вектора $\boldsymbol{a}$.

5.5. *Нулевой вектор* $\mathbf{0}$ — это класс эквивалентности нулевого направленного отрезка.

5.6. *Длина* вектора $\boldsymbol{a}$ — это длина любого направленного отрезка, являющегося его представителем; обозначение $|\boldsymbol{a}|$. Аналогичным образом, мы можем говорить о сонаправленных и противоположно направленных векторах, о коллинеарных и компланарных векторах и т. п.

5.7. Понятие векторной величины возникло в связи с рассмотрением в физике таких понятий, как перемещение, скорость, ускорение и т. п. Понятие вектора в геометрии несколько отличается от понятия векторной величины в физике.

[1]Две прямые называются *коллинеарными*, если они параллельны или совпадают.

Равенство направленных отрезков можно определить иначе, нежели в п. 5.2; нужно лишь позаботиться о том, чтобы рассматриваемое отношение равенства было отношением эквивалентности. Рассматривая классы эквивалентности относительно этого нового отношения эквивалентности, получим другие разновидности векторов, обладающие иными свойствами. Например, если считать, что два направленных отрезка равны тогда и только тогда, когда совпадают их начала и совпадают их концы, то получим определение *связанного вектора*; именно таким вектором является скорость материальной точки в механике: вектор скорости «прикреплён» к частице, и его начальная точка перемещается вместе с частицей.

Если же считать два направленных отрезка равными в том и только том случае, когда они лежат на одной прямой, имеют одинаковое направление и равные длины, то мы приходим к понятию *скользящего вектора*. Скользящие векторы нужны, например, в статике при изучении равновесия тел: силу нельзя перемещать параллельно самой себе, но можно переносить вдоль линии её действия.

B. Линейные операции над векторами. В механике устанавливается, что действие двух сил на материальную точку эквивалентно действию одной силы, которая определяется по правилу параллелограмма. Этот экспериментальный факт инспирирует следующее определение.

5.8. *Суммой* двух векторов $\boldsymbol{a} = \overrightarrow{OA}$ и $\boldsymbol{b} = \overrightarrow{OB}$ называется вектор $\boldsymbol{a}+\boldsymbol{b}$, являющийся диагональю $\overrightarrow{OC}$ параллелограмма со сторонами $\overrightarrow{OA}$ и $\overrightarrow{OB}$.

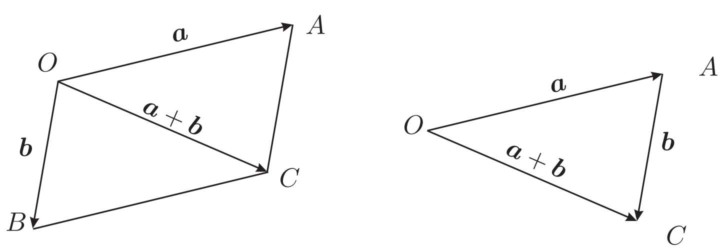


Рис. 5.2. Сумма векторов

Так как $\boldsymbol{b} = \overrightarrow{AC}$, определение суммы векторов можно записать в виде

$$\overrightarrow{OA} + \overrightarrow{AC} = \overrightarrow{OC}.$$

Эта формула выражает другое определение суммы векторов, называемое *правилом треугольника*.

Формула $\overrightarrow{AA} + \overrightarrow{AB} = \overrightarrow{AB}$ показывает, что вектор $\mathbf{0}$ является нулём сложения, т.е.
$$\mathbf{0} + \boldsymbol{a} = \boldsymbol{a}$$
для любого вектора $\boldsymbol{a}$. Переставив концевые точки вектора $\boldsymbol{a} = \overrightarrow{AB}$, получим вектор $\overrightarrow{BA}$, который обозначается символом $-\boldsymbol{a}$. Формула $\overrightarrow{AB} + \overrightarrow{BA} = \overrightarrow{AA}$ показывает, что вектор $-\boldsymbol{a}$ является противоположным вектору $\boldsymbol{a}$ по отношению к сложению, т.е.
$$\boldsymbol{a} + (-\boldsymbol{a}) = \mathbf{0}$$
для любого вектора $\boldsymbol{a}$.

5.9. *Произведением вектора* $\boldsymbol{a}$ *на вещественное число* α называется вектор $\alpha\boldsymbol{a}$, длина которого равна длине вектора $\boldsymbol{a}$, умноженной на $|\alpha|$, а направление совпадает с направлением вектора $\boldsymbol{a}$, если $\alpha > 0$, и противоположно ему, если $\alpha < 0$. Случаи $\alpha = 0$ и $\boldsymbol{a} = \mathbf{0}$ не исключаются.

Определение умножения вектора на число согласовано с определением сложения векторов в том смысле, что для любого натурального n справедливы равенства
$$n\boldsymbol{a} = \underbrace{\boldsymbol{a} + \dots + \boldsymbol{a}}_{n \text{ слагаемых}}.$$
Кроме того, очевидно,
$$(-1)\boldsymbol{a} = -\boldsymbol{a}.$$

Операции сложения векторов и умножения вектора на число, называются *линейными операциями*.

5.10. Предложение. *Линейные операции над векторами обладают следующими свойствами:*

(1) *коммутативность сложения: для любых векторов* $\boldsymbol{a}$ *и* $\boldsymbol{b}$
$$\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{b} + \boldsymbol{a};$$

(2) *ассоциативность сложения: для любых векторов* $\boldsymbol{a}$, $\boldsymbol{b}$ *и* $\boldsymbol{c}$
$$(\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c} = \boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c});$$

(3) *свойство нулевого вектора: для любого вектора* $\boldsymbol{a}$
$$\boldsymbol{a} + \mathbf{0} = \boldsymbol{a};$$

(4) *существование противоположного вектора: для любого вектора* $\boldsymbol{a}$ *существует такой вектор* $-\boldsymbol{a}$, *что*
$$\boldsymbol{a} + (-\boldsymbol{a}) = \mathbf{0};$$

(5) *свойство единицы: для любого вектора* $\boldsymbol{a}$
$$1\,\boldsymbol{a} = \boldsymbol{a};$$

(6) *ассоциативность умножения на число*: *для любого вектора* $\boldsymbol{a}$ *и любых чисел* α *и* β

$$(\alpha\beta)\,\boldsymbol{a} = \alpha\,(\beta\boldsymbol{a})\,;$$

(7) *дистрибутивность относительно сложения векторов*: *для любых векторов* $\boldsymbol{a}$ *и* $\boldsymbol{b}$ *и любого числа* α

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

(8) *дистрибутивность относительно сложения чисел*: *для любого вектора* $\boldsymbol{a}$ *и любых чисел* α *и* β

$$(\alpha + \beta)\,\boldsymbol{a} = \alpha\boldsymbol{a} + \beta\boldsymbol{a}.$$

Доказательство состоит в прямой проверке указанных соотношений с использованием определений операций сложения векторов и умножения вектора на число.

Докажем, например, коммутативность сложения (рис. 5.3(а)). Пусть $\boldsymbol{a} = \overrightarrow{OA}$, $\boldsymbol{b} = \overrightarrow{OB}$, $\boldsymbol{a} + \boldsymbol{b} = \overrightarrow{OC}$. Тогда $\boldsymbol{a} = \overrightarrow{BC}$ и поэтому

$$\boldsymbol{b} + \boldsymbol{a} = \overrightarrow{OB} + \overrightarrow{BC} = \overrightarrow{OC} = \boldsymbol{a} + \boldsymbol{b}.$$

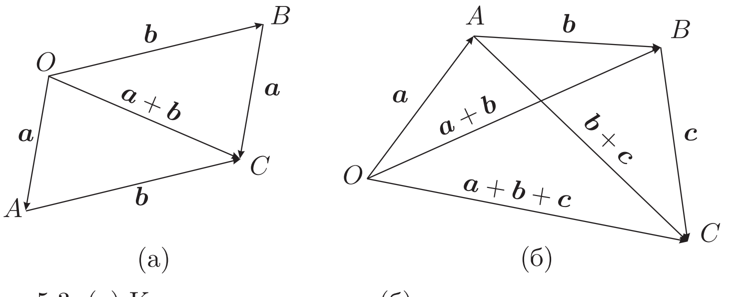


Рис. 5.3. (а) Коммутативность и (б) ассоциативность операции сложения векторов

Докажем ассоциативность сложения (см. рис. 5.3(б)). Пусть $\boldsymbol{a} = \overrightarrow{OA}$, $\boldsymbol{b} = \overrightarrow{AB}$, $\boldsymbol{c} = \overrightarrow{BC}$. Тогда

$$\boldsymbol{a} + \big(\boldsymbol{b} + \boldsymbol{c}\big) = \overrightarrow{OA} + \overrightarrow{AC} = \overrightarrow{OC}, \quad \big(\boldsymbol{a} + \boldsymbol{b}\big) + \boldsymbol{c} = \overrightarrow{OB} + \overrightarrow{BC} = \overrightarrow{OC}. \qquad \square$$

2. Определение векторного пространства

A. Определение.

5.11. Свойства линейных операций над матрицами (предложение 3.12) и над векторами (направленными отрезками, предложение 5.10) выглядят совершенно одинаково; это наводит на мысль, что матрицы и векторы являются просто различными реализациями одного и того же абстрактного объекта, а именно, множества с двумя заданными на нём операциями, которые обладают свойствами, сформулированными в указанных предложениях.

Обратим теперь точку зрения и примем свойства линейных операций над матрицами и векторами, установленные выше, за аксиомы.

5.12. Определение. Рассмотрим непустое множество $\mathcal{V}$, элементы которого будем называть *векторами* и обозначать полужирными строчными латинскими буквами $\boldsymbol{x}, \boldsymbol{y}, \ldots$. Предположим, что любым двум векторам $\boldsymbol{x} \in \mathcal{V}$ и $\boldsymbol{y} \in \mathcal{V}$ поставлен в соответствие третий вектор, обозначаемый $\boldsymbol{x} + \boldsymbol{y}$ и называемый *суммой* векторов $\boldsymbol{x}$ и $\boldsymbol{y}$. Кроме того, предположим, что любому вещественному числу $\alpha \in \mathbb{R}$ и любому вектору $\boldsymbol{x} \in \mathcal{V}$ поставлен в соответствие вектор, обозначаемый $\alpha \cdot \boldsymbol{x}$ или $\alpha\boldsymbol{x}$ и называемый произведением вектора $\boldsymbol{x}$ на число α. Множество $\mathcal{V}$ с введёнными двумя операциями называется *векторным* (или *линейным*) пространством, если эти операции обладают следующими свойствами:

ВП1: *коммутативность сложения*:
$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} \quad \boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x};$$
ВП2: *ассоциативность сложения*:
$$\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{V} \quad (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z});$$
ВП3: *существование нулевого вектора*:
$$\exists \mathbf{0} \in \mathcal{V}\ \forall \boldsymbol{x} \in \mathcal{V} \quad \mathbf{0} + \boldsymbol{x} = \boldsymbol{x};$$
ВП4: *существование противоположного вектора*:
$$\forall \boldsymbol{x} \in \mathcal{V}\ \exists {-\boldsymbol{x}} \in \mathcal{V} \quad \boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0};$$
ВП5: *свойство единицы*:
$$\forall \boldsymbol{x} \in \mathcal{V} \quad 1 \cdot \boldsymbol{x} = \boldsymbol{x};$$
ВП6: *ассоциативность умножения на число*[1]:
$$\forall \boldsymbol{x} \in \mathcal{V},\ \forall \alpha, \beta \in \mathbb{R}:\ \big(\alpha \cdot \beta\big) \cdot \boldsymbol{x} = \alpha \cdot \big(\beta \cdot \boldsymbol{x}\big);$$
ВП7: *дистрибутивность относительно сложения векторов*:
$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V},\ \forall \alpha \in \mathbb{R}:\ \alpha\big(\boldsymbol{x} + \boldsymbol{y}\big) = \alpha\boldsymbol{x} + \alpha\boldsymbol{y};$$
ВП8: *дистрибутивность относительно сложения чисел*[2]:
$$\forall \boldsymbol{x} \in \mathcal{V},\ \forall \alpha, \beta \in \mathbb{R}:\ \big(\alpha + \beta\big)\boldsymbol{x} = \alpha\boldsymbol{x} + \beta\boldsymbol{x}.$$

---

[1] Обратите внимание, что знак $\cdot$ имеет разный смысл в разных частях этой формулы: в скобках в левой части $\cdot$ обозначает умножение чисел, а во всех остальных случаях — умножение вектора на число.

[2] Здесь также знак $+$ в левой части равенства обозначает сложение чисел, а в правой части — сложение векторов.

5.13. Подчеркнём, что в этом определении не накладывается никаких ограничений ни на природу элементов множества $\mathcal{V}$, ни на конкретную реализацию операций сложения векторов и умножения вектора на число. Поэтому существуют много различных векторных пространств.

Раздел математики, изучающий векторные пространства и функции на них, называется *линейной алгеброй*. Систематический курс линейной алгебры изучается во втором семестре; сейчас же мы обсудим лишь простейшие понятия и факты линейной алгебры, необходимые для изложения геометрии на основе векторно-точечной аксиоматики Вейля. Аксиомы **ВП1–ВП8** векторного пространства составляют *первую группу аксиом Вейля.*

5.14. В определении 5.12 тот факт, что числа, на которые умножаются элементы векторного пространства, являются вещественными, несуществен: вместо вещественных чисел можно использовать, например, рациональные, комплексные или вообще числа из любого числового поля $\mathbb{K}$. В этом случае получается определение *векторного пространства над числовым полем* $\mathbb{K}$.

Б. Примеры векторных пространств.

5.15. ***Нулевое векторное пространство.*** Простейшее векторное пространство состоит из единственного нулевого вектора; будем обозначать его $\{\mathbf{0}\}$.

5.16. ***Пространства направленных отрезков.*** Направленные отрезки позволяют построить несколько различных векторных пространств: можно рассматривать векторы (направленные отрезки) на прямой (соответствующее векторное пространство будем обозначать $\mathbb{V}_1$), на плоскости ($\mathbb{V}_2$), в пространстве ($\mathbb{V}_3$).

5.17. ***Пространство матриц*** $\mathbb{K}^{m\times n}$***.*** Множество $\mathbb{K}^{m\times n}$ матриц фиксированного размера (m строк и n столбцов) с элементами из числового поля $\mathbb{K}$, а линейные операции описаны в определении 3.10 (см. с. 50), является векторным пространством. На протяжении всего курса аналитической геометрии и линейной алгебры мы будем иметь дело с этим пространством.

5.18. ***Арифметическое пространство*** $\mathbb{K}^n$***.*** Частный случай пространства матриц — пространство $\mathbb{K}^n$, элементами которого являются матрицы, состоящие из единственного столбца $(x_1, \ldots, x_n)^T$. Аналогичное пространство, элементами которого являются матрицы-строки, обозначаются $\mathbb{K}^{*n}$ (см. пп. 3.4 и 3.5, с. 47).

5.19. ***Функциональные векторные пространства.*** Пусть $\mathcal{X}$ — произвольное множество, $\mathcal{F}(\mathcal{X})$ — множество всех функций $f : \mathcal{X} \to \mathbb{R}$, определённых на $\mathcal{X}$ и принимающих значения в $\mathbb{R}$. Определим сумму $f+g$ двух функций f и g и произведение αf функции f на число α «поточечно», т.е. соотношениями

$$\forall x \in \mathcal{X} : \quad (f+g)(x) = f(x) + g(x), \quad (\alpha f)(x) = \alpha \cdot f(x).$$

Аксиомы **ВП1–ВП8** проверяются без труда. Таким образом, здесь «векторами» являются функции. Этот пример объясняет, почему в математическом анализе понятие векторного пространства играет не менее важную роль, чем в геометрии.

5.20. *Пространства многочленов.* В частности, можно рассматривать векторное пространство $\mathbb{K}[t]$, элементами которого являются многочлены от переменной t, а линейные операции заданы, как в примере 5.19. Кроме того, множества $\mathbb{K}[t]_n$ многочленов степени не выше n также являются векторными пространствами. (Объясните, почему множество многочленов степени, *равной* n, *не* является векторным пространством.)

5.21. Рассмотрим пространство, векторами которого являются положительные вещественные числа; положительное вещественное число x, рассматриваемое как вектор нашего пространства, будем обозначать $\boldsymbol{x}$. Сложение векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ (будем обозначать его знаком $\boxplus$) определим как умножение соответствующих положительных чисел:

$$\boldsymbol{x} \boxplus \boldsymbol{y} \stackrel{\text{def}}{=\!=} x \cdot y,$$

а умножение вектора $\boldsymbol{x}$ на число α (обозначаем его знаком $\boxdot$) — как возведение положительного числа x в степень α:

$$\alpha \boxdot \boldsymbol{x} \stackrel{\text{def}}{=\!=} x^{\alpha}.$$

Читателю предлагается проверить выполнение аксиом **ВП1–ВП8** самостоятельно. Этот пример показывает, что операции сложения векторов и умножения вектора на число в абстрактном векторном пространстве могут совершенно не соответствовать интуитивному представлению об этих операциях, инспирированному их названиями.

B. Простейшие следствия аксиом векторного пространства. Докажем, основываясь на аксиомах векторного пространства, несколько простых утверждений, геометрическая очевидность которых не вызывает никаких сомнений.

5.22. Предложение.

1. *Нулевой вектор* $\mathbf{0} \in \mathcal{V}$ *единствен.*
2. *Для любого вектора* $\boldsymbol{x} \in \mathcal{V}$ *противоположный вектор* $-\boldsymbol{x}$ *единствен.*
3. *Для любых двух векторов* $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{V}$ *уравнение* $\boldsymbol{a} + \boldsymbol{x} = \boldsymbol{b}$ *имеет единственное решение* $\boldsymbol{x} = \boldsymbol{b} + (-\boldsymbol{a})$, *что кратко записывается в виде* $\boldsymbol{x} = \boldsymbol{b} - \boldsymbol{a}$.
4. *Для любого вектора* $\boldsymbol{x} \in \mathcal{V}$ *имеем* $0 \cdot \boldsymbol{x} = \mathbf{0}$.
5. *Для любого вектора* $\boldsymbol{x} \in \mathcal{V}$ *его противоположный* $-\boldsymbol{x}$ *выражается формулой* $-\boldsymbol{x} = (-1) \cdot \boldsymbol{x}$.
6. *Для любого вектора* $\boldsymbol{x} \in \mathcal{V}$ *и любого числа* $\alpha \in \mathbb{K}$ *имеют место соотношения* $-(\alpha \cdot \boldsymbol{x}) = (-\alpha) \cdot \boldsymbol{x} = \alpha \cdot (-\boldsymbol{x})$.
7. *Для любого числа* $\alpha \in \mathbb{K}$ *имеем* $\alpha \cdot \mathbf{0} = \mathbf{0}$.

Доказательство.

1. Пусть $\mathbf{0}$ и $\mathbf{0}'$ — два различных нулевых вектора. Тогда $\mathbf{0}' = \mathbf{0}' + \mathbf{0} = \mathbf{0}$, т.е. $\mathbf{0}' = \mathbf{0}$.

2. Пусть $\boldsymbol{y}$ и $\boldsymbol{y}'$ — два вектора, противоположных к вектору $\boldsymbol{x}$:

$$\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}=\mathbf{0}, \quad \boldsymbol{x}+\boldsymbol{y}'=\boldsymbol{y}'+\boldsymbol{x}=\mathbf{0}.$$

В силу ассоциативности операции сложения векторов имеем

$$\boldsymbol{y}=\boldsymbol{y}+\mathbf{0}=\boldsymbol{y}+(\boldsymbol{x}+\boldsymbol{y}')=(\boldsymbol{y}+\boldsymbol{x})+\boldsymbol{y}'=\mathbf{0}+\boldsymbol{y}'=\boldsymbol{y}'.$$

3. Предположим, что вектор $\boldsymbol{x}$ является решением уравнения $\boldsymbol{a}+\boldsymbol{x}=\boldsymbol{b}$. Тогда

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

чем доказана единственность решения $\boldsymbol{x}$. Существование доказывается непосредственной проверкой:

$$\boldsymbol{a}+\big(\boldsymbol{b}+(-\boldsymbol{a})\big)=\big(\boldsymbol{a}+(-\boldsymbol{a})\big)+\boldsymbol{b}=\mathbf{0}+\boldsymbol{b}=\boldsymbol{b}.$$

4. Из цепочки равенств

$$\boldsymbol{x}+0\cdot\boldsymbol{x}\overset{\textbf{ВП5}}{=\!=}1\cdot\boldsymbol{x}+0\cdot\boldsymbol{x}\overset{\textbf{ВП8}}{=\!=}(1+0)\cdot\boldsymbol{x}=1\cdot\boldsymbol{x}\overset{\textbf{ВП5}}{=\!=}\boldsymbol{x}$$

следует, что $0\cdot\boldsymbol{x}=\mathbf{0}$.

5. Аналогично предыдущему, имеем

$$(-1)\cdot\boldsymbol{x}+\boldsymbol{x}\overset{\textbf{ВП5}}{=\!=}(-1)\cdot\boldsymbol{x}+1\cdot\boldsymbol{x}\overset{\textbf{ВП8}}{=\!=}(-1+1)\cdot\boldsymbol{x}=0\cdot\boldsymbol{x}\overset{\text{п. 4}}{=\!=}\mathbf{0},$$

т.е. $(-1)\cdot\boldsymbol{x}$ — вектор, противоположный $\boldsymbol{x}$.

6. Имеем

$$-(\alpha\boldsymbol{x})\overset{\text{п. 5}}{=\!=}(-1)\cdot(\alpha\boldsymbol{x})\overset{\textbf{ВП6}}{=\!=}(-1\cdot\alpha)\cdot\boldsymbol{x}=(-\alpha)\cdot\boldsymbol{x},$$

$$\begin{aligned}-(\alpha\boldsymbol{x})&\overset{\text{п. 5}}{=\!=}(-1)\cdot(\alpha\boldsymbol{x})\overset{\textbf{ВП6}}{=\!=}(-1\cdot\alpha)\cdot\boldsymbol{x}=\\&=\big(\alpha\cdot(-1)\big)\cdot\boldsymbol{x}\overset{\textbf{ВП6}}{=\!=}\alpha\cdot\big((-1)\cdot\boldsymbol{x}\big)\overset{\text{п. 5}}{=\!=}\alpha\cdot(-\boldsymbol{x}).\end{aligned}$$

7. Имеем

$$\alpha\cdot\boldsymbol{x}+\alpha\cdot\mathbf{0}\overset{\textbf{ВП7}}{=\!=}\alpha\cdot(\boldsymbol{x}+\mathbf{0})\overset{\textbf{ВП3}}{=\!=}\alpha\cdot\boldsymbol{x}\quad\Rightarrow\quad\alpha\cdot\mathbf{0}=\mathbf{0}.$$ □

3. Подпространства и линейные оболочки

А. Линейные комбинации.

5.23. Определение. Пусть дано семейство векторов $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, …, $\boldsymbol{x}_r\in\mathcal{V}$ и семейство чисел α_1, α_2, …, $\alpha_r\in\mathbb{K}$. *Линейной комбинацией* векторов $\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_r$ с коэффициентами $\alpha_1,\alpha_2,\ldots,\alpha_r$ называется вектор

$$\alpha_1\boldsymbol{x}_1+\alpha_2\boldsymbol{x}_2+\cdots+\alpha_r\boldsymbol{x}_r\equiv\sum_{k=1}^{r}\alpha_k\boldsymbol{x}_k; \tag{5.1}$$

про этот вектор будем говорить, что он *линейно выражается* через векторы $\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_r$.

5.24. В определении векторного пространства говорится лишь о сумме *двух* векторов. Однако определение 5.23 линейной комбинации корректно благодаря ассоциативности операции сложения векторов (см. п. IV.10 и предложение IV.11, с. 336).

5.25. Линейная комбинация называется *тривиальной*, если все её коэффициенты равны нулю:

$$\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0.$$

Ясно, что тривиальная линейная комбинация произвольных векторов равна нулевому вектору.

5.26. Мы не будем рассматривать линейные комбинации бесконечных последовательностей векторов. Для того чтобы можно было говорить о бесконечных линейных комбинациях, нужно сначала определить понятие суммы бесконечного числа слагаемых; эти вопросы выходят за рамки линейной алгебры и рассматриваются в курсе функционального анализа.

Б. Линейные оболочки.

5.27. Определение. *Линейной оболочкой* векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ называется множество всех линейных комбинаций этих векторов:

$$L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r) \overset{\text{def}}{=\!=} \Big\{ \alpha_1 \boldsymbol{x}_1 + \cdots + \alpha_r \boldsymbol{x}_r \;\Big|\; \alpha_1, \ldots, \alpha_r \in \mathbb{K} \Big\}.$$

Например, линейная оболочка двух не параллельных направленных отрезков $\boldsymbol{a}$ и $\boldsymbol{b}$ состоит из всех направленных отрезков, лежащих в одной плоскости с $\boldsymbol{a}$ и $\boldsymbol{b}$ (или параллельных этой плоскости).

5.28. Предложение (транзитивность линейной комбинации). *Пусть векторы $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p$ принадлежат линейной оболочке векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q$. Тогда*

$$L(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p) \subset L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q).$$

Иными словами, если вектор $\boldsymbol{x}$ линейно выражается через векторы $\boldsymbol{y}_1, \ldots \boldsymbol{y}_p$, каждый из которых, в свою очередь, линейно выражается через векторы $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q$, то $\boldsymbol{x}$ линейно выражается через $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q$.

Доказательство. Убедимся, что любой вектор $\boldsymbol{x}$, лежащий в линейной оболочке $L(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p)$, лежит также и в $L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q)$. Итак, пусть $\boldsymbol{x} \in L(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p)$, т.е.

$$\boldsymbol{x} = \beta_1 \boldsymbol{y}_1 + \cdots + \beta_p \boldsymbol{y}_p.$$

Поскольку $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p \in L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q)$, имеем

$$\boldsymbol{y}_1 = \alpha_{11} \boldsymbol{x}_1 + \cdots + \alpha_{1q} \boldsymbol{x}_q, \quad \ldots, \quad \boldsymbol{y}_p = \alpha_{p1} \boldsymbol{x}_1 + \cdots + \alpha_{pq} \boldsymbol{x}_q.$$

Таким образом,

$$\begin{aligned}\boldsymbol{x} &= \beta_1\boldsymbol{y}_1 + \dots + \beta_p\boldsymbol{y}_p = \\ &= \beta_1(\alpha_{11}\boldsymbol{x}_1 + \dots + \alpha_{1q}\boldsymbol{x}_q) + \dots + \beta_p(\alpha_{p1}\boldsymbol{x}_1 + \dots + \alpha_{pq}\boldsymbol{x}_q) = \\ &= (\beta_1\alpha_{11} + \dots + \beta_p\alpha_{p1})\boldsymbol{x}_1 + \dots + (\beta_1\alpha_{1q} + \dots + \beta_p\alpha_{pq})\boldsymbol{x}_q,\end{aligned}$$

а это означает, что $\boldsymbol{x} \in L(\boldsymbol{x}_1, \dots, \boldsymbol{x}_q)$, т.е.

$$L(\boldsymbol{y}_1, \dots, \boldsymbol{y}_p) \subset L(\boldsymbol{x}_1, \dots, \boldsymbol{x}_q). \qquad \square$$

В. Подпространства.

5.29. Определение. Подмножество $\mathcal{P}$ векторного пространства $\mathcal{V}$ называется *подпространством*, если для любых векторов $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{P}$ их произвольная линейная комбинация $\alpha\boldsymbol{x} + \beta\boldsymbol{y}$ также принадлежит этому множеству:

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{P},\ \forall \alpha, \beta \in \mathbb{K} \quad \alpha\boldsymbol{x} + \beta\boldsymbol{y} \in \mathcal{P}.$$

Запись $\mathcal{P} \Subset \mathcal{V}$ означает, что $\mathcal{P}$ является подпространством в $\mathcal{V}$.

Очевидно, само пространство $\mathcal{V}$ и множество $\{\mathbf{0}\}$, состоящее из одного нулевого вектора, являются подпространствами; они называются *тривиальными*.

5.30. Каждое подпространство $\mathcal{P}$ произвольного векторного пространства $\mathcal{V}$ само является векторным пространством, поскольку $\mathcal{P}$ замкнуто относительно обеих линейных операций, заданных в $\mathcal{V}$.

5.31. Примеры.

1. Для пространств направленных отрезков имеем $\mathbb{V}_1 \Subset \mathbb{V}_2 \Subset \mathbb{V}_3$.
2. Подмножество в $\mathbb{K}^n$, состоящее из всех столбцов, первый элемент которых равен нулю, образует подпространство в $\mathbb{K}^n$. Подмножество же, состоящее из столбцов, первый элемент которых равен 1, подпространством не является.
3. Подмножество в $\mathbb{K}^n$, состоящее из всех решений однородной линейной системы $AX = O$, является подпространством в $\mathbb{K}^n$ в силу принципа суперпозиции для однородных систем (теорема 4.7, с. 65).
4. Подмножество $C(\mathcal{X})$, состоящее из функций, непрерывных на открытом числовом множестве $\mathcal{X} \subset \mathbb{R}$, является подпространством в векторном пространстве $\mathcal{F}(\mathcal{X})$ всех функций, определённых на $\mathcal{X}$: $C(\mathcal{X}) \Subset \mathcal{F}(\mathcal{X})$. Аналогично, подмножество $C^1(\mathcal{X})$, состоящее из функций, имеющих непрерывную производную в любой точке $x \in \mathcal{X}$, является подпространством в $C(\mathcal{X})$: $C^1(\mathcal{X}) \Subset C(\mathcal{X}) \Subset \mathcal{F}(\mathcal{X})$.
5. Векторное пространство $\mathbb{R}[t]$, состоящее из всех многочленов от переменной t, является подпространством векторного пространства $\mathcal{F}(\mathbb{R})$ всех функций, определённых на всей вещественной оси: $\mathbb{R}[t] \Subset \mathcal{F}(\mathbb{R})$.

6. Векторное пространство $\mathbb{K}[t]_n$, состоящее из всех многочленов степени не выше n от переменной t, является подпространством векторного пространства $\mathbb{K}[t]_m$, где $m > n$; таким образом,
$$\mathbb{K} = \mathbb{K}[t]_0 \Subset \mathbb{K}[t]_1 \Subset \mathbb{K}[t]_2 \Subset \cdots \Subset \mathbb{K}[t].$$

5.32. Теорема. *Линейная оболочка $L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$ произвольных векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ векторного пространства $\mathcal{V}$ является подпространством в $\mathcal{V}$.*

Доказательство. Рассмотрим произвольные векторы $\boldsymbol{y}$ и $\boldsymbol{z}$ из указанной линейной оболочки; эти векторы линейно выражаются следующим образом:
$$\boldsymbol{y} = \sum_{k=1}^{r} \alpha_k \boldsymbol{x}_k, \quad \boldsymbol{z} = \sum_{k=1}^{r} \beta_k \boldsymbol{x}_k,$$
где α_k, β_k — некоторые числа. Рассмотрим произвольную линейную комбинацию векторов $\boldsymbol{y}$ и $\boldsymbol{z}$ и убедимся, что она также лежит в линейной оболочке $L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$:
$$\alpha \boldsymbol{y} + \beta \boldsymbol{z} = \alpha \left(\sum_{k=1}^{r} \alpha_k \boldsymbol{x}_k \right) + \beta \left(\sum_{k=1}^{r} \beta_k \boldsymbol{x}_k \right) = \sum_{k=1}^{r} (\alpha \alpha_k + \beta \beta_k) \boldsymbol{x}_k. \quad \square$$

5.33. Предложение. *Пересечение $\mathcal{P} \cap \mathcal{Q}$ двух подпространств $\mathcal{P}$ и $\mathcal{Q}$ произвольного векторного пространства $\mathcal{V}$ само является подпространством в $\mathcal{V}$.*

Доказательство. Для произвольных векторов $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{P} \cap \mathcal{Q}$ и произвольных чисел $\alpha, \beta \in \mathbb{K}$ имеем[1]:
$$\begin{aligned} &\boldsymbol{x}, \boldsymbol{y} \in \mathcal{P} \cap \mathcal{Q} \Rightarrow (\boldsymbol{x}, \boldsymbol{y} \in P) \wedge (\boldsymbol{x}, \boldsymbol{y} \in Q) \Rightarrow \\ &\qquad \Rightarrow (\alpha \boldsymbol{x} + \beta \boldsymbol{y} \in P) \wedge (\alpha \boldsymbol{x} + \beta \boldsymbol{y} \in Q) \Rightarrow \alpha \boldsymbol{x} + \beta \boldsymbol{y} \in P \cap Q. \quad \square \end{aligned}$$

Г. Способы задания подпространств в $\mathbb{K}^n$.

5.34. Фактически существуют всего два способа задать подпространство в $\mathbb{K}^n$:

(i) либо как линейную оболочку некоторого семейства столбцов (теорема 5.32),

(ii) либо как множество решений некоторой однородной системы линейных уравнений (теорема 4.7, с. 65).

[1]Знак $\wedge$ означает логическую связку «*и*» (см. определение I.14, с. 294).

5.35. Чтобы перейти от задания подпространства способом (i) к заданию способом (ii), нужно составить однородную систему уравнений, имеющую заданное множество решений; это делается с помощью алгоритма, описанного в п. 4.3.В (см. с. 87).

5.36. Чтобы перейти от задания подпространства способом (ii) к заданию способом (i), требуется решить заданную систему однородных уравнений; выражение общего решения системы является представлением произвольного вектора рассматриваемого подпространства.

4. Линейная зависимость. Размерность

А. Линейно зависимые и линейно независимые семейства векторов.

5.37. Определение. Семейство векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ называется *линейно зависимым*, если существует нетривиальная линейная комбинация этих векторов, равная нулевому вектору.

Семейство векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ называется *линейно независимым*, если равенство линейной комбинации этих векторов нулевому вектору возможно лишь в том случае, если эта линейная комбинация тривиальна.

5.38. Подчеркнём, что понятие линейной зависимости или независимости относится не к отдельным векторам, а к их семействам (упорядоченным наборам). Заметим, однако, что свойство семейства векторов быть линейно зависимым или линейно независимым не зависит от нумерации векторов в нём. Допуская некоторую вольность формулировок, часто говорят не о «линейно зависимых (независимых) семействах векторов», а о «линейно зависимых (независимых) векторах».

5.39. Примеры.

1. В векторном пространстве $\{\mathbf{0}\}$ нет линейно независимых семейств: любое семейство векторов этого пространства (состоящее, разумеется, лишь из нулевых векторов) линейно зависимо.
2. Если в векторном пространстве $\mathcal{V}$ имеется ненулевой вектор $\boldsymbol{x}$, то одноэлементное семейство $(\boldsymbol{x})$ линейно независимо, а двухэлементное семейство $(\boldsymbol{x}, -\boldsymbol{x})$ линейно зависимо.
3. На прямой $\mathbb{V}_1$ любой ненулевой направленный отрезок образует линейно независимое семейство, а любые два направленных отрезка линейно зависимы.
4. На плоскости $\mathbb{V}_2$ любые два ненулевых непараллельных направленных отрезка линейно независимы, а любые три направленных отрезка линейно зависимы.
5. В пространстве $\mathbb{V}_3$ любые три ненулевых направленных отрезка, не параллельных одной плоскости, линейно независимы, а любые четыре направленных отрезка линейно зависимы.

6. В пространстве $\mathbb{K}[t]$ всех многочленов с коэффициентами из поля $\mathbb{K}$ многочлены 1, t, t^2, …, t^n линейно независимы при любом $n \in \mathbb{N}$.

5.40. Ясно, что множество векторов, состоящее из единственного элемента, линейно независимо тогда и только тогда, когда оно состоит из ненулевого вектора. Действительно, если $\boldsymbol{a} \neq \mathbf{0}$, то равенство $\alpha\boldsymbol{a} = \mathbf{0}$ возможно лишь при $\alpha = 0$. Линейно зависимое семейство двух векторов называется *коллинеарным*, а трёх — *компланарным*; про сами векторы, образующие такие семейства, также говорят, что они коллинеарны (соответственно, компланарны).

5.41. Векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ коллинеарны тогда и только тогда, когда один из них линейно выражается через другой, т.е. эти векторы пропорциональны и, следовательно, — в наглядных геометрических терминах — параллельны одной и той же прямой.

Компланарность векторов $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ означает, что один из них может быть линейно выражен через два других, например, $\boldsymbol{c} = \alpha\boldsymbol{a} + \beta\boldsymbol{b}$; с наглядно-геометрической точки зрения это означает, что вектор $\boldsymbol{c}$ лежит в одной плоскости с векторами $\boldsymbol{a}$ и $\boldsymbol{b}$.

5.42. Предложение. *Однородная линейная система $AX = O$ имеет нетривиальное решение тогда и только тогда, когда столбцы $A_1, \ldots, A_n$ её основной матрицы A линейно зависимы.*

Доказательство. 1. Пусть однородная система имеет нетривиальное решение $x_0^1, x_0^2, \ldots, x_0^n$, т.е.

$$A_1x_0^1 + A_2x_0^2 + \cdots + A_nx_0^n = O,$$

где хотя бы один из коэффициентов x_0^k отличен от нуля; это и означает линейную зависимость столбцов $A_1, A_2, \ldots, A_n$.

2. Пусть столбцы основной матрицы системы линейно зависимы, т.е.

$$A_1x_0^1 + A_2x_0^2 + \cdots + A_nx_0^n = O,$$

где хотя бы один из коэффициентов x_0^k ненулевой. Этот набор коэффициентов и образует нетривиальное решение однородной системы. □

Б. Свойства линейно зависимых и линейно независимых семейств векторов.

5.43. Теорема.

1. *Любое семейство с повторениями*[1] *линейно зависимо.*
2. *Если в семействе векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ имеется нулевой вектор $\mathbf{0}$, то это семейство линейно зависимо.*

[1] Это означает, что среди членов семейства имеются одинаковые.

3. *Семейство векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ *линейно зависимо тогда и только тогда, когда хотя бы один из этих векторов можно представить в виде линейной комбинации остальных.*
4. *Если в семействе векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r, \boldsymbol{x}_{r+1}, \ldots, \boldsymbol{x}_s$ *имеется линейно зависимое подсемейство* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$*, то и всё семейство также линейно зависимо.*
5. *Любая часть линейно независимого семейства векторов является линейно независимым подсемейством.*
6. *Если семейство векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ *линейно независимо, а семейство векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r, \boldsymbol{x}$ *линейно зависимо, то* $\boldsymbol{x}$ *является линейной комбинацией векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$.
7. *Если семейство векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ *линейно независимо, а вектор* $\boldsymbol{x}$ *нельзя через них выразить, то семейство векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r, \boldsymbol{x}$ *также линейно независимо.*

Доказательство. 1. Пусть, например, в семействе векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ первый и второй векторы совпадают, $\boldsymbol{x}_1 = \boldsymbol{x}_2$. Тогда нетривиальная линейная комбинация

$$1 \cdot \boldsymbol{x}_1 + (-1) \cdot \boldsymbol{x}_2 + 0 \cdot \boldsymbol{x}_3 + \cdots + 0 \cdot \boldsymbol{x}_r$$

равна нулевому вектору.

2. Пусть в семействе векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ один вектор нулевой, например, $\boldsymbol{x}_r = \boldsymbol{0}$. Нетривиальная линейная комбинация

$$0 \cdot \boldsymbol{x}_1 + 0 \cdot \boldsymbol{x}_2 + \cdots + 1 \cdot \boldsymbol{x}_r$$

равна, очевидно, нулевому вектору.

3. Пусть семейство векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ линейно зависимо; тогда существует их нетривиальная линейная комбинация, равная нулевому вектору:

$$\alpha_1 \boldsymbol{x}_1 + \cdots + \alpha_r \boldsymbol{x}_r = \boldsymbol{0}.$$

Среди коэффициентов этой линейной комбинации имеются ненулевые; для определённости будем считать, что $\alpha_r \neq 0$; тогда

$$\boldsymbol{x}_r = -\frac{\alpha_1}{\alpha_r} \boldsymbol{x}_1 - \cdots - \frac{\alpha_{r-1}}{\alpha_r} \boldsymbol{x}_{r-1},$$

что и требовалось. Обратно, если

$$\boldsymbol{x}_r = \beta_1 \boldsymbol{x}_1 + \cdots + \beta_{r-1} \boldsymbol{x}_{r-1},$$

то

$$(-1)\boldsymbol{x}_r + \beta_1 \boldsymbol{x}_1 + \cdots + \beta_{r-1} \boldsymbol{x}_{r-1} = \boldsymbol{0};$$

это нетривиальная линейная комбинация, равная нулевому вектору, т.е. семейство $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ линейно зависимо.

4. Если подсемейство $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ линейно зависимо, то существует линейная комбинация

$$\alpha_1 \boldsymbol{x}_1 + \cdots + \alpha_r \boldsymbol{x}_r = \boldsymbol{0},$$

в которой имеется хотя бы один ненулевой коэффициент. Если теперь к этой линейной комбинации добавить тривиальную линейную комбинацию векторов $\boldsymbol{x}_{r+1}, \ldots, \boldsymbol{x}_s$, то получится нетривиальная линейная комбинация

$$\overbrace{\underbrace{\alpha_1 \boldsymbol{x}_1 + \cdots + \alpha_r \boldsymbol{x}_r}_{\text{нетрив. л.к.}} \;+\; \underbrace{0 \cdot \boldsymbol{x}_{r+1} + \ldots 0 \cdot \boldsymbol{x}_s}_{\text{трив. л.к.}}}^{\text{нетрив. л.к.}} = \mathbf{0},$$

что и требовалось.

Утверждение 5 непосредственно следует из предыдущего по контрапозиции (см. п. I.45, с. 301).

6. Поскольку семейство векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r, \boldsymbol{x}$ линейно зависимо, выполняется соотношение

$$\alpha_1 \boldsymbol{x}_1 + \cdots + \alpha_r \boldsymbol{x}_r + \alpha \boldsymbol{x} = \mathbf{0}. \tag{5.2}$$

Докажем, что $\alpha \neq 0$. Действительно, в противном случае получим

$$\alpha_1 \boldsymbol{x}_1 + \cdots + \alpha_r \boldsymbol{x}_r = \mathbf{0}$$

и, следовательно, $\alpha_1 = \cdots = \alpha_r = 0$ в силу линейной независимости семейства векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$. Значит, равенство (5.2) выполняется лишь при нулевых значениях всех коэффициентов, что противоречит линейной зависимости семейства векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r, \boldsymbol{x}$. Итак, в (5.2) имеем $\alpha \neq 0$, откуда

$$\boldsymbol{x} = -\frac{\alpha_1}{\alpha} \boldsymbol{x}_1 - \cdots - \frac{\alpha_r}{\alpha} \boldsymbol{x}_r.$$

Утверждение 7 непосредственно следует из предыдущего по контрапозиции. □

Следующий факт о линейной зависимости векторов менее очевиден.

5.44. Теорема. *Если векторы $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s$ принадлежат линейной оболочке $L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$ и $s > r$, то они линейно зависимы.*

Доказательство. Каждый из векторов $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s$ можно представить в виде линейной комбинации векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$:

$$\begin{aligned} \boldsymbol{y}_1 &= a_1^1 \boldsymbol{x}_1 + a_1^2 \boldsymbol{x}_2 + \cdots + a_1^r \boldsymbol{x}_r, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \boldsymbol{y}_s &= a_s^1 \boldsymbol{x}_1 + a_s^2 \boldsymbol{x}_2 + \cdots + a_s^r \boldsymbol{x}_r, \end{aligned}$$

где a^i_j — некоторые числа. Введя в рассмотрение матрицы[1]

$$\underset{1\times s}{\boldsymbol{Y}} = \left\|\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_s\right\|, \qquad \underset{1\times r}{\boldsymbol{X}} = \left\|\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r\right\|, \qquad \underset{r\times s}{A} = \begin{pmatrix} a^1_1 & a^1_2 & \ldots & a^1_s \\ a^2_1 & a^2_2 & \ldots & a^2_s \\ \vdots & \vdots & \ddots & \vdots \\ a^r_1 & a^r_2 & \ldots & a^r_s \end{pmatrix},$$

мы можем записать эти равенства в виде одного матричного соотношения

$$\boldsymbol{Y} = \boldsymbol{X}A. \tag{5.3}$$

Рассмотрим следующую систему r однородных линейных уравнений с s неизвестными:

$$\begin{cases} a^1_1 z^1 + a^1_2 z^2 + \cdots + a^1_s z^s = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a^r_1 z^1 + a^r_2 z^2 + \cdots + a^r_s z^s = 0 \end{cases} \quad \Leftrightarrow \quad AZ = O, \quad \text{где } Z = \begin{pmatrix} z^1 \\ z^2 \\ \vdots \\ z^s \end{pmatrix}.$$

Поскольку по условию $s > r$, т.е. количество неизвестных превышает количество уравнений, согласно предложению 4.38 (см. с. 77) эта однородная система имеет нетривиальное решение $Z_0 = (z^1_0, z^2_0, \ldots, z^s_0)^T$. Составим линейную комбинацию векторов $Y_1, \ldots, Y_s$ с коэффициентами $z^1_0, z^2_0, \ldots, z^s_0$:

$$z^1_0 Y_1 + z^2_0 Y_2 + \cdots + z^s_0 Y_s \equiv \boldsymbol{Y} Z_0.$$

В силу (5.3) имеем

$$\boldsymbol{Y} Z_0 = (\boldsymbol{X}A)Z_0 = \boldsymbol{X}(AZ_0) = \boldsymbol{X}O = \mathbf{0},$$

что и доказывает линейную зависимость векторов $Y_1, \ldots, Y_s$. □

Контрапозицией теоремы 5.44 является следующая теорема.

5.45. Теорема. *Если векторы $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s$, принадлежащие линейной оболочке $L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$, линейно независимы, то $s \leqslant r$.*

Доказательство. Теорема 5.45 допускает прямое доказательство, не опирающееся на теорию линейных систем.

Применим метод математической индукции по числу r. При $r = 1$ утверждение очевидно: если векторы $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s$ принадлежат линейной оболочке $L(\boldsymbol{x}_1)$, то они пропорциональны вектору $\boldsymbol{x}_1$, а следовательно и между собой, и потому линейно независимое семейство могут образовывать только если их не более одного, т.е. $s = 1$.

---

[1] Обратите внимание, что коэффициенты разложения вектора $\boldsymbol{y}_1$ располагаются в первом столбце, а не в первой строке матрицы A (для остальных векторов аналогично).

Предположим, что утверждение теоремы верно для $r-1$ векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{r-1}$, и докажем его справедливость для случая r векторов. Итак, пусть в линейной оболочке $L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$ имеются линейно независимые векторы

$$\boldsymbol{y}_1 = a_1^1\boldsymbol{x}_1 + \cdots + a_1^r\boldsymbol{x}_r, \quad \ldots, \quad \boldsymbol{y}_s = a_s^1\boldsymbol{x}_1 + \cdots + a_s^r\boldsymbol{x}_r. \tag{5.4}$$

Требуется доказать, что $s \leqslant r$.

Если все коэффициенты при $\boldsymbol{x}_r$ равны нулю, то утверждение доказано, поскольку в этом случае, по предположению индукции, имеем $s \leqslant r-1$, так что подавно $s \leqslant r$.

Предположим, что хотя бы один из коэффициентов при $\boldsymbol{x}_r$ отличен от нуля, например, a_s^r. Выразим из последнего равенства (5.4) вектор $\boldsymbol{x}_r$:

$$\boldsymbol{x}_r = \frac{1}{a_s^r}\boldsymbol{y}_s - \frac{a_s^1}{a_s^r}\boldsymbol{x}_1 - \cdots - \frac{a_s^{r-1}}{a_s^r}\boldsymbol{x}_{r-1}.$$

Подставляя полученное выражение в каждое из предыдущих равенств (5.4), приводя подобные слагаемые и обозначая образующиеся коэффициенты буквой b с индексами, получаем следующие равенства:

$$\boldsymbol{y}_1 - \frac{a_1^r}{a_s^r}\boldsymbol{y}_s = b_1^1\boldsymbol{x}_1 + \cdots + b_1^{r-1}\boldsymbol{x}_{r-1}, \quad \ldots, \quad \boldsymbol{y}_{s-1} - \frac{a_{s-1}^r}{a_s^r}\boldsymbol{y}_s = b_{s-1}^1\boldsymbol{x}_1 + \cdots + b_{s-1}^{r-1}\boldsymbol{x}_{r-1}.$$

Эти равенства означают, что векторы

$$\boldsymbol{y}_1' = \boldsymbol{y}_1 - \frac{a_1^r}{a_s^r}\boldsymbol{y}_s, \quad \ldots, \quad \boldsymbol{y}_{s-1}' = \boldsymbol{y}_{s-1} - \frac{a_{s-1}^r}{a_s^r}\boldsymbol{y}_s$$

принадлежат линейной оболочке $L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{r-1})$. Если мы установим, что они линейно независимы, отсюда по предположению индукции будет следовать, что $s-1 \leqslant r-1$, т.е. $s \leqslant r$.

Докажем линейную независимость векторов $\boldsymbol{y}_1', \ldots, \boldsymbol{y}_{s-1}'$. Рассмотрим равенство

$$\lambda^1\boldsymbol{y}_1' + \cdots + \lambda^{s-1}\boldsymbol{y}_{s-1}' = \mathbf{0},$$

т.е.

$$\lambda^1\left(\boldsymbol{y}_1 - \frac{a_1^r}{a_s^r}\boldsymbol{y}_s\right) + \cdots + \lambda^{s-1}\left(\boldsymbol{y}_{s-1} - \frac{a_{s-1}^r}{a_s^r}\boldsymbol{y}_s\right) = \mathbf{0}.$$

Раскрывая скобки, получим

$$\lambda^1\boldsymbol{y}_1 + \cdots + \lambda^{s-1}\boldsymbol{y}_{s-1} - \left(\lambda^1\frac{a_1^r}{a_s^r} + \cdots + \lambda^{s-1}\frac{a_{s-1}^r}{a_s^r}\right)\boldsymbol{y}_s = \mathbf{0}.$$

В силу линейной независимости векторов $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s$ все коэффициенты в этом равенстве равны нулю; в частности, $\lambda^1 = \cdots = \lambda^{s-1} = 0$, так что векторы $\boldsymbol{y}_1', \ldots, \boldsymbol{y}_{s-1}'$ линейно независимы. □

B. Сохранение линейной зависимости и независимости при элементарных преобразованиях строк матрицы.

5.46. Теорема. *Пусть матрица A' получена из матрицы A элементарными преобразованиями строк.*

1. *Если столбцы матрицы A линейно независимы, то столбцы матрицы A' также линейно независимы.*

2. *Если между столбцами матрицы A имеется линейная зависимость*
$$\alpha_1 A_1 + \alpha_2 A_2 + \cdots + \alpha_n A_n = O,$$
то столбцы матрицы A' связаны такой же линейной зависимостью:
$$\alpha_1 A'_1 + \alpha_2 A'_2 + \cdots + \alpha_n A'_n = O.$$

Доказательство. Обозначив через C матрицу выполненных элементарных преобразований, можем записать $A'_j = CA_j$. Отметим, что матрица C обратима (см. предложение 4.43, с. 82).

1. Пусть столбцы $A_1, \ldots, A_n$ матрицы A линейно независимы. Предположим, что столбцы преобразованной матрицы A', напротив, линейно зависимы, т.е. существуют такие коэффициенты $\alpha_1, \ldots, \alpha_n$, среди которых имеется хотя бы один ненулевой, что
$$\alpha_1 A'_1 + \alpha_2 A'_2 + \cdots + \alpha_n A'_n = O.$$
Заменяя A'_j на CA_j и вынося за скобки матрицу C, получаем
$$\alpha_1 CA_1 + \alpha_2 CA_2 + \cdots + \alpha_n CA_n = O$$
т.е.
$$C\Big(\alpha_1 A_1 + \alpha_2 A_2 + \cdots + \alpha_n A_n\Big) = O.$$
Умножая обе части равенства слева на C^{-1}, находим
$$\alpha_1 A_1 + \alpha_2 A_2 + \cdots + \alpha_n A_n = O,$$
что означает линейную зависимость столбцов матрицы A, противоречие.

2. Умножая слева на C обе части равенства
$$\alpha_1 A_1 + \alpha_2 A_2 + \cdots + \alpha_n A_n = O$$
и замечая, что $CA_j = A'_j$, получаем
$$\alpha_1 A'_1 + \alpha_2 A'_2 + \cdots + \alpha_n A'_n = O. \qquad \square$$

5. Размерность и базис векторного пространства

A. Размерность векторного пространства.

5.47. Определение. Если для любого $n \in \mathbb{N}$ в векторном пространстве $\mathcal{V}$ найдётся линейно независимое семейство, состоящее более чем из n векторов, то пространство $\mathcal{V}$ называется *бесконечномерным.*

Пример 5.39.6 показывает, что пространство $\mathbb{K}[t]$ бесконечномерно.

5.48. С использованием кванторов определение 5.47 записывается следующим образом: векторное пространство $\mathcal{V}$ называется бесконечномерным, если

$$\forall n \in \mathbb{N}\ \exists k,\ k > n,\ \exists \boldsymbol{x}_1, \dots, \boldsymbol{x}_k \text{ — линейно независимые.}$$

Сформулируем отрицание этого определения (см. п. I.30, с. 298):

$$\exists n \in \mathbb{N}\ \forall k,\ k > n,\ \forall \boldsymbol{x}_1, \dots, \boldsymbol{x}_k \text{ — линейно зависимые.} \tag{5.5}$$

5.49. Определение. Векторное пространство $\mathcal{V}$ называется *конечномерным*, а именно, *n-мерным*, если

(i) в $\mathcal{V}$ существует линейно независимое семейство, состоящее из n векторов;
(ii) любое семейство в $\mathcal{V}$, состоящее более чем из n векторов, линейно зависимо.

Число n называется *размерностью*[1] пространства $\mathcal{V}$ и обозначается $\dim \mathcal{V}$.

5.50. Пример 5.39.1 означает, что $\dim\{\mathbf{0}\} = 0$. Примеры 5.39.3–5 показывают, что $\dim \mathbb{V}_1 = 1$, $\dim \mathbb{V}_2 = 2$ и $\dim \mathbb{V}_3 = 3$.

5.51. Геометрически очевидные факты $\dim \mathbb{V}_1 = 1$, $\dim \mathbb{V}_2 = 2$ и $\dim \mathbb{V}_3 = 3$ легко доказываются в рамках системы аксиом Гильберта. Однако в аксиоматике Вейля эти утверждения не являются следствиями аксиом **ВП1–ВП8** векторного пространства. Поэтому для построения геометрии на базе систем аксиом Вейля мы принимаем их за аксиомы, называемые *аксиомами размерности*:

Р1: Размерность векторного пространства, используемого в геометрии прямой, равна 1: $\dim \mathbb{V}_1 = 1$.
Р2: Размерность векторного пространства, используемого в планиметрии, равна 2: $\dim \mathbb{V}_2 = 2$.
Р3: Размерность векторного пространства, используемого в стереометрии, равна 3: $\dim \mathbb{V}_3 = 3$.

Условимся причислять аксиомы размерности к первой группе аксиом Вейля (см. с. 96).

Нас в первую очередь интересуют именно конечномерные векторные пространства. Немногочисленные примеры бесконечномерных пространств всегда будут специально оговариваться.

Б. Базис и координаты вектора. Пусть $\mathcal{V}$ — конечномерное векторное пространство.

[1] Обратите внимание, что это наименьшее из чисел, обладающих свойством (5.5).

5.52. Определение. *Базисом* векторного пространства $\mathcal{V}$ называется линейно независимое семейство $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$ элементов этого пространства, через которые может быть линейно выражен любой вектор из $\mathcal{V}$.

5.53. Подчеркнём, что базис — это не множество, а именно семейство, т.е. *упорядоченный набор векторов.* Переставив местами векторы некоторого базиса, мы снова получим базис, но другой.

5.54. Предложение. *Все базисы конечномерного векторного пространства $\mathcal{V}$ состоят из одинакового числа векторов. Это число равно размерности* $\dim \mathcal{V}$ *векторного пространства $\mathcal{V}$.*

Доказательство. Рассмотрим в векторном пространстве $\mathcal{V}$ два различных базиса $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$ и $\boldsymbol{F} = \big(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m\big)$. Поскольку $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m \in L(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$, согласно теореме 5.45 имеем $m \leqslant n$. Аналогично получаем $n \leqslant m$, откуда $n = m$.

Поскольку в $\mathcal{V}$ имеются линейно независимые семейства, состоящие из n векторов (например, базисы), а любое семейство, состоящее из $n+1$ векторов, линейно зависимо по теореме 5.44, число n и является размерностью этого пространства согласно определению 5.49. □

5.55. Ясно, что если $\big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$ — базис в векторном пространстве $\mathcal{V}$, то $\mathcal{V} = L(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$.

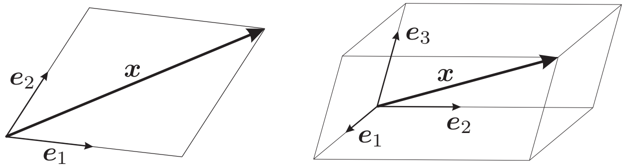
5.56. Определение. Пусть $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$ — произвольный базис n-мерного векторного пространства $\mathcal{V}$. Тогда для любого вектора $\boldsymbol{x} \in \mathcal{V}$ существуют такие числа $x^1, \ldots, x^n$, что

$$\boldsymbol{x} = x^1 \boldsymbol{e}_1 + \cdots + x^n \boldsymbol{e}_n \equiv \sum_{k=1}^{n} x^k \boldsymbol{e}_k. \tag{5.6}$$

Числа $x^1, \ldots, x^n$ называются *координатами* вектора $\boldsymbol{x}$ в базисе $\boldsymbol{E}$, а формула (5.6) — *разложением вектора* $\boldsymbol{x}$ по базису $\boldsymbol{E}$.

5.57. Пусть $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$ — строка из базисных векторов, а $X = \big(x^1, \ldots, x^n\big)^T$ — столбец координат вектора $\boldsymbol{x}$ в этом базисе (обратите внимание, что нумерация векторов базиса (нижние индексы) и координат вектора (верхние индексы) согласована с правилом нумерации элементов матриц, см. с. 46). Тогда формула (5.6) может быть записана в векторном виде:

$$\boldsymbol{x} = \boldsymbol{E} X. \tag{5.7}$$



(а) (б)

Рис. 5.4. Разложение вектора по базису
(а) на плоскости; (б) в пространстве

5.58. Предложение. *Разложение вектора по базису единственно, т.е. набор координат векторов в данном базисе определен однозначно.*

Доказательство. Предположим, что в базисе $\big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$ вектор $\boldsymbol{x}$ имеет два различных набора координат:

$$\boldsymbol{x} = x^1\boldsymbol{e}_1 + \cdots + x^n\boldsymbol{e}_n = y^1\boldsymbol{e}_1 + \cdots + y^n\boldsymbol{e}_n.$$

Вычитая второе разложение из первого, получим

$$\mathbf{0} = (x^1 - y^1)\boldsymbol{e}_1 + \cdots + (x^n - y^n)\boldsymbol{e}_n.$$

Так как векторы базиса линейно независимы, то это равенство возможно лишь при нулевых коэффициентах, т.е.

$$x^1 = y^1, \quad \ldots, \quad x^n = y^n. \qquad \square$$

5.59. Пусть $\boldsymbol{x}$ и $\boldsymbol{y}$ — два вектора, а $x^1, \ldots, x^n$ и $y^1, \ldots, y^n$ — их координаты в некотором базисе $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$. Тогда

$$\boldsymbol{x} = x^1\boldsymbol{e}_1 + \cdots + x^n\boldsymbol{e}_n, \quad \boldsymbol{y} = y^1\boldsymbol{e}_1 + \cdots + y^n\boldsymbol{e}_n.$$

Поэтому

$$\boldsymbol{x} + \boldsymbol{y} = x^1\boldsymbol{e}_1 + \cdots + x^n\boldsymbol{e}_n + y^1\boldsymbol{e}_1 + \cdots + y^n\boldsymbol{e}_n = \\ = (x^1 + y^1)\boldsymbol{e}_1 + \cdots + (x^n + y^n)\boldsymbol{e}_n,$$

т.е. *при сложении векторов их координаты складываются.* Аналогично,

$$\alpha\boldsymbol{x} = \alpha(x^1\boldsymbol{e}_1 + \cdots + x^n\boldsymbol{e}_n) = (\alpha x^1)\boldsymbol{e}_1 + \cdots + (\alpha x^n)\boldsymbol{e}_n,$$

т.е. *при умножении вектора на число его координаты умножаются на это число.*

5.60. Предложение (теорема о монотонности размерности)**.**

1. *Пусть $\mathcal{P}$ и $\mathcal{Q}$ — два подпространства в $\mathcal{V}$, причём $\mathcal{Q} \subset \mathcal{P}$. Тогда $\dim\mathcal{Q} \leqslant \dim\mathcal{P}$.*

2. *Пусть* $\mathcal{P}$ *и* $\mathcal{Q}$ *— два подпространства в* $\mathcal{V}$*, причём* $\mathcal{Q} \subset \mathcal{P}$ *и* $\dim \mathcal{P} = \dim \mathcal{Q}$*. Тогда* $\mathcal{P} = \mathcal{Q}$*.*

Доказательство. 1. Пусть $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ — базис в $\mathcal{P}$, т.е. $\dim \mathcal{P} = r$, а $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s$ — базис в $\mathcal{Q}$, т.е. $\dim \mathcal{Q} = s$. По теореме 5.45 имеем $s \leqslant r$, что и требовалось.

2. Пусть $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s$ — базис в $\mathcal{Q}$, т.е. $\dim \mathcal{Q} = s$. Предположим, что $\mathcal{Q} \subset \mathcal{P}$, но $\mathcal{Q} \neq \mathcal{P}$. Тогда существует такой вектор $\boldsymbol{z} \in \mathcal{P}$, что $\boldsymbol{z} \notin \mathcal{Q}$. Поскольку $\boldsymbol{z}$ не может быть представлен в виде линейной комбинации векторов $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s$, делаем вывод, что семейство $\boldsymbol{z}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_s$ линейно независимо (см. утверждение 7 теоремы 5.43, с. 103). Таким образом, $\mathcal{P}$ содержит не менее $s+1$ линейно независимых векторов, т.е. $\dim \mathcal{P} \geqslant s+1$. Получили противоречие с условием $\dim \mathcal{P} = \dim \mathcal{Q}$. □

B. Правило суммирования Эйнштейна. Векторы базиса нумеруются нижними индексами, а координаты вектора верхними именно для того, чтобы для записи сумм однотипных слагаемых вида (5.6) можно было использовать компактные матричные обозначения типа (5.7). Легко видеть, что выбранный способ нумерации полностью согласуется с правилом нумерации элементов матриц: верхний индекс нумерует строку, а нижний — столбец.

5.61. В алгебре используется и другой способ записи сумм вида (5.6), предложенный А. Эйнштейном и состоящий в следующем. Пусть имеется выражение (одночлен), представляющее собой произведение некоторых величин, снабжённых верхними и нижними индексами. Если некоторый индекс встречается в этом выражении дважды, причём один раз сверху и один раз снизу, то это выражение обозначает *сумму* одночленов, в которой повторяющийся индекс принимает некоторый (заранее оговорённый) диапазон значений, хотя знак суммирования $\sum$ отсутствует.

5.62. Так, выражение $a_k b^k$ означает сумму $a_1 b^1 + a_2 b^2 + \cdots + a_n b^n$ (должно быть ясно из контекста или оговорено в явном виде, что индекс суммирования k пробегает значения от 1 до n). Выражение для элементов произведения матриц

$$\{AB\}^k_j = \sum_{l=1}^{s} a^k_l\, b^l_j$$

также можно записать в обозначениях Эйнштейна:

$$\{AB\}^k_j = a^k_l\, b^l_j, \quad l = 1, \ldots, s.$$

Выражение $a_k b^k c^k$ *не* обозначает сумму, поскольку индекс суммирования встречается в нём не дважды, а трижды. Однако в выражениях вида $(a_k + b_k)c^k$, напоминающих по форме свойство дистрибутивности умножения относительно сложения, предполагается суммирование:

$$\begin{aligned}(a_k + b_k)c^k &= (a_1 + b_1)c^1 + (a_2 + b_2)c^2 + \dots + (a_n + b_n)c^n = \\ &= a_1 c^1 + a_2 c^2 + \dots + a_n c^n + b_1 c^1 + b_2 c^2 + \dots + b_n c^n = \\ &= a_k c^k + b_k c^k.\end{aligned}$$

Читатель быстро привыкнет к этим обозначениям и оценит их компактность и удобство; сомнительные же и трудные случаи мы всегда будем особо оговаривать.

5.63. Напомним (см. формулу 3.1, с. 53), что *символ Кронекера* — это обозначение для элементов единичной матрицы:

$$\delta_k^j = \begin{cases} 1, & \text{если} \quad j = k, \\ 0, & \text{если} \quad j \neq k. \end{cases}$$

Часто встречаются суммы вида $a_j \delta_k^j$, $b^k \delta_k^j$ и т. п. В развёрнутом виде первая из этих сумм имеет вид

$$a_1 \delta_k^1 + a_2 \delta_k^2 + \dots + a_k \delta_k^k + \dots + a_n \delta_k^n.$$

Из n слагаемых в этой сумме отлично от нуля лишь одно, а именно k-е, поэтому вся сумма равна a_k. Таким образом,

$$a_j \delta_k^j = a_k,$$

т.е. *суммирование с символом Кронекера сводится просто к замене индекса.*

Г. Примеры.

5.64. Нулевое векторное пространство $\{\mathbf{0}\}$ не имеет базиса.

5.65. В качестве базиса пространства $\mathbb{V}_1$ может быть взят произвольный ненулевой вектор из $\mathbb{V}_1$. Аналогично, базисами в $\mathbb{V}_2$ и $\mathbb{V}_3$ могут служить произвольные упорядоченные пары (тройки) неколлинеарных (некомпланарных) векторов.

5.66. ***Базис арифметического пространства*** $\mathbb{K}^n$**.** Легко проверить, что векторы (столбцы)

$$\boldsymbol{e}_1 = \big(1, 0, \dots, 0\big)^T, \quad \dots, \quad \boldsymbol{e}_n = \big(0, 0, \dots, 1\big)^T \tag{5.8}$$

линейно независимы[1], а любой другой столбец является их линейной комбинацией:

$$\boldsymbol{x} = \left(x^1,\ x^2, \ldots, x^n\right)^T = x^1\boldsymbol{e}_1 + x^2\boldsymbol{e}_2 + \cdots + x^n\boldsymbol{e}_n.$$

Поэтому столбцы (5.8) образуют базис в $\mathbb{K}^n$, называемый *стандартным*, а $\dim \mathbb{K}^n = n$.

Отметим, что компоненты (элементы) столбца $\left(x^1, \ldots, x^n\right)^T$ являются координатами этого столбца в стандартном базисе пространства $\mathbb{K}$; в другом базисе координаты будут другие. Таким образом, нужно различать термины «компоненты столбца» и «координаты столбца»: первый термин имеет инвариантный (т.е. не зависящий от выбора базиса) смысл, а второй привязан к базису.

5.67. ***Фундаментальное семейство решений*** $X_1, \ldots, X_s$ однородной линейной системы $AX = O$ является базисом в множестве всех её решений, которое, напомним (см. пример 5.31.3), представляет собой подпространство пространства столбцов.

5.68. ***Пространство комплексных столбцов*** $\mathbb{C}^n$ ***над полем*** $\mathbb{R}$**.** Множество $\mathbb{C}^n$ можно рассматривать как векторное пространство над полем вещественных чисел; это означает, что компонентами столбцов являются комплексные числа, но коэффициентами при образовании линейных комбинаций (в том числе и координатами векторов) могут служить лишь вещественные числа. Будем обозначать это векторное пространство $\mathbb{C}^n(\mathbb{R})$. Стандартный базис этого пространства состоит из векторов

$$\begin{aligned} \boldsymbol{e}_1 &= \left(1, 0, \ldots, 0\right)^T, \qquad \ldots, \qquad \boldsymbol{e}_n = \left(0, 0, \ldots, 1\right)^T, \\ \boldsymbol{e}_{n+1} &= \left(i, 0, \ldots, 0\right)^T, \qquad \ldots, \qquad \boldsymbol{e}_{2n} = \left(0, 0, \ldots, i\right)^T. \end{aligned}$$

Действительно, любой вектор-столбец с комплексными компонентами может быть представлен в виде следующей вещественной линейной комбинации столбцов $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{2n}$:

$$\begin{pmatrix} x^1 + iy^1 \\ x^2 + iy^2 \\ \vdots \\ x^n + iy^n \end{pmatrix} = x^1\boldsymbol{e}_1 + \cdots + x^n\boldsymbol{e}_n + y^1\boldsymbol{e}_{n+1} + \cdots + y^n\boldsymbol{e}_{2n}.$$

Отсюда вытекает, что $\dim \mathbb{C}^n(\mathbb{R}) = 2n$.

5.69. ***Базис пространства многочленов*** $\mathbb{K}[t]_n$**.** Докажем, что в пространстве $\mathbb{K}[t]_n$ многочленов степени не выше n многочлены

$$\boldsymbol{e}_0 = 1, \quad \boldsymbol{e}_1 = t, \quad \boldsymbol{e}_2 = t^2, \quad \ldots, \quad \boldsymbol{e}_n = t^n \tag{5.9}$$

---

[1]Действительно, равенство линейной комбинации этих столбцов с коэффициентами $\alpha_1, \ldots, \alpha_n$

$$\alpha_1\boldsymbol{e}_1 + \cdots + \alpha_n\boldsymbol{e}_n \equiv \left(\alpha_1, \ldots, \alpha_n\right)^T$$

нулевому столбцу возможно лишь в случае, когда $\alpha_1 = \cdots = \alpha_n = 0$.

линейно независимы. Действительно, приравняем нулю их линейную комбинацию:

$$\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_n t^n = 0.$$

Предположив, что хотя бы один коэффициент α_k отличен от нуля, получаем уравнение относительно t степени не выше n, которое имеет лишь конечное число корней (см. предложение 2.65, с. 38), т.е. указанное равенство не может выполняться тождественно для всех t.

Далее, поскольку произвольный многочлен

$$x(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n$$

можно представить в виде

$$x(t) = a_0 \boldsymbol{e}_0 + a_1 \boldsymbol{e}_1 + a_2 \boldsymbol{e}_2 + \dots + a_n \boldsymbol{e}_n,$$

видим, что векторы (многочлены) (5.9) образуют базис в векторном пространстве $\mathbb{K}[t]_n$ и $\dim \mathbb{K}[t]_n = n + 1$; этот базис назовём *стандартным*.

6. Изоморфизм

Свойства координат, установленные в п. 5.59, могут быть обобщены следующим образом.

5.70. Определение. Пусть $\mathcal{V}$ и $\mathcal{W}$ — два векторных пространства над одним и тем же числовым полем $\mathbb{K}$. Взаимно однозначное отображение $g\colon \mathcal{V} \to \mathcal{W}$ называется *изоморфизмом*, если оно обладает следующим свойством (свойством линейности):

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V},\ \forall \alpha, \beta \in \mathbb{K}: \quad g(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha g(\boldsymbol{x}) + \beta g(\boldsymbol{y}).$$

Два векторных пространства $\mathcal{V}$ и $\mathcal{W}$, для которых существует хотя бы один изоморфизм $g\colon \mathcal{V} \to \mathcal{W}$, называются *изоморфными* (обозначение $\mathcal{V} \approx \mathcal{W}$).

5.71. В аксиоматической теории векторных пространств мы интересуемся только теми их свойствами, которые выражаются в терминах операций сложения векторов и умножения векторов на числа. У изоморфных пространств такие свойства одинаковы, поэтому в рамках аксиоматической теории изоморфные пространства неразличимы (хотя конкретные реализации изоморфных пространств могут быть совершенно различными). Таким образом, утверждения, справедливые для какого-либо векторного пространства, имеют место и во всех изоморфных ему пространств. На этом обстоятельстве и основан метод координат, лежащий в основе аналитической геометрии.

5.72. Координаты $x^1, \dots, x^n$ вектора $\boldsymbol{x} \in \mathcal{V}$ в (заранее зафиксированном) базисе $\boldsymbol{E} = \big(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\big)$ составляют упорядоченную последовательность $(x^1, \dots, x^n)$ чисел, т.е. элемент арифметического пространства $\mathbb{K}^n$ (напомним, что элементами пространства $\mathbb{K}^n$ мы договорились считать столбцы). Поэтому формула

$$g(\boldsymbol{x}) = \left(x^1, \ldots, x^n\right)^T \tag{5.10}$$

определяет взаимно однозначное отображение $g\colon \mathcal{V} \to \mathbb{K}^n$. Установленные выше (см. п. 5.59) свойства координат означают, что это отображение является изоморфизмом.

5.73. Определение. Изоморфизм, заданный формулой (5.10), называется *координатным изоморфизмом*, определяемым базисом $\left(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\right)$.

Из факта существования координатных изоморфизмов вытекает следующее утверждение.

5.74. Теорема.

1. *Каждое n-мерное векторное пространство $\mathcal{V}$ над числовым полем $\mathbb{K}$ изоморфно пространству $\mathbb{K}^n$.*
2. *Все векторные пространства над одним и тем же числовым полем, имеющие одинаковую размерность, изоморфны между собой.*

5.75. Легко проверить, что отношение изоморфности векторных пространств является *отношением эквивалентности* (см. с. 308). Таким образом, класс всех векторных пространств над одним и тем же числовым полем разбивается на непересекающиеся классы эквивалентности, каждый из которых характеризуется натуральным числом — размерностью.

Различных векторных пространств одной размерности существует необозримо много, но классов изоморфных пространств имеется лишь счётное[1] множество, причём для каждого $n \in \mathbb{N}_0$ имеется единственный класс, состоящий из всех n-мерных пространств.

5.76. Изоморфизм $g\colon \mathcal{V} \to \mathcal{W}$ векторных пространств можно задать следующим образом. Зафиксируем базис $\boldsymbol{E} = \left(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\right)$ пространства $\mathcal{V}$ и базис $\boldsymbol{F} = \left(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n\right)$ пространства $\mathcal{W}$ и рассмотрим линейное отображение $g\colon \mathcal{V} \to \mathcal{W}$, заданное на базисных векторах формулой $g(\boldsymbol{e}_k) = \boldsymbol{f}_k$, $k = 1, \ldots, n$. Ясно, что это линейное отображение взаимно однозначно, т.е. является изоморфизмом, переводящим вектор $\boldsymbol{x} \in \mathcal{V}$ в вектор $\boldsymbol{y} \in \mathcal{W}$, имеющий в базисе $\boldsymbol{F}$ те же координаты, что и вектор $\boldsymbol{x}$ в базисе $\boldsymbol{E}$. Поэтому говорят, что изоморфизм g устанавливается *по равенству координат.*

[1] Множество X называется *счётным*, если существует взаимно однозначное отображение этого множества на множество натуральных чисел; см. определение III.44, с. 328.

ГЛАВА 6

Определители

ОПРЕДЕЛИТЕЛЬ, или детерминант, — одно из важнейших понятий алгебры, имеющее разнообразные применения во всех разделах математики. Идеи, приводящие к понятию определителя, возникли в конце XVII в. в работах Г. Ф. Лейбница[1] в связи с исследованием систем линейных уравнений, однако само понятие определителя, его свойства и методы вычисления появились лишь в середине XVIII в. в работах Г. Крамера[2]. Методы Крамера сразу же получили широкое распространение и быстро нашли многочисленные приложения как в самой математике, так и в родственных науках — астрономии и механике. Термин «определитель» («детерминант») был введён К. Ф. Гауссом в XIX в.

1. Определители второго порядка

A. Система двух уравнений с двумя неизвестными.

6.1. Рассмотрим систему уравнений

$$\begin{cases} ax + by = p, \\ cx + dy = q, \end{cases} \tag{6.1}$$

где a, b, c, d, p, q — заданные числа, x, y — неизвестные. Решим систему методом исключения неизвестных.

Умножая первое уравнение на d, второе на $-b$ и складывая полученные уравнения, получим

$$(ad - bc)x = pd - qb.$$

Аналогично, умножая первое уравнение на $-c$, второе на a и складывая полученные уравнения, получим

$$(ad - bc)y = qa - pc.$$

[1] Г. Ф. Лейбниц (G. W. Leibniz) — немецкий математик, физик, логик, философ, лингвист, юрист, историк, дипломат и изобретатель (1646–1716).

[2] Г. Крамер (G. Cramer) — швейцарский математик (1704–1752)

Если $ad - bc \neq 0$, то получаем
$$x = \frac{pd - qb}{ad - bc}, \quad y = \frac{qa - pc}{ad - bc}.$$
Прямая проверка показывает, что эти числа действительно представляют собой решение системы (6.1) в случае, когда $ad - bc \neq 0$. Кроме того, ясно, что это решение единственно[1].

6.2. Для облегчения запоминания этих формул запишем коэффициенты системы (6.1) в виде матрицы
$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$
называемой *основной матрицей* системы. Поставим в соответствие этой матрице число $ad - bc$; оно называется *определителем* (или *детерминантом*) матрицы A и обозначается одним из следующих способов:
$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \det A = |A| = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \stackrel{\text{def}}{=\!=} ad - bc.$$
Такой определитель называют *определителем второго порядка* (по количеству его строк и столбцов). Это понятие уже встречалось нам при обращении (2×2)-матриц (см. с. 57).

6.3. Часто используются термины «элемент (строка, столбец) определителя»; они означают соответственно элемент (строку, столбец) матрицы A, для которой рассматривается определитель.

6.4. С помощью определителей формулы для решения системы могут быть записаны в виде
$$x = \frac{\begin{vmatrix} p & b \\ q & d \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}} = \frac{\det A_x}{\det A}, \qquad y = \frac{\begin{vmatrix} a & p \\ c & q \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}} = \frac{\det A_y}{\det A}, \tag{6.2}$$
где A_x (соответственно, A_y) — матрица, полученная из A заменой первого (соответственно, второго) столбца на столбец, состоящий из свободных членов уравнений. Полученные формулы называются *формулами Крамера*.

Итак, мы доказали следующее утверждение.

[1] Действительно, поскольку каждое из уравнений $(ad - bc)y = qa - pc$ и $(ad-bc)y = qa-pc$ является следствием системы (6.1), любое решение исходной системы является одновременно и решением полученных уравнений.

6.5. Предложение. *Если определитель* $\det A$ *основной матрицы системы* (6.1) *отличен от нуля, то существует единственное решение этой системы, определяемое формулами Крамера* (6.2).

6.6. Если $\det A = 0$, то система (6.1) может либо не иметь решений, либо иметь их бесконечно много, как показывают следующие примеры:

$$\begin{cases} x + y = 1, \\ x + y = 2 \end{cases} \quad \det A = 0, \quad \begin{cases} \det A_x = -1, \\ \det A_y = 1, \end{cases} \quad \text{(решений нет);}$$

$$\begin{cases} 2x + 2y = 2, \\ 3x + 3y = 3 \end{cases} \quad \det A = 0, \quad \begin{cases} \det A_x = 0, \\ \det A_y = 0, \end{cases} \quad \text{(решений бесконечно много).}$$

Пример несовместной системы

$$\begin{cases} 0x + 0y = 1, \\ 0x + 0y = 2 \end{cases}$$

рассеивает распространённое заблуждение (содержащееся даже и в учебниках!), будто в случае $\det A = \det A_x = \det A_y = 0$ система имеет бесконечно много решений.

Б. Свойства определителя второго порядка. Для дальнейшего изложения удобно считать, что структурными элементами определителя второго порядка являются не четыре числа, а два двухэлементных столбца:

$$\boldsymbol{a} = \begin{pmatrix} a \\ c \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} b \\ d \end{pmatrix}.$$

В этом случае будем обозначать определитель следующим образом:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \det \|\boldsymbol{a},\ \boldsymbol{b}\| = |\boldsymbol{a},\ \boldsymbol{b}|.$$

Следующие свойства определителя второго порядка проверяются непосредственно.

6.7. Теорема (основные свойства определителя). *Определитель второго порядка обладает следующими свойствами*:

(1) *линейность по первому столбцу*:

$$\det \|\alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2,\ \boldsymbol{b}\| = \alpha_1 \det \|\boldsymbol{a}_1,\ \boldsymbol{b}\| + \alpha_2 \det \|\boldsymbol{a}_2,\ \boldsymbol{b}\|$$

или, в развёрнутом виде,

$$\begin{vmatrix} \alpha_1 a_1 + \alpha_2 a_2 & b \\ \alpha_1 c_1 + \alpha_2 c_2 & d \end{vmatrix} = \alpha_1 \begin{vmatrix} a_1 & b \\ c_1 & d \end{vmatrix} + \alpha_2 \begin{vmatrix} a_2 & b \\ c_2 & d \end{vmatrix};$$

также имеет место аналогичное свойство линейности по второму столбцу:

$$\det \left\| \boldsymbol{a},\ \beta_1 \boldsymbol{b}_1 + \beta_2 \boldsymbol{b}_2 \right\| = \beta_1 \det \left\| \boldsymbol{a},\ \boldsymbol{b}_1 \right\| + \beta_2 \det \left\| \boldsymbol{a},\ \boldsymbol{b}_2 \right\|;$$

(2) *кососимметричность*:

$$\det \left\| \boldsymbol{a},\ \boldsymbol{b} \right\| = -\det \left\| \boldsymbol{b},\ \boldsymbol{a} \right\|;$$

в развёрнутом виде

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = -\begin{vmatrix} b & a \\ d & c \end{vmatrix}$$

(*при перестановке столбцов определитель меняет знак*);

(3) *нормировка*:

$$\det \left\| \boldsymbol{e}_1, \boldsymbol{e}_2 \right\| = 1,$$

где $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ — *векторы стандартного базиса арифметического пространства* $\mathbb{K}^2$ (*см. пример* 5.66, *с.* 113), *или*

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$$

Эти три свойства определителя называются *основными*, или *характеристическими*. Из них можно вывести ряд новых (*производных*) свойств, полезных при вычислениях.

6.8. Предложение (производные свойства определителя)**.**

1. *Определитель с двумя одинаковыми столбцами равен нулю*:

$$\det \left\| \boldsymbol{a}, \boldsymbol{a} \right\| = 0 \quad \iff \quad \begin{vmatrix} a & a \\ c & c \end{vmatrix} = 0.$$

2. *Определитель не изменится, если к одному его столбцов прибавить другой столбец, умноженный на произвольное число*:

$$\det \left\| \boldsymbol{a} + \beta \boldsymbol{b},\ \boldsymbol{b} \right\| = \det \left\| \boldsymbol{a}, \boldsymbol{b} \right\|.$$

Доказательство. 1. При перестановке столбцов знак определителя должен измениться в силу кососимметричности, но с другой стороны, определитель не изменится, поскольку переставляемые столбцы одинаковы.

2. Действительно,

$$\det \left\| \boldsymbol{a} + \beta \boldsymbol{b},\ \boldsymbol{b} \right\| = \det \left\| \boldsymbol{a}, \boldsymbol{b} \right\| + \beta \underbrace{\det \left\| \boldsymbol{b}, \boldsymbol{b} \right\|}_{=0} = \det \left\| \boldsymbol{a}, \boldsymbol{b} \right\|. \qquad \square$$

6.9. Производные свойства, перечисленные в предложении 6.8, можно вывести непосредственно из определения, но мы вывели их из основных свойств, установленных в теореме 6.7. В дальнейшем основные свойства будут приняты за аксиомы, и производные свойства придётся выводить именно таким образом.

Кроме того, имеет место следующее свойство (оно доказывается непосредственной проверкой).

6.10. Предложение. *Определитель второго порядка не меняется при транспонировании*:

$$\det A^T = \det A \quad \Leftrightarrow \quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix}.$$

Это означает, что строки и столбцы определителя равноправны: любое утверждение, справедливое для столбцов, будет справедливым и для строк.

6.11. Предложение (критерий равенства определителя нулю)**.** *Определитель второго порядка равен нулю тогда и только тогда, когда его столбцы* (*строки*) *линейно зависимы.*

Доказательство. 1. Пусть определитель равен нулю. Имеем:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = 0 \quad \Leftrightarrow \quad ad = bc \quad \Leftrightarrow \quad \frac{a}{c} = \frac{b}{d} = \alpha \quad \Leftrightarrow \quad \begin{pmatrix} a \\ b \end{pmatrix} = \alpha \begin{pmatrix} c \\ d \end{pmatrix}$$

(самостоятельно проанализируйте случай, когда среди элементов определителя имеются нулевые).

2. Пусть столбцы определителя линейно зависимы (т.е. пропорциональны); тогда

$$\begin{pmatrix} a \\ c \end{pmatrix} = \alpha \begin{pmatrix} b \\ d \end{pmatrix} \quad \Rightarrow \quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} \alpha b & b \\ \alpha d & d \end{vmatrix} = 0. \qquad \square$$

2. Определители порядка n

Теперь попытаемся построить понятие определителя порядка n таким образом, чтобы основные его свойства и приложения были аналогичны свойствам и приложениям определителей второго порядка.

A. Определение.

6.12. Рассмотрим квадратную $(n \times n)$-матрицу $A \in \mathbb{K}^{n\times n}$. Как и в случае определителей второго порядка, будем считать, что структурными элементами этой матрицы являются не отдельные её элементы, а целые столбцы:

$$A = \begin{pmatrix} a_1^1 & a_2^1 & \dots & a_n^1 \\ a_1^2 & a_2^2 & \dots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^n & a_2^n & \dots & a_n^n \end{pmatrix} = \Big\| A_1,\ A_2,\ \dots, A_n \Big\|,$$

где

$$A_1 = \begin{pmatrix} a_1^1 \\ a_1^2 \\ \vdots \\ a_1^n \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_2^1 \\ a_2^2 \\ \vdots \\ a_2^n \end{pmatrix}, \quad \dots, \quad A_n = \begin{pmatrix} a_n^1 \\ a_n^2 \\ \vdots \\ a_n^n \end{pmatrix}.$$

6.13. Определение. *Определителем*, или *детерминантом*, квадратной матрицы A называется числовая функция столбцов этой матрицы

$$\det : \underbrace{\mathbb{K}^n \times \dots \times \mathbb{K}^n}_{n} \to \mathbb{K},$$

обозначаемая

$$\det A = \big|A\big| = \det \Big\| A_1, A_2, \dots, A_n \Big\| = \Big| A_1, A_2, \dots, A_n \Big|$$

и обладающая следующими свойствами[1]:

(1) *полилинейность*, т.е. линейность по каждому аргументу; например, свойство линейности по первому аргументу заключается в выполнении соотношения

$$\Big|\alpha' A_1' + \alpha'' A_1'',\ A_2,\ \dots,\ A_n\Big| = \\ = \alpha' \Big|A_1',\ A_2,\ \dots,\ A_n\Big| + \alpha'' \Big|A_1'',\ A_2,\ \dots,\ A_n\Big|;$$

(2) *кососимметричность*: при перестановке любых двух *соседних* столбцов определитель меняет знак на противоположный,

$$\Big|A_1,\ \dots,\ A_k,\ A_{k+1},\ \dots,\ A_n\Big| = -\Big|A_1,\ \dots,\ A_{k+1},\ A_k,\ \dots,\ A_n\Big|;$$

---

[1] Обратите внимание, что эти свойства скопированы с основных свойств определителя второго порядка.

(3) *нормировка*: определитель единичной матрицы $\mathbb{1} \in \mathbb{K}^{n\times n}$ равен единице,

$$\det \mathbb{1} = 1.$$

Перечисленные свойства называются *основными* (*характеристическими*) свойствами определителя.

Разумеется, сформулированное определение не даёт нам основания считать, что такая функция существует. Кроме того, мы должны прояснить вопрос о её единственности и разработать алгоритмы вычисления её значений для любой заданной матрицы. Однако уже сейчас можно получить некоторые следствия характеристических свойств, которыми должен обладать определитель, если он существует.

6.14. Предложение (производные свойства определителя)**.**

1. *При перестановке* любых *двух столбцов* (*не обязательно соседних*) *определитель меняет знак.*
2. *Определитель с двумя одинаковыми столбцами равен нулю.*
3. *Определитель не изменится, если к любому его столбцу прибавить произвольную линейную комбинацию остальных столбцов.*

Доказательство. 1. Смена знака при перестановке *соседних* столбцов — это определение свойства кососимметричности. Допустим теперь, что между переставляемыми столбцами имеется m столбцов, и докажем утверждение индукцией по m.

Базой индукции является утверждение при $m = 0$: перестановка соседних столбцов приводит к смене знака определителя.

Предположим, что утверждение справедливо при некотором значении m, т.е. при перестановке двух столбцов, между которыми расположены m столбцов, знак определителя меняется. Посмотрим, что произойдёт при перестановке двух столбцов, между которыми имеется $m+1$ столбцов (переставить требуется столбцы A_j и A_k, они подчёркнуты):

$$\left| A_1,\ \ldots,\ \underline{A_j},\ \underbrace{A_{j+1},\ A_{j+2},\ \ldots,\ A_{k-1}}_{m+1 \text{ столбцов}},\ \underline{A_k},\ \ldots,\ A_n \right| =$$

сначала переставим соседние A_j и A_{j+1}; при этом знак определителя изменится:

$$= -\left| A_1,\ \ldots,\ A_{j+1},\ \underline{A_j},\ \underbrace{A_{j+2},\ \ldots,\ A_{k-1}}_{m \text{ столбцов}},\ \underline{A_k},\ \ldots,\ A_n \right| =$$

переставим A_j и A_k, между которыми имеется m столбцов; в силу предположения индукции при этом знак определителя изменится:

$$= \left| A_1,\ \ldots,\ \underline{A_{j+1}},\ \underline{A_k},\ A_{j+2},\ \ldots,\ A_{k-1},\ A_j,\ \ldots,\ A_n \right| =$$

наконец, переставим соседние A_{j+1} и A_k; при этом знак определителя изменится ещё раз и мы получим окончательный результат:

$$= -\left| A_1,\ \ldots,\ A_k,\ \underbrace{A_{j+1},\ A_{j+2},\ \ldots,\ A_{k-1}}_{m+1 \text{ столбцов}},\ A_j,\ \ldots,\ A_n \right|.$$

2. Допустим, что у матрицы A имеются одинаковые столбцы. Если их переставить местами, то, с одной стороны, знак определителя $|A|$ изменится на противоположный в силу свойства кососимметричности, а с другой стороны — определитель не изменится, поскольку переставленные столбцы одинаковы. Таким образом, $|A| = -|A|$, откуда[1] $2|A| = 0$ и $|A| = 0$.

3. Проверим, что определитель не изменится, если к его первому столбцу прибавить линейную комбинацию остальных столбцов:

$$\left| A_1 + \alpha^2 A_2 + \cdots + \alpha^n A_n,\ A_2,\ \ldots,\ A_n \right| = \left| A_1,\ A_2,\ \ldots,\ A_n \right| +$$
$$+\alpha^2 \underbrace{\left| A_2,\ A_2,\ \ldots,\ A_n \right|}_{=0} + \cdots + \alpha^n \underbrace{\left| A_n,\ A_2,\ \ldots,\ A_n \right|}_{=0} =$$
$$= \left| A_1,\ A_2,\ \ldots,\ A_n \right|. \qquad \square$$

Б. Формулы Крамера. Мы также можем получить формулы для решения системы n линейных уравнений с n неизвестными (формулы Крамера), ещё не зная способов вычисления определителя и не будучи уверенными в его существовании!

6.15. Рассмотрим квадратную систему n линейных уравнений с n неизвестными:

$$\begin{cases} a_1^1 x^1 + a_2^1 x^2 + \cdots + a_n^1 x^n = b^1, \\ a_1^2 x^1 + a_2^2 x^2 + \cdots + a_n^2 x^n = b^2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_1^n x^1 + a_2^n x^2 + \cdots + a_n^n x^n = b^n. \end{cases} \tag{6.3}$$

[1]Отметим, что в поле характеристики 2 (см. определение IV.53, с. 345) из равенства $2|A| = 0$ не вытекает равенство $|A| = 0$. По этой причине для такого поля $\mathbb{K}$ кососимметричность определяется как свойство равенства нулю определителя с двумя одинаковыми столбцами.

Введя в рассмотрение основную и расширенную матрицы этой системы

$$A = \begin{pmatrix} a_1^1 & a_2^1 & \dots & a_n^1 \\ a_1^2 & a_2^2 & \dots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^n & a_2^n & \dots & a_n^n \end{pmatrix}, \quad \tilde{A} = \Big\| A \,\big|\, B \Big\| = \left(\begin{array}{cccc|c} a_1^1 & a_2^1 & \dots & a_n^1 & b^1 \\ a_1^2 & a_2^2 & \dots & a_n^2 & b^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1^n & a_2^n & \dots & a_n^n & b^n \end{array}\right),$$

их столбцы

$$A_1 = \begin{pmatrix} a_1^1 \\ a_1^2 \\ \vdots \\ a_1^n \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_2^1 \\ a_2^2 \\ \vdots \\ a_2^n \end{pmatrix}, \quad \dots, \quad A_n = \begin{pmatrix} a_n^1 \\ a_n^2 \\ \vdots \\ a_n^n \end{pmatrix}, \quad B = \begin{pmatrix} b^1 \\ b^2 \\ \vdots \\ b^n \end{pmatrix}$$

и столбец неизвестных $X = (x^1, x^2, \dots, x^n)^T$, мы можем записать систему в виде

$$x^1 A_1 + x^2 A_2 + \dots + x^n A_n = B \tag{6.4}$$

или

$$AX = B. \tag{6.5}$$

Рассмотрим определитель основной матрицы системы

$$\Delta = \Big| A_1,\ A_2,\ \dots,\ A_n \Big|$$

и определитель, полученный из предыдущего заменой первого столбца A_1 на столбец свободных членов:

$$\Delta_1 = \Big| B,\ A_2,\ \dots,\ A_n \Big|.$$

Если столбец $X = (x^1, x^2, \dots, x^n)^T$ является решением системы, то в силу соотношения (6.4) имеем

$$\begin{aligned} \Delta_1 &= \Big| B,\ A_2,\ \dots,\ A_n \Big| = \\ &= \Big| x^1 A_1 + x^2 A_2 + \dots + x^n A_n,\ A_2,\ \dots,\ A_n \Big| = \\ &= x^1 \cdot \underbrace{\Big| A_1,\ A_2,\ \dots,\ A_n \Big|}_{=\Delta} + x^2 \cdot \underbrace{\Big| A_2,\ A_2,\ \dots,\ A_n \Big|}_{=0} + \\ &\qquad + \dots + x^n \cdot \underbrace{\Big| A_n,\ A_2,\ \dots,\ A_n \Big|}_{=0}, \end{aligned}$$

откуда, при условии $\Delta \neq 0$, получаем

$$x^1 = \frac{\Delta_1}{\Delta}.$$

Аналогично получаются формулы для остальных неизвестных:

$$x^k = \frac{\Delta_k}{\Delta}, \quad k = 1, 2, \ldots, n, \tag{6.6}$$

где определители Δ_k получены из определителя основной матрицы системы Δ заменой k-го столбца на столбец свободных членов уравнений системы. Эти формулы называются *формулами Крамера.*

6.16. Таким образом, мы доказали *единственность* решения системы (6.3) в предположении, что определитель основной матрицы отличен от нуля. Доказательство же *существования* решения будет проведено позже (см. п. 6.61, с. 144).

6.17. Формулы Крамера дают решение в случае, когда определитель основной матрицы системы отличен от нуля, и при этом доказывают единственность этого решения. Если же $\Delta = 0$, то формулы Крамера неприменимы; в этом случае система может либо не иметь решений, либо иметь более одного решения.

B. Перестановки.

6.18. Определение. Пусть N — n-элементное множество; будем считать, что $N = \{1, 2, \ldots, n\}$. Будем называть *перестановкой* множества N называется взаимно однозначное отображение $\sigma\colon N \to N$ множества N в себя. Перестановку можно задать таблицей значений отображения σ, в явном виде указывая образы всех элементов множества N:

$$\begin{bmatrix} 1 & 2 & \ldots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(n) \end{bmatrix};$$

во второй строке таблицы каждое число из множества N встречается ровно по одному разу. Множество всех перестановок n-множества N обозначается S_n.

6.19. Если переставить местами столбцы этой таблицы, сохраняя все соответствия $i \mapsto \sigma(i)$, то получим другую запись той же перестановки, например

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 2 & 1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 4 & 2 \\ 2 & 4 & 3 & 1 \end{bmatrix}.$$

Удобно записывать перестановки так, чтобы числа верхнего ряда таблицы образовывали возрастающую последовательность $1, 2, \ldots, n$.

6.20. Перестановка ε, заданная формулой $\varepsilon(i) = i$ для любого $i \in N$, т.е.

$$\varepsilon = \begin{bmatrix} 1 & 2 & \dots & n \\ 1 & 2 & \dots & n \end{bmatrix},$$

называется *тождественной.*

6.21. Предложение (см. п. III.32, с. 322)**.** *Количество всех перестановок n-элементного множества N равно*

$$n! = 1 \cdot 2 \cdots (n-1) \cdot n. \tag{6.7}$$

Доказательство. Подсчитаем количество различных вариантов заполнить вторую строку таблицы значений перестановки. В качестве $\sigma(1)$ может быть выбрано любое из чисел $1, 2, \dots, n$ (всего n вариантов); в качестве $\sigma(2)$ — любое из оставшихся (всего $n-1$ вариантов) и т. д. Перемножая числа возможных вариантов, приходим к (6.7). □

6.22. Так, множество S_1 состоит из единственной перестановки ε, S_2 состоит из $2! = 2$ перестановок:

$$\varepsilon = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix},$$

а множество S_3 — из $3! = 6$ элементов:

$$\varepsilon = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}.$$

6.23. Ещё раз подчеркнём, что каждый элемент множества S_n является отображением $N \to N$. Введём на S_n бинарную операцию композиции отображений: если $\sigma\colon N \to N$ и $\tau\colon N \to N$ — перестановки (элементы множества S_n), то их *произведением* назовём перестановку $\tau\sigma$ (σ — первый «сомножитель», τ — второй), определённую правилом

$$(\tau\sigma)(i) = \tau\bigl(\sigma(i)\bigr) \quad \forall i \in N.$$

Например, если

$$\sigma = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}, \quad \tau = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix},$$

то

$$\tau\sigma = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix},$$

$$\sigma\tau = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix},$$

поскольку

$$\begin{array}{rccc} & 1 & 2 & 3 \\ \sigma\colon & \downarrow & \downarrow & \downarrow \\ & 3 & 2 & 1 \\ \tau\colon & \downarrow & \downarrow & \downarrow \\ & 3 & 1 & 2 \end{array} \qquad\qquad \begin{array}{rccc} & 1 & 2 & 3 \\ \tau\colon & \downarrow & \downarrow & \downarrow \\ & 2 & 1 & 3 \\ \sigma\colon & \downarrow & \downarrow & \downarrow \\ & 2 & 3 & 1 \end{array}$$

Введённая таким образом операция умножения перестановок *некоммутативна* (это видно из приведённого примера), но, как и любая композиция отображений, обладает свойством *ассоциативности*. Тождественная перестановка является, очевидно, нейтральным элементом этой операции. Кроме того, нетрудно видеть, что любая перестановка имеет *обратную* относительно введённой операции: для её построения достаточно поменять местами строки таблицы, задающей перестановку, после чего переставить столбцы местами так, чтобы элементы верхней строки стояли в естественном порядке $1, 2, \ldots, n$. Например,

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{bmatrix}, \quad \sigma^{-1} = \begin{bmatrix} 3 & 1 & 4 & 2 \\ 1 & 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{bmatrix}.$$

6.24. Множество S_n всех перестановок n-множества, снабжённое операцией композиции (умножения) перестановок, является группой (см. определение IV.25, с. 338), которая называется *симметрической группой*.

6.25. Определение. Будем говорить, что два различных элемента i, j множества $N = \{1, 2, \ldots, n\}$ *образуют инверсию* в перестановке $\sigma \in S_n$, если разности $i - j$ и $\sigma(i) - \sigma(j)$ имеют разные знаки.

6.26. Если перестановка задана как двухрядная таблица, верхний ряд которой записан в возрастающем порядке, то каждая инверсия обнаруживается в виде пары чисел нижнего ряда, в которой большее число расположено левее меньшего. Например, в тождественной перестановке инверсий нет, а в перестановке

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 2 & 1 & 4 \end{bmatrix}$$

имеется 6 инверсий, обнаруживаемых как следующие пары чисел второго ряда предыдущей таблицы:

$$(3, 2),\ (3, 1),\ (5, 2),\ (5, 1),\ (5, 4),\ (2, 1).$$

6.27. Определение. Перестановка $\sigma \in S_n$ называется *чётной* (*нечётной*), если она содержит чётное (нечётное) число инверсий. *Знаком перестановки* σ называется число

$$\operatorname{sgn}\sigma = \begin{cases} 1, & \text{если } \sigma \text{ — чётная перестановка,} \\ -1, & \text{если } \sigma \text{ — нечётная перестановка.} \end{cases}$$

Обратите внимание, что знак перестановки — это число (1 или -1), а не символ $+$ или $-$.

6.28. Ясно, что знак перестановки σ равен произведению знаков чисел $\dfrac{i-j}{\sigma(i)-\sigma(j)}$, соответствующих всевозможным парам $\{i,j\}$ различных элементов множества N:

$$\operatorname{sgn}(\sigma) = \prod_{\substack{\{i,j\}\subset N \\ i\neq j}} \operatorname{sign}\frac{i-j}{\sigma(i)-\sigma(j)},$$

где символ sign означает знак вещественного числа (см. с. 33). Очевидно, что для любых $x, y \in \mathbb{R}$ имеем $\operatorname{sign}(xy) = \operatorname{sign} x\cdot\operatorname{sign} y$. Отсюда следует, что знак перестановки не зависит от способа её записи, т.е. определение 6.27 корректно.

6.29. Предложение. *Знак произведения двух перестановок равен произведению знаков этих перестановок:*

$$\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}\sigma\cdot\operatorname{sgn}\tau. \tag{6.8}$$

Доказательство. Перестановку σ можно представить в виде

$$\sigma = \begin{bmatrix} 1 & 2 & \ldots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(n) \end{bmatrix} = \begin{bmatrix} \tau(1) & \tau(2) & \ldots & \tau(n) \\ \sigma(\tau(1)) & \sigma(\tau(2)) & \ldots & \sigma(\tau(n)) \end{bmatrix},$$

поэтому

$$\operatorname{sgn}\sigma = \prod_{\substack{\{i,j\}\subset N \\ i\neq j}} \operatorname{sign}\frac{\tau(i)-\tau(j)}{\sigma(\tau(i))-\sigma(\tau(j))}.$$

Следовательно, имеем

$$\operatorname{sgn}\sigma\cdot\operatorname{sgn}\tau =$$

$$= \prod_{\substack{\{i,j\}\subset N \\ i\neq j}} \operatorname{sign}\frac{\tau(i)-\tau(j)}{\sigma(\tau(i))-\sigma(\tau(j))}\cdot \prod_{\substack{\{i,j\}\subset N \\ i\neq j}} \operatorname{sign}\frac{i-j}{\tau(i)-\tau(j)} =$$

$$= \prod_{\substack{\{i,j\}\subset N \\ i\neq j}} \operatorname{sign}\frac{\tau(i)-\tau(j)}{\sigma(\tau(i))-\sigma(\tau(j))}\cdot\operatorname{sign}\frac{i-j}{\tau(i)-\tau(j)} =$$

$$= \prod_{\substack{\{i,j\}\subset N\\ i\neq j}} \operatorname{sign}\left(\frac{\tau(i)-\tau(j)}{\sigma(\tau(i))-\sigma(\tau(j))}\cdot\frac{i-j}{\tau(i)-\tau(j)}\right) =$$

$$= \prod_{\substack{\{i,j\}\subset N\\ i\neq j}} \operatorname{sign}\left(\frac{i-j}{\sigma(\tau(i))-\sigma(\tau(j))}\right) = \operatorname{sgn}(\sigma\tau). \qquad \square$$

6.30. Предложение 6.29 означает, что отображение $\operatorname{sgn} : S_n \to \{\pm 1\}$ является гомоморфизмом групп (см. определение IV.59, с. 346).

6.31. Следствие. *Взаимно обратные перестановки имеют один и тот же знак.*

Доказательство. Поскольку $\sigma^{-1}\sigma = \varepsilon$ и $\operatorname{sgn}\varepsilon = 1$, имеем

$$1 = \operatorname{sgn}\varepsilon = \operatorname{sgn}(\sigma^{-1}\sigma) = \operatorname{sgn}\sigma^{-1}\cdot\operatorname{sgn}\sigma. \qquad \square$$

Г. Существование и единственность определителя. Формула полного развёртывания определителя.

6.32. Допуская некоторую вольность формулировок, будем отождествлять квадратную матрицу A с упорядоченным набором её столбцов и считать термины «функция квадратной матрицы» и «функция столбцов квадратной матрицы» синонимами.

6.33. Найдём сначала общее выражение произвольной полилинейной кососимметрической функции $F(A_1, A_2, \ldots, A_n)$ столбцов квадратной матрицы A через элементы этих столбцов.

Разложим каждый столбец матрицы A по стандартному базису пространства $\mathbb{K}^n$ (см. пример 5.66, с. 113) и воспользуемся свойством полилинейности:

$$F\Big(A_1,\, A_2,\, \ldots,\, A_n\Big) =$$
$$= F\left(\sum_{\sigma_1=1}^{n} a_1^{\sigma_1}\boldsymbol{e}_{\sigma_1},\ \sum_{\sigma_2=1}^{n} a_2^{\sigma_2}\boldsymbol{e}_{\sigma_2},\ \ldots,\ \sum_{\sigma_n=1}^{n} a_n^{\sigma_n}\boldsymbol{e}_{\sigma_n}\right) =$$
$$= \sum_{\sigma_1=1}^{n}\sum_{\sigma_2=1}^{n}\cdots\sum_{\sigma_1=1}^{n} a_1^{\sigma_1}\cdot a_2^{\sigma_2}\cdots a_n^{\sigma_n}\cdot\Big|\boldsymbol{e}_{\sigma_1},\ \boldsymbol{e}_{\sigma_2},\ \ldots,\ \boldsymbol{e}_{\sigma_n}\Big|.$$

Полученное разложение состоит из n^n слагаемых, но многие из них обращаются в нуль: именно, если в выражении $F\Big(\boldsymbol{e}_{\sigma_1},\ \boldsymbol{e}_{\sigma_2},\ \ldots,\ \boldsymbol{e}_{\sigma_n}\Big)$ имеются одинаковые столбцы, то оно равен нулю в силу свойства кососимметричности. Таким образом, отличными от нуля могут быть только те слагаемые, в которых все индексы σ_n различны; в этом случае последовательность номеров $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ (точнее, отображение $\sigma : j \mapsto \sigma_j$) представляет собой перестановку[1]. Поскольку количество различных перестановок из n элементов равно $n!$, число слагаемых в последней сумме уменьшается с n^n до $n!$. Кроме того,

$$F\Big(\boldsymbol{e}_{\sigma_1},\ \boldsymbol{e}_{\sigma_2},\ \ldots,\ \boldsymbol{e}_{\sigma_n}\Big) = \operatorname{sgn}\sigma \cdot F\Big(\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \ldots,\ \boldsymbol{e}_n\Big),$$

где $\operatorname{sgn}\sigma$ — знак перестановки σ. Обозначив $F\Big(\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \ldots,\ \boldsymbol{e}_n\Big)$ через c, получаем

$$F\Big(A_1,\ A_2,\ \ldots,\ A_n\Big) = c \cdot \sum_{\sigma \in S_n} \operatorname{sgn}\sigma \cdot a_1^{\sigma_1} \cdot a_2^{\sigma_2} \cdots a_n^{\sigma_n}, \tag{6.9}$$

где суммирование производится по всевозможным перестановкам $\sigma \in S_n$, т.е. сумма состоит из $n!$ слагаемых. Таким образом, *если существует* полилинейная кососимметрическая функция столбцов квадратной матрицы, то она *единственна* и выражается формулой (6.9).

6.34. Для доказательства *существования* достаточно убедиться, что функция $F(A_1, A_2, \ldots, A_n)$, заданная выражением (6.9), является полилинейной и кососимметрической.

Проверим, например, *линейность* по первому аргументу:

$$\begin{aligned}
&F\Big(\alpha' A_1' + \alpha'' A_1'',\ A_2,\ \ldots,\ A_n\Big) = \\
&\qquad = c \cdot \sum_{\sigma \in S_n} \operatorname{sgn}\sigma \cdot \Big(\alpha' a_1'^{\sigma_1} + \alpha'' a_1''^{\sigma_1}\Big) \cdot a_2^{\sigma_2} \cdots a_n^{\sigma_n} = \\
&\qquad\qquad = c\alpha' \sum_{\sigma \in S_n} \operatorname{sgn}\sigma \cdot a_1'^{\sigma_1} \cdot a_2^{\sigma_2} \cdots a_n^{\sigma_n} + \\
&\qquad\qquad + c\alpha'' \sum_{\sigma \in S_n} \operatorname{sgn}\sigma \cdot a_1''^{\sigma_1} \cdot a_2^{\sigma_2} \cdots a_n^{\sigma_n} = \\
&\qquad = \alpha' \cdot F\Big(A_1',\ A_2,\ \ldots,\ A_n\Big) + \alpha'' \cdot F\Big(A_1'',\ A_2,\ \ldots,\ A_n\Big).
\end{aligned}$$

Для любого другого аргумента проверка совершенно аналогична.

Чтобы проверить *кососимметричность*, переставим в матрице A два соседних столбца A_k и A_{k+1} (будем выделять переставляемые столбцы подчёркиванием) и запишем выражение

[1]Сейчас нам удобнее писать σ_k вместо $\sigma(k)$ для обозначения значения отображения σ на элементе k. Кроме того, опуская первую строку таблицы значений функции σ, в которой числа $1, 2, \ldots, n$ расположены в возрастающем порядке, будем для обозначения перестановки использовать запись $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$.

для $F(A_1, \ldots, \underline{A_{k+1}}, \underline{A_k}, \ldots, A_n)$:

$$F\Big(A_1,\ \ldots,\ A_{k-1},\ \underline{A_{k+1}},\ \underline{A_k},\ A_{k+2},\ \ldots,\ A_n\Big) =$$
$$= c\sum_{\sigma\in S_n} \operatorname{sgn}(\sigma_1, \ldots, \underline{\sigma_k}, \underline{\sigma_{k+1}}, \ldots, \sigma_n)\cdot a_1^{\sigma_1}\cdots \underline{a_{k+1}^{\sigma_k}}\cdot \underline{a_k^{\sigma_{k+1}}}\cdots a_n^{\sigma_n}. \tag{6.10}$$

Рассмотрим какое-либо слагаемое из этой суммы:

$$\operatorname{sgn}(\sigma_1, \ldots, \underline{\sigma_k}, \underline{\sigma_{k+1}}, \ldots, \sigma_n)\cdot a_1^{\sigma_1}\cdots \underline{a_{k+1}^{\sigma_k}}\cdot \underline{a_k^{\sigma_{k+1}}}\cdots a_n^{\sigma_n}.$$

Переставив местами подчёркнутые сомножители, получим

$$\operatorname{sgn}(\sigma_1, \ldots, \underline{\sigma_k}, \underline{\sigma_{k+1}}, \ldots, \sigma_n)\cdot a_1^{\sigma_1}\cdots \underline{a_k^{\sigma_{k+1}}}\cdot \underline{a_{k+1}^{\sigma_k}}\cdots a_n^{\sigma_n};$$

это произведение входит в сумму (6.9) с противоположным знаком, поскольку

$$\operatorname{sgn}(\sigma_1, \ldots, \underline{\sigma_{k+1}}, \underline{\sigma_k}, \ldots, \sigma_n) = -\operatorname{sgn}(\sigma_1, \ldots, \underline{\sigma_k}, \underline{\sigma_{k+1}}, \ldots, \sigma_n).$$

Последнее равенство объясняется тем, что в перестановке изменилось лишь взаимное расположение элементов σ_k и σ_{k+1}, т.е. изменился факт наличия или отсутствия инверсии только между указанными элементами, что и приводит к смене знака перестановки.

Итак, каждое слагаемое из формулы (6.10), входит в формулу (6.9) с противоположным знаком, что завершает доказательство кососимметричности.

6.35. Определитель квадратной матрицы A — это полилинейная кососимметрическая функция $F(A)$ столбцов матрицы A, удовлетворяющая условию нормировки $F(\mathbb{1}) = 1$. Полагая $c = 1$ в формуле (6.9), получаем для определителя следующую *формулу полного развёртывания*, выражающую $\det A$ через элементы матрицы A:

$$\det A = \Big|A_1,\ A_2,\ \ldots,\ A_n\Big| = \sum_{\sigma\in S_n} \operatorname{sgn}\sigma\cdot a_1^{\sigma_1}\cdot a_2^{\sigma_2}\cdots a_n^{\sigma_n}. \tag{6.11}$$

(Условие нормировки вытекает из очевидного факта, что в случае единичной матрицы сумма в (6.11) содержит единственное ненулевое слагаемое $\operatorname{sgn}(1, 2, \ldots, n)\cdot \delta_1^1\delta_2^2\cdots\delta_n^n = 1$.)

Итак, мы доказали, что функция столбцов квадратной матрицы A, обладающая перечисленными в определении 6.13 свойствами, существует, единственна и выражается через элементы матрицы A формулой (6.11). Кроме того, мы установили следующий факт.

6.36. Предложение. *Если $F(A)$ — какая-либо полилинейная кососимметрическая функция столбцов квадратной матрицы A, то*

$$F(A) = F(\mathbb{1}) \cdot \det A.$$

6.37. Формула полного развёртывания определителя (6.11) для практических вычислений непригодна. Действительно, попробуем оценить время, необходимое для вычисления определителя с помощью этой формулы. Правая часть в (6.11) содержит $n!$ слагаемых, каждое из которых является произведением n сомножителей, т.е. требует $n-1$ попарных умножений. Таким образом, для вычисления определителя этим способом требуется выполнить $n!(n-1)$ умножений и некоторое количество сложений, которые для простоты не будем учитывать как менее трудоёмкие операции. Оценивая количество операций с помощью формулы Стирлинга (III.1) (см. с. 321), получаем для $n = 100$ число $100! \cdot 99 > 10^{153}$. Суперкомпьютеру, имеющему производительность 10 петафлопс (т.е. выполняющему 10^{16} операций в секунду), потребуется для такого вычисления время, превышающее

$$\frac{10^{153}}{10^{16}} = 10^{137} \text{ с} \sim 10^{130} \text{ лет},$$

что значительно превышает возраст Вселенной (около $1{,}4 \cdot 10^{10}$ лет).

Д. Инвариантность определителя относительно транспонирования. Формула полного развёртывания определителя позволяет доказать свойство инвариантности определителя относительно операции транспонирования матрицы.

6.38. Теорема. $\det A^T = \det A$.

Доказательство. Обозначив элементы матрицы A^T через $\tilde{a}^k_j$, запишем и преобразуем выражение для её определителя:

$$\det A^T = \sum_{\sigma \in S_n} \operatorname{sgn}\sigma \cdot \tilde{a}_1^{\sigma_1} \cdot \tilde{a}_2^{\sigma_2} \cdots \tilde{a}_n^{\sigma_n} =$$

в силу определения транспонированной матрицы $\tilde{a}^k_j = a^j_k$

$$= \sum_{\sigma \in S_n} \operatorname{sgn}\sigma \cdot a^1_{\sigma_1} \cdot a^2_{\sigma_2} \cdots a^n_{\sigma_n} =$$

переставим сомножители так, чтобы нижние индексы (номера столбцов) расположились в возрастающем порядке; тогда порядок верхних индексов будет определяться перестановкой $\tau = \sigma^{-1}$, обратной к перестановке σ:

$$= \sum_{\sigma \in S_n} \operatorname{sgn}\sigma \cdot a_1^{\sigma_1^{-1}} \cdot a_2^{\sigma_2^{-1}} \cdots a_n^{\sigma_n^{-1}} =$$

поскольку $\operatorname{sgn}\sigma^{-1} = \operatorname{sgn}\sigma$ (см. следствие 6.31, с. 130) условие $\sigma \in S_n$ эквивалентно условию $\tau = \sigma^{-1} \in S_n$, получаем

$$= \sum_{\tau\in S_n} \operatorname{sgn}\tau \cdot a_1^{\tau_1} \cdot a_2^{\tau_2} \cdots a_n^{\tau_n} = \det A. \qquad \square$$

6.39. В частности, отсюда следует, что *определитель представляет собой полилинейную кососимметрическую функцию строк матрицы.*

E. Определители специального вида.

6.40. Предложение. *Определитель треугольной матрицы равен произведению её диагональных элементов.*

Доказательство. Докажем утверждение для нижнетреугольной матрицы:

$$\begin{vmatrix} a_1^1 & 0 & 0 & \dots & 0 \\ a_1^2 & a_2^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1^n & a_2^n & a_3^n & \dots & a_n^n \end{vmatrix} = a_1^1 \cdot a_2^2 \cdots a_n^n.$$

Произведение диагональных элементов матрицы $a_1^1 \cdot a_2^2 \cdots a_n^n$ входит в сумму (6.11) со знаком плюс, так как соответствует тривиальной перестановке. Все остальные слагаемые в этой сумме равны нулю. Действительно, для нижнетреугольной матрицы, у которой $a_k^j = 0$ при $j < k$, из $a_1^{\sigma_1} a_2^{\sigma_2} \dots a_n^{\sigma_n} \neq 0$ следует

$$\sigma_1 \geqslant 1, \quad \sigma_2 \geqslant 2, \quad \dots, \quad \sigma_n \geqslant n.$$

Поскольку $\sigma_1 + \sigma_2 + \dots + \sigma_n = 1 + 2 + \dots + n$, это возможно лишь при $\sigma_1 = 1$, $\sigma_2 = 2$, …, $\sigma_n = n$. $\square$

6.41. Предложение. *Пусть квадратная матрица A имеет блочную структуру*

$$A = \begin{Vmatrix} B & D \\ O & C \end{Vmatrix},$$

где B и C — квадратные блоки. Тогда

$$\begin{vmatrix} B & D \\ O & C \end{vmatrix} = |B| \cdot |C|. \tag{6.12}$$

Доказательство. Определитель $\det A$ является полилинейной кососимметрической функцией столбцов матрицы A, так что при фиксированных матрицах D и C он является полилинейной кососимметрической функцией первых m столбцов матрицы A (здесь

m — размер B) или, эквивалентно, полилинейной кососимметрической функцией столбцов матрицы B:

$$\det A = \begin{vmatrix} B & D \\ O & C \end{vmatrix} = F(B).$$

Согласно предложению 6.36 такую функцию можно представить в виде

$$F(B) = F(\mathbb{1}) \cdot |B|,$$

где, очевидно,

$$F(\mathbb{1}) = \begin{vmatrix} \mathbb{1} & D \\ O & C \end{vmatrix}.$$

Этот определитель при фиксированной матрице D является полилинейной кососимметрической функцией своих последних строк, т.е. полилинейной кососимметрической функцией

$$F(\mathbb{1}) = \begin{vmatrix} \mathbb{1} & D \\ O & C \end{vmatrix} = G(C)$$

строк матрицы C и потому, снова в силу предложения 6.36,

$$G(C) = G(\mathbb{1}) \cdot |C|.$$

Здесь

$$G(\mathbb{1}) = \begin{vmatrix} \mathbb{1} & D \\ O & \mathbb{1} \end{vmatrix}$$

— определитель треугольной матрицы с единицами на диагонали, который, разумеется, равен единице. Итак,

$$\det A = F(B) = F(\mathbb{1}) \cdot |B| = G(C) \cdot |B| = \\ = G(\mathbb{1}) \cdot |C| \cdot |B| = |C| \cdot |B|. \qquad \square$$

3. Разложение определителя по столбцам и строкам

А. Алгебраические дополнения элементов.

6.42. Пусть A — квадратная $(n \times n)$-матрица, $A_1, \ldots, A_n$ — её столбцы. Обозначим через $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ векторы стандартного базиса арифметического пространства $\mathbb{K}^n$:

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \boldsymbol{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

(см. пример 5.66, с. 113).

6.43. Разложив первый столбец A_1 матрицы A в линейную комбинацию столбцов $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, для определителя $\det A$ мы можем записать

$$\begin{aligned}\det A &= \left| A_1,\ A_2,\ \ldots,\ A_n \right| = \\ &= \Big| \underbrace{a_1^1 \boldsymbol{e}_1 + a_1^2 \boldsymbol{e}_2 + \cdots + a_1^n \boldsymbol{e}_n}_{=A_1},\ A_2,\ \ldots,\ A_n \Big| = \\ &= a_1^1 \underline{\left| \boldsymbol{e}_1,\ A_2,\ \ldots,\ A_n \right|} + \cdots + a_1^n \underline{\left| \boldsymbol{e}_n,\ A_2,\ \ldots,\ A_n \right|}.\end{aligned}$$

Подчеркнутые определители называются *алгебраическими дополнениями* элементов $a_1^1, \ldots, a_1^n$ первого столбца матрицы A; обозначим их $\mathcal{A}_1^1, \ldots, \mathcal{A}_1^n$ соответственно. Очевидно, эти величины *не зависят от элементов первого столбца матрицы* A.

6.44. Полученная формула, называемая *разложением определителя по первому столбцу*, может служить средством вычисления определителя, если известен способ нахождения алгебраических дополнений:

$$\det A = \sum_{k=1}^{n} a_1^k \mathcal{A}_1^k.$$

6.45. Разумеется, вместо первого столбца можно выбрать любой другой:

$$\begin{aligned}\det A &= \left| A_1,\ \ldots,\ A_j,\ \ldots,\ A_n \right| = \\ &= \Big| A_1,\ \ldots,\ \underbrace{a_j^1 \boldsymbol{e}_1 + a_j^2 \boldsymbol{e}_2 + \cdots + a_j^n \boldsymbol{e}_n}_{=A_j},\ \ldots,\ A_n \Big| = \\ &= a_j^1 \cdot \underbrace{\left| A_1,\ \ldots,\ \boldsymbol{e}_1,\ \ldots,\ A_n \right|}_{=\mathcal{A}_j^1} + \\ &+ a_j^2 \cdot \underbrace{\left| A_1,\ \ldots,\ \boldsymbol{e}_2,\ \ldots,\ A_n \right|}_{=\mathcal{A}_j^2} + \cdots + \\ &+ a_j^n \cdot \underbrace{\left| A_1,\ \ldots,\ \boldsymbol{e}_n,\ \ldots,\ A_n \right|}_{=\mathcal{A}_j^n} = \sum_{k=1}^{n} a_j^k \mathcal{A}_j^k.\end{aligned}$$

Ясно, что алгебраические дополнения элементов j-го столбца не зависят от элементов самого этого столбца.

6.46. Конечно, благодаря теореме 6.38 имеют место и аналогичные формулы *разложения определителя по строкам*:

$$\det A = \sum_{j=1}^{n} a_j^k \mathcal{A}_j^k.$$

6.47. Для алгебраического дополнения $\mathcal{A}_1^1$ элемента a_1^1 по формуле (6.12) для определителя с нулевым углом немедленно получаем

$$\mathcal{A}_1^1 = \left| \boldsymbol{e}_1,\ A_2,\ \ldots,\ A_n \right| = \left| \begin{array}{c|ccc} 1 & a_2^1 & \ldots & a_n^1 \\ \hline 0 & a_2^2 & \ldots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_2^n & \ldots & a_n^n \end{array} \right| = \begin{vmatrix} a_2^2 & \ldots & a_n^2 \\ \vdots & \ddots & \vdots \\ a_2^n & \ldots & a_n^n \end{vmatrix}.$$

Последний определитель представляет собой детерминант порядка $n-1$, полученный из $\det A$ вычёркиванием первой строки и первого столбца.

6.48. Найдём теперь алгебраическое дополнение $\mathcal{A}_j^k$ элемента a_j^k. По определению

$$\mathcal{A}_j^k = \left| A_1,\ \ldots,\ A_{j-1},\ \boldsymbol{e}_k,\ A_{j+1},\ \ldots,\ A_n \right| =$$

$$= \begin{vmatrix} a_1^1 & \ldots & a_{j-1}^1 & 0 & a_{j+1}^1 & \ldots & a_n^1 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_1^{k-1} & \ldots & a_{j-1}^{k-1} & 0 & a_{j+1}^{k-1} & \ldots & a_n^{k-1} \\ a_1^k & \ldots & a_{j-1}^k & 1 & a_{j+1}^k & \ldots & a_n^k \\ a_1^{k+1} & \ldots & a_{j-1}^{k+1} & 0 & a_{j+1}^{k+1} & \ldots & a_n^{k+1} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_1^n & \ldots & a_{j-1}^n & 0 & a_{j+1}^n & \ldots & a_n^n \end{vmatrix}.$$

Многократно переставляя местами j-й столбец этого определителя, т.е. столбец $\boldsymbol{e}_k$, с соседним столбцом слева, т.е. сначала с A_{j-1}, затем с A_{j-2} и т. д., переместим его на первое место; всего при этом будет сделано $j-1$ перестановок столбцов, в результате каждой из которых определитель изменит знак, т.е. приобретёт множитель $(-1)^{j-1}$:

$$\mathcal{A}_j^k = (-1)^{j-1}\cdot\begin{vmatrix} 0 & a_1^1 & \dots & a_{j-1}^1 & a_{j+1}^1 & \dots & a_n^1 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_1^{k-1} & \dots & a_{j-1}^{k-1} & a_{j+1}^{k-1} & \dots & a_n^{k-1} \\ 1 & a_1^k & \dots & a_{j-1}^k & a_{j+1}^k & \dots & a_n^k \\ 0 & a_1^{k+1} & \dots & a_{j-1}^{k+1} & a_{j+1}^{k+1} & \dots & a_n^{k+1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_1^n & \dots & a_{j-1}^n & a_{j+1}^n & \dots & a_n^n \end{vmatrix}.$$

Далее подобным образом переместим k-ю строку полученного определителя на первое место; всего будет сделано $k-1$ перестановок, в результате чего определитель приобретёт множитель $(-1)^{k-1}$:

$$\mathcal{A}_j^k = \underbrace{(-1)^{j-1}\cdot(-1)^{k-1}}_{=(-1)^{j+k}}\cdot\begin{vmatrix} 1 & a_1^k & \dots & a_{j-1}^k & a_{j+1}^k & \dots & a_n^k \\ 0 & a_1^1 & \dots & a_{j-1}^1 & a_{j+1}^1 & \dots & a_n^1 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_1^{k-1} & \dots & a_{j-1}^{k-1} & a_{j+1}^{k-1} & \dots & a_n^{k-1} \\ 0 & a_1^{k+1} & \dots & a_{j-1}^{k+1} & a_{j+1}^{k+1} & \dots & a_n^{k+1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_1^n & \dots & a_{j-1}^n & a_{j+1}^n & \dots & a_n^n \end{vmatrix}.$$

Применяя к последнему определителю формулу (6.12) для определителя с нулевым углом, получим определитель порядка $n-1$, который образуется из исходного при вычёркивании k-й строки и j-го столбца. Этот определитель называется *дополнительным минором* элемента a_j^k и обозначается $\overline{M}_j^k$ (линии изображают вычеркнутые строку и столбец):

$$\overline{M}_j^k = \left|\begin{array}{ccc|ccc} a_1^1 & \dots & a_{j-1}^1 & a_{j+1}^1 & \dots & a_n^1 \\ \vdots & & \vdots & \vdots & & \vdots \\ a_1^{k-1} & \dots & a_{j-1}^{k-1} & a_{j+1}^{k-1} & \dots & a_n^{k-1} \\ \hline a_1^{k+1} & \dots & a_{j-1}^{k+1} & a_{j+1}^{k+1} & \dots & a_n^{k+1} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_1^n & \dots & a_{j-1}^n & a_{j+1}^n & \dots & a_n^n \end{array}\right|.$$

Таким образом, мы нашли выражение алгебраического дополнения $\mathcal{A}_j^k$ элемента a_j^k через соответствующий дополнительный минор $\overline{M}_j^k$:

$$\mathcal{A}_j^k = (-1)^{j+k}\overline{M}_j^k.$$

Доказана следующая теорема.

6.49. Теорема (разложение определителя по столбцу (строке))**.** *Определитель равен сумме попарных произведений элементов произвольно выбранного столбца* (*или строки*) *на их алгебраические дополнения*:

$$\det A = \sum_{k=1}^{n} a_j^k \mathcal{A}_j^k = \sum_{k=1}^{n} (-1)^{k+j} a_j^k\, \overline{M}_j^k; \tag{6.13}$$

$$\det A = \sum_{j=1}^{n} a_j^k \mathcal{A}_j^k = \sum_{j=1}^{n} (-1)^{k+j} a_j^k\, \overline{M}_j^k. \tag{6.14}$$

Каждая из формул (6.13), (6.14) приводит вычисление определителя n-го порядка к вычислению n определителей, порядок каждого из которых на единицу меньше.

Б. Фальшивое разложение определителя. Рассмотрим формулу (6.13) разложения определителя матрицы A по j-му столбцу. Заменив в ней элементы a_j^k $(k = 1, 2, \dots, n)$ j-го столбца матрицы A на произвольные числа $b^1, \dots, b^n$, получим, очевидно, разложение по j-му столбцу определителя новой матрицы A', которая образуется из матрицы A заменой j-го столбца $A_j = (a_j^1, a_j^2, \dots, a_j^n)^T$ на столбец $B = (b^1, b^2, \dots, b^n)$. В частности, если в качестве столбца B взять любой столбец матрицы A, кроме A_j, то получится разложение определителя с двумя одинаковыми столбцами, который, конечно, равен нулю. Таким образом, получено следующее утверждение.

6.50. Теорема (фальшивое разложение определителя)**.** *Сумма попарных произведений элементов произвольно выбранного столбца на алгебраические дополнения элементов другого столбца равна нулю*:

$$\sum_{k=1}^{n} a_j^k \mathcal{A}_p^k = 0, \quad \textit{если } j \neq p.$$

Аналогичное утверждение имеет место и для строк:

$$\sum_{j=1}^{n} a_j^k \mathcal{A}_j^q = 0, \quad \textit{если } k \neq q.$$

6.51. Объединяя эти формулы с (6.13) и (6.14), получаем:

$$\sum_{k=1}^{n} a_j^k \mathcal{A}_p^k = \det A \cdot \delta_{jp} = \begin{cases} \det A, & \text{если } j = p, \\ 0, & \text{если } j \neq p; \end{cases} \tag{6.15}$$

$$\sum_{j=1}^{n} a_j^k \mathcal{A}_j^q = \det A \cdot \delta_{kq} = \begin{cases} \det A, & \text{если } k = q, \\ 0, & \text{если } k \neq q. \end{cases} \tag{6.16}$$

4. Важные теоремы об определителях

А. Критерий равенства определителя нулю.

6.52. Теорема. *Определитель матрицы A равен нулю тогда и только тогда, когда её столбцы* (*строки*) *линейно зависимы.*

Доказательство.
Достаточность. Пусть столбцы матрицы линейно зависимы. По теореме 5.43.3 (см. с. 103) один из них равен линейной комбинации остальных. Вычитая из этого столбца указанную линейную комбинацию, получим определитель с нулевым столбцом, который, очевидно, равен нулю.

Необходимость докажем индукцией по числу n столбцов матрицы. Базой индукции является утверждение при $n = 2$: это предложение 6.11 (см. с. 121).

Итак, пусть $\det A = 0$. Если один из столбцов матрицы A нулевой, то её столбцы линейно зависимы (см. теорему 5.43.2). Предположим, что все столбцы ненулевые; тогда без ограничения общности можно считать, что $a_1^1 \neq 0$.

Обозначим строки матрицы A через $A^1, A^2, \ldots, A^n$. Вычтем из каждой строки первую строку, умноженную на такие коэффициенты, чтобы в первом столбце образовались нулевые элементы:

$$0 = \det \begin{Vmatrix} A^1 \\ A^2 \\ \vdots \\ A^n \end{Vmatrix} = \begin{vmatrix} a_1^1 & a_2^1 & \dots & a_n^1 \\ a_1^2 & a_2^2 & \dots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^n & a_2^n & \dots & a_n^n \end{vmatrix} =$$

$$= \det \begin{Vmatrix} A^1 \\ A^2 - \dfrac{a_1^2}{a_1^1} A^1 \\ \vdots \\ A^n - \dfrac{a_1^n}{a_1^1} A^1 \end{Vmatrix} = \begin{vmatrix} a_1^1 & a_2^1 & \dots & a_n^1 \\ 0 & a_2^2 - \dfrac{a_1^2}{a_1^1} a_2^1 & \dots & a_n^2 - \dfrac{a_1^2}{a_1^1} a_n^1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_2^n - \dfrac{a_1^n}{a_1^1} a_2^1 & \dots & a_n^n - \dfrac{a_1^n}{a_1^1} a_n^1 \end{vmatrix}. \tag{6.17}$$

Разлагая этот определитель по первому столбцу и учитывая, что $a_1^1 \neq 0$, получаем соотношение

$$\begin{vmatrix} a_2^2 - \dfrac{a_1^2}{a_1^1} a_2^1 & \dots & a_n^2 - \dfrac{a_1^2}{a_1^1} a_n^1 \\ \vdots & \ddots & \vdots \\ a_2^n - \dfrac{a_1^n}{a_1^1} a_2^1 & \dots & a_n^n - \dfrac{a_1^n}{a_1^1} a_n^1 \end{vmatrix} = 0.$$

В силу предположения индукции строки этого определителя порядка $(n-1)$ линейно зависимы, а потому линейно зависимы и строки определителя, полученного в (6.17):

$$B^2 = \left(0,\ a_2^2 - \frac{a_1^2}{a_1^1} a_2^1,\ \dots,\ a_n^2 - \frac{a_1^2}{a_1^1} a_n^1\right) = A^2 - \frac{a_1^2}{a_1^1} A^1,$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$B^n = \left(0,\ a_2^n - \frac{a_1^n}{a_1^1} a_2^1,\ \dots,\ a_n^2 - \frac{a_1^n}{a_1^1} a_n^1\right) = A^n - \frac{a_1^n}{a_1^1} A^1.$$

Таким образом, существуют такие не равные одновременно нулю коэффициенты $\alpha_2, \dots, \alpha_n$, что

$$\alpha_2 B^2 + \dots + \alpha_n B^n = \mathbf{0}.$$

Но тогда

$$\alpha_2 B^2 + \dots + \alpha_n B^n = \alpha_2 \left(A^2 - \frac{a_1^2}{a_1^1} A^1 \right) + \dots + \alpha_n \left(A^n - \frac{a_1^n}{a_1^1} A^1 \right) =$$
$$= -\left(\alpha_2 \frac{a_1^2}{a_1^1} + \dots + \alpha_n \frac{a_1^n}{a_1^1} \right) A^1 + \alpha_2 A^2 + \dots + \alpha_n A^n = \mathbf{0},$$

т.е. строки $A^1, A^2, \dots, A^n$ также линейно зависимы. □

Б. Совместность квадратной однородной системы.

6.53. Предложение. *Однородная квадратная система* $AX = O$ *имеет нетривиальные решения тогда и только тогда, когда определитель её основной матрицы равен нулю*: $\det A = 0$.

Доказательство. Согласно предложению 5.42 (см. с. 103) система $AX = O$ имеет нетривиальные решения тогда и только тогда, когда столбцы её основной матрицы линейно зависимы. Но линейная зависимость столбцов квадратной матрицы эквивалентна обращению в нуль её определителя. □

В. Определитель произведения матриц.

6.54. Теорема. *Если* A *и* B — *квадратные матрицы одного порядка, то*

$$\det(AB) = \det A \cdot \det B.$$

Доказательство. Пусть $C = AB$; согласно предложению 3.26 (см. с. 56), k-й столбец матрицы C равен линейной комбинации столбцов матрицы A с коэффициентами, равными элементам k-го столбца матрицы B:

$$C_k = \sum_{l=1}^{n} A_l \; b_k^l,$$

где n — порядок матриц. Для определителя $\det C$ получаем

$$\det C = \det \left\| C_1, \; \dots, \; C_n \right\| = \det \left\| \sum_{l_1=1}^{n} A_{l_1} \; b_1^{l_1}, \; \dots, \; \sum_{l_n=1}^{n} A_{l_n} \; b_n^{l_n} \right\| =$$
$$= \sum_{l_1=1}^{n} \cdots \sum_{l_n=1}^{n} b_1^{l_1} \cdots b_n^{l_n} \cdot \det \left\| A_{l_1}, \; \dots, \; A_{l_n} \right\|.$$

В этой сумме, состоящей из n^n слагаемых, обращаются в нуль все слагаемые, в состав которых в качестве множителя входят определители $\det \left\| A_{l_1}, \; \dots, \; A_{l_n} \right\|$ с одинаковыми столбцами. Ненулевыми могут быть лишь те из указанных определителей, в которых все индексы $l_1, \ldots, l_n$ различны, т.е. образуют некоторую перестановку σ чисел $1, \ldots, n$. Число слагаемых в сумме уменьшается до $n!$, а сама сумма записывается в виде

$$\det C = \sum_{\sigma \in S_n} b_1^{\sigma_1} \cdots b_n^{\sigma_n} \cdot \det \left\| A_{\sigma_1}, \ \ldots, \ A_{\sigma_n} \right\|.$$

Поскольку

$$\det \left\| A_{\sigma_1}, \ \ldots, \ A_{\sigma_n} \right\| = \operatorname{sgn} \sigma \cdot \det \left\| A_1, \ \ldots, \ A_n \right\|,$$

получаем далее

$$\det C = \underbrace{\det \left\| A_1, \ \ldots, \ A_n \right\|}_{= \det A} \cdot \sum_{\sigma \in S_n} \operatorname{sgn} \sigma \cdot b_1^{\sigma_1} \cdots b_n^{\sigma_n} = \det A \cdot \det B. \quad \square$$

5. Обратная матрица

6.55. Определение. Квадратная матрица называется *невырожденной*, или *неособой*, если её определитель отличен от нуля.

6.56. В силу теоремы 6.52 столбцы (строки) невырожденной матрицы линейно независимы.

6.57. Теорема. *Квадратная матрица обратима тогда и только тогда, когда она невырождена.*

Доказательство. Необходимость. Пусть матрица A обратима и A^{-1} — её обратная. По теореме 6.54 об определителе произведения матриц имеем

$$|A \cdot A^{-1}| = |\mathbb{1}| = 1 \quad \Rightarrow \quad |A| \neq 0.$$

Достаточность. Рассмотрим следующую матрицу, составленную из алгебраических дополнений элементов матрицы A^T:

$$A^{\vee} = \begin{pmatrix} \mathcal{A}_1^1 & \mathcal{A}_1^2 & \ldots & \mathcal{A}_1^n \\ \mathcal{A}_2^1 & \mathcal{A}_2^2 & \ldots & \mathcal{A}_2^n \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{A}_n^1 & \mathcal{A}_n^2 & \ldots & \mathcal{A}_n^n \end{pmatrix};$$

она называется *присоединённой* (или *союзной*) матрицей для матрицы A; иногда используется обозначение $\operatorname{adj} A$ (adjoint).

Найдём произведение матриц A и $A^\vee$: для любого элемента этого произведения имеем[1]

$$\left\{A \cdot A^\vee\right\}_j^k = \sum_{p=1}^{n} \{A\}_p^k \, \{A^\vee\}_j^p = \sum_{p=1}^{n} a_p^k \, \mathcal{A}_p^j = \det A \cdot \delta_{kj}$$

(мы воспользовались формулой (6.16)). Таким образом,

$$A \cdot A^\vee = |A| \cdot \mathbb{1}, \tag{6.18}$$

так что при $|A| \neq 0$ имеем

$$A^{-1} = \frac{1}{|A|} A^\vee, \tag{6.19}$$

т.е. мы не только доказали существование обратной матрицы для A, но и получили формулу для её вычисления. □

6.58. Поскольку для любого числа λ выполняется равенство $|\lambda A| = \lambda^n |A|$ (здесь n — порядок матрицы A), из (6.18) и теоремы об определителе произведения матриц получаем $|A| \cdot |A^\vee| = |A|^n$, откуда

$$|A^\vee| = |A|^{n-1}. \tag{6.20}$$

6.59. Отметим, что в случае, когда $|A| = 0$, произведение матрицы A на её присоединённую равно нулевой матрице:

$$\det A = 0 \quad \Rightarrow \quad A \cdot A^\vee = O.$$

6.60. Формула (6.19) для вычисления обратной матрицы на практике применяется редко, поскольку требует большого объёма вычислений: чтобы обратить матрицу порядка n, требуется найти один определитель порядка n и n определителей порядка $n-1$. Удобным и экономным (в смысле объёма вычислений) является метод, основанный на алгоритме Гаусса—Жордана (см. п. 4.48, с. 85).

6.61. Теперь мы можем легко доказать существование решения квадратной системы линейных уравнений $AX = B$ (см. (6.3) или (6.5), с. 124). Действительно, если $\det A \neq 0$, то существует обратная матрица A^{-1}, и, умножая обе части матричного уравнения $AX = B$ слева на A^{-1}, получаем $X = A^{-1}B$. Непосредственная проверка показывает, что этот столбец действительно является решением системы (6.5):

$$AX = A(A^{-1}B) = (AA^{-1})B = \mathbb{1}B = B.$$

В силу единственности то же самое решение определяется также и формулами Крамера (6.6).

[1] По поводу обозначения $\{A\}_p^k$ см. п. 3.2, с. 47.

6. Ранг матрицы

A. Пространства строк и столбцов матрицы.

6.62. Рассмотрим матрицу

$$A = \begin{pmatrix} a_1^1 & a_2^1 & \dots & a_n^1 \\ a_1^2 & a_2^2 & \dots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^m & a_2^m & \dots & a_n^m \end{pmatrix} \in \mathbb{K}^{m\times n}$$

и введём обычные обозначения для её столбцов

$$A_1 = \begin{pmatrix} a_1^1 \\ a_1^2 \\ \vdots \\ a_1^m \end{pmatrix}, \quad \dots, \quad A_n = \begin{pmatrix} a_n^1 \\ a_n^2 \\ \vdots \\ a_n^m \end{pmatrix}$$

и строк:

$$A^1 = \left(a_1^1,\ a_2^1,\ \dots,\ a_n^1\right), \quad \dots, \quad A^m = \left(a_1^m,\ a_2^m,\ \dots,\ a_n^m\right).$$

Рассмотрим линейные оболочки столбцов и строк матрицы A:

$$L(A_1, A_2, \dots, A_n) \Subset \mathbb{K}^m, \quad L(A^1, A^2, \dots, A^m) \Subset \mathbb{K}^{*n}.$$

6.63. Теорема. *Если матрица A' получена из матрицы A при помощи элементарных преобразований строк, то*
(1) $\dim L(A'_1, A'_2, \dots, A'_n) = \dim L(A_1, A_2, \dots, A_n)$;
(2) $\dim L(A'^1, A'^2, \dots, A'^m) = \dim L(A^1, A^2, \dots, A^m)$.

Доказательство. (1) Утверждение, касающееся линейных оболочек столбцов, фактически было доказано в гл. 5 (см. теорему 5.46, с. 107). Действительно, каждой (в том числе максимальной) линейно независимой системе столбцов одной матрицы будет отвечать линейно независимая система столбцов (с теми же номерами!) другой матрицы, чем и устанавливается требуемое равенство.

(2) Мы докажем даже больше, а именно, что линейная оболочка строк матрицы A' совпадает с линейной оболочкой строк матрицы A:

$$L(A'^1, A'^2, \dots, A'^m) = L(A^1, A^2, \dots, A^m),$$

отсюда сразу же получим равенство размерностей этих линейных оболочек. Действительно, поскольку каждая строка A'^i преобразованной матрицы A' является линейной комбинацией строк исходной матрицы A, по теореме 5.60 имеем $L(A'^1, A'^2, \dots, A'^m) \subset L(A^1, A^2, \dots, A^m)$. Так как все элементарные преобразования обратимы, то имеет место также и обратное вложение $L(A^1, A^2, \ldots, A^m) \subset L(A'^1, A'^2, \ldots, A'^m)$, откуда и следует требуемое равенство линейных оболочек, а следовательно, и равенство их размерностей. $\square$

6.64. Обратите внимание, что после выполнения элементарных преобразований строк матрицы линейная оболочка столбцов изменяется, но сохраняются номера базисных столбцов. Напротив, линейная оболочка строк остаётся неизменной, но номера базисных строк преобразованной матрицы не совпадают, вообще говоря, с номерами базисных строк исходной матрицы.

6.65. До настоящего момента термины «базисный столбец» и «базисная строка» означали строки и столбцы специального вида, образующиеся при приведении матрицы к упрощённому виду (см. определение на с. 73). Теперь становится понятным, что *базисные строки и столбцы матрицы образуют соответственно базисы в линейных оболочках строк и столбцов*, чем и объясняется их название.

Б. Понятие ранга матрицы.

6.66. Теорема. *Для любой матрицы A размерность линейной оболочки её столбцов равна размерности линейной оболочки строк*:

$$\dim L(A_1, A_2, \ldots, A_n) = \dim L(A^1, A^2, \ldots, A^m);$$

эта размерность называется рангом *матрицы и обозначается* $\operatorname{rk} A$, $\operatorname{rank} A$, $\operatorname{rg} A$, $\operatorname{rang} A$, $r(A)$ *и т. п.*

Доказательство. Приведём матрицу A к упрощённому виду при помощи алгоритма Гаусса—Жордана:

$$\begin{pmatrix}
\boxed{1} & * & 0 & * & 0 & * & \cdots & 0 & * & * \\
0 & 0 & \boxed{1} & * & 0 & * & \cdots & 0 & * & * \\
0 & 0 & 0 & 0 & \boxed{1} & * & \cdots & 0 & * & * \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & * & * \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & \boxed{1} & * & * \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0
\end{pmatrix}.$$

Очевидно, количество базисных строк (равное максимальному количеству линейно независимых строк матрицы, т.е. размерности линейной оболочки строк) совпадает с количеством базисных столбцов, что и требовалось. $\square$

6.67. Для нахождения ранга матрицы можно использовать как алгоритм Гаусса—Жордана, так и алгоритм Гаусса. При использовании первого из них получаем не только значение ранга, но и линейные зависимости между столбцами, т.е. номера базисных столбцов и разложения всех остальных столбцов в линейные комбинации базисных столбцов. При использовании алгоритма Гаусса можно найти лишь номера базисных столбцов матрицы и их количество, но разложение каждого столбца в линейную комбинацию базисных столбцов останется неизвестным.

B. Совместность неоднородных систем.

6.68. Теорема (теорема Кронекера—Капелли[1]). *Неоднородная система $AX = B$ совместна тогда и только тогда, когда ранги её основной и расширенной матриц совпадают:*

$$\operatorname{rk} A = \operatorname{rk} \|A \mid B\|.$$

Доказательство. Поскольку линейная оболочка столбцов основной матрицы системы содержится в линейной оболочке столбцов расширенной матрицы, $L(A_1, \dots, A_n) \subset L(A_1, \dots, A_n, B)$, имеем неравенство $\operatorname{rk} A \leqslant \operatorname{rk} \|A \mid B\|$, справедливое для любой системы (совместной или несовместной).

Записав систему $AX = B$ в виде

$$x^1 A_1 + x^2 A_2 + \dots + x^n A_n = B,$$

видим, что совместность рассматриваемой системы эквивалентна тому факту, что столбец B принадлежит линейной оболочке столбцов матрицы A, т.е. $L(A_1, \dots, A_n, B) \subset L(A_1, \dots, A_n)$ (см. теорему 5.60) и потому $\operatorname{rk} \|A \mid B\| \leqslant \operatorname{rk} A$. Таким образом, для совместных систем и только для них $\operatorname{rk} \|A \mid B\| = \operatorname{rk} A$. $\square$

6.69. При решении неоднородных систем методом Гаусса—Жордана использовать теорему Кронекера—Капелли нет необходимости, поскольку из упрощённого вида матрицы системы факт совместности или несовместности системы очевиден. У несовместных систем несовпадение рангов основной и расширенной матриц проявляется в появлении строк вида $(0, 0, \dots, 0, 1)$, отвечающих несовместным уравнениям $0 \cdot x^1 + \dots + 0 \cdot x^n = 1$.

---

[1]А. Капелли (A. Capelli) — итальянский математик (1855–1910).

Г. Ранг произведения матриц.

6.70. Предложение. *Если существует произведение матриц A и B, то*

$$\operatorname{rk}(AB) \leqslant \operatorname{rk} A, \quad \operatorname{rk}(AB) \leqslant \operatorname{rk} B, \tag{6.21}$$

т.е. ранг произведения матриц не превышает ранга каждого из сомножителей.

Доказательство. Докажем первое неравенство. Столбцы матрицы

$$\underset{m\times s}{C} = \underset{m\times n}{A} \ \underset{n\times s}{B}$$

суть линейные комбинации столбцов матрицы A (см. теорему 3.26, с. 56), так что

$$L(C_1, C_2, \ldots, C_s) \subset L(A_1, A_2, \ldots, A_n)$$

(см. утверждение 1 теоремы 5.60, с. 111). Поэтому

$$\dim L(C_1, C_2, \ldots, C_s) \leqslant \dim L(A_1, A_2, \ldots, A_n). \qquad \square$$

7. Теорема о базисном миноре

А. Подматрицы и миноры.

6.71. Выберем в прямоугольной матрице $A \in \mathbb{K}^{m\times n}$ строки с номерами $i_1 < i_2 < \cdots < i_k$ (количество выбранных строк равно k) и столбцы с номерами $j_1 < j_2 < \cdots < j_l$ (количество выбранных столбцов равно l) и образуем из элементов на их пересечении матрицу размера $k \times l$:

$$\begin{pmatrix} a_1^1 & \ldots & a_{j_1}^1 & \ldots & a_{j_2}^1 & \ldots & a_{j_l}^1 & \ldots & a_n^1 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_1^{i_1} & \ldots & a_{j_1}^{i_1} & \ldots & a_{j_2}^{i_1} & \ldots & a_{j_l}^{i_1} & \ldots & a_n^{i_1} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_1^{i_2} & \ldots & a_{j_1}^{i_2} & \ldots & a_{j_2}^{i_2} & \ldots & a_{j_l}^{i_2} & \ldots & a_n^{i_2} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_1^{i_k} & \ldots & a_{j_1}^{i_k} & \ldots & a_{j_2}^{i_k} & \ldots & a_{j_l}^{i_k} & \ldots & a_n^{i_k} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_1^m & \ldots & a_{j_1}^m & \ldots & a_{j_2}^m & \ldots & a_{j_l}^m & \ldots & a_n^m \end{pmatrix} \Rightarrow \begin{pmatrix} a_{j_1}^{i_1} & \ldots & a_{j_2}^{i_1} & \ldots & a_{j_l}^{i_1} \\ \vdots & & \vdots & & \vdots \\ a_{j_1}^{i_2} & \ldots & a_{j_2}^{i_2} & \ldots & a_{j_l}^{i_2} \\ \vdots & & \vdots & & \vdots \\ a_{j_1}^{i_k} & \ldots & a_{j_2}^{i_k} & \ldots & a_{j_l}^{i_k} \end{pmatrix};$$

будем называть её *подматрицей* матрицы A, образованной выбранными строками и столбцами, и обозначать

$$A\begin{pmatrix} i_1\, i_2\, \ldots\, i_k \\ j_1\, j_2\, \ldots\, j_l \end{pmatrix} = \begin{pmatrix} a^{i_1}_{j_1} & \ldots & a^{i_1}_{j_2} & \ldots & a^{i_1}_{j_l} \\ \vdots & & \vdots & & \vdots \\ a^{i_2}_{j_1} & \ldots & a^{i_2}_{j_2} & \ldots & a^{i_2}_{j_l} \\ \vdots & & \vdots & & \vdots \\ a^{i_k}_{j_1} & \ldots & a^{i_k}_{j_2} & \ldots & a^{i_k}_{j_l} \end{pmatrix} \in \mathbb{K}^{k\times l}.$$

Матрица порядка $(m-k)\times(n-l)$, которая получится, если из матрицы A *удалить* выбранные строки и столбцы, называется *дополнительной подматрицей* и обозначается

$$\overline{A}\begin{pmatrix} i_1\, i_2\, \ldots\, i_k \\ j_1\, j_2\, \ldots\, j_l \end{pmatrix} \in \mathbb{K}^{(m-k)\times(n-l)}.$$

6.72. Если введённые в п. 6.71 подматрицы квадратные, то их определители называются соответственно *минором* и *дополнительным минором* матрицы A, соответствующими выбранным строкам и столбцам, и обозначаются[1]

$$M\begin{pmatrix} i_1\, i_2\, \ldots\, i_k \\ j_1\, j_2\, \ldots\, j_k \end{pmatrix} = \det A\begin{pmatrix} i_1\, i_2\, \ldots\, i_k \\ j_1\, j_2\, \ldots\, j_k \end{pmatrix},$$
$$\overline{M}\begin{pmatrix} i_1\, i_2\, \ldots\, i_k \\ j_1\, j_2\, \ldots\, j_l \end{pmatrix} = \det \overline{A}\begin{pmatrix} i_1\, i_2\, \ldots\, i_k \\ j_1\, j_2\, \ldots\, j_l \end{pmatrix}.$$

Напомним, что в п. 6.48 дополнительным минором $\overline{M}^i_j$ элемента a^i_j матрицы мы назвали определитель подматрицы, полученной из A *вычёркиванием* i-й строки и j-го столбца.

Б. Теорема о базисном миноре.

6.73. Определение. Минор порядка r матрицы A называется *базисным*, если он отличен от нуля, а любой минор большего порядка равен нулю.

6.74. Разумеется, у матрицы может быть не один базисный минор. Например, у матрицы

$$\begin{pmatrix} 1 & 1 & 3 & 4 \\ 2 & 2 & 4 & 6 \end{pmatrix}$$

[1] В обозначении *дополнительного* минора учтён тот факт, что число k вычеркнутых строк может не совпадать с числом l вычеркнутых столбцов: важно лишь, чтобы дополнительная подматрица оказалась квадратной, т.е. $m-k=n-l$.

имеется три базисных минора второго порядка:

$$\begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} = -2, \quad \begin{vmatrix} 1 & 4 \\ 2 & 6 \end{vmatrix} = -2, \quad \begin{vmatrix} 3 & 4 \\ 4 & 6 \end{vmatrix} = 2.$$

6.75. Будем говорить, что подматрица $\hat{\hat{A}}$ матрицы A *окаймляет* подматрицу $\hat{A}$, если $\hat{A}$ получается из $\hat{\hat{A}}$ вычёркиванием одной крайней (т.е. первой или последней) строки и одного крайнего столбца.

6.76. Теорема (теорема о базисном миноре)**.**

1. *Столбцы (строки) матрицы, образующие её базисный минор, линейно независимы.*
2. *Любой столбец (строка) матрицы является линейной комбинацией столбцов (строк), образующих базисный минор.*

Доказательство. 1. Предположение о том, что столбцы, образующие базисный минор, линейно зависимы, приводит к противоречию, поскольку определитель с линейно зависимыми столбцами равен нулю (см. теорему 6.52, с. 140).

2. Не ограничивая общности, будем считать, что базисный минор образован первыми r строками и столбцами матрицы, т.е. расположен в левом верхнем углу:

$$\begin{pmatrix} a_1^1 & \cdots & a_r^1 & a_{r+1}^1 & \cdots & a_n^1 \\ \vdots & & \vdots & \vdots & & \vdots \\ a_1^r & \cdots & a_r^r & a_{r+1}^r & \cdots & a_n^r \\ a_1^{r+1} & \cdots & a_r^{r+1} & a_{r+1}^{r+1} & \cdots & a_n^{r+1} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_1^m & \cdots & a_r^m & a_{r+1}^m & \cdots & a_n^m \end{pmatrix}.$$

Пусть A_k — произвольный столбец матрицы A. Покажем, что он является линейной комбинацией столбцов $A_1, \ldots, A_r$. При $k \leqslant r$ утверждение очевидно, так как

$$A_k = 0 \cdot A_1 + \cdots + 0 \cdot A_{k-1} + A_k + 0 \cdot A_{k+1} + \cdots + 0 \cdot A_r.$$

Если же $k > r$, то рассмотрим минор, полученный окаймлением базисного минора столбцом A_k и какой-либо строкой A^s:

$$\begin{vmatrix} a_1^1 & \cdots & a_r^1 & a_k^1 \\ \vdots & & \vdots & \vdots \\ a_1^r & \cdots & a_r^r & a_k^r \\ a_1^s & \cdots & a_r^s & a_k^s \end{vmatrix}. \tag{6.22}$$

Этот минор равен нулю, так как при $s \leqslant r$ он содержит две одинаковые строки, а при $s > r$ является минором бо́льшего порядка, нежели базисный минор. Разложив определитель (6.22) по элементам последней строки, получим

$$a_1^s \mathcal{M}_1 + \dots + a_r^s \mathcal{M}_r + a_k^s \mathcal{M} = 0,$$

где $\mathcal{M}_1, \dots, \mathcal{M}_r$ — алгебраические дополнения первых r элементов последней строки, не зависящие от номера s этой строки, а $\mathcal{M}$ — базисный минор. Отсюда получаем

$$a_k^s = -\frac{\mathcal{M}_1}{\mathcal{M}} a_1^s - \dots - \frac{\mathcal{M}_r}{\mathcal{M}} a_r^s,$$

так что

$$A_k = -\frac{\mathcal{M}_1}{\mathcal{M}} A_1 - \dots - \frac{\mathcal{M}_r}{\mathcal{M}} A_r. \qquad \square$$

6.77. Таким образом, столбцы (строки), образующие базисный минор, образуют базис в линейной оболочке столбцов (строк) матрицы (т.е. являются базисными столбцами (строками)), а порядок базисного минора матрицы равен размерности каждой из этих линейных оболочек, т.е. рангу матрицы.

6.78. ***Метод окаймляющих миноров.*** Для нахождения ранга матрицы можно использовать следующий приём. Пусть в матрице A найден ненулевой минор M порядка k. Рассмотрим все миноры порядка $k+1$, окаймляющие минор M; если все они равны нулю, то ранг матрицы равен k. В противном случае среди окаймляющих миноров найдется ненулевой, и вся процедура повторяется.

Использование метода окаймляющих миноров для решения практических задач неудобно, поскольку связано с вычислением большого количества определителей.

8. Ориентация векторного пространства

Все результаты данного раздела относятся только к *вещественным* векторным пространствам.

А. Наглядное представление об ориентации. Понятие ориентации является абстрактным обобщением наглядных представлений о направлении движения на прямой (вправо или влево), направлении вращения на плоскости (по часовой стрелке или против) и о винтовом движении в пространстве.

6.79. Например, выбор ориентации на прямой означает выбор одного из двух возможных направлений на этой прямой. Если прямая изображена на чертеже горизонтальной прямой, то на ней обычно выбирают направление слева направо, а для вертикальных прямых — снизу вверх и называют такое направление положительным. Однако подобный выбор направления не имеет никакого инвариантного (т.е. не зависящего от чертежа) смысла: при повороте чертежа положительное направление переходит в отрицательное и наоборот. Прямую, на которой зафиксировано направление (одно из двух возможных), обычно называют *осью.*

6.80. На плоскости задание ориентации сводится к заданию одного из двух возможных направлений вращения: по часовой стрелке или против. Базис на плоскости, состоящий из двух неколлинеарных направленных отрезков, считается «правым», если кратчайшее вращение, совмещающее первый вектор базиса со вторым, осуществляется против часовой стрелки, и «левым» в противоположном случае. Инвариантного смысла такое соглашение также не имеет: если посмотреть на плоскость с другой стороны, то ориентация сменится.

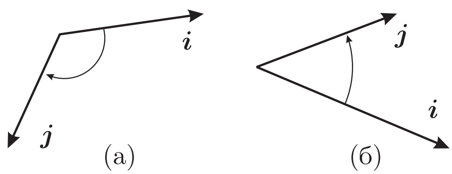


Рис. 6.1. Базис на плоскости: (а) левый, (б) правый

6.81. Аналогично, в пространстве «правая» ориентация определяется одним из следующих трёх эквивалентных соглашений:

(1) *«правило правой руки»*: базис $(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$ считается правым, если его векторы расположены так же, как большой, указательный и средний пальцы правой руки;

(2) *«правило буравчика»*: базис $(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$ считается правым, если вращение буравчика в направлении, переводящем (кратчайшим образом) вектор $\boldsymbol{i}$ в вектор $\boldsymbol{j}$, приводит к поступательному движению буравчика в направлении вектора $\boldsymbol{k}$;

(3) базис $(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$ считается правым, если при наблюдении из конца вектора $\boldsymbol{k}$ кратчайший поворот, переводящий вектор $\boldsymbol{i}$ в вектор $\boldsymbol{j}$, осуществляется против часовой стрелки.

6.82. Термин «правая (левая) тройка векторов» может использоваться в отношении любых трёх некомпланарных векторов в пространстве безотносительно к тому, рассматриваются ли эти векторы как базис (т.е. используются ли для разложения других векторов; конечно, базис они образуют).

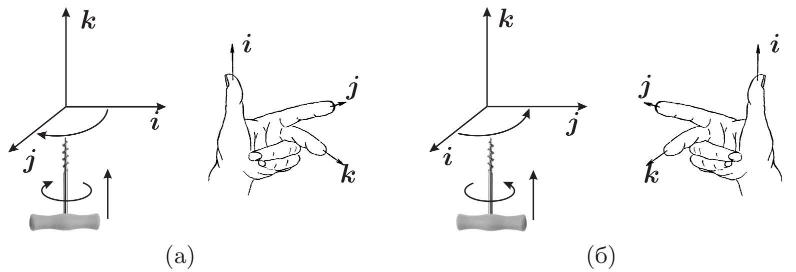


Рис. 6.2. Базис в пространстве: (а) левый, (б) правый

6.83. Если $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ — правая тройка векторов, то тройки $\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{b}$ и $\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{a}$, полученные из исходной циклической перестановкой векторов (т.е. перестановкой «по кругу»), также являются правыми, а тройки $\boldsymbol{b}, \boldsymbol{a}, \boldsymbol{c}$ (полученная из исходной перестановкой первых двух векторов), $\boldsymbol{c}, \boldsymbol{b}, \boldsymbol{a}$ и $\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{b}$ (полученные из тройки $\boldsymbol{b}, \boldsymbol{a}, \boldsymbol{c}$ циклической перестановкой) — левыми.

6.84. Наконец, отметим, что термины «правый» и «левый» по отношению к пространству имеют «почти инвариантный» смысл: мы не можем посмотреть на пространство «с другой стороны» или совместить левую руку с правой. Однако любой базис в пространстве и его отражение в зеркале имеют противоположные ориентации.

6.85. Никакого «внутреннего», *чисто математического* способа различить «правую» и «левую» ориентации не существует; можно лишь установить, что таковых ориентаций имеется две.

Вопрос о том, существуют ли *физические процессы*, позволяющие выбрать ориентацию пространства, был решён положительно, ко всеобщему удивлению, в 1956 г.: американские физики Ц. Ли (Tsung-Dao Lee) и Ч. Янг (Chen-Ning Yang) теоретически предсказали возможность нарушения так называемого свойства чётности в слабых взаимодействиях, а годом спустя это было подтверждено в экспериментах Ц. Ву (Chien-Shiung Wu). В 1957 г. Ли и Янг были удостоены за это открытие Нобелевской премии; мадам Ву осталась без награды, что было воспринято многими как проявление сексизма.

Б. Абстрактное понятие ориентации.

6.86. Рассмотрим в n-мерном вещественном векторном пространстве $\mathcal{V}$ два базиса: $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$ и $\boldsymbol{F} = \big(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n\big)$. Каждый вектор второго базиса можно разложить по первому:

$$\begin{aligned} \boldsymbol{f}_1 &= c_1^1 \boldsymbol{e}_1 + c_1^2 \boldsymbol{e}_2 + \dots + c_1^n \boldsymbol{e}_n, \\ \boldsymbol{f}_2 &= c_2^1 \boldsymbol{e}_1 + c_2^2 \boldsymbol{e}_2 + \dots + c_2^n \boldsymbol{e}_n, \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ \boldsymbol{f}_n &= c_n^1 \boldsymbol{e}_1 + c_n^2 \boldsymbol{e}_2 + \dots + c_n^n \boldsymbol{e}_n. \end{aligned} \tag{6.23}$$

Эти разложения можно записать компактно, используя правило суммирования Эйнштейна:

$$\boldsymbol{f}_k = c_k^j \boldsymbol{e}_j, \quad j, k = 1, \dots, n.$$

Коэффициенты этих разложений естественно записать в виде квадратной матрицы

$$C = \begin{pmatrix} c_1^1 & c_2^1 & \dots & c_n^1 \\ c_1^2 & c_2^2 & \dots & c_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_1^n & c_2^n & \dots & c_n^n \end{pmatrix},$$

которая называется *матрицей перехода* от базиса $\boldsymbol{E}$ к базису $\boldsymbol{F}$. Обратите внимание на индексы, нумерующие элементы матрицы перехода: нижний индекс является номером того вектора второго базиса, который подвергается разложению, а верхний — номером координаты (в соответствии с соглашением о нумерации координат; см. с. 110). Таким образом, координаты вектора $\boldsymbol{f}_k$ в базисе $\boldsymbol{E}$ расположены в k*-м столбце* матрицы перехода (а не в k-й строке!); соглашение о нумерации элементов матрицы (см. с. 46) также выполнено.

6.87. Формулы (6.23) можно записать в матричном виде:

$$\boldsymbol{F} = \boldsymbol{E}C. \tag{6.24}$$

6.88. Определение. Базисы $\boldsymbol{E}$ и $\boldsymbol{F}$ называются *одноимёнными* (обозначение $\boldsymbol{F} \simeq \boldsymbol{E}$), если матрица перехода C от $\boldsymbol{E}$ к $\boldsymbol{F}$ имеет положительный определитель, и *разноимёнными* в противном случае.

6.89. Предложение. *Отношение одноимённости базисов является отношением эквивалентности* (см. определение II.33, с. 308)*, т.е. обладает следующими свойствами:*

(i) *рефлексивность*: $\boldsymbol{E} \simeq \boldsymbol{E}$*;*

(ii) *симметричность*: *если* $\boldsymbol{E} \simeq \boldsymbol{F}$*, то* $\boldsymbol{F} \simeq \boldsymbol{E}$;

(iii) *транзитивность*: *если* $\boldsymbol{E} \simeq \boldsymbol{F}$ *и* $\boldsymbol{F} \simeq \boldsymbol{G}$*, то* $\boldsymbol{E} \simeq \boldsymbol{G}$.

Поэтому совокупность всех базисов данного векторного пространства распадается на два непересекающихся класса (*см. теорему* II.36*, с.* 308)*. К одному из этих классов относятся все базисы, одноимённые с некоторым заранее зафиксированным базисом* $\boldsymbol{E}_0$*, а к другому — все базисы, разноимённые с базисом* $\boldsymbol{E}_0$ (*но одноимённые между собой*).

Доказательство. 1. Поскольку $\boldsymbol{E} = \boldsymbol{E} \cdot \mathbb{1}$ и $\det \mathbb{1} = 1$, получаем $\boldsymbol{E} \simeq \boldsymbol{E}$.

2. Факт $\boldsymbol{E} \simeq \boldsymbol{F}$ означает, что $\boldsymbol{F} = \boldsymbol{E} \cdot C$, где $\det C > 0$. Но тогда $\boldsymbol{E} = \boldsymbol{F} \cdot C^{-1}$, где $\det C^{-1} = (\det C)^{-1} > 0$, т.е. $\boldsymbol{F} \simeq \boldsymbol{E}$.

3. Если $\boldsymbol{E} \simeq \boldsymbol{F}$ и $\boldsymbol{F} \simeq \boldsymbol{G}$, то $\boldsymbol{F} = \boldsymbol{E} \cdot C$, $\boldsymbol{G} = \boldsymbol{F} \cdot D$, где $\det C > 0$ и $\det D > 0$. Тогда $\boldsymbol{G} = (\boldsymbol{E} \cdot C) \cdot D = \boldsymbol{E} \cdot (CD)$, где $\det(CD) = \det C \cdot \det D > 0$, т.е. $\boldsymbol{E} \simeq \boldsymbol{G}$. □

6.90. Определение. Каждый из двух классов одноимённых базисов называется *ориентацией* векторного пространства; одна из них называется *положительной*, другая *отрицательной*. Векторное пространство называется *ориентированным*, если в нём зафиксирована одна из двух возможных ориентаций.

6.91. В случае пространств размерностей 1, 2 и 3 абстрактное понятие ориентации имеет наглядную интерпретацию, которая обсуждалась выше.

Отметим, что абстрактные термины «положительная ориентация» и «отрицательная ориентация» основаны лишь на алгебраических свойствах матриц перехода, в то время как термины «правый базис» и «левый базис» связаны с наглядными представлениями о предметах окружающего нас мира.

В. Деформации базисов. С интуитивной точки зрения два базиса одинаково ориентированы, если один из них может быть преобразован в другой непрерывной деформацией. Придадим этому наглядному описанию точный математический смысл.

Рассмотрим функцию

$$\boldsymbol{x}\colon [0,1] \to \mathcal{V}, \quad t \mapsto \boldsymbol{x}(t), \tag{6.25}$$

определённую на промежутке $[0,1]$ и принимающую значения в векторном пространстве $\mathcal{V}$; такую функцию естественно называть *вектор-функцией числового аргумента* t. Если в $\mathcal{V}$ задан произвольный базис $\boldsymbol{E} = \big(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\big)$, то при каждом значении $t \in [0,1]$ вектор $\boldsymbol{x}(t)$ можно разложить по этому базису, в результате чего получим n числовых функций $x^1(t), \dots, x^n(t)$, которые являются координатами вектор-функции $t \mapsto \boldsymbol{a}(t)$:

$$\boldsymbol{x}(t) = x^1(t)\boldsymbol{e}_1 + \dots + x^n(t)\boldsymbol{e}_n. \tag{6.26}$$

6.92. Определение. Вектор-функция (6.25) называется *непрерывной*, если все её координаты (6.26) относительно некоторого базиса являются непрерывными функциями. (Всюду под непрерывностью подразумевается непрерывность на $[0,1]$, т.е. непрерывность во внутренних точках и односторонняя непрерывность в граничных точках этого промежутка.)

6.93. Предложение. *Определение* 6.92 *корректно, т.е. не зависит от выбора базиса. Иными словами, если координаты вектор-функции относительно некоторого базиса являются непрерывными функциями, то координаты той же вектор-функции относительно любого другого базиса также будут непрерывными функциями.*

Доказательство. Пусть $\boldsymbol{F}$ — другой базис в $\mathcal{V}$, C — матрица перехода от базиса $\boldsymbol{E}$ к базису $\boldsymbol{F}$, так что $\boldsymbol{F} = \boldsymbol{E}C$. Разложения вектор-функции $\boldsymbol{x}(t)$ по базисам $\boldsymbol{E}$ и $\boldsymbol{F}$ имеют вид

$$\boldsymbol{x} = \boldsymbol{E}X(t), \quad \boldsymbol{x} = \boldsymbol{F}Y(t),$$

где $X(t) = \big(x^1(t), \dots, x^n(t)\big)^T$ и $Y = \big(y^1(t), \dots, y^n(t)\big)^T$ — столбцы координат. Таким образом, $\boldsymbol{E}X(t) = \boldsymbol{E}CY(t)$. Пользуясь единственностью разложения вектора по базису, заключаем, что $X(t) = CY(t)$. Таким образом, координаты вектора в одном базисе являются линейными комбинациями с постоянными коэффициентами его же координат в другом базисе и потому согласно известной теореме анализа также представляют собой непрерывные функции. $\square$

6.94. Определение. Будем говорить, что базис $\boldsymbol{F} = \bigl(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n\bigr)$ векторного пространства $\mathcal{V}$ получен из базиса $\boldsymbol{E} = \bigl(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\bigr)$ *непрерывной деформацией*, если существует такое семейство непрерывных вектор-функций $\boldsymbol{A}(t) = \bigl(\boldsymbol{a}_1(t), \ldots, \boldsymbol{a}_n(t)\bigr)$, что

(i) при каждом $t \in [0, 1]$ семейство $\boldsymbol{A}(t)$ является базисом,
(ii) $\boldsymbol{A}(0) = \boldsymbol{E}$, $\boldsymbol{A}(1) = \boldsymbol{F}$.

6.95. Предложение. *Если базис $\boldsymbol{F}$ получен из базиса $\boldsymbol{E}$ непрерывной деформацией, то эти два базиса одноимённы.*

Доказательство. Пусть $\Delta(t)$ — определитель матрицы перехода от базиса $\boldsymbol{E} = \boldsymbol{A}(0)$ к базису $\boldsymbol{A}(t)$. Функция $\Delta(t)$ непрерывна, $\Delta(0) = 1$ и $\Delta(t) \neq 0$ для всех $t \in [0, 1]$. Из курса анализа известно, что непрерывная функция, не обращающаяся в нуль на отрезке, сохраняет постоянный знак. Поэтому $\Delta(1) > 0$ и, следовательно, базисы $\boldsymbol{E}' = \boldsymbol{A}(1)$ и $\boldsymbol{E}$ одноимённы. $\square$

6.96. Условие одноимённости базисов, выраженное предложением 6.95, является не только достаточным, но и необходимым: если два базиса одноимённы, то каждый из них может быть получен из другого непрерывной деформацией. Заинтересованный читатель может найти доказательство этого утверждения, например, в замечательном учебнике М. М. Постникова [19].

ГЛАВА 7

Аффинные пространства

Аффинное пространство состоит из точек, которые можно соединять векторами; в таком пространстве уже возможно построить содержательную геометрию. Однако в отличие от пространства, изучаемого в школьном курсе элементарной геометрии, в аффинном пространстве не определены такие понятия, как расстояние между точками, длины линий, площади и объёмы фигур, величины углов и перпендикулярность. При исследовании фигур в аффинном пространстве рассматриваются лишь те их свойства, которые не зависят от этих метрических понятий.

1. Определение аффинного пространства

A. Аксиомы аффинного пространства.

7.1. Определение. Пусть $\mathcal{A}$ — непустое множество, элементы которого будем называть *точками* и обозначать прописными латинскими буквами $A, B, C, \ldots$, и $\mathcal{V}$ — векторное пространство над числовым полем $\mathbb{K}$. Предположим, что каждой упорядоченной паре точек (A, B) множества $\mathcal{A}$ поставлен в соответствие некоторый вектор из $\mathcal{V}$, который будем обозначать $\overrightarrow{AB}$, а точки A и B называть его началом и концом соответственно. Другими словами, задано отображение

$$\Phi\colon \mathcal{A} \times \mathcal{A} \to \mathcal{V}, \quad (A, B) \mapsto \overrightarrow{AB}. \tag{7.1}$$

Множество $\mathcal{A}$ называется *аффинным пространством*, если выполнены следующие условия, называемые *аксиомами аффинного пространства*:

АП1: для любой точки $A \in \mathcal{A}$ и любого вектора $\boldsymbol{v} \in \mathcal{V}$ существует единственная точка $B \in \mathcal{A}$, для которой $\overrightarrow{AB} = \boldsymbol{v}$; будем говорить, что точка B получена из точки A *параллельным переносом* на вектор $\boldsymbol{v}$ и писать $B = A + \boldsymbol{v}$;

АП2: для любых трёх точек $A, B, C \in \mathcal{A}$ имеет место равенство

$$\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC},$$

называемое *правилом треугольника*.

Векторное пространство $\mathcal{V}$ называют *ассоциированным* с аффинным пространством $\mathcal{A}$, а само $\mathcal{A}$ — аффинным пространством над числовым полем $\mathbb{K}$.

В большинстве случаев для обозначения аффинного пространства достаточно одного символа $\mathcal{A}$. В ситуациях, когда требуется указать ассоциированное векторное пространство и/или числовое поле, будем писать $(\mathcal{A}, \mathcal{V})$ или $(\mathcal{A}, \mathcal{V}, \mathbb{K})$.

Аксиомы **АП1** и **АП2** образуют *вторую группу аксиом Вейля* векторно-точечной аксиоматики геометрии.

7.2. С наглядной точки зрения аксиома **АП1** означает возможность отложить любой вектор от любой точки. Полагая в аксиоме **АП2** $A = B = C$, получим $\overrightarrow{AA} = \mathbf{0}$. Если же положить $C = A$, то получим $\overrightarrow{BA} = -\overrightarrow{AB}$.

7.3. Определение. Пусть $(\mathcal{A}, \mathcal{V})$ — аффинное пространство. Размерность $\dim \mathcal{V}$ его ассоциированного векторного пространства $\mathcal{V}$ называется *размерностью аффинного пространства* $\mathcal{A}$ и обозначается $\dim \mathcal{A}$.

7.4. Пример. Пусть $\mathcal{V}$ — произвольное векторное пространство. Определим аффинное пространство, полагая $\mathcal{A} = \mathcal{V}$ и $\overrightarrow{\boldsymbol{ab}} = \boldsymbol{b} - \boldsymbol{a}$ для любых $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{V}$. Проверим выполнение аксиом аффинного пространства:

АП1: для любой «точки» $A = \boldsymbol{a} \in \mathcal{A}$ и любого вектора $\boldsymbol{c} \in \mathcal{V}$ «точка» $\boldsymbol{b} = \boldsymbol{a} + \boldsymbol{c}$ является, очевидно, единственной точкой, для которой $\overrightarrow{\boldsymbol{ab}} = \boldsymbol{c}$;

АП2: для любых трёх точек $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathcal{A}$ имеет место равенство

$$\underbrace{\boldsymbol{c} - \boldsymbol{a}}_{\overrightarrow{\boldsymbol{ac}}} = \underbrace{(\boldsymbol{b} - \boldsymbol{a})}_{\overrightarrow{\boldsymbol{ab}}} + \underbrace{(\boldsymbol{c} - \boldsymbol{b})}_{\overrightarrow{\boldsymbol{bc}}}.$$

Таким образом, *любое векторное пространство $\mathcal{V}$ можно рассматривать как аффинное пространство*; в таком качестве будем обозначать его $\mathcal{V}_{\text{афф}}$. Главное отличие пространств $\mathcal{V}$ и $\mathcal{V}_{\text{афф}}$ в том, что в $\mathcal{V}$ имеется выделенный элемент $\mathbf{0}$ — «начало координат», а в $\mathcal{V}_{\text{афф}}$ все элементы равноправны, что позволяет рассматривать аффинное пространство как векторное пространство с «забытым» началом координат. Отсутствие «центра мира» (на современном языке — однородность пространства) было провозглашено Галилеем[1] в его главном труде — «Диалоге о двух главнейших системах мира: птолемеевой и коперниковой» (1632 г.).

[1]Г. Галилей (G. Galilei) — итальянский физик, механик, астроном, философ и математик (1564–1642).

7.5. В частности, арифметическое пространство $\mathbb{R}^n$ естественным образом является аффинным пространством, с которым ассоциировано векторное пространство $\mathbb{R}^n$. Таким образом, символ $\mathbb{R}^n$ обозначает два разных объекта: векторное пространство и аффинное пространство. Чтобы их различать, можно писать, например, $\mathbb{R}^n_{\text{век}}$ и $\mathbb{R}^n_{\text{афф}}$. Если $A = (a^1, \dots, a^n)^T$, $B = (b^1, \dots, b^n)^T$ — две точки из $\mathbb{R}^n_{\text{афф}}$, то вектор $\overrightarrow{AB} \in \mathbb{R}^n_{\text{век}}$ определяется формулой

$$\overrightarrow{AB} = (b^1 - a^1,\ \dots,\ b^n - a^n)^T$$

(ср. с п. 1.26, с. 17).

7.6. Раздел математики, изучающий аффинные пространства, называется *аффинной геометрией*. При выбранном нами подходе в этой геометрии имеются два первоначальных неопределяемых понятия — точка и вектор — и три основных неопределяемых отношения между этими понятиями:

(i) отношение между тремя векторами $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$, состоящее в том, что $\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b}$;

(ii) отношение между двумя векторами $\boldsymbol{a}$, $\boldsymbol{b}$ и числом α, состоящее в том, что $\boldsymbol{b} = \alpha \cdot \boldsymbol{a}$;

(iii) отношение между двумя точками A, B и вектором $\boldsymbol{a}$, состоящее в том, что $\boldsymbol{a} = \overrightarrow{AB}$.

Эти отношения должны удовлетворять восьми «векторным» аксиомам **ВП1**–**ВП8** (см. с. 95) и двум «аффинным» аксиомам **АП1**–**АП2**. Кроме того, если аффинная геометрия предназначается для описания окружающего нас мира, то должна быть указана размерность этого мира, т.е. должна выполняться одна из аксиом размерности **Р1**–**Р3** (см. с. 109).

7.7. Отметим, что в аффинном пространстве отсутствуют понятия длины вектора и угла между векторами. Аффинная геометрия представляет собой ту часть элементарной геометрии, в которой не используются измерения длин и углов. Эта часть довольно велика и содержит множество важных результатов. Однако некоторые виды измерений в аффинном пространстве производить всё-таки можно; этим вопросам посвящён раздел 6.

Б. Аффинная система координат.

7.8. Определение. *Аффинная система координат*, или *репéр*, в аффинном пространстве $(\mathcal{A}, \mathcal{V})$ — это пара, состоящая из точки $O \in \mathcal{A}$ (*начала координат*) и базиса $\boldsymbol{E} = (\boldsymbol{e}_1, \dots, \boldsymbol{e}_n)$ ассоциированного векторного пространства $\mathcal{V}$; обозначение $O\boldsymbol{E}$ или $O\boldsymbol{e}_1 \dots \boldsymbol{e}_n$.

На рис. 7.1 изображены аффинные системы координат на прямой, на плоскости и в пространстве.

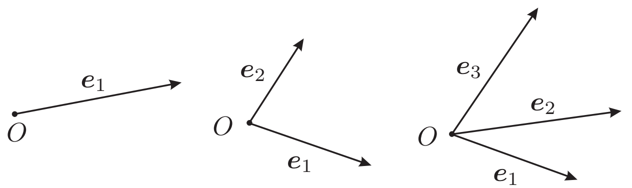


Рис. 7.1. Аффинные системы координат на прямой, на плоскости и в пространстве

7.9. Определение. *Радиус-вектором* точки $A \in \mathcal{A}$ аффинного пространства в аффинной системе координат $\boldsymbol{OE} = O\boldsymbol{e}_1 \ldots \boldsymbol{e}_n$ называется вектор $\overrightarrow{OA}$. *Координатами* точки $A \in \mathcal{A}$ в аффинной системе координат $\boldsymbol{OE}$ называются координаты радиус-вектора $\overrightarrow{OA}$ в базисе $\boldsymbol{E}$ векторного пространства $\mathcal{V}$.

7.10. Как и для векторного пространства, координаты точки принято записывать в виде столбца $X = (x^1, \ldots, x^n)^T$. Впрочем, в случаях, когда не используется матричная техника, вполне допустимо записывать координаты точек и векторов в виде строки. В дальнейшем записи $A(a^1, \ldots, a^n)$ и $\boldsymbol{a}(a^1, \ldots, a^n)$ означают, что точка A и вектор $\boldsymbol{a}$ имеют координаты $(a^1, \ldots, a^n)$, а запись $A(\boldsymbol{r})$ означает, что $\boldsymbol{r}$ является радиус-вектором точки A.

2. Прямые и плоскости в аффинном пространстве

В первую очередь нас интересуют аффинные пространства размерности $\leqslant 3$ над числовым полем $\mathbb{K}$, однако в рамках аксиоматики Вейля все геометрические конструкции совершенно идентичны в пространствах любой размерности.

7.11. Одномерное аффинное пространство $\mathcal{A}_1$ с ассоциированным векторным пространством $\mathbb{V}_1$ (см. пример 5.16, с. 96) естественно называть *аффинной прямой*. Аналогично, двумерное аффинное пространство $(\mathcal{A}_2, \mathbb{V}_2)$ называется *аффинной плоскостью*. Для трёхмерного аффинного пространства $(\mathcal{A}_3, \mathbb{V}_3)$ отдельного термина нет, что создаёт некоторые неудобства, усугубляемые ещё и тем, что слово «плоскость» может обозначать как двумерное аффинное пространство, так и множество точек в трёхмерном аффинном пространстве.

А. Мотивировка определений. Сначала проведём рассмотрение на наглядном интуитивном уровне, опираясь на школьные знания.

7.12. Начнём с прямых на плоскости или в пространстве. Любая прямая l полностью определяется своей произвольной точкой M_0 и произвольным ненулевым вектором $\boldsymbol{a}$, параллельным этой прямой (см. рис. 7.2). Условие, что некоторая точка M принадлежит прямой l, состоит в том, что вектор $\overrightarrow{M_0M}$ коллинеарен вектору $\boldsymbol{a}$, т.е. существует такое число t, что $\overrightarrow{M_0M} = t\boldsymbol{a}$.

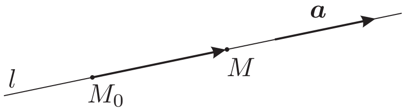


Рис. 7.2.

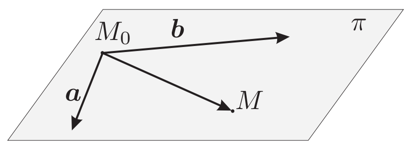


Рис. 7.3.

7.13. Аналогично, любая плоскость π в пространстве полностью определяется своей произвольной точкой M_0 и парой неколлинеарных векторов $\boldsymbol{a}$ и $\boldsymbol{b}$, параллельных этой плоскости (см. рис. 7.3). Условие, что некоторая точка M принадлежит плоскости π, состоит в том, что векторы $\overrightarrow{M_0M}$, $\boldsymbol{a}$ и $\boldsymbol{b}$ компланарны, т.е. существуют такие числа s и t, что $\overrightarrow{M_0M} = s\boldsymbol{a} + t\boldsymbol{b}$.

При аксиоматическом построении геометрии на базе системы аксиом Вейля мы положим эти свойства в основу определения.

Б. Прямые и отрезки.

7.14. Определение. *Прямой* в аффинном пространстве $\mathcal{A}$, задаваемой точкой $M_0 \in \mathcal{A}$ (*опорной точкой*) и ненулевым вектором $\boldsymbol{a} \in \mathcal{V}$ (*направляющим вектором*), называется множество всех точек $M \in \mathcal{A}$, для которых вектор $\overrightarrow{M_0M}$ лежит в линейной оболочке $L(\boldsymbol{a})$ направляющего вектора. Иными словами, для любой точки M прямой существует такое число $t \in \mathbb{R}$, что

$$\overrightarrow{M_0M} = t\boldsymbol{a}. \tag{7.2}$$

Прямую с опорной точкой M_0 и направляющим вектором $\boldsymbol{a}$ будем обозначать $l(M_0, \boldsymbol{a})$:

$$l(M_0, \boldsymbol{a}) \overset{\text{def}}{=\!=} \Big\{ M : \overrightarrow{M_0M} = t\boldsymbol{a},\ t \in \mathbb{R} \Big\}.$$

7.15. Если в аффинном пространстве $\mathcal{A}$ выбрано произвольным образом начало координат O, то соотношение (7.2), определяющее прямую, можно записать в виде

$$\boldsymbol{r} = \boldsymbol{r}_0 + t\boldsymbol{a}, \tag{7.3}$$

где $\boldsymbol{r}_0 = \overrightarrow{OM_0}$, $\boldsymbol{r} = \overrightarrow{OM}$ — радиус-векторы точек M_0, M относительно начала координат O. Равенство (7.3) называется *векторным параметрическим уравнением прямой.*

7.16. Предложение. *Каждая прямая $l(M_0, \boldsymbol{a})$ в произвольном аффинном пространстве $(\mathcal{A}, \mathcal{V})$ сама является одномерным аффинным пространством, ассоциированное векторное пространство которого — линейная оболочка $L(\boldsymbol{a})$ направляющего вектора прямой.*

Доказательство. Действительно, все условия определения 7.1 выполнены. □

7.17. Предложение. *Пусть N_0 — произвольная точка прямой $l(M_0, \boldsymbol{a})$, $\boldsymbol{b}$ — произвольный вектор, коллинеарный вектору $\boldsymbol{a}$. Тогда $l(N_0, \boldsymbol{b}) = l(M_0, \boldsymbol{a})$. Иными словами, в качестве опорной точки прямой можно выбрать любую её точку, а в качестве направляющего вектора — любой ненулевой вектор, коллинеарный прямой.*

Доказательство. Условие $N_0 \in l(M_0, \mathcal{P})$ означает, что $\overrightarrow{M_0N_0} = t\boldsymbol{a}$. Требуется доказать совпадение множеств

$$l(M_0, \boldsymbol{a}) = \Big\{ M \in \mathcal{A} \;\Big|\; \overrightarrow{M_0M} = t\boldsymbol{a},\ t \in \mathbb{R} \Big\},$$
$$l(N_0, \boldsymbol{b}) = \Big\{ M \in \mathcal{A} \;\Big|\; \overrightarrow{N_0M} = s\boldsymbol{b},\ s \in \mathbb{R} \Big\}.$$

Линейные оболочки векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ совпадают; обозначим их через $\mathcal{P} = L(\boldsymbol{a}) = L(\boldsymbol{b})$. Пусть $M \in l(M_0, \boldsymbol{a})$, т.е. $\overrightarrow{M_0M} \in \mathcal{P}$. Тогда

$$\overrightarrow{N_0M} = \underbrace{\overrightarrow{M_0M}}_{\in \mathcal{P}} - \underbrace{\overrightarrow{M_0N_0}}_{\in \mathcal{P}} \in \mathcal{P},$$

т.е. $M \in l(N_0, \boldsymbol{b})$. Обратно, если $M \in l(N_0, \boldsymbol{b})$, т.е. $\overrightarrow{N_0M} \in \mathcal{P}$, то

$$\overrightarrow{M_0M} = \underbrace{\overrightarrow{M_0N_0}}_{\in \mathcal{P}} + \underbrace{\overrightarrow{N_0M}}_{\in \mathcal{P}} \in \mathcal{P} \quad \Rightarrow \quad M \in l(M_0, \boldsymbol{a}).$$ □

Легко доказать следующее утверждение, которое в рамках аксиоматики Гильберта представляет собой одну из аксиом.

7.18. Предложение. *Через любые две различные точки M_0 и M_1 аффинного пространства проходит единственная прямая; её обозначают (M_0M_1).*

Доказательство. Существование такой прямой гарантируется определением: если взять в качестве опорной точку M_0, а в качестве направляющего вектора — вектор $\overrightarrow{M_0M_1}$, то множество всех точек M искомой прямой описывается соотношением $\overrightarrow{M_0M} = t\overrightarrow{M_0M_1}$, $t \in \mathbb{K}$. В частности, сами точки M_0 и M_1 получаются при $t = 0$ и $t = 1$ соответственно. Единственность прямой вытекает из предложения 7.17. □

7.19. Все предыдущие рассмотрения справедливы для аффинного пространства над любым числовым полем $\mathbb{K}$. В вещественном случае, т.е. $\mathbb{K} = \mathbb{R}$, можно ввести следующие определения.

7.20. Определение. Говорят, что точка M прямой (M_0M_1) *лежит между точками* M_0 *и* M_1, если соответствующее этой точке значение параметра t в равенстве $\overrightarrow{M_0M} = t\overrightarrow{M_0M_1}$ удовлетворяет условию $0 < t < 1$.

7.21. Отношение «лежать между» является одним из основных (неопределяемых) отношений в системе аксиом Гильберта (отношение инцидентности); в рамках аксиоматики Вейля оно порождается отношением порядка $<$ на множестве вещественных чисел.

7.22. Определение. Множество всех точек прямой (M_0M_1), лежащих между точками M_0 и M_1, вместе с самими этими точками называется *отрезком* с концами M_0 и M_1 и обозначается $[M_0M_1]$:

$$[M_0M_1] \stackrel{\text{def}}{=\!=} \left\{M \;\middle|\; \overrightarrow{M_0M} = t\overrightarrow{M_0M_1},\ 0 \leqslant t \leqslant 1\right\}.$$

B. Плоскости.

7.23. Определение. *Плоскостью* в аффинном пространстве $\mathcal{A}$, задаваемой точкой M_0 (*опорной точкой*) и двумя неколлинеарными векторами $\boldsymbol{a}$, $\boldsymbol{b}$ (*направляющими векторами*), называется множество всех точек $M \in \mathcal{A}$, для которых вектор $\overrightarrow{M_0M}$ лежит в линейной оболочке $L(\boldsymbol{a}, \boldsymbol{b})$ направляющих векторов. Иными словами, для любой точки M плоскости существует такие числа $s, t \in \mathbb{R}$, что

$$\overrightarrow{M_0M} = s\boldsymbol{a} + t\boldsymbol{b}. \tag{7.4}$$

Плоскость с опорной точкой M_0 и направляющими векторами $\boldsymbol{a}$, $\boldsymbol{b}$ будем обозначать $\pi(M_0, \boldsymbol{a}, \boldsymbol{b})$:

$$\pi(M_0, \boldsymbol{a}, \boldsymbol{b}) = \left\{M : \overrightarrow{M_0M} = s\boldsymbol{a} + t\boldsymbol{b},\ s, t \in \mathbb{R}\right\}.$$

7.24. Если в аффинном пространстве $\mathcal{A}$ выбрано произвольным образом начало координат O, то соотношение (7.4), определяющее плоскость, можно записать в виде

$$\boldsymbol{r} = \boldsymbol{r}_0 + s\boldsymbol{a} + t\boldsymbol{b}, \tag{7.5}$$

где $\boldsymbol{r}_0 = \overrightarrow{OM_0}$, $\boldsymbol{r} = \overrightarrow{OM}$ — радиус-векторы точек M_0, M относительно начала координат O. Равенство (7.5) называется *векторным параметрическим уравнением плоскости.*

7.25. Предложение (ср. предложение 7.16). *Каждая плоскость $\pi(M_0, \boldsymbol{a}, \boldsymbol{b})$ в произвольном аффинном пространстве $(\mathcal{A}, \mathcal{V})$ сама является двумерным аффинным пространством, ассоциированное векторное пространство которого — линейная оболочка $L(\boldsymbol{a}, \boldsymbol{b})$ направляющих векторов плоскости.*

7.26. Для плоскостей справедлив аналог предложения 7.17: *если N_0 — произвольная точка плоскости $\pi(M_0, \boldsymbol{a}, \boldsymbol{b})$ и $\boldsymbol{a}', \boldsymbol{b}'$ — произвольные линейно независимые векторы линейной оболочки $L(\boldsymbol{a}, \boldsymbol{b})$, то $\pi(M_0, \boldsymbol{a}', \boldsymbol{b}') = \pi(N_0, \boldsymbol{a}, \boldsymbol{b})$.*

3. Многомерные плоскости

А. p-Плоскости.

7.27. Определение. Пусть $M_0 \in \mathcal{A}$ — некоторая фиксированная точка n-мерного аффинного пространства $\mathcal{A}$, $\mathcal{P} \Subset \mathcal{V}$ — подпространство его ассоциированного векторного пространства, $\dim \mathcal{P} = p$. Множество всех точек $M \in \mathcal{A}$, для которых вектор $\overrightarrow{M_0M}$ принадлежит подпространству $\mathcal{P}$, называется *плоскостью размерности p* (*p-плоскостью*), проходящей через точку M_0 параллельно подпространству $\mathcal{P}$, и обозначается $\pi(M_0, \mathcal{P})$:

$$\pi(M_0, \mathcal{P}) \stackrel{\text{def}}{=\!=} \left\{ M \in \mathcal{A} \;\middle|\; \overrightarrow{M_0M} \in \mathcal{P} \right\}. \tag{7.6}$$

Подпространство $\mathcal{P}$ называется *направляющим подпространством* плоскости π, а точка M_0 — *опорной точкой.* Число $n - p$ называется *коразмерностью* плоскости π и обозначается $\operatorname{codim} \pi$.

7.28. Плоскости размерности 1 называются *прямыми.* Плоскости коразмерности 1 называются *гиперплоскостями.*

Согласно определению 7.27 точки пространства $\mathcal{A}$ представляют собой 0-плоскости.

7.29. Предложение. *Каждая p-плоскость в аффинном пространстве сама является аффинным пространством размерности p.*

7.30. Предложение. *Если $N_0 \in \pi(M_0, \mathcal{P})$, то $\pi(N_0, \mathcal{P}) = \pi(M_0, \mathcal{P})$, т.е. в качестве опорной точки может быть выбрана любая точка p-плоскости.*

Доказательства предложений 7.29 и 7.30 совершенно аналогичны доказательствам предложений 7.25 и 7.17 соответственно.

Б. Параллельные плоскости.

7.31. Определение. Пусть $\pi_1(M_1, \mathcal{P}_1)$ и $\pi_2(M_2, \mathcal{P}_2)$ — две многомерные плоскости в аффинном пространстве $\mathcal{A}$, имеющие размерности p_1 и p_2 соответственно, $p_1 \leqslant p_2$. Плоскости π_1 и π_2 называются *параллельными*, если $\mathcal{P}_1 \Subset \mathcal{P}_2$.

7.32. Возможна ситуация, когда параллельные плоскости не имеют общих точек. Если же общие точки имеются, то плоскость меньшей размерности целиком содержится в плоскости большей размерности (в случае равных размерностей эти плоскости совпадают).

7.33. Предложение. *Пусть параллельные плоскости* π_1 *и* π_2 *имеют общую точку, причём* $\dim \pi_1 \leqslant \dim \pi_2$. *Тогда* $\pi_1 \subseteq \pi_2$.

Доказательство. Обозначим через A общую точку параллельных плоскостей $\pi_1 = \pi_1(M_1, \mathcal{P}_1)$ и $\pi_2 = \pi_2(M_2, \mathcal{P}_2)$:

$$A \in \pi_1 \;\Leftrightarrow\; \overrightarrow{M_1A} \in \mathcal{P}_1, \qquad A \in \pi_2 \;\Leftrightarrow\; \overrightarrow{M_2A} \in \mathcal{P}_2.$$

Параллельность означает, что $\mathcal{P}_1 \Subset \mathcal{P}_2$. Пусть X — произвольная точка из π_1, т.е. $\overrightarrow{M_1X} \in \mathcal{P}_1$. Тогда

$$\overrightarrow{AX} = \underbrace{\overrightarrow{M_1X}}_{\in \mathcal{P}_1} - \underbrace{\overrightarrow{M_1A}}_{\in \mathcal{P}_1} \in \mathcal{P}_1 \Subset \mathcal{P}_2 \quad \Rightarrow \quad \overrightarrow{AX} \in \mathcal{P}_2,$$

откуда

$$\overrightarrow{M_2X} = \underbrace{\overrightarrow{M_2A}}_{\in \mathcal{P}_2} + \underbrace{\overrightarrow{AX}}_{\in \mathcal{P}_2} \in \mathcal{P}_2 \quad \Rightarrow \quad X \in \pi_2. \qquad \square$$

B. Пересекающиеся плоскости.

7.34. Предложение. *Если две плоскости пересекаются, то их пересечением является плоскость* (см. рис. 7.4).

Доказательство. Рассмотрим две пересекающиеся плоскости $\pi_1(M_1, \mathcal{P}_1)$ и $\pi_2(M_2, \mathcal{P}_2)$ и обозначим через π их пересечение: $\pi = \pi_1 \cap \pi_2$. Пусть $M_3 \in \pi$. Если π не содержит других точек, кроме M_3, то утверждение доказано: π является 0-плоскостью. Предположим, что кроме M_3 имеются и другие точки пересечения, и проверим, что π является плоскостью с опорной точкой M_3 и направляющим подпространством $\mathcal{P}_1 \cap \mathcal{P}_2$, т.е. для любой точки $X \in \pi$ вектор $\overrightarrow{M_3X}$ лежит в $\mathcal{P}_1 \cap \mathcal{P}_2$. Поскольку $X \in \pi$, получаем

$$X \in \pi_1 \;\Leftrightarrow\; \overrightarrow{M_1X} \in \mathcal{P}_1,$$
$$X \in \pi_2 \;\Leftrightarrow\; \overrightarrow{M_2X} \in \mathcal{P}_2.$$

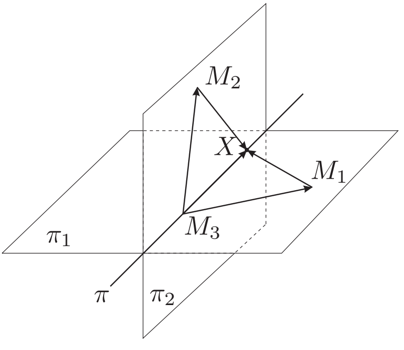


Рис. 7.4. Пересекающиеся плоскости

Далее,

$$\overrightarrow{M_3X} = \overrightarrow{M_3M_1} + \overrightarrow{M_1X} \in \mathcal{P}_1, \quad \overrightarrow{M_3X} = \overrightarrow{M_3M_2} + \overrightarrow{M_2X} \in \mathcal{P}_2,$$

откуда $\overrightarrow{M_3X} \in \mathcal{P}_1 \cap \mathcal{P}_2$. $\square$

Г. Скрещивающиеся плоскости. Скрещивающиеся плоскости — это плоскости, которые не пересекаются и не параллельны.

7.35. Определение. Рассмотрим в аффинном пространстве $(\mathcal{A}, \mathcal{V})$ две непараллельные[1] плоскости $\pi_1(M_1, \mathcal{P}_1)$ и $\pi_2(M_2, \mathcal{P}_2)$ размерностей p_1 и p_2 соответственно, причём $p_1 \leqslant p_2$. Если эти плоскости не имеют общих точек, будем говорить, что они *скрещиваются вдоль подпространства* $\mathcal{P}_1 \cap \mathcal{P}_2$.

7.36. В двумерном аффинном пространстве (т.е. на плоскости) не бывает скрещивающихся 1-плоскостей (прямых). В трёхмерном аффинном пространстве существуют 1-плоскости, скрещивающиеся вдоль нульмерных подпространств — обычные скрещивающиеся прямые.

Д. Параметрическое уравнение плоскости.

7.37. Пусть в аффинном пространстве $\mathcal{A}$ выбрана произвольная точка O — начало координат. Для того чтобы записать уравнение p-плоскости $\pi(M_0, \mathcal{P})$ с опорной точкой M_0 и направляющим подпространством $\mathcal{P}$, выберем в $\mathcal{P}$ произвольный базис $\big(\boldsymbol{a}_1, \dots, \boldsymbol{a}_p\big)$ (векторы этого базиса назовём *направляющими векторами* плоскости π) и обозначим через $\boldsymbol{r}_0$ и $\boldsymbol{r}$ радиус-векторы точек M_0 и M относительно начала координат O: $\boldsymbol{r}_0 = \overrightarrow{OM_0}$, $\boldsymbol{r} = \overrightarrow{OM}$. Принимая во внимание тот факт, что $\overrightarrow{M_0M} = \boldsymbol{r} - \boldsymbol{r}_0$, можем записать определяющее соотношение $\overrightarrow{M_0M} \in \mathcal{P}$ плоскости π (см. (7.6)) в виде

$$\boldsymbol{r} = \boldsymbol{r}_0 + t^1\boldsymbol{a}_1 + \dots + t^p\boldsymbol{a}_p. \tag{7.7}$$

При изменении параметров $t^1, \dots, t^p$ в поле $\mathbb{K}$ точка M с радиус-вектором $\boldsymbol{r}$, задаваемым формулой (7.7), пробегает всю рассматриваемую p-плоскость, поэтому равенство (7.7) называется *векторным параметрическим уравнением плоскости.*

7.38. Если ввести в аффинном пространстве аффинный репер $O\boldsymbol{e}_1 \dots \boldsymbol{e}_n$, то векторное параметрическое уравнение плоскости (7.7) примет вид

$$X = X_0 + t^1A_1 + \dots + t^pA_p, \tag{7.8}$$

где X, X_0 — столбцы координат точек M и M_0 (или радиус-векторов $\boldsymbol{r}$ и $\boldsymbol{r}_0$), $A_1, \dots, A_p$ — столбцы координат направляющих векторов $\boldsymbol{a}_1, \dots, \boldsymbol{a}_p$:

$$X = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix}, \quad X_0 = \begin{pmatrix} x_0^1 \\ x_0^2 \\ \vdots \\ x_0^n \end{pmatrix}, \quad A_1 = \begin{pmatrix} a_1^1 \\ a_1^2 \\ \vdots \\ a_1^n \end{pmatrix}, \quad \dots, \quad A_p = \begin{pmatrix} a_p^1 \\ a_p^2 \\ \vdots \\ a_p^n \end{pmatrix}.$$

В развёрнутой координатной форме уравнение (7.8) имеет вид системы

$$\begin{cases} x^1 = x_0^1 + t^1a_1^1 + \dots + t^ka_k^1, \\ x^2 = x_0^2 + t^1a_1^2 + \dots + t^ka_k^2, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ x^n = x_0^n + t^1a_1^n + \dots + t^ka_k^n, \end{cases} \qquad t^j \in \mathbb{K}, \ j = 1, 2, \dots, p.$$

[1] Непараллельность означает, что $\mathcal{P}_1 \notin \mathcal{P}_2$.

E. Задание плоскости системой уравнений.

7.39. Структура параметрического уравнения плоскости (7.7) или (7.8) в точности совпадает со структурой общего решения 4.3 (см. с. 67) неоднородной системы линейных уравнений, где $\boldsymbol{r}_0$ играет роль частного решения неоднородной системы, а линейная комбинация $t^1\boldsymbol{a}_1 + \dots + t^p\boldsymbol{a}_p \in \mathcal{P}$ — роль общего решения однородной системы, определяющей направляющее подпространство $\mathcal{P}$ нашей p-плоскости. Это совпадение не случайно.

7.40. Предложение. *В n-мерном аффинном пространстве $\mathcal{A}$ в любых аффинных координатах всякая p-плоскость может быть задана неоднородной системой линейных уравнений ранга*[1] $r = n - p$.

Обратите внимание, что коразмерность плоскости совпадает с рангом задающей её неоднородной системы уравнений.

Доказательство. Пусть $O\boldsymbol{E}$ — система аффинных координат в аффинном пространстве $(\mathcal{A}, \mathcal{V})$, где O — произвольная точка в $\mathcal{A}$, $\boldsymbol{E}$ — произвольный базис пространства $\mathcal{V}$. Рассмотрим p-плоскость $\pi(M_0, \mathcal{P})$, проходящую через точку M_0 в направлении подпространства $\mathcal{P}$. В указанной системе координат плоскость π задаётся параметрическими уравнениями (7.7) или (7.8). Рассмотрим другую систему координат $M_0\boldsymbol{E}$, начало которой расположено в точке M_0. Радиус-векторы $\boldsymbol{r}'$ и $\boldsymbol{r}$ произвольной точки $M \in \pi$ относительно начал координат M_0 и O связаны соотношением

$$\boldsymbol{r}' = \overrightarrow{M_0M} = \overrightarrow{OM} - \overrightarrow{OM_0} = \boldsymbol{r} - \boldsymbol{r}_0,$$

т.е. столбцы их координат — соотношением

$$X' = X - X_0,$$

где X', X и X_0 — столбцы координат векторов $\boldsymbol{r}'$, $\boldsymbol{r}$, $\boldsymbol{r}_0$ в базисе $\boldsymbol{E}$. По определению плоскости π вектор $\boldsymbol{r}' = \overrightarrow{M_0M}$ лежит в направляющем подпространстве $\mathcal{P}$ плоскости π. Поскольку любое p-мерное подпространство n-мерного векторного пространства может быть задано как множество решений некоторой однородной системы уравнений ранга $r = n - p$, заключаем, что столбец X' удовлетворяет однородной системе $AX' = 0$, где $A \in \mathbb{K}^{r \times n}$ — матрица этой системы. Подставляя сюда вместо X' его выражение $X' = X - X_0$, находим

$$A(X - X_0) = 0 \;\Leftrightarrow\; AX = B, \quad \text{где } B = AX_0.$$

Поскольку X — столбец координат произвольной точки плоскости π в системе координат $O\boldsymbol{E}$, получаем, что найденная система $AX = B$ и определяет плоскость π. □

7.41. Таким образом, аналогично подпространствам векторных пространств (см. п. 5.34, с. 101), плоскости в аффинных пространствах можно задавать двумя взаимно двойственными способами:

(i) задавая опорную точку и базис направляющего подпространства;

(ii) задавая неоднородную систему линейных уравнений.

7.42. Гиперплоскости — это плоскости коразмерности 1, поэтому они могут быть заданы неоднородной системой ранга 1, т.е. одним уравнением вида

$$A_1x^1 + A_2x^2 + \dots + A_nx^n = B. \tag{7.9}$$

---

[1] Под рангом системы подразумевается ранг её матрицы, основной или расширенной (эти ранги для совместной системы совпадают в силу теоремы Кронекера—Капелли).

Ж. Полупространства. В этом разделе мы рассматриваем *вещественное* аффинное пространство $(\mathcal{A}, \mathbb{R})$.

7.43. Будем говорить, что гиперплоскость π *разделяет* две (различные) точки $M_1, M_2 \in \mathcal{A}$, если отрезок $[M_1M_2]$ имеет общие точки с гиперплоскостью π. Аналогично, гиперплоскость π *не разделяет* две точки M_1 и M_2, если эти точки либо совпадают, либо отрезок $[M_1M_2]$ не имеет общих точек с гиперплоскостью π.

7.44. Найдём аналитическое условие того, что данная гиперплоскость π разделяет данные точки. Пусть в некоторой аффинной системе координат гиперплоскость π задана уравнением (7.9), а точки M_1 и M_2 имеют координаты $M_1(x_1^1, \dots, x_1^n)$ и $M_2(x_2^1, \dots, x_2^n)$. Для сокращения записи введём обозначение

$$F(M) \equiv F(x^1, \dots, x^n) = A_1x^1 + A_2x^2 + \dots + A_nx^n - B,$$

подразумевая, что точка M имеет координаты $(x^1, \dots, x^n)$.

7.45. Предложение. *Точки M_1 и M_2, не принадлежащие гиперплоскости* (7.9)*, разделены этой гиперплоскостью тогда и только тогда, когда $F(M_1)F(M_2) \leqslant 0$.*

Доказательство. Уравнение прямой (M_1M_2) имеет вид $\boldsymbol{r} = \boldsymbol{r}_1 + t(\boldsymbol{r}_2 - \boldsymbol{r}_1)$; запишем его в виде $M = M_1 + t(M_2 - M_1)$. Найдём общие точки этой прямой и гиперплоскости $F(M) = 0$:

$$0 = F(M) = F\Big(M_1 + t(M_2 - M_1)\Big) = F(M_1) + t\Big(F(M_2) - F(M_1)\Big)$$

$$t_0 = \frac{F(M_1)}{F(M_1) - F(M_2)}, \tag{7.10}$$

т.е. общая точка прямой и гиперплоскости есть $M = M_1 + t_0(M_2 - M_1)$. (Заметим, что в случае $F(M_1) = F(M_2)$ решения не существует, т.е. прямая (M_1M_2) параллельна гиперплоскости (7.9).)

По определению, точки M_1 и M_2 разделены гиперплоскостью (7.9), если число (7.10) существует и лежит в промежутке $[0; 1]$, т.е.

$$0 \leqslant \frac{F(M_1)}{F(M_1) - F(M_2)} \leqslant 1 \;\Leftrightarrow\; \begin{cases} \dfrac{F(M_1)}{F(M_1) - F(M_2)} \geqslant 0, \\[2ex] \dfrac{F(M_2)}{F(M_1) - F(M_2)} \leqslant 0. \end{cases}$$

При $F(M_1) > F(M_2)$ это возможно тогда и только тогда, когда $F(M_1) \geqslant 0$ и $F(M_2) \leqslant 0$; при $F(M_1) < F(M_2)$ — тогда и только тогда, когда $F(M_1) \leqslant 0$ и $F(M_2) \geqslant 0$. В обоих случаях выполняется неравенство $F(M_1)F(M_2) \leqslant 0$. $\square$

7.46. Из доказанной теоремы вытекает, что отношение *неразделённости* точек, заданное на множестве $\mathcal{A} \setminus \pi$ (т.е. на множестве всех точек,не лежащих на гиперплоскости π), является отношением эквивалентности. Для данной точки M_1 её класс эквивалентности состоит из всех точек M, для которых $F(M)F(M_1) > 0$. Таким образом, по отношению к данной гиперплоскости π все точки пространства, не лежащие на π, разбиваются на два класса эквивалентности, которые называются *открытыми полупространствами* с границей π, или *сторонами гиперплоскости* π. Множества

$$\pi_+ = \Big\{M \in \mathcal{A} \;\Big|\; F(M) \geqslant 0\Big\}, \quad \pi_- = \Big\{M \in \mathcal{A} \;\Big|\; F(M) \leqslant 0\Big\}$$

называются *замкнутыми полупространствами*; гиперплоскость π, заданная уравнением $F(M) = 0$, называется их *границей*. Говорят также, что гиперплоскость π разбивает пространство $\mathcal{A}$ на два полупространства π_+ и π_-.

7.47. В частности, точка разбивает прямую на две *полупрямые* (их также называют *лучами*), а прямая разбивает плоскость на две *полуплоскости*.

4. Двумерная аффинная геометрия

В этом разделе будем использовать терминологию элементарной геометрии: двумерное аффинное пространство называть плоскостью, а плоскости размерности 1 — прямыми.

В двумерном случае удобно использовать следующие обозначения для координат точек и векторов: (x, y) вместо (x^1, x^2), (a_x, a_y) вместо (a^1, a^2) и т. п. Базисные векторы координатной системы будем обозначать $\boldsymbol{i}, \boldsymbol{j}$.

А. Различные виды уравнений прямой на плоскости.

7.48. Выбрав в двумерном аффинном пространстве $\mathcal{A}$ произвольную аффинную систему координат $O\boldsymbol{ij}$ и введя координаты (x, y), (x_0, y_0), (a_x, a_y) векторов $\boldsymbol{r}$, $\boldsymbol{r}_0$, $\boldsymbol{a}$ соответственно, перепишем векторное параметрическое уравнение (7.3) в виде

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + t \begin{pmatrix} a_x \\ a_y \end{pmatrix} \quad \Longleftrightarrow \quad \begin{cases} x = x_0 + ta_x, \\ y = y_0 + ta_y. \end{cases} \tag{7.11}$$

Полученная система называется *системой параметрических уравнений прямой*. Часто говорят «*параметрическое уравнение прямой*», помня, что на самом деле уравнений два. При решении задач параметрическое уравнение прямой — один из наиболее удобных типов уравнений.

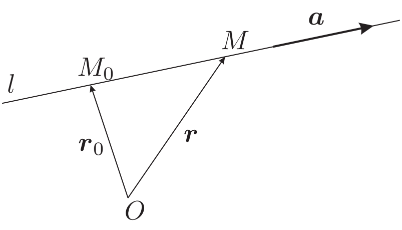


Рис. 7.5.

7.49. Попытаемся исключить из параметрического уравнения параметр t. Если обе координаты a_x и a_y направляющего вектора $\boldsymbol{a}$ отличны от нуля, то очевидным образом получаем

$$\frac{x - x_0}{a_x} = \frac{y - y_0}{a_y}. \tag{7.12}$$

Это уравнение называется *каноническим уравнением прямой* на плоскости. Разумеется, его можно всячески преобразовывать по обычным алгебраическим правилам, но чаще всего при решении практических задач бывает удобно перейти от данного канонического уравнения к параметрическому уравнению и в дальнейшем пользоваться именно им:

$$\frac{x-x_0}{a_x} = \frac{y-y_0}{a_y} \;\Leftrightarrow\; \frac{x-x_0}{a_x} = \frac{y-y_0}{a_y} = t \;\Leftrightarrow\; \begin{cases} x = x_0 + ta_x, \\ y = y_0 + ta_y. \end{cases}$$

7.50. Если же одна из координат направляющего вектора $\boldsymbol{a}$ равна нулю, например, $a_x = 0$, то имеем

$$\begin{cases} x = x_0, \\ y = y_0 + ta_y. \end{cases}$$

Эту систему кратко записывают в виде

$$\frac{x-x_0}{0} = \frac{y-y_0}{a_y}.$$

Естественно, здесь не имеется в виду деление на нуль; данное уравнение подлежит преобразованию в параметрическое уравнение по рецепту, указанному выше:

$$\frac{x-x_0}{0} = \frac{y-y_0}{a_y} \;\Leftrightarrow\; \frac{x-x_0}{0} = \frac{y-y_0}{a_y} = t \;\Leftrightarrow\; \begin{cases} x = x_0, \\ y = y_0 + ta_y. \end{cases}$$

Можно поступить и иначе:

$$\frac{x-x_0}{0} = \frac{y-y_0}{a_y} \;\Leftrightarrow\; a_y(x-x_0) = 0(y-y_0) \;\Leftrightarrow\; \begin{cases} x = x_0, \\ y \text{ — любой.} \end{cases}$$

Итак, каноническое уравнение прямой допускается записывать и в случае, когда одна из координат направляющего вектора прямой равна нулю.

7.51. Условие коллинеарности векторов $\overrightarrow{M_0M}$ и $\boldsymbol{a}$ может быть записано как равенство нулю определителя второго порядка, составленного из координат этих векторов (см. предложение 6.11, с. 121):

$$\begin{vmatrix} x-x_0 & y-y_0 \\ a_x & a_y \end{vmatrix} = 0. \tag{7.13}$$

Это уравнение также называют *каноническим уравнением* прямой на плоскости.

7.52. Чтобы составить уравнение прямой, проходящей через две заданные точки $M_1(\boldsymbol{r}_1)$ и $M_2(\boldsymbol{r}_2)$, заметим, что в качестве опорной точки прямой можно взять любую из них (возьмём M_1), а в качестве направляющего вектора прямой — вектор $\overrightarrow{M_1M_2} = \boldsymbol{r}_2 - \boldsymbol{r}_1$.

Тогда получим параметрические уравнения

$$\boldsymbol{r} = \boldsymbol{r}_1 + t(\boldsymbol{r}_2 - \boldsymbol{r}_1), \qquad \begin{cases} x = x_1 + t(x_2 - x_1), \\ y = y_1 + t(y_2 - y_1) \end{cases}$$

и канонические уравнения

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}, \qquad \begin{vmatrix} x - x_1 & y - y_1 \\ x_2 - x_1 & y_2 - y_1 \end{vmatrix} = 0,$$

где (x_1, y_1) и (x_2, y_2) — координаты векторов $\boldsymbol{r}_1$ и $\boldsymbol{r}_2$.

7.53. Каноническое уравнение прямой можно преобразовать следующим образом:

$$\frac{x - x_0}{a_x} = \frac{y - y_0}{a_y} \qquad \Leftrightarrow \qquad a_y(x - x_0) = a_x(y - y_0);$$

введя обозначения $A = a_y$, $B = -a_x$, получим уравнение

$$A(x - x_0) + B(y - y_0) = 0,$$

называемое *общим уравнением прямой, проходящей через заданную точку* (x_0, y_0). Раскрыв скобки и введя обозначение $D = Ax_0 + By_0$, получим так называемое *общее уравнение прямой на плоскости*

$$Ax + By = D. \tag{7.14}$$

7.54. Согласно п. 7.42 это уравнение представляет собой уравнение гиперплоскости: прямая в двумерном аффинном пространстве (т.е. на плоскости) является гиперплоскостью.

7.55. Если $D = 0$, то, очевидно, прямая проходит через начало координат. Если же $D \neq 0$, то, разделив обе части последнего уравнения на D, получим уравнение вида

$$\frac{x}{a} + \frac{y}{b} = 1,$$

где $a = D/A$, $b = D/B$. Геометрический смысл коэффициентов a и b достаточно очевиден: прямая проходит через точки $(a, 0)$ и $(0, b)$, отчего последнее уравнение носит название *уравнения прямой в отрезках* (a и b — отрезки, отсекаемые нашей прямой на осях координат[1]).

---

[1] *Осями координат* называют прямые, проходящие через начало координат и параллельные базисным векторам.

Б. Взаимное расположение прямых на плоскости.

7.56. Пусть на плоскости, т.е. в двумерном аффинном пространстве $\mathcal{A}$, имеются две прямые, заданные векторными параметрическими уравнениями

$$\boldsymbol{r} = \boldsymbol{r}_1 + t\boldsymbol{a}_1, \quad \boldsymbol{r} = \boldsymbol{r}_2 + s\boldsymbol{a}_2;$$

радиус-векторы всех точек рассматриваются относительно некоторого зафиксированного начала координат O. Исследуем вопрос об общих точках этих прямых. Приравнивая правые части уравнений, получаем

$$\boldsymbol{r}_1 + t\boldsymbol{a}_1 = \boldsymbol{r}_2 + s\boldsymbol{a}_2 \ \Leftrightarrow \ t\boldsymbol{a}_1 - s\boldsymbol{a}_2 = \boldsymbol{r}_2 - \boldsymbol{r}_1.$$

7.57. Если векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ линейно независимы (неколлинеарны), то они образуют базис в ассоциированном векторном пространстве $\mathcal{V}$, и потому вектор $\boldsymbol{r}_2 - \boldsymbol{r}_1$ в правой части последнего уравнения можно разложить по этому базису. Подставляя однозначно определённые координаты t_0 и $-s_0$ вектора $\boldsymbol{r}_2 - \boldsymbol{r}_1$ в базисе $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ в параметрические уравнения прямых, найдём координаты единственной общей точки рассматриваемых прямых.

7.58. Если векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ линейно зависимы (коллинеарны), то прямые называются *коллинеарными*. Если при этом $\boldsymbol{r}_2 - \boldsymbol{r}_1 \in L(\boldsymbol{a}_1) = L(\boldsymbol{a}_2)$, то эти коллинеарные прямые совпадают, поскольку имеют общую точку (см. предложение 7.18, с. 162); если же $\boldsymbol{r}_2 - \boldsymbol{r}_1 \notin L(\boldsymbol{a}_1) = L(\boldsymbol{a}_2)$, то уравнение $t\boldsymbol{a}_1 - s\boldsymbol{a}_2 = \boldsymbol{r}_2 - \boldsymbol{r}_1$ решений не имеет, т.е. прямые не имеют общих точек; в этом случае их называют *параллельными*.

Итак, доказано следующее утверждение.

7.59. Предложение (взаимное расположение двух прямых на плоскости)**.** *Для двух прямых $\boldsymbol{r} = \boldsymbol{r}_1 + t\boldsymbol{a}_1$ и $\boldsymbol{r} = \boldsymbol{r}_2 + s\boldsymbol{a}_2$ на плоскости* (*т.е. в двумерном аффинном пространстве*) *возможны следующие варианты взаимного расположения*:

(а) *прямые могут иметь единственную общую точку*; *этот случай имеет место тогда и только тогда, когда векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ неколлинеарны* (*линейно независимы*);

(б) *прямые могут быть параллельными* (*т.е. не иметь общих точек*); *этот случай имеет место тогда и только тогда, когда векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ коллинеарны, а векторы $\boldsymbol{r}_2 - \boldsymbol{r}_1$ и $\boldsymbol{a}_1$ неколлинеарны*;

(в) *прямые могут совпадать*; *этот случай имеет место тогда и только тогда, когда векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ коллинеарны и векторы $\boldsymbol{r}_2 - \boldsymbol{r}_1$ и $\boldsymbol{a}_1$ также коллинеарны.*

7.60. Анализ взаимного расположения двух прямых на плоскости можно провести на основе уравнений общего вида. Пусть две прямые заданы уравнениями

$$A_1x + B_1y = D_1, \quad A_2x + B_2y = D_2.$$

Согласно формулам Крамера, система, состоящая из этих двух уравнений, имеет единственное решение (т.е. прямые пересекаются в единственной точке), если отличен от нуля определитель основной матрицы этой системы:

$$\begin{vmatrix} A_1 & B_2 \\ A_2 & B_2 \end{vmatrix} \neq 0 \quad \Longleftrightarrow \quad \frac{A_1}{A_2} \neq \frac{B_1}{B_2}.$$

Если же этот определитель равен нулю, т.е. его строки пропорциональны, то при

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} \neq \frac{D_1}{D_2}$$

система уравнений несовместна (прямые параллельны), а при

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{D_1}{D_2}$$

каждое из уравнений является следствием другого, т.е. прямые совпадают.

B. Пучок прямых.

7.61. *Пучком прямых* на плоскости называется множество всех прямых на этой плоскости, проходящих через некоторую точку C, называемую *центром* пучка. Пучок с центром C будем обозначать $\Pi(C)$. В некоторых задачах аналитической геометрии требуется получить уравнение некоторой прямой пучка (направление которой описано каким-либо способом), если известны две прямые этого пучка. Оказывается, это можно сделать, не находя координаты центра пучка.

7.62. Предложение. *Пусть $A_1x + B_1y = D_1$ и $A_2x + B_2y = D_2$ — уравнения двух прямых, пересекающихся в единственной точке $C(x_0, y_0)$. Прямая $Ax + By = D$ принадлежит пучку $\Pi(C)$ тогда и только тогда, когда она задаётся уравнением*

$$\alpha(A_1x + B_1y) + \beta(A_2x + B_2y) = \alpha D_1 + \beta D_2, \tag{7.15}$$

где α и β — числа, не равные одновременно нулю. Необходимое и достаточное условие принадлежности прямой $Ax + By = D$ пучку $\Pi(C)$ можно записать в следующем эквивалентном виде:

$$\begin{vmatrix} A & B & D \\ A_1 & B_1 & D_1 \\ A_2 & B_2 & D_2 \end{vmatrix} = 0. \tag{7.16}$$

Доказательство. Достаточность. Рассмотрим уравнение (7.15). Поскольку α и β не равны одновременно нулю, хотя бы одно из чисел $\alpha A_1+\beta A_2$ и $\alpha B_1+\beta B_2$ отлично от нуля, так что (7.15) действительно представляет собой уравнение прямой. Далее, если точка (x_0, y_0) является решением каждого из уравнений $A_1x + B_1y = D_1$ и $A_2x + B_2y = D_2$, то она будет также и решением уравнения (7.15). Следовательно, при любых α и β, не равных одновременно нулю, уравнение (7.15) определяет некоторую прямую, принадлежащую пучку с центром $C(x_0, y_0)$.

Необходимость. Остаётся доказать, что α и β всегда можно подобрать так, чтобы получающееся уравнение определяло любую наперёд заданную прямую рассматриваемого пучка. Итак, рассмотрим прямую данного пучка, проходящую через точку $M_1(x_1, y_1) \neq C(x_0, y_0)$. Покажем, что эту прямую можно задать уравнением (7.15) с подходящими α и β.

Подставляя в (7.15) координаты (x_1, y_1) точки M_1, получим

$$\alpha(A_1x_1 + B_1y_1) + \beta(A_2x_1 + B_2y_1) = \alpha D_1 + \beta D_2 \;\Leftrightarrow$$
$$\Leftrightarrow\; \alpha\bigl[A_1x_1 + B_1y_1 - D_1\bigr] + \beta\bigl[A_2x_1 + B_2y_1 - D_2\bigr] = 0.$$

Поскольку точка $M_1(x_1, y_1)$ не может лежать одновременно на прямых $A_1x + B_1y = D_1$ и $A_2x + B_2y = D_2$, по крайней мере одно из выражений в квадратных скобках отлично от нуля, а это означает, что один из коэффициентов α, β может быть выражен через другой. Таким образом, задавая один из этих коэффициентов произвольно и вычисляя второй указанным способом, получаем уравнение, которое описывает прямую, проходящую через точку M_1.

Теперь докажем эквивалентность условий (7.15) и (7.16). Если выполнено (7.15), то

$$A = \alpha A_1 + \beta A_2, \quad B = \alpha B_1 + \beta B_2, \quad D = \alpha D_1 + \beta DA_2, \tag{7.17}$$

и определитель в (7.16) равен нулю в силу линейной зависимости строк. Наоборот, если выполняется (7.16), то строки определителя линейно зависимы. Предположив, что коэффициент при первой строке равен нулю, видим, что тогда линейно зависимы вторая и третья строки, т.е. соответствующие уравнения описывают либо совпадающие, либо параллельные прямые, что противоречит условию. Таким образом, коэффициент линейной зависимости при первой строке отличен от нуля, т.е. имеют место соотношения (7.17), а следовательно, и (7.15). □

7.63. Назовём *несобственным пучком* совокупность всех прямых на плоскости, параллельных какой-либо одной прямой, в отличие от собственных пучков, рассмотренных выше. Соответствие между собственными пучками и точками плоскости

собственный пучок $\mapsto$ его центр

взаимно однозначно. Можно пополнить плоскость «несобственными» или «бесконечно удалёнными» точками, поставив в соответствие каждому несобственному пучку «несобственную точку» и объявив её единственной общей точкой всех прямых данного несобственного пучка. Таким образом, каждый несобственный пучок получает свой «центр» — «несобственную» («бесконечно удалённую») точку, в которой «пересекаются» все (параллельные между собой) прямые этого пучка. Совокупность всех несобственных точек плоскости объявляется несобственной (бесконечно удалённой) прямой. Плоскость, множество точек которой пополнено всеми несобственными точками, а множество всех прямых — одной несобственной прямой, состоящей из всех несобственных точек, называется *проективной плоскостью.* На проективной плоскости любые две прямые пересекаются в одной точке: собственной, если на обычной плоскости эти прямые не параллельны, и в их единственной несобственной точке, если эти прямые параллельны на обычной плоскости. Наконец, если одна из двух прямых несобственная, то она пересекается со второй прямой в единственной несобственной точке этой последней. Вопросы, затронутые здесь, являются предметом отдельного раздела геометрии — *проективной геометрии.*

5. Трехмерная аффинная геометрия

А. Уравнения плоскостей. В этом разделе мы рассмотрим некоторые вопросы аффинной геометрии в трёхмерном аффинном пространстве. Как и в предыдущем разделе, координаты радиус-векторов $\boldsymbol{r}$, $\boldsymbol{r}_0$ и т. д. обозначаем (x, y, z), (x_0, y_0, z_0) и т. д., а координаты векторов $\boldsymbol{a}$ и т. д. — (a_x, a_y, a_z).

7.64. Если в трёхмерном аффинном пространстве $\mathcal{A}$ выбрана некоторая аффинная система координат $O\boldsymbol{ijk}$, то векторное параметрическое уравнение плоскости (7.5) (см. с. 163)

$$\boldsymbol{r} = \boldsymbol{r}_0 + s\boldsymbol{a} + t\boldsymbol{b} \tag{7.5}$$

записывается в виде

$$\begin{cases} x = x_0 + sa_x + tb_x, \\ y = y_0 + sa_y + tb_y, \\ z = z_0 + sa_z + tb_z; \end{cases} \tag{7.18}$$

эта система называется *системой параметрических уравнений плоскости.*

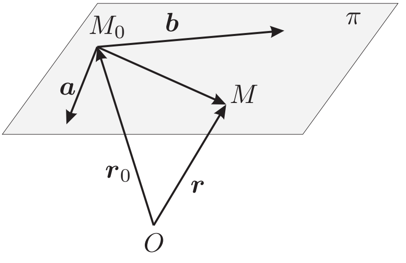


Рис. 7.6.

7.65. Факт линейной зависимости векторов $\overrightarrow{M_0M} = \boldsymbol{r} - \boldsymbol{r}_0$, $\boldsymbol{a}$, $\boldsymbol{b}$, лежащий в основе определения плоскости, может быть записан в виде равенства нулю определителя, составленного из координат этих векторов (см. теорему 6.52, с. 140):

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix} = 0. \tag{7.19}$$

Уравнение (7.19) будем по аналогии с уравнением (7.13) называть *каноническим уравнением плоскости.*

7.66. Раскрывая определитель (7.19) по первой строке и вводя обозначения

$$A = \begin{vmatrix} a_y & a_z \\ b_y & b_z \end{vmatrix}, \quad B = -\begin{vmatrix} a_x & a_z \\ b_x & b_z \end{vmatrix}, \quad C = \begin{vmatrix} a_x & a_y \\ b_x & b_y \end{vmatrix}$$

для алгебраических дополнений элементов первой строки, получим уравнение вида

$$A(x-x_0)+B(y-y_0)+C(z-z_0)=0, \tag{7.20}$$

называемое *общим уравнением плоскости, проходящей через заданную точку* $M_0(x_0,y_0,z_0)$. Раскрывая скобки и вводя обозначение $D=Ax_0+By_0+Cz_0$, получаем *общее уравнение плоскости*:

$$Ax+By+Cz=D. \tag{7.21}$$

Очевидно, если $D=0$, то плоскость проходит через начало координат. При $D\neq 0$, разделив обе части последнего уравнения на D и вводя обозначения $a=D/A$, $b=D/B$, $c=D/C$, приходим к *уравнению плоскости в отрезках*:

$$\frac{x}{a}+\frac{y}{b}+\frac{z}{c}=1.$$

Геометрический смысл чисел a, b, c, очевиден: плоскость проходит через точки $(a,0,0)$, $(0,b,0)$, $(0,0,c)$, т.е. отсекает на координатных осях Ox, Oy, Oz отрезки a, b, c соответственно.

7.67. Согласно п. 7.42 уравнение (7.21) представляет собой уравнение гиперплоскости: плоскость в трёхмерном аффинном пространстве является гиперплоскостью.

Б. Свойства плоскостей в пространстве.

7.68. Предложение. *Существует единственная плоскость, проходящая через три произвольные неколлинеарные (т.е. не лежащие на одной прямой) точки M_1, M_2, M_3.*

Доказательство. В качестве опорной точки можно взять любую из данных точек (возьмём M_1), а в качестве направляющих векторов — векторы $\overrightarrow{M_1M_2}$ и $\overrightarrow{M_1M_3}$, которые линейно независимы (неколлинеарны) в силу условия, что точки M_1, M_2, M_3 не лежат на одной прямой. Тогда уравнение

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ x_3-x_1 & y_3-y_1 & z_3-z_1 \end{vmatrix}=0$$

является уравнением искомой плоскости. (Здесь (x_1,y_1,z_1) — координаты точки M_1 и т. п.) □

Отметим что в рамках аксиоматики Гильберта это предложение представляет собой одну из аксиом.

7.69. Предложение. *Существует единственная плоскость, проходящая через через две заданные точки M_1 и M_2 параллельно вектору $\boldsymbol{a}$, который не коллинеарен вектору $\overrightarrow{M_1M_2}$.*

Доказательство. В качестве опорной точки плоскости можно взять любую из данных точек (возьмём M_1), а в качестве направляющих векторов — векторы $\boldsymbol{a}$ и $\overrightarrow{M_1M_2}$. В результате уравнение искомой плоскости можно записать в виде

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ a_x & a_y & a_z \end{vmatrix} = 0. \qquad \square$$

В. Уравнения прямых в пространстве.

7.70. Векторное параметрическое уравнение прямой $\boldsymbol{r} = \boldsymbol{r}_0 + t\boldsymbol{a}$ в координатах принимает вид

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + t \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} \quad \Longleftrightarrow \quad \begin{cases} x = x_0 + ta_x, \\ y = y_0 + ta_y, \\ z = z_0 + ta_z \end{cases}$$

и называется системой *параметрических уравнений прямой.*

7.71. Исключение параметра t из параметрических уравнений приводит к *системе канонических уравнений прямой*:

$$\frac{x-x_0}{a_x} = \frac{y-y_0}{a_y} = \frac{z-z_0}{a_z}.$$

Ясно, что это не одно уравнение, а система уравнений, однако по аналогии с планиметрическим случаем эту систему часто называют *каноническим уравнением* (в единственном числе). Нули в знаменателях трактуются так же, как в планиметрическом случае (см. с. 170). В условиях задач прямые часто задаются каноническими уравнениями, однако при решении удобно преобразовать каноническое уравнение в параметрическое следующим образом:

$$\frac{x-x_0}{a_x} = \frac{y-y_0}{a_y} = \frac{z-z_0}{a_x} = t \quad \Leftrightarrow \quad \begin{cases} x = x_0 + ta_x, \\ y = y_0 + ta_y, \\ z = z_0 + ta_z. \end{cases}$$

Г. Взаимное расположение плоскостей в пространстве.

7.72. Рассмотрим плоскости

$$A_1x + B_1y + C_1z = D_1,$$
$$A_2x + B_2y + C_2z = D_2$$

и проанализируем систему, состоящую из этих двух уравнений. Если левые части уравнений пропорциональны, т.е.

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2},$$

то в зависимости от соотношения свободных членов уравнений возможны следующие ситуации:

(а) если

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2} \neq \frac{D_1}{D_2},$$

то система несовместна, т.е. плоскости не имеют общих точек (параллельны);

(б) если

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2} = \frac{D_1}{D_2},$$

то каждое из уравнений является следствием другого, т.е. плоскости совпадают.

7.73. Если же левые части уравнений не пропорциональны, т.е., например,

$$\frac{A_1}{A_2} \neq \frac{B_1}{B_2},$$

то, переписав систему в виде

$$\begin{cases} A_1x + B_1y = D_1 - C_1z, \\ A_2x + B_2y = D_2 - C_2z, \end{cases} \qquad \begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} \neq 0,$$

видим, что при любом $z \in \mathbb{R}$ она имеет решение, которое можно найти, например, по формулам Крамера:

$$x = \frac{\begin{vmatrix} D_1 - C_1z & B_1 \\ D_2 - C_2z & B_2 \end{vmatrix}}{\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}}, \quad y = \frac{\begin{vmatrix} A_1 & D_1 - C_1z \\ A_2 & D_2 - C_2z \end{vmatrix}}{\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}}.$$

Введя подходящие обозначения для коэффициентов, можно записать решение в виде

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ 0 \end{pmatrix} + t \begin{pmatrix} a_x \\ a_y \\ 1 \end{pmatrix},$$

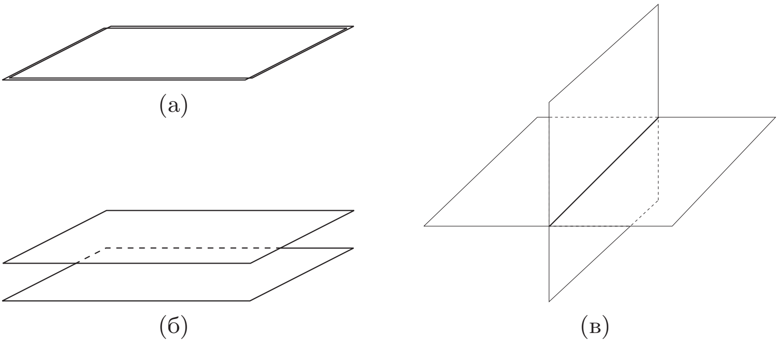


Рис. 7.7. Взаимное расположение двух плоскостей в пространстве

представляющем собой параметрическое уравнение прямой. Итак, если две плоскости пересекаются, то их пересечением является прямая[1].

7.74. Таким образом, прямая в пространстве может быть ещё задана как пересечение двух непараллельных плоскостей:

$$\begin{cases} A_1x + B_1y + C_1z = D_1, \\ A_2x + B_2y + C_2z = D_2. \end{cases}$$

Чтобы получить параметрическое уравнение этой прямой, нужно найти общее решение указанной неоднородной системы уравнений.

Д. Взаимное расположение прямой и плоскости в пространстве.

7.75. Пусть плоскость задана общим уравнением

$$Ax + By + Cz = D,$$

а прямая — параметрическими уравнениями

$$\begin{cases} x = x_0 + ta_x, \\ y = y_0 + ta_y, \\ z = z_0 + ta_z. \end{cases} \tag{7.22}$$

Подставляя (7.22) в уравнение плоскости, получим

$$A(x_0 + ta_x) + B(y_0 + ta_y) + C(z_0 + ta_z) = D \;\Leftrightarrow$$
$$\Leftrightarrow\; t(Aa_x + Ba_y + Ca_z) = D - (Ax_0 + By_0 + Cz_0).$$

[1]Ср. с предложением 7.34, с. 165.

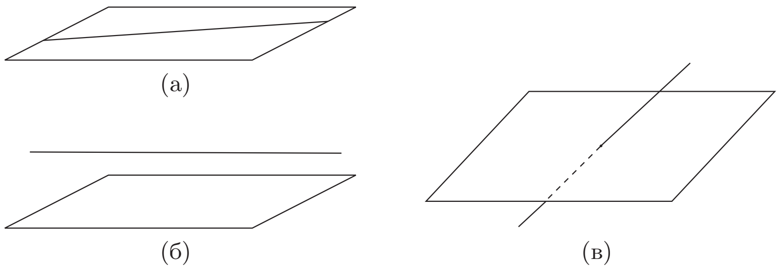


Рис. 7.8. Взаимное расположение прямой и плоскости в пространстве

7.76. В случае $Aa_x + Ba_y + Ca_z \neq 0$ это уравнение имеет единственное решение

$$t_0 = \frac{D - (Ax_0 + By_0 + Cz_0)}{Aa_x + Ba_y + Ca_z},$$

подставив которое в (7.22), найдём координаты единственной точки пересечения прямой и плоскости.

7.77. Если же $Aa_x + Ba_y + Ca_z = 0$, то возможны следующие ситуации:

(а) при $D - (Ax_0 + By_0 + Cz_0) = 0$ решением является любое значение t; это означает, что прямая содержится в плоскости (обратите внимание, что равенство $D - (Ax_0 + By_0 + Cz_0) = 0$ означает, что точка (x_0, y_0, z_0) лежит в плоскости);

(б) при $D - (Ax_0 + By_0 + Cz_0) \neq 0$ уравнение для t не имеет решений, т.е. прямая и плоскость не имеют общих точек; в этом случае их называют *параллельными*.

Е. Взаимное расположение прямых в пространстве.

7.78. Выясним вопрос о взаимном расположении в пространстве двух прямых, заданные уравнениями $\boldsymbol{r} = \boldsymbol{p} + s\boldsymbol{a}$ и $\boldsymbol{r} = \boldsymbol{q} - t\boldsymbol{b}$ (знак «$-$» взят для удобства). Задача сводится к анализу уравнения

$$s\boldsymbol{a} + t\boldsymbol{b} = \boldsymbol{q} - \boldsymbol{p}. \tag{7.23}$$

7.79. Если векторы $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{q} - \boldsymbol{p}$ некомпланарны (линейно независимы), то у уравнения (7.23) решений нет, т.е. прямые не имеют общих точек, но и не являются параллельными (поскольку их направляющие векторы не коллинеарны). Такие прямые называют *скрещивающимися*.

7.80. Если векторы $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{q}-\boldsymbol{p}$ компланарны, но при этом $\boldsymbol{a}$ и $\boldsymbol{b}$ неколлинеарны, а $\boldsymbol{q}-\boldsymbol{p}$ лежит в их линейной оболочке, то уравнение (7.23) имеет единственное решение (s_0, t_0): эта пара чисел является разложением вектора $\boldsymbol{q}-\boldsymbol{p}$ по векторам $\boldsymbol{a}$ и $\boldsymbol{b}$. Геометрически это означает, что прямые лежат в одной плоскости и пересекаются в единственной точке.

7.81. Если векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ коллинеарны, а вектор $\boldsymbol{q}-\boldsymbol{p}$ им не коллинеарен, то уравнение (7.23) не имеет решений, прямые параллельны.

7.82. Наконец, если все три вектора $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{q}-\boldsymbol{p}$ коллинеарны, то решением уравнения (7.23) является любая пара чисел (s,t); это означает, что прямые совпадают.

7.83. Полученные результаты можно выразить в терминах рангов основной и расширенной матриц системы уравнений (7.23) (см. теорему Кронекера—Капелли, с. 147). Пусть

$$\boldsymbol{a} = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix}, \quad \boldsymbol{q}-\boldsymbol{p} = \begin{pmatrix} c_x \\ c_y \\ c_z \end{pmatrix}.$$

Составим матрицы

$$A = \big\|\boldsymbol{a},\ \boldsymbol{b}\big\| = \begin{pmatrix} a_x & b_x \\ a_y & b_y \\ a_z & b_z \end{pmatrix}, \quad \tilde{A} = \big\|\boldsymbol{a},\ \boldsymbol{b},\ \boldsymbol{q}-\boldsymbol{p}\big\| = \left(\begin{array}{cc|c} a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & c_z \end{array}\right)$$

и обозначим через r и R их ранги: $r = \operatorname{rk} A$, $R = \operatorname{rk} \tilde{A}$. Тогда имеем:

(i) при $r = 2$, $R = 3$ прямые скрещиваются;
(ii) при $r = R = 2$ прямые пересекаются в единственной точке;
(iii) при $r = 1$, $R = 2$ прямые параллельны;
(iv) при $r = R = 1$ то прямые совпадают.

Ж. Пучок плоскостей.

7.84. *Пучком плоскостей* в пространстве называется множество всех плоскостей, проходящих через некоторую прямую l, называемую *осью* пучка. Пучок с центром l будем обозначать $\Pi(l)$.

7.85. Предложение. *Пусть*

$$A_1x + B_1y + C_1z = D_1, \quad A_2x + B_2y + C_2z = D_2$$

— уравнения двух (непараллельных) плоскостей, пересекающихся по прямой l. Плоскость принадлежит пучку $\Pi(l)$ тогда и только тогда, когда она задаётся уравнением

$$\alpha\big(A_1x + B_1y + C_1z\big) + \beta\big(A_2x + B_2y + C_2z\big) = \alpha D_1 + \beta D_2, \tag{7.24}$$

где α и β — числа, не равные одновременно нулю.

Доказательство совершенно аналогично доказательству предложения 7.62 (см. с. 173).

7.86. Здесь уместно было бы привести рассуждения, приводящие к понятию проективного пространства, однако ограничение объёма не позволяет нам этого сделать.

6. Аффинная мера

А. Наглядные соображения.

7.87. Процесс измерения состоит в сравнении измеряемой величины с эталоном. Выбрав в векторном пространстве $\mathcal{V}$ некоторый вектор $\boldsymbol{a}$ в качестве эталона, мы сможем сравнивать с ним все коллинеарные векторы: если $\boldsymbol{x} = \lambda\boldsymbol{a}$, то число λ и является результатом измерения. Однако векторы, не коллинеарные вектору $\boldsymbol{a}$, сравнить с ним таким способом невозможно.

7.88. Как известно из школьного курса, в рамках гильбертова подхода площадь параллелограмма на плоскости равна произведению основания на высоту или произведению сторон на синус угла между ними: $S = ab\sin\varphi$.

Будем рассматривать так называемую *ориентированную площадь*: для параллелограмма, стороны которого образованы векторами $\boldsymbol{a}$ и $\boldsymbol{b}$ (обозначим этот параллелограмм $\Pi(\boldsymbol{a},\boldsymbol{b})$), она равна обычной площади, если эта пара векторов правая, и отличается знаком от обычной площади в противном случае. Если считать угол φ между сторонами параллелограмма ориентированным, то выражение $ab\sin\varphi$ (в силу нечётности синуса) как раз и даёт значение ориентированной площади. Ориентированную площадь параллелограмма $\Pi(\boldsymbol{a},\boldsymbol{b})$ обозначим $\bigl|\Pi(\boldsymbol{a},\boldsymbol{b})\bigr|$.

7.89. Пусть φ — ориентированный угол между векторами $\boldsymbol{a}$ и $\boldsymbol{b}$. Найдём ориентированную площадь параллелограмма $\Pi(\boldsymbol{a},\boldsymbol{b})$, считая что координаты векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ заданы в декартовой прямоугольной системе координат (см. рис. 7.9):

$$\begin{aligned}\bigl|\Pi(\boldsymbol{a},\boldsymbol{b})\bigr| &= |\boldsymbol{a}|\cdot|\boldsymbol{b}|\cdot\sin\varphi = |\boldsymbol{a}|\cdot|\boldsymbol{b}|\cdot\sin(\beta-\alpha) = \\ &= |\boldsymbol{a}|\cdot|\boldsymbol{b}|\cdot\bigl(\sin\beta\cos\alpha - \cos\beta\sin\alpha\bigr) = \\ &= \underbrace{|\boldsymbol{a}|\cos\alpha}_{=a_x}\cdot\underbrace{|\boldsymbol{b}|\sin\beta}_{b_y} - \underbrace{|\boldsymbol{a}|\sin\alpha}_{a_y}\cdot\underbrace{|\boldsymbol{b}|\cos\beta}_{b_x} = \begin{vmatrix} a_x & b_x \\ a_y & b_y \end{vmatrix}.\end{aligned}$$

7.90. Отметим, что приведённые выше рассуждения возможны лишь при наличии понятий длины отрезка и меры угла, т.е. они неприменимы в аффинных пространствах.

7.91. Получился определитель матрицы перехода от базиса $\bigl(\boldsymbol{i},\boldsymbol{j}\bigr)$, образованного единичными векторами координатных осей, к базису[1] $\bigl(\boldsymbol{a},\boldsymbol{b}\bigr)$, ясно, что его знак указывает на ориентацию базиса $\boldsymbol{a},\boldsymbol{b}$. Таким образом, ориентированная площадь параллелограмма выражается определителем, столбцы которого состоят из координат векторов, образующих стороны параллелограмма.

7.92. Обратим внимание на следующие легко проверяемые свойства ориентированной площади $\bigl|\Pi(\boldsymbol{a},\boldsymbol{b})\bigr|$:

[1]Разумеется, в случае, когда векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ неколлинеарны: в противном случае этот определитель равен нулю.

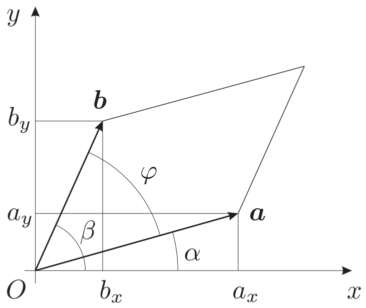


Рис. 7.9.

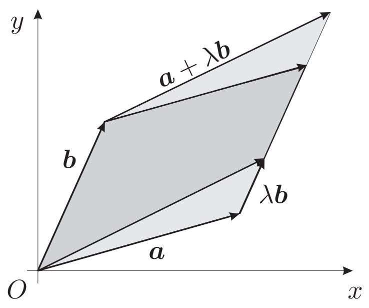


Рис. 7.10.

(1) $\left|\Pi(\boldsymbol{a}+\lambda\boldsymbol{b},\boldsymbol{b})\right| = \left|\Pi(\boldsymbol{a},\boldsymbol{b})\right|$;
(2) $\left|\Pi(\boldsymbol{a},\boldsymbol{b})\right| = -\left|\Pi(\boldsymbol{b},\boldsymbol{a})\right|$;
(3) $\left|\Pi(\boldsymbol{i},\boldsymbol{j})\right| = 1$.

О свойствах (2) и (3) было сказано выше, а свойство (1) проиллюстрировано на рис. 7.10 (оно выражает тот факт, что равносоставленные параллелограммы имеют равные площади). Очевидно, эти свойства ориентированной площади совпадают со свойствами определителя.

Б. Определение.

7.93. Рассмотрим произвольное аффинное пространство $\mathcal{A}$ и произвольный базис $\boldsymbol{E} = \left(\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n\right)$ в нём. Пусть $\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n\right)$ — линейно независимое семейство векторов и M_0 — произвольная точка. Назовём *параллелепипедом*, построенным на векторах $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n$, множество точек

$$\Pi(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n) \overset{\text{def}}{=\!=} \left\{M : \overrightarrow{M_0M} = t_1\boldsymbol{a}_1+\cdots+t_n\boldsymbol{a}_n,\ \ 0\leqslant t_j\leqslant 1,\ j=1,\ldots,n\right\}.$$

7.94. *Аффинным объёмом* (*аффинной мерой*) параллелепипеда $\Pi(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n)$ в базисе $\left(\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n\right)$ называется число, равное определителю, столбцы которого представляют собой столбцы $A_1,\ldots,A_n$ координат векторов $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n$ относительно базиса $\left(\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n\right)$:

$$\left|\Pi(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n)\right| = \det\left\|A_1,\ldots,A_n\right\|.$$

Очевидно, если векторы $\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n\right)$ линейно зависимы, то аффинный объём ориентированного параллелепипеда, построенного на этих векторах, равен нулю.

7.95. Аффинный объём ориентированного параллелепипеда зависит от выбора базиса $\left(\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n\right)$, который служит эталоном объёма. Однако в аффинном пространстве (в отличие от евклидова, см. гл. 8) нет никакого способа выбрать естественный эталон объёма.

7.96. Ясно, что понятие аффинной меры можно ввести в любой плоскости этого пространства; для этого нужно лишь в рассматриваемой плоскости (вернее, в её направляющем подпространстве) выбрать некоторый базис — эталон объёма. Однако сравнивать между собой эталоны объёма на непараллельных плоскостях (хотя бы даже и одинаковой размерности) невозможно.

ГЛАВА 8

Евклидовы пространства

ЕВКЛИДОВО пространство — это векторное или аффинное пространство, в котором имеется дополнительная операция над векторами, позволяющая ввести метрические понятия: длины и углы; эта операция называется скалярным умножением векторов.

Понятие скалярного произведения было введено У. Гамильтоном в 1846 г. одновременно с векторным произведением в связи с изучением кватернионов — объектов, являющихся обобщением вещественных и комплексных чисел. Скалярное умножение широко используется в физике; так, работа постоянной силы при прямолинейном движении материальной точки определяется скалярным произведением вектора этой силы и вектора перемещения точки.

1. Скалярное произведение в элементарной геометрии

Как обычно, начнём с того, что изучим вопрос с наглядно-интуитивной точки зрения.

8.1. В элементарной геометрии *скалярным произведением* $\boldsymbol{a} \cdot \boldsymbol{b}$ *двух векторов* $\boldsymbol{a}$ *и* $\boldsymbol{b}$ называется число, равное произведению длин $|\boldsymbol{a}|$ и $|\boldsymbol{b}|$ этих векторов на косинус угла φ между ними:

$$\boldsymbol{a} \cdot \boldsymbol{b} = |\boldsymbol{a}| \cdot |\boldsymbol{b}| \cdot \cos\varphi.$$

Ориентация угла не имеет значения, поскольку косинус — чётная функция.

8.2. Предложение. *Скалярное произведение векторов обладает следующими свойствами:*

(1) *линейность по каждому аргументу* (*билинейность*):

$$(\alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2) \cdot \boldsymbol{b} = \alpha_1\, \boldsymbol{a}_1 \cdot \boldsymbol{b} + \alpha_2\, \boldsymbol{a}_2 \cdot \boldsymbol{b},$$

$$\boldsymbol{a} \cdot (\beta_1 \boldsymbol{b}_1 + \beta_2 \boldsymbol{b}_2) = \beta_1\, \boldsymbol{a} \cdot \boldsymbol{b}_1 + \beta_2\, \boldsymbol{a} \cdot \boldsymbol{b}_2,$$

где $\boldsymbol{a}$, $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{b}$, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ *— произвольные векторы,* α_1, α_2, β_1, β_2 *— произвольные вещественные числа;*

(2) *симметричность* (*коммутативность*):

$$\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{b} \cdot \boldsymbol{a};$$

(3) *положительная определённость: для любого вектора* $\boldsymbol{a}$

$$\boldsymbol{a}\cdot\boldsymbol{a}\geqslant 0,$$

причём $\boldsymbol{a}\cdot\boldsymbol{a}=0$ *тогда и только тогда, когда* $\boldsymbol{a}=\boldsymbol{0}$.

Доказательство. Для доказательства свойств скалярного произведения его удобно выразить через ортогональные проекции.

Ортогональная проекция вектора $\overrightarrow{AB}$ *на ось*[1] $\vec{l}$ — это число

$$\mathrm{pr}_{\vec{l}}\overrightarrow{AB}\overset{\text{def}}{=\!=}\begin{cases}|\overrightarrow{A'B'}|, & \text{если } \overrightarrow{A'B'}\uparrow\uparrow\vec{l},\\ -|\overrightarrow{A'B'}|, & \text{если } \overrightarrow{A'B'}\uparrow\downarrow\vec{l},\end{cases}$$

где A' и B' — ортогональные проекции точек A и B на прямую l (см. рис. 8.1).

Очевидно, проекция $\mathrm{pr}_{\vec{l}}\boldsymbol{a}$ вектора $\boldsymbol{a}$ на ось $\vec{l}$ равна

$$\mathrm{pr}_{\vec{l}}\boldsymbol{a}=|\boldsymbol{a}|\cos\varphi,$$

где φ — угол между вектором $\boldsymbol{a}$ и осью $\vec{l}$ (или, что то же, угол между векторами $\boldsymbol{a}$ и $\boldsymbol{l}$, где $\boldsymbol{l}$ — произвольный направляющий вектор оси $\vec{l}$). Тогда скалярное произведение векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ равно произведению длины одного из векторов (например, $\boldsymbol{a}$) и проекции другого на ось, определяемую первым:

$$\boldsymbol{a}\cdot\boldsymbol{b}=|\boldsymbol{a}|\cdot\mathrm{pr}_{\boldsymbol{a}}\,\boldsymbol{b}.$$

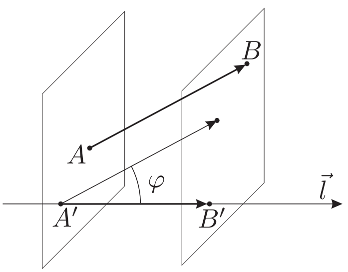


Рис. 8.1. Ортогональная проекция вектора на ось

Проекции вектора на ось обладают очевидным свойством линейности, которое легко доказывается в рамках аксиоматики Гильберта:

$$\mathrm{pr}_{\vec{l}}(\alpha\boldsymbol{a}+\beta\boldsymbol{b})=\alpha\,\mathrm{pr}_{\vec{l}}\boldsymbol{a}+\beta\,\mathrm{pr}_{\vec{l}}\boldsymbol{b}.$$

Поэтому интересующие нас свойства скалярного умножения векторов легко выводятся из определения и свойств проекций. Например,

$$(\alpha_1\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2)\cdot\boldsymbol{b}=\mathrm{pr}_{\boldsymbol{b}}\Big(\alpha_1\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2\Big)\cdot|\boldsymbol{b}|=$$
$$=\alpha_1\,\mathrm{pr}_{\boldsymbol{b}}\,\boldsymbol{a}_1\cdot|\boldsymbol{b}|+\alpha_2\,\mathrm{pr}_{\boldsymbol{b}}\,\boldsymbol{a}_2\cdot|\boldsymbol{b}|=\alpha_1\,\boldsymbol{a}_1\cdot\boldsymbol{b}+\alpha_2\,\boldsymbol{a}_2\cdot\boldsymbol{b}.\quad\square$$

[1] Напомним, что ось — это ориентированная прямая (см. с. 152).

8.3. Через скалярное произведение выражаются длина вектора и угол φ между векторами $\boldsymbol{a}$ и $\boldsymbol{b}$:

$$|\boldsymbol{a}| = \sqrt{\boldsymbol{a}\cdot\boldsymbol{a}}, \quad \varphi = \arccos\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}|\cdot|\boldsymbol{b}|}.$$

2. Евклидово пространство

А. Определения и основные свойства. В аксиоматической теории мы обращаем полученные результаты и принимаем свойства скалярного произведения, установленные в предложении 8.2, за аксиомы.

Итак, пусть $\mathcal{V}$ — произвольное *вещественное* векторное пространство.[1]

8.4. Определение. *Скалярным умножением* на векторном пространстве $\mathcal{V}$ называется функция двух векторных аргументов, принимающая значения в множестве вещественных чисел,

$$g\colon \mathcal{V}\times\mathcal{V}\to\mathbb{R}, \quad (\boldsymbol{x},\boldsymbol{y})\mapsto g(\boldsymbol{x},\boldsymbol{y}),$$

обладающая следующими свойствами:

СП1: линейность по каждому аргументу (полилинейность, точнее, билинейность):

$$g(\alpha_1\boldsymbol{x}_1+\alpha_2\boldsymbol{x}_2,\boldsymbol{y}) = \alpha_1 g(\boldsymbol{x}_1,\boldsymbol{y})+\alpha_2 g(\boldsymbol{x}_2,\boldsymbol{y}),$$
$$g(\boldsymbol{x},\beta_1\boldsymbol{y}_1+\beta_2\boldsymbol{y}_2) = \beta_1 g(\boldsymbol{x},\boldsymbol{y}_1)+\beta_2 g(\boldsymbol{x},\boldsymbol{y}_2);$$

СП2: симметричность (коммутативность):

$$g(\boldsymbol{x},\boldsymbol{y}) = g(\boldsymbol{y},\boldsymbol{x});$$

СП3: положительная определённость:

$$\forall\boldsymbol{x}\in\mathcal{V}\quad g(\boldsymbol{x},\boldsymbol{x})\geqslant 0,$$

причём $g(\boldsymbol{x},\boldsymbol{x})=0$ тогда и только тогда, когда $\boldsymbol{x}=\boldsymbol{0}$.

Значение $g(\boldsymbol{x},\boldsymbol{y})$ называется *скалярным произведением векторов* $\boldsymbol{x}$ *и* $\boldsymbol{y}$ и обозначается обычно[2] $(\boldsymbol{x},\boldsymbol{y})$.

8.5. Определение. *Евклидовым векторным пространством* называется вещественное векторное пространство $\mathcal{V}$ с заданным на нём скалярным умножением g.

---

[1]Понятие скалярного произведения может быть введено и для векторных пространств над полем комплексных чисел; это будет сделано в следующем семестре.

[2]Используются также обозначения $\boldsymbol{x}\cdot\boldsymbol{y}$ (особенно в физической литературе) и $\langle\boldsymbol{x},\boldsymbol{y}\rangle$.

Евклидовым точечным[1] *пространством* называется вещественное аффинное пространство $\mathcal{A}$, в ассоциированном векторном пространстве которого задано скалярное умножение g.

Перечисленные свойства **СП1–СП3** называются *аксиомами скалярного умножения* или *аксиомами евклидова пространства* и образуют *третью группу аксиом Вейля*.

Как векторные, так и точечные евклидовы пространства будем обозначать символом $\mathcal{E}$; тип пространства (векторное или точечное) будет ясен из контекста.

Б. Длины и углы.

8.6. Определение. *Длиной вектора* $\boldsymbol{x}$ в евклидовом векторном пространстве называется число

$$|\boldsymbol{x}| = \sqrt{(\boldsymbol{x}, \boldsymbol{x})}. \tag{8.1}$$

Расстоянием между точками A и B в евклидовом точечном пространстве называется число

$$|AB| = \bigl|\overrightarrow{AB}\bigr| = \sqrt{\bigl(\overrightarrow{AB}, \overrightarrow{AB}\bigr)}. \tag{8.2}$$

8.7. Аксиома **СП3** гарантирует, что для любого вектора $\boldsymbol{x}$ однозначно определена его длина $|\boldsymbol{x}|$, являющаяся неотрицательным вещественным числом. Нулевую длину при этом может иметь лишь нулевой вектор, а расстояние равно нулю лишь для совпадающих точек.

8.8. В векторных пространствах, элементами которых являются матрицы или функции (см. примеры 5.17, 5.19, 5.20, с. 96), термин «длина вектора» и соответствующее обозначение $|\boldsymbol{x}|$, совпадающее с обозначением модуля числа, неудобны и не используются. Вместо этого говорят о «норме вектора» и обозначают эту величину $\|\boldsymbol{x}\|$.

8.9. Определение. *Углом между векторами $\boldsymbol{x}$ и $\boldsymbol{y}$* в евклидовом векторном пространстве называется число

$$\varphi = \arccos \frac{(\boldsymbol{x}, \boldsymbol{y})}{|\boldsymbol{x}| \cdot |\boldsymbol{y}|}. \tag{8.3}$$

Углом между прямыми, имеющими направляющие векторы $\boldsymbol{a}$ и $\boldsymbol{b}$, в евклидовом точечном пространстве называется число

$$\varphi = \arccos \frac{|(\boldsymbol{a}, \boldsymbol{b})|}{|\boldsymbol{a}| \cdot |\boldsymbol{b}|}. \tag{8.4}$$

8.10. Обратите внимание на различие формул (8.3) и (8.4): угол *между векторами* принимает значения $\varphi \in [0, \pi]$, а угол *между прямыми* — значения $\varphi \in [0, \pi/2]$.

---

[1] Термин «евклидово аффинное пространство» не используется.

В. Неравенство Коши—Буняковского—Шварца и его следствия. Для формул (8.3) и (8.4) требует доказательства тот факт, что выражение под знаком арккосинуса не превышает по модулю единицы. Этот факт является следствием следующего важного утверждения.

8.11. Теорема (неравенство Коши[1]—Буняковского[2]—Шварца[3]). *Для любых* $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{E}$ *имеет место неравенство*

$$(\boldsymbol{x}, \boldsymbol{y})^2 \leqslant (\boldsymbol{x}, \boldsymbol{x}) \cdot (\boldsymbol{y}, \boldsymbol{y})$$

или, что эквивалентно,

$$\bigl|(\boldsymbol{x}, \boldsymbol{y})\bigr| \leqslant |\boldsymbol{x}| \cdot |\boldsymbol{y}|.$$

Доказательство. Пусть $\boldsymbol{x}$, $\boldsymbol{y}$ — произвольные векторы, t — произвольное вещественное число. Имеем:

$$0 \overset{\textbf{СП3}}{\leqslant} (t\boldsymbol{x} + \boldsymbol{y}, t\boldsymbol{x} + \boldsymbol{y}) = t^2(\boldsymbol{x}, \boldsymbol{x}) + 2t(\boldsymbol{x}, \boldsymbol{y}) + (\boldsymbol{y}, \boldsymbol{y}).$$

Квадратный трёхчлен может принимать только неотрицательные значения лишь в случае, когда его дискриминант неположителен:

$$D = 4(\boldsymbol{x}, \boldsymbol{y})^2 - 4(\boldsymbol{x}, \boldsymbol{x}) \cdot (\boldsymbol{y}, \boldsymbol{y}) \leqslant 0. \qquad \square$$

8.12. Предложение.

1. *Для любых* $\boldsymbol{x} \in \mathcal{E}$ *и* $\alpha \in \mathbb{R}$ *имеет место формула*[4]
$$|\alpha\boldsymbol{x}| = |\alpha| \cdot |\boldsymbol{x}|.$$
2. *Для любых* $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{E}$ *справедливы следующие неравенства, называемые неравенствами треугольника:*
$$\bigl||\boldsymbol{x}| - |\boldsymbol{y}|\bigr| \leqslant |\boldsymbol{x} + \boldsymbol{y}| \leqslant |\boldsymbol{x}| + |\boldsymbol{y}|.$$

Доказательство. 1. Имеем:

$$|\alpha\boldsymbol{x}| = \sqrt{(\alpha\boldsymbol{x}, \alpha\boldsymbol{x})} = \sqrt{\alpha^2(\boldsymbol{x}, \boldsymbol{x})} = |\alpha| \cdot |\boldsymbol{x}|.$$

2. Имеем:

$$|\boldsymbol{x} + \boldsymbol{y}|^2 = (\boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y}) = |\boldsymbol{x}|^2 + 2(\boldsymbol{x}, \boldsymbol{y}) + |\boldsymbol{y}|^2 \leqslant \\ \leqslant |\boldsymbol{x}|^2 + 2|\boldsymbol{x}||\boldsymbol{y}| + |\boldsymbol{y}|^2 = \bigl(|\boldsymbol{x}| + |\boldsymbol{y}|\bigr)^2,$$

---

[1] О. Л. Коши (A. L. Cauchy) — выдающийся французский математик и механик (1789–1857).

[2] В. Я. Буняковский — русский математик (1804–1889).

[3] К. Шварц (K. Schwarz) — немецкий математик (1843–1921).

[4] Обратите внимание, что $|\alpha|$ означает модуль вещественного числа, а $|\boldsymbol{x}|$ — длину (норму) вектора в смысле определения 8.6.

откуда $|\boldsymbol{x}+\boldsymbol{y}| \leqslant |\boldsymbol{x}| + |\boldsymbol{y}|$. Аналогично,

$$|\boldsymbol{x}+\boldsymbol{y}|^2 = |\boldsymbol{x}|^2 + 2(\boldsymbol{x},\boldsymbol{y}) + |\boldsymbol{y}|^2 \geqslant |\boldsymbol{x}|^2 - 2|\boldsymbol{x}||\boldsymbol{y}| + |\boldsymbol{y}|^2 = \big(|\boldsymbol{x}| - |\boldsymbol{y}|\big)^2,$$

откуда $|\boldsymbol{x}+\boldsymbol{y}| \geqslant \big||\boldsymbol{x}| - |\boldsymbol{y}|\big|$. □

Г. Примеры евклидовых векторных пространств. В следующих примерах самостоятельно проверьте выполнение аксиом **СП1–СП3**.

8.13. Важнейшими для геометрии примерами являются векторные пространства направленных отрезков, в которых скалярное умножение введено, как в п. 8.1.

8.14. Арифметическое векторное пространство $\mathbb{R}^n$ становится евклидовым, если для векторов (столбцов)

$$\boldsymbol{x} = X = \big(x^1, \ldots, x^n\big)^T, \quad \boldsymbol{y} = Y = \big(y^1, \ldots, y^n\big)^T$$

определить скалярное произведение по формуле

$$(\boldsymbol{x},\boldsymbol{y}) = \sum_{j=1}^{n} x^j y^j = X^T Y. \tag{8.5}$$

8.15. В векторном пространстве $\mathbb{R}^{m\times n}$ прямоугольных матриц можно ввести скалярное произведение по формуле

$$(X,Y) = \operatorname{tr}(X^T Y). \tag{8.6}$$

8.16. В сущности, формулы (8.5) и (8.6) реализуют одну и ту же идею скалярного произведения упорядоченных наборов чисел как суммы попарных произведений одноимённых компонент этих наборов.

8.17. В пространстве многочленов $\mathbb{R}[t]_n$ степени не выше n можно ввести скалярное произведение векторов (многочленов)

$$\boldsymbol{x} = x(t) = a_0 + a_1 t + \cdots + a_n t^n, \quad \boldsymbol{y} = y(t) = b_0 + b_1 t + \cdots + b_n t^n$$

следующим образом:

$$(\boldsymbol{x},\boldsymbol{y}) = \int\limits_{\alpha}^{\beta} x(t)y(t)dt,$$

где α, β — произвольные вещественные числа.

3. Ортонормированные базисы

А. Ортогональные векторы. Пусть $\mathcal{E}$ — евклидово векторное пространство.

8.18. Определение.

1. Векторы $\boldsymbol{x},\boldsymbol{y} \in \mathcal{E}$ называются *ортогональными*, если $(\boldsymbol{x},\boldsymbol{y}) = 0$; обозначение $\boldsymbol{x} \perp \boldsymbol{y}$.

2. Пусть $\mathcal{P} \Subset \mathcal{E}$ — подпространство в $\mathcal{E}$. Вектор $\boldsymbol{x}$ называется *ортогональным подпространству* $\mathcal{P} \Subset \mathcal{E}$, если он ортогонален любому вектору из $\mathcal{P}$:
$$\boldsymbol{x} \perp \mathcal{P} \quad \Leftrightarrow \quad \forall \boldsymbol{y} \in \mathcal{P}\ \boldsymbol{x} \perp \boldsymbol{y}.$$
3. Два подпространства $\mathcal{P} \Subset \mathcal{E}$, $\mathcal{Q} \Subset \mathcal{E}$ в $\mathcal{E}$ называются *ортогональными*, если любой вектор $\boldsymbol{x} \in \mathcal{P}$ ортогонален любому вектору $\boldsymbol{y} \in \mathcal{Q}$:
$$\mathcal{P} \perp \mathcal{Q} \quad \Leftrightarrow \quad \forall \boldsymbol{x} \in \mathcal{P},\ \forall \boldsymbol{y} \in \mathcal{Q}\ \boldsymbol{x} \perp \boldsymbol{y}.$$

8.19. Предложение.

1. $\boldsymbol{x} \perp \boldsymbol{x}$ *тогда и только тогда, когда* $\boldsymbol{x} = \mathbf{0}$.
2. *Если* $\boldsymbol{x} \in \mathcal{P}$ *и* $\boldsymbol{x} \perp \mathcal{P}$, *то* $\boldsymbol{x} = \mathbf{0}$.
3. *Если* $\boldsymbol{y} \perp \boldsymbol{x}_1, \ \ldots, \ \boldsymbol{y} \perp \boldsymbol{x}_p$, *то* $\boldsymbol{y} \perp L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$.
4. *Теорема Пифагора: равенство*
$$|\boldsymbol{x} + \boldsymbol{y}|^2 = |\boldsymbol{x}|^2 + |\boldsymbol{y}|^2$$
выполняется тогда и только тогда, когда $\boldsymbol{x} \perp \boldsymbol{y}$.
5. *Если ненулевые векторы* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ *попарно ортогональны, т.е.* $\boldsymbol{x}_j \perp \boldsymbol{x}_k$, $j \neq k$, *то они линейно независимы.*

Доказательство. Утверждение 1 является просто переформулировкой аксиомы **СП3**.

2. Пусть $\boldsymbol{x} \in \mathcal{P}$ и $\boldsymbol{x} \perp \mathcal{P}$; тогда, в частности, $\boldsymbol{x} \perp \boldsymbol{x}$, так что (в силу утверждения 1) $\boldsymbol{x} = \mathbf{0}$.

3. Пусть $\boldsymbol{y} \perp \boldsymbol{x}_1, \ \ldots, \ \boldsymbol{y} \perp \boldsymbol{x}_p$. Любой вектор $\boldsymbol{x} \in L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ может быть представлен в виде $\boldsymbol{x} = \alpha^j \boldsymbol{x}_j$ (суммирование по $j = 1, \ldots, p$), поэтому
$$(\boldsymbol{y}, \boldsymbol{x}) = \big(\boldsymbol{y},\ \alpha^j \boldsymbol{x}_j\big) = \alpha^j \underbrace{(\boldsymbol{y}, \boldsymbol{x}_j)}_{=0} = 0 \ \Leftrightarrow\ \boldsymbol{y} \perp L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p).$$

4. Имеем
$$|\boldsymbol{x} + \boldsymbol{y}|^2 = (\boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{x}) + 2(\boldsymbol{x}, \boldsymbol{y}) + (\boldsymbol{y}, \boldsymbol{y});$$
это выражение равно $|\boldsymbol{x}|^2 + |\boldsymbol{y}|^2$ в том и только том случае, если $(\boldsymbol{x}, \boldsymbol{y}) = 0$.

5. Рассмотрим линейную комбинацию векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$, равную нулевому вектору:
$$\alpha^1 \boldsymbol{x}_1 + \cdots + \alpha^j \boldsymbol{x}_j + \cdots + \alpha^p \boldsymbol{x}_p = \mathbf{0}.$$
Умножая скалярно это равенство на вектор $\boldsymbol{x}_j$, получаем
$$\alpha^1 \underbrace{(\boldsymbol{x}_1, \boldsymbol{x}_j)}_{=0} + \cdots + \alpha^j \underbrace{(\boldsymbol{x}_j, \boldsymbol{x}_j)}_{=1} + \cdots + \alpha^p \underbrace{(\boldsymbol{x}_p, \boldsymbol{x}_j)}_{=0} = 0,$$

откуда получаем $\alpha^j = 0$. Таким образом, все коэффициенты линейной комбинации равны нулю, что означает линейную независимость векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$. □

8.20. Определение. *Ортогональной проекцией* вектора $\boldsymbol{x} \in \mathcal{E}$ на подпространство $\mathcal{P} \Subset \mathcal{E}$ (см. рис. 8.2) называется такой вектор $\boldsymbol{x}' \in \mathcal{P}$, что разность $\boldsymbol{x} - \boldsymbol{x}'$ ортогональна подпространству $\mathcal{P}$; обозначение $\boldsymbol{x}' = \operatorname{Pr}^{\perp}_{\mathcal{P}} \boldsymbol{x}$ (если ясно, что речь идёт об ортогональной проекции, используется сокращённое обозначение $\operatorname{Pr}_{\mathcal{P}} \boldsymbol{x}$).

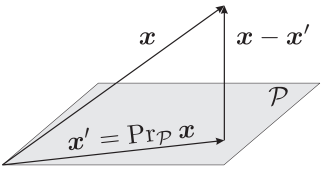


Рис. 8.2. Ортогональная проекция вектора на подпространство

Б. Ортонормированный базис.

8.21. Определение. Семейство векторов $(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_p)$ евклидова векторного пространства называется *ортонормированным*, если

(i) длина каждого вектора равна 1,

(ii) любые два различных вектора ортогональны,

т.е. если

$$(\boldsymbol{i}_j, \boldsymbol{i}_k) = \delta_{jk} = \begin{cases} 1, & \text{если } j = k, \\ 0, & \text{если } j \neq k. \end{cases} \tag{8.7}$$

8.22. В частности, можно рассматривать *ортонормированные базисы*, которые очень удобны для вычислений.

8.23. Предложение. *Пусть $\boldsymbol{I} = (\boldsymbol{i}_1, \ldots, \boldsymbol{i}_n)$ — произвольный ортонормированный базис в евклидовом пространстве $\mathcal{E}$.*

1. *Скалярное произведение двух векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ выражается через их координаты относительно базиса $\boldsymbol{I}$ формулой*

$$(\boldsymbol{x}, \boldsymbol{y}) = \delta_{jk} x^j y^k = \sum_{j=1}^{n} x^j y^j = X^T Y, \tag{8.8}$$

где X и Y — столбцы координат векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ в базисе $\boldsymbol{I}$.

2. *Координаты произвольного вектора $\boldsymbol{x}$ относительно ортонормированного базиса могут быть вычислены по формулам*

$$x^k = (\boldsymbol{x}, \boldsymbol{i}_k), \tag{8.9}$$

называемым формулами Гиббса[1].

[1]Дж. У. Гиббс (J. W. Gibbs) — американский физик, математик и механик (1839–1903).

Доказательство. 1. Разложим векторы $\boldsymbol{x}$ и $\boldsymbol{y}$ по базису $\boldsymbol{I}$:

$$\boldsymbol{x} = x^j \boldsymbol{i}_j, \quad \boldsymbol{y} = y^k \boldsymbol{i}_k,$$

где подразумевается суммирование по $j, k = 1, \ldots, n = \dim \mathcal{E}$. Имеем:

$$(\boldsymbol{x}, \boldsymbol{y}) = \left(x^j \boldsymbol{i}_j,\ y^k \boldsymbol{i}_k\right) = x^j \left(\boldsymbol{i}_j,\ y^k \boldsymbol{i}_k\right) =$$

$$= x^j y^k \underbrace{(\boldsymbol{i}_j, \boldsymbol{i}_k)}_{=\delta_{jk}} = \sum_{j=1}^{n} x^j y^j,$$

поскольку в сумме остаются только те слагаемые, в которых $j = k$ (см. с. 113).

2. Умножив скалярно обе части разложения $\boldsymbol{x} = x^j \boldsymbol{i}_j$ вектора $\boldsymbol{x}$ по базису $\boldsymbol{I}$ на вектор $\boldsymbol{i}_k$, получим

$$(\boldsymbol{x}, \boldsymbol{i}_k) = (x^j \boldsymbol{i}_j, \boldsymbol{i}_k) = x^j (\boldsymbol{i}_j, \boldsymbol{i}_k) = x^j \delta_{jk} = x^k. \qquad \square$$

8.24. Обратите внимание на положение индексов в формуле Гиббса (8.9): слева в этой формуле индекс k — нижний, а справа — верхний. Такое перемещение индекса из одной позиции в другую характерно для формул, записанных в ортонормированных базисах; причина этого явления и его последствия будут обсуждаться во втором семестре.

8.25. Предложение. *Пусть* $\left(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_p\right)$ *— ортонормированное семейство векторов в евклидовом пространстве* $\mathcal{E}$, $p \leqslant \dim \mathcal{E}$. *Ортогональная проекция вектора* $\boldsymbol{x} \in \mathcal{E}$ *на подпространство* $\mathcal{P} = L(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_p)$ *равна*

$$\mathrm{Pr}^{\perp}_{\mathcal{P}}\, \boldsymbol{x} = (\boldsymbol{x}, \boldsymbol{i}_1)\boldsymbol{i}_1 + \cdots + (\boldsymbol{x}, \boldsymbol{i}_p)\boldsymbol{i}_p = \sum_{j=1}^{p} (\boldsymbol{x}, \boldsymbol{i}_j)\boldsymbol{i}_j. \tag{8.10}$$

Обратите внимание, что последнюю сумму нельзя записать с помощью правила суммирования Эйнштейна (объясните почему).

Доказательство. Достаточно показать, что вектор

$$\boldsymbol{y} = \boldsymbol{x} - \sum_{j=1}^{p} (\boldsymbol{x}, \boldsymbol{i}_j)\boldsymbol{i}_j \tag{8.11}$$

ортогонален каждому из векторов $\boldsymbol{i}_k$, $k = 1, \ldots, p$. Действительно,

$$(\boldsymbol{y}, \boldsymbol{i}_k) = (\boldsymbol{x}, \boldsymbol{i}_k) - \sum_{j=1}^{p} (\boldsymbol{x}, \boldsymbol{i}_j) \underbrace{(\boldsymbol{i}_j, \boldsymbol{i}_k)}_{=\delta_{jk}} = (\boldsymbol{x}, \boldsymbol{i}_k) - (\boldsymbol{x}, \boldsymbol{i}_k) = 0. \qquad \square$$

B. Существование ортонормированных семейств.

8.26. Предложение. *В любом евклидовом пространстве $\mathcal{E}$ размерности n существуют ортонормированные семейства векторов.*

Доказательство. Докажем утверждение методом математической индукции. Для произвольного вектора $\boldsymbol{a} \in \mathcal{E}$ вектор $\boldsymbol{i} = \boldsymbol{a}/|\boldsymbol{a}|$ имеет единичную длину и, следовательно, образует ортонормированное семейство, состоящее из одного члена.

Допустим, что уже доказано существование ортонормированных семейств, состоящих из p элементов ($p < n$, где $n = \dim \mathcal{E}$). Пусть $\left(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_p\right)$ — одно из таких семейств и $\boldsymbol{x} \notin L(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_p) \equiv \mathcal{P}$ — произвольный вектор; такой вектор существует, поскольку $\dim L(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_p) = p < n$ (напомним, что попарно ортогональные векторы линейно независимы). Вектор $\boldsymbol{y} = \boldsymbol{x} - \mathrm{Pr}_{\mathcal{P}}\, \boldsymbol{x}$ ортогонален подпространству $\mathcal{P}$ (т.е. каждому из векторов $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_p$), а вектор $\boldsymbol{i}_{p+1} = \boldsymbol{y}/|\boldsymbol{y}|$, кроме того, имеет единичную длину. □

8.27. Пошаговый метод построения ортонормированного семейства векторов, использованный в доказательстве предложения 8.26, называется *методом ортогонализации Грама*[1]*—Шмидта*[2]; более подробно мы обсудим его в следующем семестре.

4. Векторное и смешанное произведения векторов в трехмерном евклидовом пространстве

Мы живём в трёхмерном евклидовом мире — это экспериментальный факт, зафиксированный в аксиоме размерности **P3**. Однако многие физические задачи допускают плоскую, двумерную формулировку. Поэтому из всех евклидовых (векторных и точечных) пространств в первую очередь интересны для нас двумерное и трёхмерное.

В трёхмерном евклидовом векторном пространстве помимо скалярного умножения имеются ещё две операции над векторами, специфичные именно для размерности 3: векторное и смешанное умножения. Эти операции находят широкое применение во всех приложениях векторного исчисления.

Изложение в этом разделе вынужденно использует некоторые наглядные представления, поскольку в определениях векторного и смешанного произведения содержатся упоминания правых и левых троек векторов (см. п. 6.81, с. 152). Чисто аксиоматическое введение понятий векторного и смешанного произведений возможно, но тогда наглядные представления придётся привлекать при использовании этих понятий в естественнонаучных приложениях.

---

[1] Й. Грам (J. P. Gram) — датский математик (1850–1916).

[2] Э. Шмидт (E. Schmidt) — немецкий математик (1876–1959).

В этом разделе мы используем традиционные для трёхмерной ситуации обозначения: векторы произвольного ортонормированного базиса евклидова пространства обозначаем $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$, координаты векторов — a_x, a_y, a_z.

A. Векторное произведение векторов.

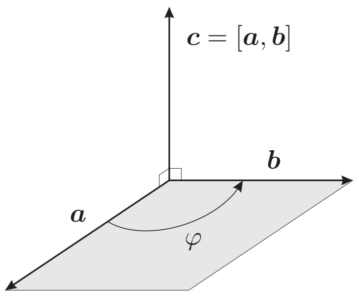


Рис. 8.3.

8.28. Определение. *Векторным произведением* упорядоченной пары векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ называется вектор $\boldsymbol{c} \overset{\text{def}}{=\!=} [\boldsymbol{a}, \boldsymbol{b}]$, удовлетворяющий следующим требованиям:

(i) $|\boldsymbol{c}| = |\boldsymbol{a}| \cdot |\boldsymbol{b}| \cdot \sin\varphi$, где φ — (неориентированный) угол между векторами $\boldsymbol{a}$ и $\boldsymbol{b}$;

(ii) вектор $\boldsymbol{c}$ ортогонален каждому из векторов $\boldsymbol{a}$ и $\boldsymbol{b}$;

(iii) векторы $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ (в указанном порядке) образуют правую тройку (см. рис. 8.3).

Векторное произведение векторов обозначается также, особенно в физической литературе, символом $\boldsymbol{a} \times \boldsymbol{b}$ (в англоязычной литературе используется даже термин «cross-product» — умножение «крестом»).

8.29. Векторное произведение двух векторов равно нулевому вектору, если один из сомножителей есть нулевой вектор либо если перемножаемые векторы коллинеарны. Для неколлинеарных векторов условие (i) означает, что длина вектора $[\boldsymbol{a}, \boldsymbol{b}]$ численно равна площади $S_{\boldsymbol{ab}}$ параллелограмма, построенного на векторах $\boldsymbol{a}$ и $\boldsymbol{b}$:

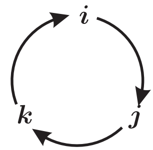
$$S_{\boldsymbol{ab}} = \Big|[\boldsymbol{a}, \boldsymbol{b}]\Big|.$$

8.30. Предложение (критерий коллинеарности двух векторов). *Векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ коллинеарны тогда и только тогда, когда $[\boldsymbol{a}, \boldsymbol{b}] = \boldsymbol{0}$.*

Доказательство. Векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ коллинеарны тогда и только тогда, когда либо $\boldsymbol{a} = \boldsymbol{0}$, либо $\boldsymbol{b} = \boldsymbol{0}$, либо $\sin\varphi = 0$. Это равносильно тому, что $\Big|[\boldsymbol{a}, \boldsymbol{b}]\Big| = 0$, т.е. $[\boldsymbol{a}, \boldsymbol{b}] = \boldsymbol{0}$. □

8.31. Таблица векторного умножения для векторов *правого* ортонормированного базиса приведена ниже. Справа от таблицы изображена диаграмма, облегчающая запоминание: произведение любых двух векторов равно третьему вектору, взятому со знаком $+$, если движение по диаграмме от первого сомножителя ко второму происходит по стрелке, и со знаком $-$ в противном случае:

I множ. \ II множ.	$\boldsymbol{i}$	$\boldsymbol{j}$	$\boldsymbol{k}$
$\boldsymbol{i}$	0	$\boldsymbol{k}$	$-\boldsymbol{j}$
$\boldsymbol{j}$	$-\boldsymbol{k}$	0	$\boldsymbol{i}$
$\boldsymbol{k}$	$\boldsymbol{j}$	$-\boldsymbol{i}$	0



(8.12)

8.32. Для векторов *левого* ортонормированного базиса все знаки $+$ и $-$ должны быть изменены на противоположные. В дальнейшем мы будем пользоваться исключительно правыми базисами.

8.33. Предложение. *Операция векторного умножения антикоммутативна, т.е. для любых векторов* $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{E}$ *имеем*

$$[\boldsymbol{a}, \boldsymbol{b}] = -[\boldsymbol{b}, \boldsymbol{a}].$$

Доказательство. Утверждение очевидно для коллинеарных векторов (в этом случае $[\boldsymbol{a}, \boldsymbol{b}] = -[\boldsymbol{b}, \boldsymbol{a}] = \boldsymbol{0}$). Если векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ не коллинеарны, то длины (ненулевых) векторов $[\boldsymbol{a}, \boldsymbol{b}]$ и $[\boldsymbol{b}, \boldsymbol{a}]$ равны, оба эти вектора перпендикулярны плоскости, содержащей векторы $\boldsymbol{a}$ и $\boldsymbol{b}$. Поскольку тройка векторов $\boldsymbol{a}, \boldsymbol{b}, [\boldsymbol{a}, \boldsymbol{b}]$ правая, то тройка $\boldsymbol{b}, \boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{a}]$ также правая, так что тройка $\boldsymbol{a}, \boldsymbol{b}, [\boldsymbol{b}, \boldsymbol{a}]$ — левая. □

Б. Смешанное произведение векторов.

8.34. Определение. *Смешанным произведением* $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ трёх векторов $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ называется число, равное скалярному произведению вектора $\boldsymbol{a}$ на векторное произведение векторов $\boldsymbol{b}$ и $\boldsymbol{c}$:

$$(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \stackrel{\text{def}}{=\!=} \big(\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big).$$

8.35. Предложение (критерий компланарности трёх векторов)**.** *Векторы* $\boldsymbol{a}$, $\boldsymbol{b}$ *и* $\boldsymbol{c}$ *компланарны тогда и только тогда, когда* $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \boldsymbol{0}$.

Доказательство. Необходимость. Пусть векторы $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ компланарны. Будем считать, что $\boldsymbol{a} \neq \boldsymbol{0}$ и векторы $\boldsymbol{b}$ и $\boldsymbol{c}$ неколлинеарны (очевидно, в каждом из этих случаев $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = 0$). Тогда векторы $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ параллельны плоскости π, содержащей векторы $\boldsymbol{b}$ и $\boldsymbol{c}$, причём вектор $[\boldsymbol{b}, \boldsymbol{c}]$ ортогонален этой плоскости. Следовательно, $\big(\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big) = 0$.

Достаточность. Пусть $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = 0$. Тогда либо $|\boldsymbol{a}| = 0$, либо $[\boldsymbol{b}, \boldsymbol{c}] = \boldsymbol{0}$, либо $\cos\alpha = 0$, где α — угол между векторами $\boldsymbol{a}$ и $[\boldsymbol{b}, \boldsymbol{c}]$. Это означает, что либо $\boldsymbol{a} = \boldsymbol{0}$, либо $\boldsymbol{b}$ и $\boldsymbol{c}$ коллинеарны, либо $\boldsymbol{a}$

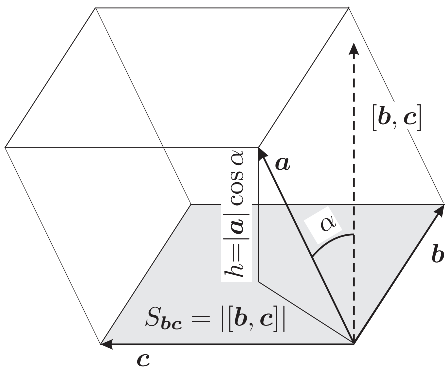


Рис. 8.4. Объём параллелепипеда

параллелен плоскости π, содержащей векторы $\boldsymbol{b}$ и $\boldsymbol{c}$. Во всех трёх случаях векторы $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ компланарны. □

8.36. Предложение. *Смешанное произведение некомпланарных векторов $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ равно ориентированному объёму $\big|\Pi(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})\big|$ параллелепипеда, построенного на этих векторах, отложенных от одной точки*:

$$(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=\big|\Pi(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})\big|=\begin{cases}\ \ V(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}), & \text{если тройка } \boldsymbol{a},\boldsymbol{b},\boldsymbol{c} \text{ правая},\\ -V(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}), & \text{если тройка } \boldsymbol{a},\boldsymbol{b},\boldsymbol{c} \text{ левая},\end{cases}$$

где символом $V(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$ обозначен объём параллелепипеда в смысле элементарной геометрии.

Доказательство. Отложив векторы $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ от одной точки, получим параллелепипед, объём которого можно найти по формуле

$$V(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=S(\boldsymbol{b},\boldsymbol{c})\cdot h,$$

где $S(\boldsymbol{b},\boldsymbol{c})$ — площадь основания, численно равная длине вектора $\Big|[\boldsymbol{b},\boldsymbol{c}]\Big|$, а h — высота параллелепипеда, равная $|\boldsymbol{a}|\cdot|\cos\alpha|$, где α — угол между векторами $\boldsymbol{a}$ и $[\boldsymbol{b},\boldsymbol{c}]$. Итак,

$$V(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=\Big|[\boldsymbol{b},\boldsymbol{c}]\Big|\cdot|\boldsymbol{a}|\cdot|\cos\alpha|=\Big|\big(\boldsymbol{a},[\boldsymbol{b},\boldsymbol{c}]\big)\Big|=\Big|\big(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}\big)\Big|.$$

Знак смешанного произведения $(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$ определяется только знаком $\cos\alpha$, но $\cos\alpha>0$ тогда и только тогда, когда векторы $\boldsymbol{a}$ и $[\boldsymbol{b},\boldsymbol{c}]$ направлены в одну сторону от плоскости $\pi(\boldsymbol{b},\boldsymbol{c})$, содержащей векторы $\boldsymbol{b}$ и $\boldsymbol{c}$, т.е. тогда и только тогда, когда тройка $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ — правая. □

8.37. Предложение. *Для любых векторов* $\boldsymbol{a}$, $\boldsymbol{b}$ *и* $\boldsymbol{c}$

$$\Big(\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\Big) = \Big([\boldsymbol{a}, \boldsymbol{b}], \boldsymbol{c}\Big). \tag{8.13}$$

Доказательство. Для компланарных векторов утверждение очевидно. Если векторы $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ некомпланарны, то тройки векторов $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ и $\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{b}$ одинаково ориентированы. Для определённости будем считать, что обе тройки правые. Тогда

$$\Big([\boldsymbol{a}, \boldsymbol{b}], \boldsymbol{c}\Big) = \Big(\boldsymbol{c}, [\boldsymbol{a}, \boldsymbol{b}]\Big) = \big(\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{b}\big) = V_{\boldsymbol{cab}} =$$
$$= V_{\boldsymbol{abc}} = \big(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\big) = \Big(\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\Big). \quad \square$$

8.38. Следствие. *Для любых векторов* $\boldsymbol{a}$, $\boldsymbol{b}$ *и* $\boldsymbol{c}$

$$\begin{aligned}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) &= (\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{a}) = (\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{b}) = \\ &= -(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{b}) = -(\boldsymbol{b}, \boldsymbol{a}, \boldsymbol{c}) = -(\boldsymbol{c}, \boldsymbol{b}, \boldsymbol{a}).\end{aligned} \tag{8.14}$$

Доказательство. Например, имеем

$$\big(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\big) = \Big(\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\Big) = \Big(\boldsymbol{a}, -[\boldsymbol{c}, \boldsymbol{b}]\Big) = -(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{b}).$$

Можно рассуждать иначе: каждое из шести смешанных произведений совпадает (без учёта знака) с объёмом параллелепипеда, построенного на векторах $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$, знак же зависит от ориентации тройки, которая сохраняется при циклической перестановке векторов и меняется на противоположный при обычной перестановке. $\square$

8.39. Следствие. *Смешанное произведение линейно по каждому из трёх сомножителей, например,*

$$\Big(\alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2,\ \boldsymbol{b},\ \boldsymbol{c}\Big) = \alpha_1 \Big(\boldsymbol{a}_1,\ \boldsymbol{b},\ \boldsymbol{c}\Big) + \alpha_2 \Big(\boldsymbol{a}_2,\ \boldsymbol{b},\ \boldsymbol{c}\Big).$$

Доказательство. Линейность по первому аргументу является следствием линейности скалярного произведения; линейность по остальным сомножителям вытекает из формул (8.14). $\square$

8.40. Предложение. *Векторное произведение линейно по каждому из сомножителей, например,*

$$\big[\alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2,\ \boldsymbol{b}\big] = \alpha_1 [\boldsymbol{a}_1, \boldsymbol{b}] + \alpha_2 [\boldsymbol{a}_2, \boldsymbol{b}].$$

Доказательство. Докажем линейность по первому сомножителю. Введём обозначение

$$\boldsymbol{d} = [\alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2, \boldsymbol{b}] - \alpha_1 [\boldsymbol{a}_1, \boldsymbol{b}] - \alpha_2 [\boldsymbol{a}_2, \boldsymbol{b}].$$

Имеем

$$(\boldsymbol{d},\boldsymbol{d}) = \Big([\alpha_1\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2,\boldsymbol{b}]-\alpha_1[\boldsymbol{a}_1,\boldsymbol{b}]-\alpha_2[\boldsymbol{a}_2,\boldsymbol{b}],\ \boldsymbol{d}\Big) = \\ = \big(\alpha_1\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2,\boldsymbol{b},\boldsymbol{d}\big)-\alpha_1\big(\boldsymbol{a}_1,\boldsymbol{b},\boldsymbol{d}\big)-\alpha_2\big(\boldsymbol{a}_2,\boldsymbol{b},\boldsymbol{d}\big) =$$

(используем линейность смешанного произведения)

$$= \alpha_1\big(\boldsymbol{a}_1,\boldsymbol{b},\boldsymbol{d}\big)+\alpha_2\big(\boldsymbol{a}_2,\boldsymbol{b},\boldsymbol{d}\big)-\alpha_1\big(\boldsymbol{a}_1,\boldsymbol{b},\boldsymbol{d}\big)-\alpha_2\big(\boldsymbol{a}_2,\boldsymbol{b},\boldsymbol{d}\big) = 0,$$

откуда в силу аксиомы **СП3** скалярного произведения следует $\boldsymbol{d}=\boldsymbol{0}$, что и требовалось. □

B. Вычисление векторного и смешанного произведений в ортонормированном базисе.

8.41. Пусть векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ заданы координатами относительно *правого* ортонормированного базиса $\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}$:

$$\boldsymbol{a} = a_x\boldsymbol{i}+a_y\boldsymbol{j}+a_z\boldsymbol{k}, \quad \boldsymbol{b} = b_x\boldsymbol{i}+b_y\boldsymbol{j}+b_z\boldsymbol{k},$$

Пользуясь линейностью векторного произведения (предложение 8.40) и таблицей (8.12) векторного умножения для векторов ортонормированного базиса, для векторного произведения $[\boldsymbol{a},\boldsymbol{b}]$ получаем:

$$\begin{aligned} [\boldsymbol{a},\boldsymbol{b}] &= \Big[a_x\boldsymbol{i}+a_y\boldsymbol{j}+a_z\boldsymbol{k},\ b_x\boldsymbol{i}+b_y\boldsymbol{j}+b_z\boldsymbol{k}\Big] = \\ &= a_xb_x\underbrace{[\boldsymbol{i},\boldsymbol{i}]}_{=\boldsymbol{0}}+a_xb_y\underbrace{[\boldsymbol{i},\boldsymbol{j}]}_{=\boldsymbol{k}}+a_xb_z\underbrace{[\boldsymbol{i},\boldsymbol{k}]}_{=-\boldsymbol{j}}+ \\ &+ a_yb_x\underbrace{[\boldsymbol{j},\boldsymbol{i}]}_{=-\boldsymbol{k}}+a_yb_y\underbrace{[\boldsymbol{j},\boldsymbol{j}]}_{=\boldsymbol{0}}+a_yb_z\underbrace{[\boldsymbol{j},\boldsymbol{k}]}_{=\boldsymbol{i}}+ \\ &+ a_zb_x\underbrace{[\boldsymbol{k},\boldsymbol{i}]}_{=\boldsymbol{j}}+a_zb_y\underbrace{[\boldsymbol{k},\boldsymbol{j}]}_{=-\boldsymbol{i}}+a_zb_z\underbrace{[\boldsymbol{k},\boldsymbol{k}]}_{=\boldsymbol{0}} \\ &= \boldsymbol{i}\Big(a_yb_z-a_zb_y\Big)+\boldsymbol{j}\Big(a_zb_x-a_xb_z\Big)+\boldsymbol{k}\Big(a_xb_y-a_yb_z\Big). \end{aligned}$$

8.42. Эту формулу удобно записать с помощью определителей:

$$[\boldsymbol{a},\boldsymbol{b}] = \boldsymbol{i}\begin{vmatrix} a_y & a_z \\ b_y & b_z \end{vmatrix} - \boldsymbol{j}\begin{vmatrix} a_x & a_z \\ b_z & b_z \end{vmatrix} + \boldsymbol{k}\begin{vmatrix} a_x & a_y \\ b_z & b_y \end{vmatrix}. \tag{8.15}$$

8.43. Запись можно ещё сократить, введя символический определитель третьего порядка:

$$[\boldsymbol{a},\boldsymbol{b}] = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a_x & a_y & a_z \\ b_z & b_y & b_z \end{vmatrix}; \tag{8.16}$$

здесь предполагается, что определитель нужно разложить по первой строке.

8.44. В случае *левого* ортонормированного базиса все векторные произведения базисных векторов имеют противоположный знак, поэтому формула для векторного произведения имеет вид

$$[\boldsymbol{a}, \boldsymbol{b}] = -\begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a_x & a_y & a_z \\ b_z & b_y & b_z \end{vmatrix}.$$

8.45. Предложение. *Смешанное произведение векторов* $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ *выражается через их координаты относительно правого ортонормированного базиса* $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ *формулой*

$$(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix}. \tag{8.17}$$

Доказательство. Введём обозначение

$$\boldsymbol{d} = [\boldsymbol{b}, \boldsymbol{c}] = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix} = \begin{pmatrix} d_x \\ d_y \\ d_z \end{pmatrix}.$$

Имеем

$$(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = (\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]) = a_x d_x + a_y d_y + a_z d_z =$$

$$= a_x \begin{vmatrix} b_y & b_z \\ c_y & c_z \end{vmatrix} - a_y \begin{vmatrix} b_x & b_z \\ c_x & c_z \end{vmatrix} + a_z \begin{vmatrix} b_y & b_z \\ c_y & c_z \end{vmatrix} = \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix}. \quad \square$$

8.46. В случае левого ортонормированного базиса в формуле (8.17) перед определителем нужно поставить знак «−».

Г. Двойное векторное произведение. Из трёх векторов $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ можно составить произведения $\big[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big]$ или $\big[[\boldsymbol{a}, \boldsymbol{b}], \boldsymbol{c}\big]$, называемые *двойными векторными произведениями*.

8.47. Теорема (формула Лагранжа[1]). *Имеет место следующая формула*:

$$\big[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big] = \boldsymbol{b}(\boldsymbol{a}, \boldsymbol{c}) - \boldsymbol{c}(\boldsymbol{a}, \boldsymbol{b}). \tag{8.18}$$

[1] Ж. Л. Лагранж (J. L. Lagrange) — французский математик, астроном и механик (1736–1813).

Доказательство с помощью координат. Рассмотрим правый ортонормированный базис $\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{i}_3$, третий вектор $\boldsymbol{i}_3$ которого сонаправлен с вектором $\boldsymbol{c}$, а второй $\boldsymbol{i}_2$ лежит в плоскости векторов $\boldsymbol{b}$ и $\boldsymbol{c}$. В этом базисе векторы $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ имеют следующие разложения:

$$\boldsymbol{a} = a_1\boldsymbol{i}_1 + a_2\boldsymbol{i}_2 + a_3\boldsymbol{i}_3, \quad \boldsymbol{b} = b_2\boldsymbol{i}_2 + b_3\boldsymbol{i}_3, \quad \boldsymbol{c} = c_3\boldsymbol{i}_3.$$

Имеем:

$$[\boldsymbol{b}, \boldsymbol{c}] = \begin{vmatrix} \boldsymbol{i}_1 & \boldsymbol{i}_2 & \boldsymbol{i}_3 \\ 0 & b_2 & b_3 \\ 0 & 0 & c_3 \end{vmatrix} = b_2 c_3 \boldsymbol{i}_1,$$

$$[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]] = \begin{vmatrix} \boldsymbol{i}_1 & \boldsymbol{i}_2 & \boldsymbol{i}_3 \\ a_1 & a_2 & a_3 \\ b_2 c_3 & 0 & 0 \end{vmatrix} = a_3 b_2 c_3 \boldsymbol{i}_2 - a_2 b_2 c_3 \boldsymbol{i}_3. \tag{8.19}$$

Далее,

$$(\boldsymbol{a}, \boldsymbol{c}) = a_3 c_3, \quad (\boldsymbol{a}, \boldsymbol{b}) = a_2 b_2 + a_3 b_3,$$

так что

$$\boldsymbol{b}(\boldsymbol{a}, \boldsymbol{c}) = a_3 c_3 (b_2\boldsymbol{i}_2 + b_3\boldsymbol{i}_3), \quad \boldsymbol{c}(\boldsymbol{a}, \boldsymbol{b}) = (a_2 b_2 + a_3 b_3) c_3 \boldsymbol{i}_3$$

и далее

$$\boldsymbol{b}(\boldsymbol{a}, \boldsymbol{c}) - \boldsymbol{c}(\boldsymbol{a}, \boldsymbol{b}) = a_3 b_2 c_3 \boldsymbol{i}_2 - a_2 b_2 c_3 \boldsymbol{i}_3.$$

Сравнивая последнее выражение с (8.19), убеждаемся в справедливости (8.18). □

Бескоординатное доказательство. Заметим, что векторы $\boldsymbol{b}$, $\boldsymbol{c}$ и $\big[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big]$ компланарны, поскольку по определению векторного произведения все они ортогональны вектору $[\boldsymbol{b}, \boldsymbol{c}]$. Поэтому вектор $\big[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big]$ можно разложить по $\boldsymbol{b}$ и $\boldsymbol{c}$:

$$\big[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big] = \beta\boldsymbol{b} + \gamma\boldsymbol{c},$$

где β и γ — подлежащие определению коэффициенты.

Умножим обе части последнего равенства скалярно на $\boldsymbol{a}$. Так как $\big[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big] \perp \boldsymbol{a}$, скалярное произведение в левой части равенства будет равно нулю, и мы получим

$$0 = \beta(\boldsymbol{a}, \boldsymbol{b}) + \gamma(\boldsymbol{a}, \boldsymbol{c}).$$

Отсюда следует, что

$$\frac{\beta}{(\boldsymbol{a}, \boldsymbol{c})} = -\frac{\gamma}{(\boldsymbol{a}, \boldsymbol{b})} = \alpha,$$

так что

$$\beta = \alpha(\boldsymbol{a}, \boldsymbol{c}), \quad \gamma = -\alpha(\boldsymbol{a}, \boldsymbol{b}).$$

Таким образом, получаем

$$\big[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big] = \alpha\big(\boldsymbol{b}(\boldsymbol{a}, \boldsymbol{c}) - \boldsymbol{c}(\boldsymbol{a}, \boldsymbol{b})\big). \tag{8.20}$$

Докажем теперь, что коэффициент α не зависит от вектора $\boldsymbol{a}$. Предположим обратное и запишем равенство (8.20) для произвольных векторов $\boldsymbol{a}$ и $\boldsymbol{a}'$ и их суммы $\boldsymbol{a} + \boldsymbol{a}'$:

$$\big[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big] = \alpha\big(\boldsymbol{b}(\boldsymbol{a}, \boldsymbol{c}) - \boldsymbol{c}(\boldsymbol{a}, \boldsymbol{b})\big),$$

$$\big[\boldsymbol{a}', [\boldsymbol{b}, \boldsymbol{c}]\big] = \alpha'\big(\boldsymbol{b}(\boldsymbol{a}', \boldsymbol{c}) - \boldsymbol{c}(\boldsymbol{a}', \boldsymbol{b})\big),$$
$$\big[\boldsymbol{a} + \boldsymbol{a}', [\boldsymbol{b}, \boldsymbol{c}]\big] = \alpha''\big(\boldsymbol{b}(\boldsymbol{a} + \boldsymbol{a}', \boldsymbol{c}) - \boldsymbol{c}(\boldsymbol{a} + \boldsymbol{a}', \boldsymbol{b})\big).$$

Сложим два первые равенства:

$$\big[\boldsymbol{a} + \boldsymbol{a}', [\boldsymbol{b}, \boldsymbol{c}]\big] = \boldsymbol{b}\big(\alpha\boldsymbol{a} + \alpha'\boldsymbol{a}', \boldsymbol{c}\big) - \boldsymbol{c}\big(\alpha\boldsymbol{a} + \alpha'\boldsymbol{a}', \boldsymbol{b}\big),$$

Сравнивая коэффициенты при одинаковых векторах в двух последних равенствах, получим:

$$\big(\alpha''\boldsymbol{a} + \alpha''\boldsymbol{a}', \boldsymbol{c}\big) = \big(\alpha\boldsymbol{a} + \alpha'\boldsymbol{a}', \boldsymbol{c}\big),$$
$$\big(\alpha''\boldsymbol{a} + \alpha''\boldsymbol{a}', \boldsymbol{b}\big) = \big(\alpha\boldsymbol{a} + \alpha'\boldsymbol{a}', \boldsymbol{b}\big).$$

Таким образом, находим

$$\alpha''\boldsymbol{a} + \alpha''\boldsymbol{a}' = \alpha\boldsymbol{a} + \alpha'\boldsymbol{a}',$$

откуда $\alpha'' = \alpha' = \alpha$, что доказывает независимость α от первого сомножителя. Аналогично доказывается, что α не зависит также от $\boldsymbol{b}$ и $\boldsymbol{c}$.

Осталось убедиться в том, что $\alpha = 1$. Для нахождения α можно взять любые векторы, для которых произведения легко вычисляются, например, базисные векторы $\boldsymbol{i}$, $\boldsymbol{k}$, $\boldsymbol{i}$:

$$\big[\boldsymbol{i}, [\boldsymbol{k}, \boldsymbol{i}]\big] = \alpha\Big[\boldsymbol{k}\underbrace{(\boldsymbol{i}, \boldsymbol{i})}_{=1} - \boldsymbol{i}\underbrace{(\boldsymbol{i}, \boldsymbol{k})}_{=0}\Big] = \alpha\boldsymbol{k}.$$

С другой стороны, $\big[\boldsymbol{i}, [\boldsymbol{k}, \boldsymbol{i}]\big] = [\boldsymbol{i}, \boldsymbol{j}] = \boldsymbol{k}$, так что $\alpha = 1$, и мы получаем требуемую формулу. □

8.48. Предложение. *Справедливо следующее тождество, называемое тождеством Якоби*[1]:

$$[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]] + [\boldsymbol{b}, [\boldsymbol{c}, \boldsymbol{a}]] + [\boldsymbol{c}, [\boldsymbol{a}, \boldsymbol{b}]] = \boldsymbol{0}.$$

Доказательство. Складывая разложение

$$\big[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big] = \boldsymbol{b}(\boldsymbol{a}, \boldsymbol{c}) - \boldsymbol{c}(\boldsymbol{a}, \boldsymbol{b})$$

с аналогичными разложениями для $\big[\boldsymbol{b}, [\boldsymbol{c}, \boldsymbol{a}]\big]$ и $\big[\boldsymbol{c}, [\boldsymbol{a}, \boldsymbol{b}]\big]$, получаем требуемое тождество. □

Д. Аксиальные и полярные векторы.

8.49. При исследовании многих физических вопросов используется понятие *аксиального вектора*, который может быть нестрого охарактеризован как вектор, изменяющий направление при зеркальном отражении пространства; «обычные» векторы называют в этом случае *полярными*. Можно также сказать, что аксиальные векторы при зеркальном отражении ведут себя не так, как материальные объекты.

Примером аксиального вектора является вектор ω угловой скорости вращающегося тела; его величина («длина») равна углу поворота тела вокруг оси вращения за единицу времени, а направление определяется *правилом буравчика*: вектор угловой скорости ω направлен вдоль оси вращения таким образом, что для наблюдателя, находящегося на конце вектора ω, вращение выглядит

[1] К. Якоби (C. Jacobi) — немецкий математик и механик (1804–1851).

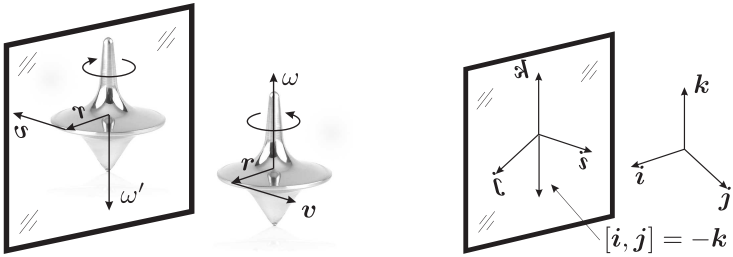


Рис. 8.5.

происходящим против часовой стрелки. При переходе от правого базиса к левому (т.е. при взгляде в зеркало) вектор угловой скорости меняет направление: векторы угловых скоростей вращающегося волчка (ω) и его зеркального изображения (ω') направлены противоположно (см. рис. 8.5), т.е. вектор ω' не является зеркальным отражением вектора ω.

Аналогичным свойством обладает векторное произведение. Так, для базисных векторов $\boldsymbol{i}$, $\boldsymbol{j}$ и $\boldsymbol{k}$ правого базиса имеем $[\boldsymbol{i}, \boldsymbol{j}] = \boldsymbol{k}$, в то время как для левого базиса $[\boldsymbol{i}, \boldsymbol{j}] = -\boldsymbol{k}$. Поскольку левый базис получается зеркальным отражением из правого, можно заключить, что векторное произведение меняет направление при зеркальном отражении пространства.

8.50. Нетрудно понять, что подобное поведение является не свойством вектора как такового, а свойством операции векторного произведения, связанным с изменением ориентации пространства. Поскольку в этом разделе все рассуждения проводились на наглядно-геометрическом уровне, мы не в состоянии сейчас объяснить все детали обнаруженного явления. Отметим лишь, что оно связано с тем фактом, что векторное произведения на самом деле представляет собой не вектор, а так называемый кососимметричный тензор ранга 2 (которому, впрочем, при определённых условиях можно поставить в соответствие вектор—направленный отрезок). К изучению подобных геометрических объектов мы приступим в следующем семестре.

5. Двумерная евклидова геометрия

В этом разделе мы установим некоторые факты геометрии двумерного евклидова точечного пространства (евклидовой плоскости) $\mathcal{E}_2$. Считаем, что в $\mathcal{E}_2$ зафиксирована *ортогональная система координат* $O\boldsymbol{I}$, состоящая из фиксированной точки O (начала отсчёта) и ортонормированного базиса $\boldsymbol{I} = \big(\boldsymbol{i}, \boldsymbol{j}\big)$ ассоциированного векторного пространства. В элементарной геометрии такая система координат называется декартовой прямоугольной системой.

А. Нормальное уравнение прямой на плоскости. Пусть в евклидовой плоскости $\mathcal{E}_2$ заданы ненулевой вектор $\boldsymbol{n}(A, B)$ и точка $M_0(x_0, y_0)$.

8.51. Предложение. *Множество всех точек M евклидовой плоскости, для которых вектор $\boldsymbol{n}$ ортогонален вектору $\overrightarrow{M_0M}$,* *является прямой l, проходящей через точку M_0. Вектор $\boldsymbol{n}$ называется* нормальным вектором *прямой l; говорят, что вектор $\boldsymbol{n}$* ортогонален (перпендикулярен) *этой прямой; обозначение $\boldsymbol{n}\perp l$.*

Доказательство. Условие ортогональности

$$\boldsymbol{n}\perp\overrightarrow{M_0M} \iff \big(\boldsymbol{n},\overrightarrow{M_0M}\big)=0$$

в декартовых прямоугольных координатах записывается в виде $A(x-x_0)+B(y-y_0)=0$, что совпадает с общим уравнением прямой. □

8.52. Таким образом, уравнение прямой на плоскости может быть записано в виде

$$(\boldsymbol{r}-\boldsymbol{r}_0,\boldsymbol{n})=0, \quad (\boldsymbol{r},\boldsymbol{n})=D \tag{8.21}$$

или в прямоугольных координатах

$$A(x-x_0)+B(y-y_0)=0, \quad Ax+By=D, \tag{8.22}$$

где $\boldsymbol{n}(A,B)$ — вектор нормали прямой. Эти уравнения, а также уравнение

$$Ax+By=D,$$

получающееся после раскрытия скобок и приведения подобных слагаемых, называются *нормальными уравнениями прямой.*

8.53. Обратите внимание, что уравнение прямой может быть записано в виде (8.21) или (8.22) в *любой* декартовой системе координат (не обязательно прямоугольной), но *лишь в прямоугольных координатах коэффициенты A и B имеют смысл координат вектора нормали прямой*, ибо только в ортонормированном базисе скалярное произведение векторов имеет вид (8.8).

Б. Некоторые метрические задачи. Геометрические задачи, связанные с необходимостью использования скалярного произведения, т.е. включающие понятия расстояния, угла, перпендикулярности называются *метрическими задачами.* Аффинные же задачи могут быть решены без привлечения указанных понятий.

8.54. Определение.

1. Две прямые на плоскости будем называть *перпендикулярными*, если их направляющие векторы ортогональны.
2. *Ортогональной проекцией* точки M_1 на прямую l называется такая точка $M_2\in l$, что $\overrightarrow{M_1M_2}\perp l$.
3. *Расстоянием* $d(M_1,l)$ *от точки M_1 до прямой l* называется длина вектора $\overrightarrow{M_1M_2}$, где M_2 — ортогональная проекция точки M_1 на прямую l. Вектор $\overrightarrow{M_1M_2}$ называется при этом *перпендикуляром*, опущенным из точки M_1 на прямую l.

4. Точка M_3 называется *симметричной* точке M_1 относительно прямой l, если вектор $\overrightarrow{M_1M_3}$ ортогонален прямой l и расстояния от каждой из этих точек до прямой l равны: $d(M_1, l) = d(M_3, l)$.

8.55. Предложение. *Пусть на евклидовой плоскости $\mathcal{E}_2$ даны точка $M_1(\boldsymbol{r}_1)$ и прямая l, заданная уравнением $(\boldsymbol{r}, \boldsymbol{n}) = D$.*

1. *Радиус-вектор $\boldsymbol{r}_2$ точки M_2, являющейся ортогональной проекцией точки $M_1(\boldsymbol{r}_1)$ на прямую l, выражается формулой*
$$\boldsymbol{r}_2 = \boldsymbol{r}_1 - \frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{(\boldsymbol{n}, \boldsymbol{n})}\boldsymbol{n}. \tag{8.23}$$
2. *Расстояние от точки $M_1(\boldsymbol{r}_1)$ до прямой l выражается формулой*
$$d(M_1, l) = \frac{|(\boldsymbol{r}_1, \boldsymbol{n}) - D|}{|\boldsymbol{n}|}. \tag{8.24}$$
3. *Радиус-вектор $\boldsymbol{r}_3$ точки M_3, симметричной точке $M_1(\boldsymbol{r}_1)$ относительно прямой l, выражается формулой*
$$\boldsymbol{r}_3 = \boldsymbol{r}_1 - 2\frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{(\boldsymbol{n}, \boldsymbol{n})}\boldsymbol{n}. \tag{8.25}$$

Доказательство. 1. Очевидно (см. рис. 8.6),
$$\overrightarrow{M_1M_2} = \overrightarrow{OM_2} - \overrightarrow{OM_1} = \lambda\boldsymbol{n}.$$
Умножим обе части последнего равенства скалярно на вектор $\boldsymbol{n}$:
$$\underbrace{(\overrightarrow{OM_2}, \boldsymbol{n})}_{=D} - \underbrace{(\overrightarrow{OM_1}, \boldsymbol{n})}_{=(\boldsymbol{r}_1, \boldsymbol{n})} = \lambda(\boldsymbol{n}, \boldsymbol{n}),$$
откуда
$$\lambda = -\frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{(\boldsymbol{n}, \boldsymbol{n})},$$
так что для радиус-вектора $\boldsymbol{r}_2$ проекции M_2 точки M_1 на прямую имеем:
$$\boldsymbol{r}_2 = \overrightarrow{OM_2} = \overrightarrow{OM_1} + \overrightarrow{M_1M_2} = \\ = \boldsymbol{r}_1 + \lambda\boldsymbol{n} = \\ = \boldsymbol{r}_1 - \frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{(\boldsymbol{n}, \boldsymbol{n})}\boldsymbol{n}.$$

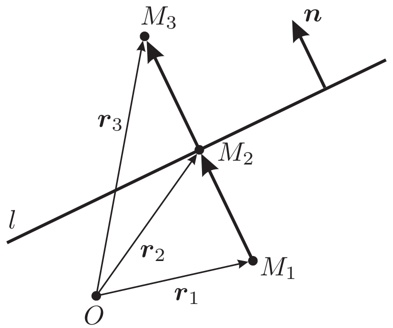


Рис. 8.6. К предложению 8.55

2. Расстояние от точки M_1 до прямой l:
$$d(M_1, l) = \left|\overrightarrow{M_1M_2}\right| = \left|\frac{(\boldsymbol{r}_1, n) - D}{(\boldsymbol{n}, \boldsymbol{n})}\boldsymbol{n}\right| = \frac{|(\boldsymbol{r}_1, n) - D|}{|\boldsymbol{n}|}.$$

3. Для радиус-вектора $\boldsymbol{r}_3$ точки M_3, симметричной точке M_1 относительно прямой, имеем:

$$\boldsymbol{r}_3 = \overrightarrow{OM_3} = \overrightarrow{OM_1} + \overrightarrow{M_1M_3} = \overrightarrow{OM_1} + 2\overrightarrow{M_1M_2} =$$
$$= \boldsymbol{r}_1 + 2\lambda\boldsymbol{n} = \boldsymbol{r}_1 - 2\frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{(\boldsymbol{n}, \boldsymbol{n})}\boldsymbol{n}. \qquad \square$$

8.56. Формула (8.24) в прямоугольных координатах выглядит следующим образом:

$$d(M_1, l) = \frac{\left|Ax_1 + By_1 - D\right|}{\sqrt{A^2 + B^2}}, \tag{8.26}$$

где (x_1, y_1) — координаты точки M_1, (A, B) — координаты нормального вектора $\boldsymbol{n}$ прямой l.

В. Нормированное уравнение прямой. Считаем, что система координат прямоугольная.

8.57. При использовании нормального уравнения прямой часто бывает удобно считать, что нормальный вектор $\boldsymbol{n}$ имеет единичную длину, $|\boldsymbol{n}| = 1$, а правая часть уравнения, традиционно обозначаемая в этом случае буквой p, неотрицательна: $p \geqslant 0$. В этом случае координаты нормального вектора равны косинусам углов, которые этот вектор составляет с осями координат: $\boldsymbol{n}(\cos\alpha, \cos\beta)$ (числа $\cos\alpha$ и $\cos\beta$ называются *направляющими косинусами* вектора $\boldsymbol{n}$; ср. п. 1.29, с. 19). В указанных обозначениях уравнение прямой имеет вид

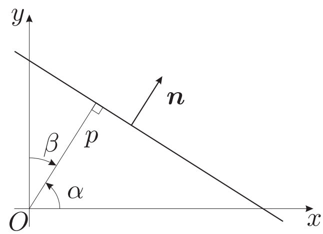


Рис. 8.7.

$$x\cos\alpha + y\cos\beta = p.$$

Так как $\beta = \pi/2 - \alpha$, уравнение можно записать в виде

$$x\cos\alpha + y\sin\alpha = p. \tag{8.27}$$

Согласно (8.26), число p равно расстоянию от начала координат до прямой l. Конечно, в случае, когда прямая проходит через начало координат, $p = 0$.

8.58. Пусть $M_1(x_1, y_1)$ — некоторая точка плоскости. Расстояние от этой точки до прямой l, заданной нормированным уравнением (8.27), согласно предложению 8.55.2, равно $d(M_1, l) = \left|x_1\cos\alpha + y_1\sin\alpha - p\right|$. Величина, которая получится, если в последней формуле убрать знак модуля, т.е.

$$\delta(M_1, l) = x_1\cos\alpha + y_1\sin\alpha - p,$$

называется *отклонением* точки $M_1(x_1, y_1)$ от прямой l; модуль этой величины, очевидно, равен расстоянию $d(M_1, l)$, а знак показывает, в какой из двух полуплоскостей, на которые плоскость разбивается прямою l, лежит точка M_1: если $\delta(M_1, l) < 0$, то точка M_1 и начало координат O лежат в одной полуплоскости, а если $\delta(M_1, l) > 0$ — то в разных. Нетрудно видеть, что отклонение точки $M_1(\boldsymbol{r}_1)$ от прямой, заданной векторным нормальным уравнением $(\boldsymbol{r}, \boldsymbol{n}) = D$, может быть найдено по формуле, аналогичной формуле для расстояния от точки до прямой (здесь $\operatorname{sign} D$ — знак числа D):

$$\delta(M_1, l) = \frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{|\boldsymbol{n}|} \cdot \operatorname{sign} D.$$

6. Трехмерная евклидова геометрия

В трёхмерном евклидовом пространстве $\mathcal{E}_3$, помимо скалярного умножения, имеются операции векторного и смешанного умножения, что значительно обогащает геометрическую структуру пространства и облегчает решение задач.

А. Перпендикулярность прямых и плоскостей.

8.59. Определение.

1. Две прямые в пространстве называются *перпендикулярными*, если их направляющие векторы ортогональны.
2. Прямая l называется *перпендикулярной* плоскости π, если направляющий вектор этой прямой ортогонален направляющему подпространству плоскости.

8.60. Перпендикулярность двух плоскостей определяется сложнее. Пусть π_1 и π_2 — две плоскости, l_1 и l_2 — перпендикулярные им прямые: $l_1 \perp \pi_1$, $l_2 \perp \pi_2$. Будем говорить, что плоскости π_1 и π_2 *перпендикулярны*, если перпендикулярны прямые l_1 и l_2.

Б. Плоскости в трехмерном евклидовом пространстве. Пусть в трёхмерном евклидовом пространстве $\mathcal{E}_3$ заданы ненулевой вектор $\boldsymbol{n}(A, B, C)$ и точка $M_0(x_0, y_0, z_0)$.

8.61. Предложение. *Множество всех точек M пространства, для которых вектор $\boldsymbol{n}$ ортогонален вектору $\overrightarrow{M_0M}$, является плоскостью, проходящей через точку M_0. Вектор $\boldsymbol{n}$ называется* нормальным вектором *этой плоскости; говорят, что вектор $\boldsymbol{n}$* ортогонален (перпендикулярен) *этой плоскости; обозначение $\boldsymbol{n} \perp \pi$.*

Доказательство. Условие ортогональности

$$\boldsymbol{n} \perp \overrightarrow{M_0M} \iff \left(\boldsymbol{n}, \overrightarrow{M_0M}\right) = 0$$

в прямоугольных координатах принимает вид

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0,$$

что совпадает с общим уравнением плоскости. □

8.62. Ясно, что две плоскости перпендикулярны (см. определение 8.60), если их нормальные векторы ортогональны.

8.63. Таким образом, уравнение плоскости в пространстве может быть записано в виде

$$(\boldsymbol{r}-\boldsymbol{r}_0,\boldsymbol{n})=0,\quad (\boldsymbol{r},\boldsymbol{n})=D, \tag{8.28}$$

не отличающемся от уравнения прямой на плоскости (8.21) (см. с. 203). Координатная форма уравнений (8.28) в прямоугольных координатах:

$$\begin{gathered} A(x-x_0)+B(y-y_0)+C(z-z_0)=0,\\ Ax+By+Cz=D. \end{gathered} \tag{8.29}$$

Уравнения в форме (8.28) и (8.29) называют *нормальными.*

8.64. Уравнение плоскости может быть записано в виде (8.28) или (8.29) в *любой* декартовой системе координат (не обязательно прямоугольной), но *лишь в прямоугольных координатах коэффициенты A, B и C имеют смысл координат вектора нормали плоскости*, ибо только в ортонормированном базисе скалярное произведение векторов имеет вид (8.8) (ср. замечание 8.53).

8.65. Переход от векторного параметрического уравнения плоскости $\boldsymbol{r}=\boldsymbol{r}_0+s\boldsymbol{a}+t\boldsymbol{b}$ в нормальному виду $(\boldsymbol{r},\boldsymbol{n})=D$ осуществляется следующим образом. В качестве вектора нормали плоскости можно взять вектор $\boldsymbol{n}=[\boldsymbol{a},\boldsymbol{b}]$, поэтому из (8.28) получаем

$$\big(\boldsymbol{r}-\boldsymbol{r}_0,[\boldsymbol{a},\boldsymbol{b}]\big)=0 \;\Leftrightarrow\; (\boldsymbol{r},[\boldsymbol{a},\boldsymbol{b}])=(\boldsymbol{r}_0,\boldsymbol{a},\boldsymbol{b}) \;\Leftrightarrow\; (\boldsymbol{r},\boldsymbol{n})=D,$$

где $\boldsymbol{n}=[\boldsymbol{a},\boldsymbol{b}]$, $D=(\boldsymbol{r}_0,\boldsymbol{a},\boldsymbol{b})$.

8.66. Определение (ср. определение 8.54, с. 203)**.**

1. *Ортогональной проекцией* точки M_1 на плоскость π называется такая точка $M_2\in\pi$, что $\overrightarrow{M_1M_2}\perp\pi$.
2. *Расстоянием $d(M_1,\pi)$ от точки M_1 до плоскости* π называется длина вектора $\overrightarrow{M_1M_2}$, где M_2 — ортогональная проекция точки M_1 на плоскость π. Вектор $\overrightarrow{M_1M_2}$ называется при этом *перпендикуляром*, опущенным из точки M_1 на плоскость π.
3. Точка M_3 называется *симметричной* точке M_1 относительно плоскости π, если вектор $\overrightarrow{M_1M_3}$ ортогонален плоскости π и расстояния от каждой из этих точек до плоскости π равны: $d(M_1,\pi)=d(M_3,\pi)$.

8.67. Предложение. *Пусть в евклидовом точечном пространстве $\mathcal{E}_3$ даны точка $M_1(\boldsymbol{r}_1)$ и плоскость π, заданная уравнением* $(\boldsymbol{r},\boldsymbol{n})=D$ (см. рис. 8.8).

1. *Радиус-вектор* $\boldsymbol{r}_2$ *точки* M_2*, являющейся ортогональной проекцией точки* $M_1(\boldsymbol{r}_1)$ *на плоскость* π*, выражается формулой*

$$\boldsymbol{r}_2 = \boldsymbol{r}_1 - \frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{(\boldsymbol{n}, \boldsymbol{n})}\boldsymbol{n}. \qquad (8.30)$$

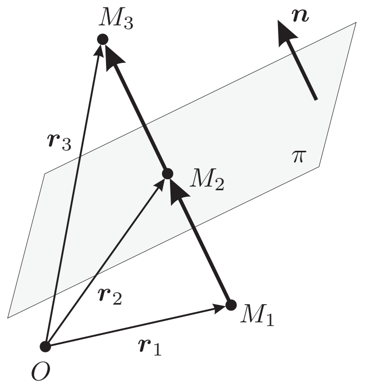


Рис. 8.8. К предложению 8.67

2. *Расстояние от точки* $M_1(\boldsymbol{r}_1)$ *до плоскости* π *выражается формулой*

$$d(M_1, \pi) = \frac{|(\boldsymbol{r}_1, \boldsymbol{n}) - D|}{|\boldsymbol{n}|}. \qquad (8.31)$$

3. *Радиус-вектор* $\boldsymbol{r}_3$ *точки* M_3*, симметричной точке* $M_1(\boldsymbol{r}_1)$ *относительно плоскости* π*, выражается формулой*

$$\boldsymbol{r}_3 = \boldsymbol{r}_1 - 2\frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{(\boldsymbol{n}, \boldsymbol{n})}\boldsymbol{n}. \qquad (8.32)$$

Доказательство предложения 8.67 дословно повторяет доказательство предложения 8.55.

8.68. Формула (8.31) в прямоугольных координатах выглядит следующим образом:

$$d(M_1, \pi) = \frac{\left|Ax_1 + By_1 + Cz_1 - D\right|}{\sqrt{A^2 + B^2 + C^2}}, \qquad (8.33)$$

где (x_1, y_1, z_1) — координаты точки M_1, (A, B, C) — координаты нормального вектора $\boldsymbol{n}$ плоскости π.

В. Нормированное уравнение плоскости. Считаем, что система координат прямоугольная.

8.69. Аналогично тому, как это было сделано в п. 8.57, запишем нормальное уравнение плоскости $(\boldsymbol{r}, \boldsymbol{n}) = D$ с *единичным* направляющим вектором[1] $\boldsymbol{n}(\cos\alpha, \cos\beta, \cos\gamma)$ и *неотрицательной* правой частью, которая традиционно обозначается в этом случае буквой p:

$$x\cos\alpha + y\cos\beta + z\cos\gamma = p, \quad p \geqslant 0. \qquad (8.34)$$

Согласно (8.33), число p равно расстоянию от начала координат до плоскости π. В случае, когда прямая проходит через начало координат, $p = 0$.

[1]Очевидно, координаты единичного вектора равны его направляющим косинусам (см. п. 1.29, с. 19).

8.70. Пусть $M_1(x_1, y_1, z_1)$ — некоторая точка пространства. Расстояние от этой точки до плоскости π, заданной нормированным уравнением (8.34), согласно формуле (8.33), равно

$$d(M_1, \pi) = \left| x_1 \cos\alpha + y_1 \cos\beta + z_1 \cos\gamma - p \right|.$$

Величина, которая получится, если в последней формуле убрать знак модуля, т.е.

$$\delta(M_1, \pi) = x_1 \cos\alpha + y_1 \cos\beta + z_1 \cos\gamma - p,$$

называется *отклонением* точки $M_1(x_1, y_1, z_1)$ от плоскости π; знак этой величины показывает, в каком из двух полупространств, на которые плоскость π разбивает пространство, лежит точка M_1: если $\delta(M_1, \pi) < 0$, то точка M_1 и начало координат O лежат в одном полупространстве, а если $\delta(M_1, \pi) > 0$ — то в разных. Легко доказать, что отклонение точки $M_1(\boldsymbol{r}_1)$ от плоскости, заданной векторным нормальным уравнением $(\boldsymbol{r}, \boldsymbol{n}) = D$, может быть найдено по формуле (здесь $\operatorname{sign} D$ — знак числа D)

$$\delta(M_1, \pi) = \frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{|\boldsymbol{n}|} \cdot \operatorname{sign} D.$$

Г. Уравнение прямой в пространстве.

8.71. Естественно, все способы задания прямой в аффинном пространстве можно использовать и в евклидовом пространстве:

(i) векторное параметрическое уравнение:

$$\boldsymbol{r} = \boldsymbol{r}_0 + t\boldsymbol{a};$$

(ii) система параметрических уравнений в координатах (в том числе в декартовых прямоугольных координатах):

$$\begin{cases} x = x_0 + ta_x, \\ y = y_0 + ta_y, \\ z = z_0 + ta_z; \end{cases}$$

(iii) каноническое уравнение (точнее, система канонических уравнений):

$$\frac{x - x_0}{a_x} = \frac{y - y_0}{a_y} = \frac{z - z_0}{a_z};$$

(iv) задание прямой как пересечение двух плоскостей:

$$\begin{cases} (\boldsymbol{r}, \boldsymbol{n}_1) = D_1, \\ (\boldsymbol{r}, \boldsymbol{n}_2) = D_2 \end{cases} \quad \text{или} \quad \begin{cases} A_1x + B_1y + C_1z = D_1, \\ A_2x + B_2y + C_2z = D_2. \end{cases}$$

8.72. Умножая векторное параметрическое уравнение прямой

$$\boldsymbol{r} = \boldsymbol{r}_0 + t\boldsymbol{a}$$

векторно на вектор $\boldsymbol{a}$, получаем

$$[\boldsymbol{r}, \boldsymbol{a}] = [\boldsymbol{r}_0, \boldsymbol{a}] + t\underbrace{[\boldsymbol{a}, \boldsymbol{a}]}_{=\boldsymbol{0}}.$$

Введём обозначение $\boldsymbol{b} = [\boldsymbol{r}_0, \boldsymbol{a}]$; отметим, что $\boldsymbol{b} \perp \boldsymbol{a}$. Получим уравнение прямой в виде

$$[\boldsymbol{r}, \boldsymbol{a}] = \boldsymbol{b}, \quad \text{где} \quad (\boldsymbol{a}, \boldsymbol{b}) = 0.$$

Это уравнение называется *уравнением прямой в форме Плюккера*[1].

8.73. Теорема. *Пусть в евклидовом точечном пространстве $\mathcal{E}_3$ даны точка $M_1(\boldsymbol{r}_1)$ и прямая l, заданная уравнением $\boldsymbol{r} = \boldsymbol{r}_0 + t\boldsymbol{a}$* (см. рис. 8.9).

1. *Радиус-вектор $\boldsymbol{r}_2$ точки M_2, являющейся ортогональной проекцией точки $M_1(\boldsymbol{r}_1)$ на прямую l, выражается формулой*

$$\boldsymbol{r}_2 = \boldsymbol{r}_0 + \frac{(\boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{a})}{(\boldsymbol{a}, \boldsymbol{a})}\boldsymbol{a}. \tag{8.35}$$

2. *Радиус-вектор $\boldsymbol{r}_3$ точки M_3, симметричной точке $M_1(\boldsymbol{r}_1)$ относительно прямой l, выражается формулой*

$$\boldsymbol{r}_3 = 2\boldsymbol{r}_0 - \boldsymbol{r}_1 + 2\frac{(\boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{a})}{(\boldsymbol{a}, \boldsymbol{a})}\boldsymbol{a}. \tag{8.36}$$

3. *Расстояние от точки $M_1(\boldsymbol{r}_1)$ до прямой l выражается формулой*

$$d(M_1, l) = \frac{\big|[\boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{a}]\big|}{|\boldsymbol{a}|}. \tag{8.37}$$

Доказательство. 1. Умножим скалярно на вектор $\boldsymbol{a}$ обе части равенства $\overrightarrow{M_1M_2} = \overrightarrow{OM_2} - \overrightarrow{OM_1}$ (напомним, что $\overrightarrow{M_1M_2} \perp \boldsymbol{a}$):

$$0 = (\overrightarrow{M_1M_2}, \boldsymbol{a}) = \underbrace{(\overrightarrow{OM_2}, \boldsymbol{a})}_{=(\boldsymbol{r}_0 + t\boldsymbol{a}, \boldsymbol{a})} - \underbrace{(\overrightarrow{OM_1}, \boldsymbol{a})}_{=(\boldsymbol{r}_1, \boldsymbol{a})} = (\boldsymbol{r}_0 - \boldsymbol{r}_1, \boldsymbol{a}) + t(\boldsymbol{a}, \boldsymbol{a}),$$

откуда находим значение t_0 параметра, отвечающее точке $M_2 \in l$:

$$t_0 = \frac{(\boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{a})}{(\boldsymbol{a}, \boldsymbol{a})}.$$

Поэтому для проекции M_2 точки M_1 на прямую l имеем:

$$\boldsymbol{r}_2 = \boldsymbol{r}_0 + t_0\boldsymbol{a} = \boldsymbol{r}_0 + \frac{(\boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{a})}{(\boldsymbol{a}, \boldsymbol{a})}\boldsymbol{a}.$$

[1] Ю. Плюккер (J. Plücker) — немецкий математик и физик (1801–1868).

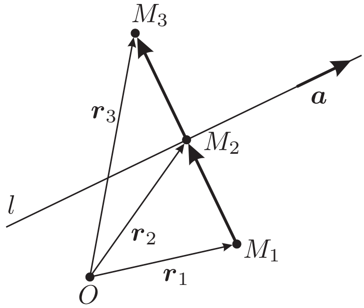


Рис. 8.9. К теореме 8.73

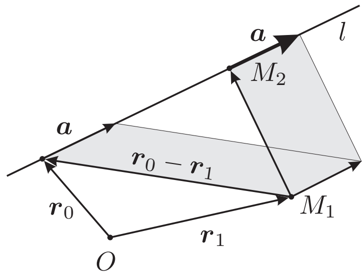


Рис. 8.10. Расстояние от точки до прямой

2. Для точки M_3, симметричной точке M_1 относительно прямой l, имеем

$$\boldsymbol{r}_3 = \overrightarrow{OM_3} = \overrightarrow{OM_1} + 2\overrightarrow{M_1M_2} = \boldsymbol{r}_1 + 2(\boldsymbol{r}_2 - \boldsymbol{r}_1) = \\ = 2\boldsymbol{r}_2 - \boldsymbol{r}_1 = 2\boldsymbol{r}_0 - \boldsymbol{r}_1 + 2\frac{(\boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{a})}{(\boldsymbol{a}, \boldsymbol{a})}\boldsymbol{a}.$$

3. Найдем расстояние от точки M_1 до прямой l:

$$d(M_1, l) = \left|\overrightarrow{M_1M_2}\right| = \left|\boldsymbol{r}_2 - \boldsymbol{r}_1\right| = \left|\boldsymbol{r}_0 - \boldsymbol{r}_1 + \frac{(\boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{a})}{(\boldsymbol{a}, \boldsymbol{a})}\boldsymbol{a}\right| = \\ = \left|\frac{(\boldsymbol{r}_0 - \boldsymbol{r}_1)(\boldsymbol{a}, \boldsymbol{a}) - (\boldsymbol{r}_0 - \boldsymbol{r}_1, \boldsymbol{a})\boldsymbol{a}}{(\boldsymbol{a}, \boldsymbol{a})}\right| = \left|\frac{[\boldsymbol{a}, [\boldsymbol{r}_0 - \boldsymbol{r}_1, \boldsymbol{a}]]}{(\boldsymbol{a}, \boldsymbol{a})}\right| = \\ = \frac{|\boldsymbol{a}| \cdot \left|[\boldsymbol{r}_0 - \boldsymbol{r}_1, \boldsymbol{a}]\right| \cdot \sin\varphi}{(\boldsymbol{a}, \boldsymbol{a})} = \frac{\left|[\boldsymbol{r}_0 - \boldsymbol{r}_1, \boldsymbol{a}]\right|}{|\boldsymbol{a}|},$$

где φ — угол между векторами $\boldsymbol{a}$ и $[\boldsymbol{r}_0 - \boldsymbol{r}_1, \boldsymbol{a}]$; здесь учтено, что векторы $\boldsymbol{a}$ и $[\boldsymbol{r}_0 - \boldsymbol{r}_1, \boldsymbol{a}]$ ортогональны, т.е. $\sin\varphi = 1$.

Эта формула может быть легко получена на основе следующих наглядных соображений (см. рис. 8.10). Найдём площадь параллелограмма, образованного векторами $\boldsymbol{r}_0 - \boldsymbol{r}_1$ и $\boldsymbol{a}$. С одной стороны, она численно равна длине вектора, представляющего собой векторное произведения этих векторов, а с другой стороны — произведению длины основания параллелограмма, равной $\boldsymbol{a}$, на высоту, равную искомому расстоянию:

$$\left|[\boldsymbol{r}_0 - \boldsymbol{r}_1, \boldsymbol{a}]\right| = |\boldsymbol{a}| \cdot d(M_1, l), \qquad \square$$

откуда получаем (8.37).

ГЛАВА 9

Эллипс, гипербола, парабола на евклидовой плоскости

Эллипс, гипербола и парабола были известны ещё античным математикам. Подробное изложение теории этих линий содержится в труде древнегреческого учёного Аполлония Пергского (III в. до н.э.); он же предложил и названия этих линий, которые ранее назывались просто *коническими сечениями*, ибо могут быть получены как сечения кругового конуса некоторой плоскостью. Эллипс, гипербола и парабола, будучи простейшими линиями, отличными от прямой, имеют многочисленные применения в различных вопросах естествознания и служат важными примерами для иллюстрации многих вопросов геометрии и анализа.

1. Канонические уравнения

А. Эллипс.

9.1. ***Фокальное определение эллипса.*** *Эллипсом* называется множество точек евклидовой плоскости, сумма расстояний от каждой из которых до двух фиксированных точек (называемых *фокусами* эллипса) есть постоянная величина; требуется чтобы эта величина была больше расстояния между фокусами.

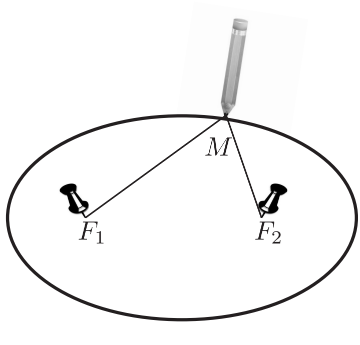


Рис. 9.1. Вычерчивание эллипса

9.2. Отметим, что сумма расстояний, о которой идёт речь в определении эллипса, *не может быть меньше* расстояния между фокусами в силу неравенства треугольника. Если же эта сумма равна расстоянию между фокусами, то получается отрезок, соединяющий фокусы. Оба эти случая исключаются из рассмотрения оговоркой в конце определения 9.1.

9.3. Из определения 9.1 вытекает следующий способ построения эллипса. Возьмём нерастяжимую нить и закрепим её концы в фокусах F_1 и F_2. Если оттянуть нить карандашом и провести линию, держа нить натянутой, то получим дугу эллипса (см. рис. 9.1).

9.4. Выведем уравнение эллипса в прямоугольной декартовой системе координат, ось абсцисс которой проходит через фокусы F_1 и F_2 (она называется *фокальной осью* эллипса), а началом является середина отрезка F_1F_2, соединяющего фокусы; такая система координат называется *канонической* для рассматриваемого эллипса.

Координаты фокусов эллипса относительно канонической системы координат суть $F_1(-c, 0)$, $F_2(c, 0)$, где $c > 0$. Рассмотрим произвольную точку эллипса $M(x, y)$. Направленные отрезки $\overrightarrow{F_1M}$, $\overrightarrow{F_2M}$ называются *фокальными радиусами* точки M; их длины вычисляются по формулам

$$|F_1M| = \sqrt{(x+c)^2 + y^2}, \quad |F_2M| = \sqrt{(x-c)^2 + y^2}.$$

Согласно определению, уравнение эллипса имеет вид

$$\sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2} = 2a,$$

где a — некоторое число, $a > c$.

Преобразуем полученное уравнение. Для этого перенесём второй радикал в правую часть и возведём обе части уравнения в квадрат:

$$(x+c)^2 + y^2 = 4a^2 - 4a\sqrt{(x-c)^2 + y^2} + (x-c)^2 + y^2.$$

После раскрытия скобок и приведения подобных членов получим

$$a\sqrt{(x-c)^2 + y^2} = a^2 - cx.$$

Возводя ещё раз в квадрат обе части этого равенства, получим после упрощений

$$(a^2 - c^2)x^2 + a^2y^2 = a^2(a^2 - c^2).$$

Так как $a > c$, то число $a^2 - c^2$ положительно; обозначим его b^2:

$$b^2 = a^2 - c^2;$$

очевидно, $b^2 < a^2$. Тогда последнее уравнение можно переписать в виде

$$b^2x^2 + a^2y^2 = a^2b^2$$

или

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad a > b > 0. \tag{9.1}$$

Это уравнение называется *каноническим уравнением эллипса.*

9.5. Мы установили, что координаты каждой точки эллипса удовлетворяют уравнению (9.1). Проверим обратное утверждение, а именно, что каждая точка $M(x, y)$, координаты которой удовлетворяют уравнению (9.1), лежит на эллипсе, т.е. что для неё выполнено условие $|F_1M| + |F_2M| = 2a$. Это не очевидно, поскольку при преобразованиях мы дважды возводили обе части уравнения в квадрат, так что могли появиться «посторонние решения» исходного уравнения.

Введём величину

$$\varepsilon = \frac{c}{a};$$

она называется *эксцентриситетом* эллипса и, поскольку $c < a$, удовлетворяет неравенству $\varepsilon < 1$. Пусть $M(x, y)$ — решение уравнения (9.1); ясно, что $|x| \leqslant a$, $|y| \leqslant b$. Тогда

$$\begin{aligned} r_1 = |F_1M| = \sqrt{(x+c)^2 + y^2} = \sqrt{(x+c)^2 + b^2\left(1 - \frac{x^2}{a^2}\right)} = \\ = \sqrt{x^2 + 2xc + c^2 + b^2 - \frac{b^2}{a^2}x^2} = \sqrt{x^2\varepsilon^2 + 2x\varepsilon a + a^2} = \\ = \sqrt{(x\varepsilon + a)^2} = |x\varepsilon + a| = a + x\varepsilon \end{aligned}$$

(при раскрытии модуля мы воспользовались тем, что $|x\varepsilon| \leqslant a$, поскольку $|x| \leqslant a$ и $\varepsilon < 1$.) Аналогично получаем

$$r_2 = |F_2M| = a - x\varepsilon.$$

Поэтому

$$|F_1M| + |F_2M| = 2a,$$

т.е. точка M с координатами (x, y) лежит на эллипсе.

Итак, доказано следующее утверждение.

9.6. Предложение (о каноническом уравнении эллипса)**.** *Линия на евклидовой плоскости является эллипсом тогда и только тогда, когда существует такая прямоугольная декартова система координат (называемая канонической системой координат), в которой эта линия имеет уравнение* (9.1).

9.7. Исследование формы эллипса. Поскольку каноническое уравнение эллипса не изменяется при замене x на $-x$, эллипс симметричен относительно оси Oy. Аналогично, поскольку уравнение не изменяется при замене y на $-y$, эллипс симметричен относительно оси Ox.

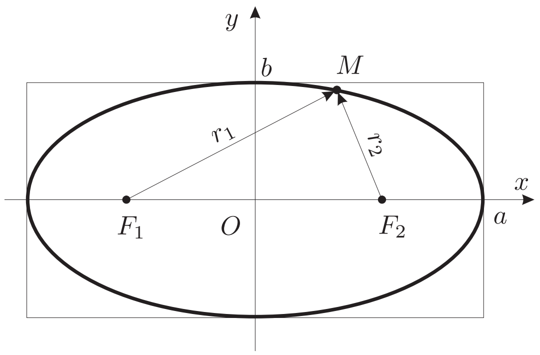


Рис. 9.2. Эллипс в канонической системе координат

Из уравнения (9.1) следует, что

$$\left|\frac{x^2}{a^2}\right| \leqslant 1, \quad \left|\frac{y^2}{b^2}\right| \leqslant 1,$$

т.е. эллипс целиком содержится в прямоугольнике, определяемом неравенствами $|x| \leqslant |a|$ и $|y| \leqslant |b|$. Эти соображения позволяют сделать выводы о внешнем виде эллипса (см. рис. 9.2).

9.8. Из выражения для длины левого фокального радиуса находим

$$r_1 = a + x\varepsilon = \varepsilon\left(x + \frac{a}{\varepsilon}\right) = \varepsilon d_1;$$

выражение в скобках представляет собой расстояние d_1 от точки $M(x, y)$ до прямой $x = -a/\varepsilon$, называемой *левой директрисой* эллипса (см. рис. 9.3).

Аналогичным образом из выражения $r_2 = a - x\varepsilon$ для длины правого фокального радиуса находим

$$r_2 = a - x\varepsilon = \varepsilon\left(\frac{a}{\varepsilon} - x\right) = \varepsilon d_2;$$

выражение в скобках представляет собой расстояние d_2 от точки $M(x, y)$ до *правой директрисы* эллипса, задаваемой уравнением $x = a/\varepsilon$. Таким образом, доказано следующее утверждение.

9.9. Теорема (директориальное свойство эллипса)**.** *Отношение расстояния r_1 (соответственно r_2) от каждой точки эллипса до его левого (соответственно правого) фокуса к расстоянию d_1 (соответственно d_2) от этой точки до левой (соответственно правой) директрисы одинаково для всех точек эллипса и равно эксцентриситету*:

$$\frac{r_1}{d_1} = \frac{r_2}{d_2} = \varepsilon < 1.$$

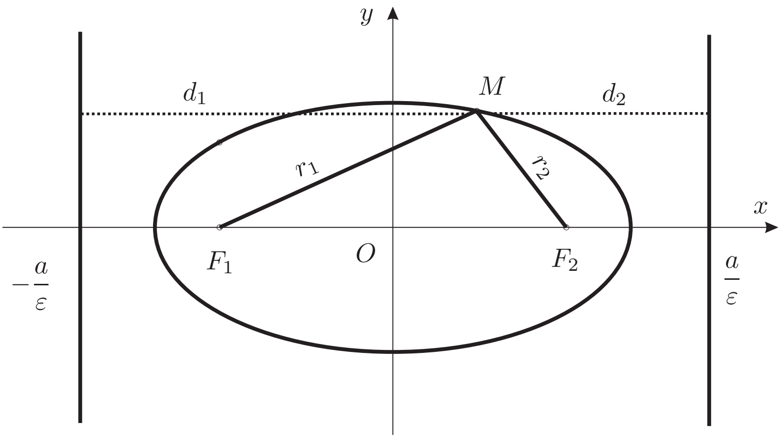


Рис. 9.3. Директрисы эллипса

Обратно, множество всех точек плоскости, обладающих указанным свойством (*т.е. таких, для которых отношение расстояния до фиксированной точки к расстоянию до фиксированной прямой одно и то же и меньше* 1), *представляет собой эллипс.*

9.10. Отметим, что эллипс может быть определён не только фокальным определением 9.1, но также своим каноническим уравнением либо директориальным свойством (тогда фокальное определение превратится в теорему — фокальное свойство эллипса).

9.11. Перечислим основные термины, связанные с эллипсом:

(i) ось Ox — большая (фокальная) ось;
(ii) ось Oy — малая ось;
(iii) точки $(\pm a, 0)$, $(0, \pm b)$ — вершины эллипса;
(iv) точки $F_1(-c, 0)$, $F_2(c, 0)$ — фокусы эллипса;
(v) точка $O(0, 0)$ — центр эллипса;
(vi) число a — большая полуось;
(vii) число b — малая полуось;
(viii) число $c = \sqrt{a^2 - b^2}$ — линейный эксцентриситет;
(ix) число $2c = |F_1F_2|$ — фокусное расстояние;
(x) число $\varepsilon = \dfrac{c}{a} < 1$ — (числовой) эксцентриситет;
(xi) прямые $x = -\dfrac{a}{\varepsilon}$, $x = \dfrac{a}{\varepsilon}$ — директрисы.

Б. Гипербола.

9.12. ***Фокальное определение гиперболы.*** *Гиперболой* называется множество точек евклидовой плоскости, модуль разности расстояний от каждой из которых до двух фиксированных точек (называемых *фокусами* гиперболы) есть постоянная величина; требуется, чтобы эта величина была меньше расстояния между фокусами и отлична от нуля.

9.13. Разность расстояний от произвольной точки M до двух фиксированных точек F_1 и F_2 *не может быть больше* расстояния $|F_1F_2|$ в силу неравенства треугольника. Эта разность *равна* расстоянию $|F_1F_2|$ тогда и только тогда, когда точка M находится на прямой (F_1F_2) вне отрезка $[F_1F_2]$. Если же разность расстояний $|MF_1| - |MF_2|$ равна нулю, т.е. точка M равноудалена от F_1 и F_2, то она лежит на серединном перпендикуляре к отрезку F_1F_2. Указанные случаи исключаем из рассмотрения оговоркой в конце определения 9.12.

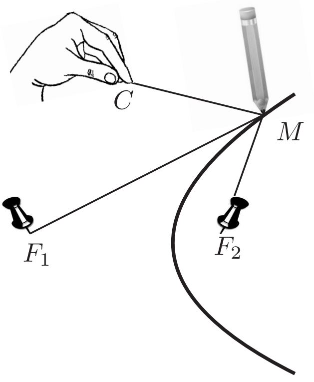


Рис. 9.4. Вычерчивание гиперболы

9.14. На определении 9.12 основано вычерчивание гиперболы при помощи двух нитей, аналогичное построению эллипса, описанному в п. 9.3. Вколем в лист бумаги две булавки (в этих точках будут находиться фокусы гиперболы F_1 и F_2). Возьмём две нити, разность длин которых $2a$ меньше расстояния между фокусами $2c$, конец каждой нити привяжем к булавкам, а оставшиеся концы свяжем узлом (точка C). Держа теперь узел в левой руке, зацепим обе нити остриём карандаша M и будем двигать карандаш на бумаге так, чтобы обе нити F_1MC и F_2MC были все время натянуты. Тогда остриё M карандаша все время находится на одной ветви гиперболы с фокусами F_1 и F_2, так как

$$|F_1M| - |F_2M| = (|F_1M| + |MC|) - (|F_2M| + |MC|) = 2a.$$

Если поменять местами булавки с привязанными к ним нитями, получится вторая ветвь гиперболы; вычерченные дуги гиперболы будут тем длиннее, чем длиннее нити.

9.15. Выведем уравнение гиперболы в прямоугольной декартовой системе координат, ось абсцисс которой проходит через фокусы F_1

и F_2, а началом является середина отрезка F_1F_2, соединяющего фокусы; такая система координат называется *канонической* для рассматриваемой гиперболы.

Пусть фокусы имеют координаты $F_1(-c, 0)$ и $F_2(c, 0)$. Рассмотрим произвольную точку $M(x, y)$ гиперболы. Направленные отрезки $\overrightarrow{F_1M}$, $\overrightarrow{F_2M}$ называются фокальными радиусами точки M; их длины вычисляются по формулам

$$r_1 = |F_1M| = \sqrt{(x+c)^2 + y^2}, \quad r_2 = |F_2M| = \sqrt{(x-c)^2 + y^2}.$$

Согласно определению, уравнение гиперболы имеет вид

$$\left|\sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2}\right| = 2a,$$

где a — некоторое число. Уничтожив радикалы, придем к уравнению

$$\frac{x^2}{a^2} - \frac{y^2}{c^2 - a^2} = 1.$$

Введя обозначение $b^2 = c^2 - a^2$, перепишем уравнение в виде

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

Это уравнение называется *каноническим уравнением гиперболы.*

9.16. Обратно, покажем, что любая точка $M(x, y)$, координаты которой являются решением этого уравнения, лежит на гиперболе.

Введём обозначение $\varepsilon = c/a$; очевидно, $\varepsilon > 1$. Число ε называется *эксцентриситетом гиперболы.* Квадрат левого фокального радиуса точки $M(x, y)$ равен

$$\begin{aligned} r_1^2 = |F_1M|^2 = (x+c)^2 + y^2 = (x+c)^2 + b^2\left(\frac{x^2}{a^2} - 1\right) = \\ = \left(\frac{b^2}{a^2} + 1\right)x^2 + 2xc + c^2 - b^2 = \frac{c^2}{a^2}x^2 + 2cx + a^2 = \\ = \varepsilon^2x^2 + 2\varepsilon^2 ax + a^2 = (\varepsilon x + a)^2. \end{aligned}$$

Поскольку $|\varepsilon x| > |x| \geqslant a$, имеем

$$r_1 = \begin{cases} x\varepsilon + a, & x > 0, \\ -x\varepsilon - a, & x < 0. \end{cases}$$

Аналогично для правого фокального радиуса получаем

$$r_2 = \begin{cases} x\varepsilon - a, & x > 0, \\ -x\varepsilon + a, & x < 0. \end{cases}$$

Следовательно,

$$r_1 - r_2 = \begin{cases} 2a, & x > 0, \\ -2a, & x < 0. \end{cases}$$

Таким образом, точка $M(x, y)$ лежит на гиперболе.

Мы доказали следующую теорему.

9.17. Предложение (о каноническом уравнении гиперболы)**.** *Линия на евклидовой плоскости является гиперболой тогда и только тогда, когда существует такая прямоугольная декартова система координат (называемая канонической системой координат), в которой эта линия имеет уравнение*

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{9.2}$$

9.18. *Исследование формы гиперболы.* Поскольку каноническое уравнение гиперболы не изменяется при замене x на $-x$, и при замене y на $-y$, гипербола симметрична относительно обеих координатных осей.

Из полученного уравнения (9.2) следует, что

$$\frac{x^2}{a^2} = 1 + \frac{y^2}{b^2} \geqslant 1 \quad \Rightarrow \quad |x| \geqslant |a|,$$

т.е. внутри полосы $|x| < |a|$ точек гиперболы нет. Поскольку при $x \geqslant 0$, $y \geqslant 0$ имеем

$$y = b\sqrt{\frac{x^2}{a^2} - 1} = b\frac{x}{a}\underbrace{\sqrt{1 - \frac{a^2}{x^2}}}_{<1} < \frac{b}{a}x,$$

часть гиперболы, расположенная в первом квадранте, лежит строго ниже прямой $y = \dfrac{b}{a}x$. Более того, применив асимптотическое разложение[1]

$$\sqrt{1 - \frac{a^2}{x^2}} = 1 - \frac{1}{2}\frac{a^2}{x^2} + o\left(\frac{1}{x^2}\right), \quad x \to \infty,$$

получаем:

$$\begin{aligned} y &= b\frac{x}{a}\left(1 - \frac{a^2}{2x^2} + o\left(\frac{1}{x^2}\right)\right) = \\ &= \frac{b}{a}x - \frac{ab}{2x} + o\left(\frac{1}{x}\right) = \frac{b}{a}x + o(1). \end{aligned}$$

[1] См. курс математического анализа.

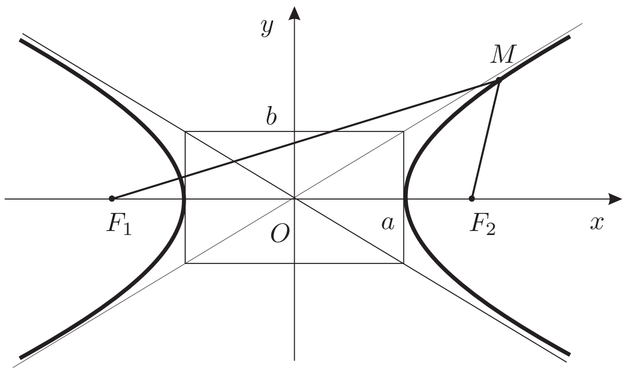


Рис. 9.5. Гипербола в канонической системе координат

Таким образом, принимая во внимание симметричность гиперболы относительно осей координат, можем сделать вывод, что прямые

$$y = \pm\frac{b}{a}x \;\Leftrightarrow\; \frac{x}{a} \pm \frac{y}{b} = 0$$

являются *наклонными асимптотами* гиперболы.

Всё вышесказанное позволяет сделать выводы о внешнем виде гиперболы (см. рис. 9.5).

9.19. Аналогично тому, как это было сделано в случае эллипса, назовём *директрисами* гиперболы прямые

$$x = -\frac{a}{\varepsilon}, \quad x = \frac{a}{\varepsilon}.$$

Справедливо следующее утверждение (читателю предлагается доказать его самостоятельно).

9.20. Теорема (директориальное свойство гиперболы, см. рис. 9.6). *Отношение расстояния r_1 (соответственно r_2) от каждой точки гиперболы до её левого (соответственно правого) фокуса к расстоянию d_1 (соответственно d_2) от этой точки до левой (соответственно правой) директрисы одинаково для всех точек гиперболы и равно эксцентриситету:*

$$\frac{r_1}{d_1} = \frac{r_2}{d_2} = \varepsilon > 1$$

Обратно, множество всех точек плоскости, обладающих указанным свойством (т.е. таких, для которых отношение расстояния до фиксированной точки к расстоянию до фиксированной прямой одно и то же и больше 1), представляет собой гиперболу.

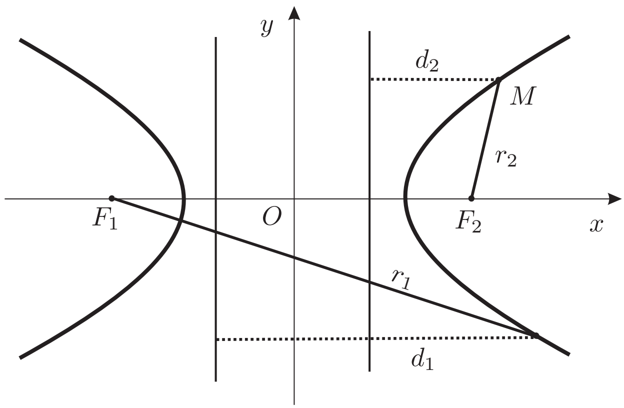


Рис. 9.6. Директрисы гиперболы

9.21. Как и в случае эллипса, гипербола может быть определена не только с помощью фокального свойства, но и с помощью канонического уравнения или директориального свойства (в этом случае фокальное определение превращается в теорему — фокальное свойство гиперболы).

9.22. Основные термины, связанные с гиперболой:

(i) ось OX — вещественная (фокальная) ось;
(ii) ось OY — мнимая (сопряжённая) ось;
(iii) точки $(\pm a, 0)$ — вершины гиперболы;
(iv) точки $F_1(-c, 0)$, $F_2(c, 0)$ — фокусы гиперболы;
(v) точка $O(0, 0)$ — центр гиперболы;
(vi) прямые $\dfrac{x}{a} + \dfrac{y}{b} = 0$, $\dfrac{x}{a} - \dfrac{y}{b} = 0$ — асимптоты гиперболы;
(vii) число a — вещественная полуось;
(viii) число b — мнимая полуось;
(ix) число $c = \sqrt{a^2 + b^2}$ — линейный эксцентриситет;
(x) число $\varepsilon = \dfrac{c}{a} > 1$ — (числовой) эксцентриситет;
(xi) число $2c$ — фокусное расстояние;
(xii) прямые $x = -\dfrac{a}{\varepsilon}$, $x = \dfrac{a}{\varepsilon}$ — директрисы;
(xiii) прямоугольник со сторонами, параллельными осям гиперболы, диагонали которого лежат на асимптотах, — основной прямоугольник гиперболы (см. рис. 9.5).

9.23. Рассмотрим уравнение

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = -1.$$

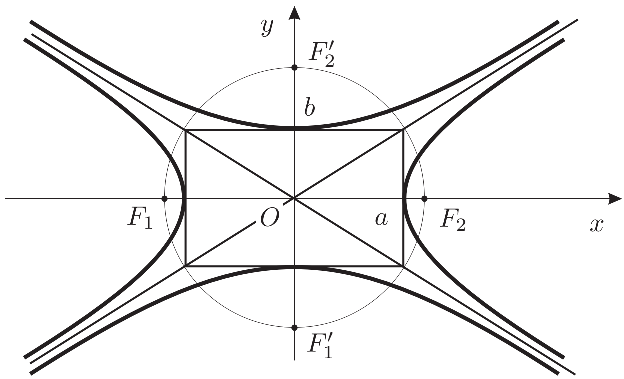


Рис. 9.7. Сопряжённые гиперболы

Перестановкой букв $x \leftrightarrow y$, $a \leftrightarrow b$ оно сводится к уравнению, рассмотренному выше, так что ясно, что оно также определяет гиперболу, вершины и фокусы которой лежат на оси Oy. Две гиперболы, заданные уравнениями

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} = -1$$

в одной и той же системе координат при одних и тех же значениях a и b, называются *взаимно сопряжёнными* (см. рис. 9.7).

Очевидно, взаимно сопряжённые гиперболы имеют общие асимптоты.

9.24. Взаимно сопряжённые гиперболы имеют одинаковый линейный эксцентриситет $c = \sqrt{a^2 + b^2}$, который к тому же равен половине диагонали их общего основного прямоугольника. Поэтому фокусы обеих взаимно сопряжённых гипербол и вершины их общего основного прямоугольника лежат на одной окружности, центр которой совпадает с общим центром обеих гипербол.

B. Парабола.

9.25. ***Фокально-директориальное определение параболы.*** *Параболой* называется множество точек евклидовой плоскости, расстояние от каждой из которых до фиксированной точки (называемой *фокусом*) равно расстоянию до фиксированной прямой (называемой *директрисой*).

9.26. Для того чтобы нарисовать параболу, потребуется линейка, угольник, нить, длина которой равна большему катету угольника, и булавки. Прикрепим один конец нити к фокусу, а другой — к вершине меньшего угла угольника. Приложим линейку к директрисе и поставим на неё угольник меньшим катетом. Карандашом натянем нить так, чтобы его остриё прижималось к большему катету. Если перемещать угольник, следя за тем, чтобы нить оставалась натянутой, то карандаш будет вычерчивать на бумаге параболу.

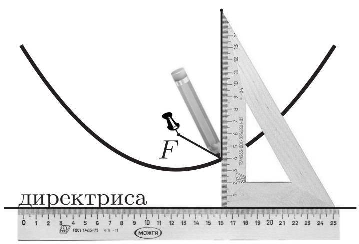


Рис. 9.8. Вычерчивание параболы

9.27. Выведем уравнение параболы в прямоугольной декартовой системе координат, ось абсцисс которой проходит через фокус F параболы перпендикулярно директрисе, а началом является середина перпендикуляра, опущенного из фокуса на директрису; такая система координат называется *канонической* для рассматриваемой параболы.

Пусть в канонической системе координат фокус параболы имеет координаты $F(p/2; 0)$; тогда уравнение директрисы имеет вид $x = -p/2$. Если $M(x, y)$ — произвольная точка параболы, то расстояние от M до директрисы равно $x + p/2$, а до фокуса $F(p/2; 0)$

$$|FM| = \sqrt{\left(x - \frac{p}{2}\right) + y^2}.$$

По определению параболы

$$x + \frac{p}{2} = \sqrt{\left(x - \frac{p}{2}\right) + y^2}.$$

Возводя в квадрат, получим *каноническое уравнение параболы*

$$y^2 = 2px.$$

9.28. Проверим, что любое решение полученного уравнения представляет точку параболы. Пусть координаты точки $M(x, y)$ являются решением уравнения; тогда

$$y^2 = 2px = \left(x + \frac{p}{2}\right)^2 - \left(x - \frac{p}{2}\right)^2,$$

$$\left|x + \frac{p}{2}\right| = \sqrt{\left(x - \frac{p}{2}\right)^2 + y^2},$$

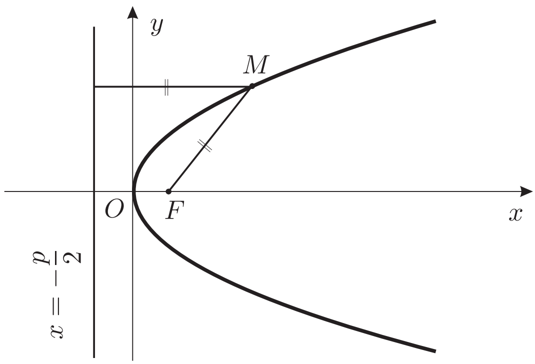


Рис. 9.9. Парабола в канонической системе координат

т.е. точка $M(x, y)$ равноудалена от прямой $x = -p/2$ и точки $(p/2, 0)$, а потому, согласно определению, лежит на параболе.

Итак, доказана следующая теорема.

9.29. Предложение (о каноническом уравнении параболы)**.** *Линия на евклидовой плоскости является параболой тогда и только тогда, когда существует такая прямоугольная декартова система координат (называемая канонической системой координат), в которой эта линия имеет уравнение*

$$y^2 = 2px.$$

9.30. Поскольку каноническое уравнение параболы эллипса не изменяется y на $-y$, парабола симметрична относительно оси Ox.

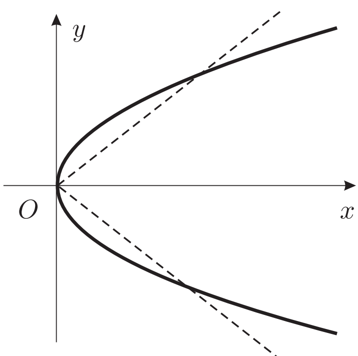


Рис. 9.10.

9.31. Рассмотрим отношение

$$\frac{|y|}{x} = \sqrt{\frac{2p}{x}};$$

оно стремится к нулю при $x \to +\infty$. Это означает, что, начиная с некоторого (достаточно большого) значения x, парабола содержится в любом симметричном угле, охватывающем положительную полуось оси абсцисс (см. рис. 9.10). Наглядно это означает, что если смотреть вдоль этой полуоси, то ветви параболы будут казаться сходящимися, хотя на самом деле они сколь угодно далеко отходят от оси абсцисс ($|y| \to +\infty$ при $x \to +\infty$).

9.32. Эксцентриситет параболы считается равным единице по определению.

9.33. Основные термины, связанные с параболой:
- (i) ось Ox — (фокальная) ось параболы;
- (ii) точка $F(p/2, 0)$ — фокус;
- (iii) прямая $x = -p/2$ — директриса;
- (iv) p — (фокальный) параметр;
- (v) $p/2$ — фокусное расстояние.

2. Касательные к эллипсу, гиперболе и параболе. Оптические свойства

Касательные к эллипсу, гиперболе и параболе можно определить так же, как это делается в анализе: как предельное положение секущей (см. курс математического анализа). Однако в курсе аналитической геометрии касательные к этим трём линиям определяются без обращения к дифференциальному исчислению.

A. Вывод уравнений касательных.

9.34. *Касательной к эллипсу* называется прямая, имеющая с эллипсом единственную общую точку.

Выведем уравнение касательной к эллипсу. Пусть (x_0, y_0) — точка касания эллипса

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

и прямой

$$\frac{x - x_0}{l} = \frac{y - y_0}{m} \qquad \Leftrightarrow \qquad x = x_0 + lt, \quad y = y_0 + mt,$$

т.е. единственная их общая точка. Найдём координаты l, m направляющего вектора касательной. Подставляя параметрические уравнения касательной в каноническое уравнение эллипса, получаем:

$$\underbrace{\frac{x_0^2}{a^2} + \frac{y_0^2}{b^2}}_{=1} + 2t\left(\frac{x_0 l}{a^2} + \frac{y_0 m}{b^2}\right) + t^2\left(\frac{l^2}{a^2} + \frac{m^2}{b^2}\right) = 1$$

или после преобразований

$$t^2\left(\frac{l^2}{a^2} + \frac{m^2}{b^2}\right) + 2t\left(\frac{x_0 l}{a^2} + \frac{y_0 m}{b^2}\right) = 0.$$

Это квадратное уравнение должно иметь один (двойной) корень, что возможно при условии

$$\frac{x_0 l}{a^2} + \frac{y_0 m}{b^2} = 0,$$

так что можно положить
$$l = \frac{y_0}{b^2}, \quad m = -\frac{x_0}{a^2}.$$
Каноническое уравнение касательной к эллипсу имеет вид
$$\frac{x - x_0}{y_0/b^2} = \frac{y - y_0}{-x_0/a^2} \;\Leftrightarrow\; \frac{x_0}{a^2}(x - x_0) + \frac{y_0}{b^2}(y - y_0) = 0,$$
откуда, учитывая соотношение $x_0^2/a^2 + y_0^2/b^2 = 1$, получаем
$$\frac{xx_0}{a^2} + \frac{yy_0}{b^2} = 1. \tag{9.3}$$

9.35. *Касательной к гиперболе* называется прямая, имеющая с гиперболой единственную общую точку и не параллельная асимптотам гиперболы. Точно так же, как в случае эллипса, получаем уравнение прямой, касающейся гиперболы $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$ в точке (x_0, y_0):
$$\frac{xx_0}{a^2} - \frac{yy_0}{b^2} = 1 \tag{9.4}$$
(проведите соответствующие выкладки самостоятельно).

9.36. *Касательная к параболе* — это прямая, имеющая с параболой единственную общую точку и не параллельная оси параболы. Уравнение касательной к параболе $y^2 = 2px$ имеет вид
$$yy_0 = p(x + x_0), \tag{9.5}$$
где (x_0, y_0) — точка касания (докажите самостоятельно).

Б. Оптические свойства эллипса, гиперболы и параболы. Оптические свойства эллипса, гиперболы и параболы лежат в основе многих технических приложений этих линий.

9.37. Теорема (оптическое свойство эллипса). *Фокальные радиусы произвольной точки M_0 эллипса составляют равные углы с касательной к эллипсу, проведённой через точку M_0.*

9.38. Этот результат имеет простую физическую интерпретацию, объясняющую название теоремы: если эллипс считать зеркалом и поместить точечный источник света в одном из фокусов эллипса, то отражённый луч проходит через другой фокус.

Доказательство. Найдем синусы углов α_1 и α_2, которые фокальные радиусы произвольной точки $M_0(x_0, y_0)$ составляют с касательной к эллипсу в точке M_0.

Расстояние $|F_1D_1|$ от фокуса $F_1(-c, 0)$ до касательной, имеющей уравнение
$$\frac{xx_0}{a^2} + \frac{yy_0}{b^2} = 1,$$

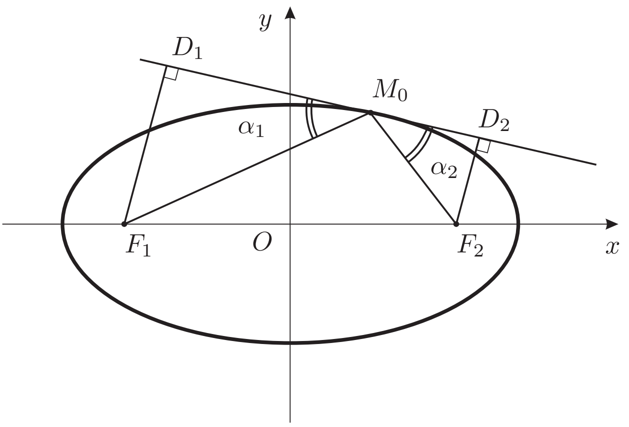


Рис. 9.11. Оптическое свойство эллипса

равно (см. формулу (8.26), с. 205)

$$|F_1D_1| = \frac{\left|\dfrac{(-c)\cdot x_0}{a^2} + \dfrac{0\cdot y_0}{b^2} - 1\right|}{\sqrt{\dfrac{x_0^2}{a^4} + \dfrac{y_0^2}{b^4}}} = \frac{\varepsilon x_0 + a}{a\sqrt{\dfrac{x_0^2}{a^4} + \dfrac{y_0^2}{b^4}}} = \frac{r_1}{a\sqrt{\dfrac{x_0^2}{a^4} + \dfrac{y_0^2}{b^4}}},$$

так что

$$\sin\alpha_1 = \frac{|F_1D_1|}{|F_1M_0|} = \frac{1}{a\sqrt{\dfrac{x_0^2}{a^4} + \dfrac{y_0^2}{b^4}}}.$$

Аналогично получаем

$$\sin\alpha_2 = \frac{|F_2D_2|}{|F_2M_0|} = \frac{1}{a\sqrt{\dfrac{x_0^2}{a^4} + \dfrac{y_0^2}{b^4}}}.$$

Из равенства $\sin\alpha_1 = \sin\alpha_2$ вытекает $\alpha_1 = \alpha_2$. □

9.39. Теорема (оптическое свойство гиперболы)**.** *Фокальные радиусы произвольной точки M_0 гиперболы составляют равные углы с касательной к гиперболе, проведённой через точку M_0.*

9.40. Теорема (оптическое свойство параболы)**.** *Фокальный радиус произвольной точки M_0 параболы и ось параболы составляют равные углы с касательной к параболе, проведённой через точку M_0.*

9.41. Свойство параболического зеркала фокусировать пучок лучей, параллельных оси параболы, используется в конструкциях

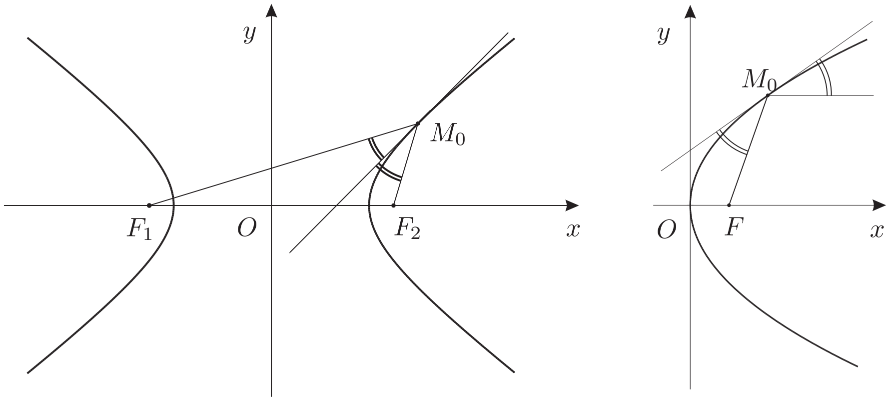


Рис. 9.12. Оптические свойства гиперболы и параболы

телескопов-рефлекторов (как в оптическом, так и инфракрасном и радиодиапазоне), узконаправленных антенн, солнечных электростанций, а обратное свойство — превращать центральный пучок лучей в параллельный — в конструкциях прожекторов, фонарей, фар и т. п.

3. Родственность эллипса, гиперболы и параболы

А. Полярные уравнения эллипса, гиперболы и параболы. Получим уравнения эллипса, гиперболы и параболы в полярной системе координат, ось которой совпадает с фокальной осью линии, а полюс находится в одном из фокусов.

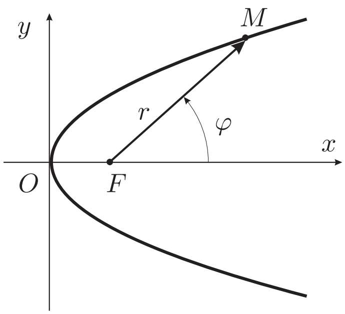


Рис. 9.13. Парабола в полярной системе координат

9.42. Наиболее просто выводится полярное уравнение параболы. Поместим полюс в фокус параболы (см. рис. 9.13). Связь декартовых и полярных координат имеет вид

$$x - \frac{p}{2} = r\cos\varphi$$

(левая часть равенства — длина ортогональной проекции вектора $\overrightarrow{FM}$ на ось Ox), а из определения параболы вытекает равенство

$$r = x + \frac{p}{2} \iff x = r - \frac{p}{2}.$$

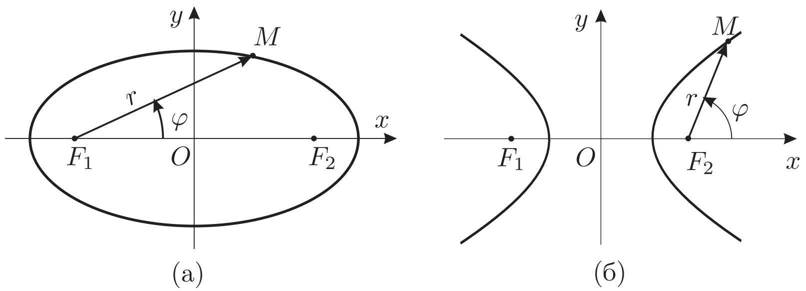


Рис. 9.14. Эллипс (а) и гипербола (б) в полярной системе координат

Таким образом,

$$r\cos\varphi = x - \frac{p}{2} = r - p \;\Leftrightarrow\; r = \frac{p}{1-\cos\varphi}.$$

Отметим, что при $\cos\varphi = 0$ (в частности, при $\varphi = \pi/2$) получаем $r = p$, т.е. число p равно половине длины хорды параболы, проходящей через фокус перпендикулярно фокальной оси, что и объясняет её название — *фокальный параметр*.

9.43. Выведем полярное уравнение эллипса. Поместим полюс в *левый* фокус эллипса (см. рис. 9.14(а)). Запишем связь декартовых и полярных координат

$$x + c = r\cos\varphi$$

и выражение для левого фокального радиуса

$$r = \varepsilon x + a.$$

Отсюда получаем

$$r = \varepsilon(r\cos\varphi - c) + a \;\Leftrightarrow\; r(1 - \varepsilon\cos\varphi) = a - \varepsilon c.$$

Как и в случае параболы, при $\cos\varphi = 0$ (т.е. $\varphi = \pi/2$) имеем $r = a - \varepsilon c$; это половина длины фокальной хорды эллипса, т.е. хорды, проходящей через фокус перпендикулярно фокальной оси. Обозначим эту величину p:

$$p = a - \varepsilon c = a - \frac{c}{a}c = \frac{a^2 - c^2}{a} = \frac{b^2}{a}.$$

Итак, уравнение эллипса в полярной системе координат, ось которой совпадает с его фокальной осью, а полюс находится в левом фокусе, имеет вид

$$r = \frac{p}{1-\varepsilon\cos\varphi}.$$

9.44. В случае гиперболы поместим полюс в *правый* фокус и ограничимся рассмотрением правой ветви гиперболы (см. рис. 9.14(б)). Как и выше, запишем связь декартовых и полярных координат и выражение для правого фокального радиуса точки, лежащей на правой ветви гиперболы:

$$x - c = r\cos\varphi, \quad r = \varepsilon x - a,$$

откуда получаем $r(1 - \varepsilon\cos\varphi) = \varepsilon c - a$. Вводя обозначение

$$p = \varepsilon c - a = \frac{c}{a}c - a = \frac{c^2 - a^2}{a} = \frac{b^2}{a}$$

и называя полученную величину *фокальным параметром гиперболы*, окончательно получаем

$$r = \frac{p}{1 - \varepsilon\cos\varphi}.$$

Геометрический смысл фокального радиуса тот же, что и ранее: это половина длины фокальной хорды гиперболы, т.е. хорды, проходящей через фокус перпендикулярно фокальной оси.

9.45. Таким образом, парабола, эллипс и гипербола (вернее, одна её ветвь) задаются в полярных координатах одним и тем же уравнением

$$r = \frac{p}{1 - \varepsilon\cos\varphi};$$

при $\varepsilon = 1$ получается парабола, при $0 < \varepsilon < 1$ — эллипс, при $\varepsilon = 0$ — окружность, а при $\varepsilon > 1$ — гипербола.

Это уравнение возникает, например, при решении задачи Кеплера об обращении тел Солнечной системы вокруг Солнца.

Б. Уравнения эллипса, гиперболы и параболы, отнесённые к вершине. Каноническая система координат для параболы устроена так, что начало координат находится в вершине параболы. Получим уравнения эллипса и гиперболы в системах координат, начало которых совмещено с одной из вершин этих линий.

9.46. Начнём с эллипса. Обозначим каноническую систему координат для эллипса $O'x'y'$, так что каноническое уравнение эллипса имеет вид

$$\frac{x'^2}{a^2} + \frac{y'^2}{b^2} = 1.$$

Рассмотрим другую систему координат Oxy, начало O которой находится в *левой* вершине эллипса, ось Ox совпадает с осью $O'x'$, а ось Oy параллельна оси $O'y'$; таким образом, система Oxy получена из канонической системы $O'x'y'$ сдвигом на a единиц *влево*, где a — большая полуось рассматриваемого эллипса. Очевидно, формулы связи координат имеют вид

$$x' = x - a, \quad y' = y.$$

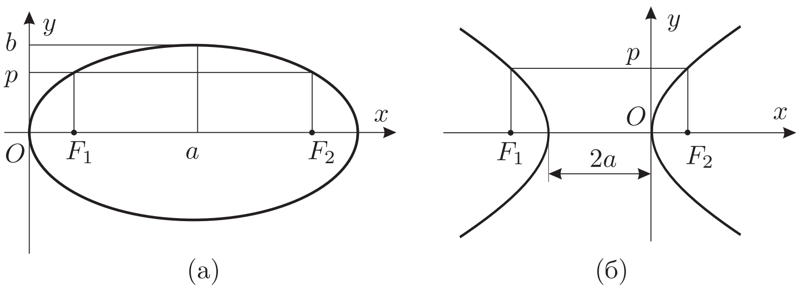


Рис. 9.15. Эллипс и гипербола в системе координат, отнесённой к вершине

Подставляя эти формулы в каноническое уравнение эллипса, получим

$$\frac{(x-a)^2}{a^2} + \frac{y^2}{b^2} = 1 \;\Leftrightarrow\; \frac{x^2}{a^2} - 2\frac{x}{a} + \frac{y^2}{b^2} = 0,$$

откуда

$$y^2 = 2\frac{b^2}{a}x - \frac{b^2}{a^2}x^2 \;\Leftrightarrow\; y^2 = 2px - \frac{p}{a}x^2,$$

поскольку $p = b^2/a$. Кроме того,

$$\frac{b^2}{a^2} = \frac{a^2 - c^2}{a^2} = 1 - \varepsilon^2,$$

так что последнее уравнение можно записать в виде

$$y^2 = 2px + (\varepsilon^2 - 1)x^2.$$

9.47. Выполним аналогичные преобразования канонического уравнения гиперболы. Обозначим каноническую систему координат гиперболы $O'x'y'$, так что каноническое уравнение имеет вид

$$\frac{x'^2}{a^2} - \frac{y'^2}{b^2} = 1.$$

Рассмотрим другую систему координат Oxy, начало O которой находится в *правой* вершине гиперболы, ось Ox совпадает с осью $O'x'$, а ось Oy параллельна оси $O'y'$; таким образом, система Oxy получена из канонической системы $O'x'y'$ сдвигом на a единиц *вправо*, где a — вещественная полуось рассматриваемой гиперболы. Подставляя формулы связи координат

$$x' = x + a, \quad y' = y$$

в каноническое уравнение гиперболы, получим

$$\frac{(x+a)^2}{a^2} - \frac{y^2}{b^2} = 1 \;\Leftrightarrow\; \frac{x^2}{a^2} + 2\frac{x}{a} - \frac{y^2}{b^2} = 0,$$

откуда

$$y^2 = 2\underbrace{\frac{b^2}{a}}_{=p}x + \frac{b^2}{a^2}x^2 \;\Leftrightarrow\; y^2 = 2px + \frac{p}{a}x^2.$$

Так как

$$\frac{b^2}{a^2} = \frac{c^2 - a^2}{a^2} = \varepsilon^2 - 1,$$

последнее уравнение можно переписать в виде

$$y^2 = 2px + (\varepsilon^2 - 1)x^2.$$

9.48. Итак, мы вновь обнаруживаем, что эллипс, гипербола и парабола задаются одним и тем же уравнением

$$y^2 = 2px + (\varepsilon^2 - 1)x^2$$

с разными значениями ε: парабола получается при $\varepsilon = 1$, эллипс — при $0 < \varepsilon < 1$, окружность — при $\varepsilon = 0$ и гипербола — при $\varepsilon > 1$.

9.49. Древнегреческим математикам были хорошо известны свойства эллипса, гиперболы и параболы, несмотря на то, что они не владели методом координат; в частности, они знали свойства этих линий, выражаемые полученным нами уравнением. «Приложением» ($\pi\alpha\rho\alpha\beta o\lambda\acute{\eta}$) данного квадрата к данному отрезку они называли задачу о построении прямоугольника с заданным основанием, равновеликого[1] данному квадрату, т.е. фактически построение высоты этого прямоугольника. Парабола, заданная уравнением $y^2 = 2px$, позволяет решить задачу приложения квадрата y^2 к основанию $2p$. Эллипс и гипербола уже не позволяют решить задачу о приложении квадрата к отрезку, поскольку в их уравнениях имеется добавочный член $(\varepsilon^2 - 1)x^2$, положительный (т.е. являющийся избытком, $\upsilon\pi\varepsilon\rho\beta o\lambda\acute{\eta}$) в случае гиперболы и отрицательный (т.е. являющийся недостатком, $\varepsilon\lambda\lambda\varepsilon\iota\psi\iota\varsigma$) в случае эллипса, что и объясняет названия соответствующих линий.

B. Эллипс, гипербола и парабола как конические сечения.

9.50. Эллипс, гипербола и парабола впервые появились в работах древнегреческого математика Менехма (IV в. до н.э.), который определял их как сечения кругового конуса плоскостью, перпендикулярной одной из образующих; в зависимости от угла при вершине в осевом сечении конуса получается либо эллипс (если этот угол острый), либо парабола (если угол прямой), либо гипербола (если угол тупой). Труды Менехма до нас не дошли, в отличие от посвящённых коническим сечениям сочинений Аполлония Пергского (III в. до н.э.), который считается одним из трёх (наряду с Евклидом и Архимедом) великих геометров античности. Аполлоний изучал сечения конуса плоскостью, которая не обязательно перпендикулярна одной из образующий, впервые рассмотрел двуполостный конус, что позволило обнаружить вторую ветвь гиперболы, и предложил названия «эллипс», «гипербола» и «парабола» для конических сечений, или коник.

9.51. Проведём через центр окружности перпендикуляр m к плоскости этой окружности и на нём возьмём точку S. Прямые, каждая из которых проходит через S и какую-либо точку окружности, образуют поверхность, которую мы будем называть *конусом* с осью m и вершиной S; сами эти прямые называются образующими конуса. Ясно, что конус имеет вид двух расширяющихся трубок, называемых полами конуса.

9.52. Плоскость, проходящая через вершину конуса, может занимать относительно этого конуса следующие три положения (см. рис. 9.16(a)):

(i) иметь с конусом лишь одну общую точку — вершину конуса,
(ii) касаться конуса вдоль его образующей,
(iii) пересекать конус по двум его различным образующим.

[1]Т.е. имеющего такую же площадь.

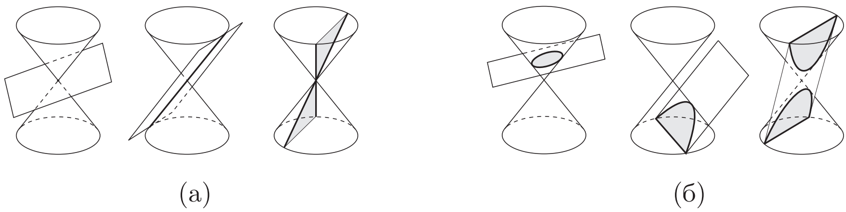


Рис. 9.16. Пересечение конуса и плоскости

Плоскость же, не проходящая через вершину конуса, может занимать относительно этого конуса следующие три положения (см. рис. 9.16(б)):

(i) пересекать все образующие конуса (в этом случае плоскость пересекает только одну полу конуса),
(ii) быть параллельной одной образующей конуса (в этом случае плоскость также пересекает только одну полу конуса),
(iii) быть параллельной двум различным образующим конуса (в этом случае плоскость пересекает обе полы конуса).

9.53. Изложим[1] подход к коническим сечениям, использованный Менехмом. Рассмотрим сечение прямого кругового конуса плоскостью, перпендикулярной к образующей этого конуса; угол при вершине в осевом сечении конуса обозначим 2α. Докажем, что при различных значениях этого угла получается эллипс, если $2\alpha < 90^\circ$, гипербола, если $2\alpha > 90^\circ$, и парабола, если $2\alpha = 90^\circ$.

9.54. Пусть S — вершина конуса, BC — диаметр его основания. Рассмотрим сечение конуса плоскостью, проходящей через точку G на образующей SC перпендикулярно SC. Эта плоскость пересекается с плоскостью SBC по прямой GK, которая является осью симметрии конического сечения (назовем её фокальной осью). Из произвольной точки L сечения опустим перпендикуляр LK на ось симметрии. Введем обозначения

$$|SG| = r, \quad |GK| = x, \quad |KL| = y.$$

Эти три отрезка взаимно перпендикулярны, поэтому

$$|SL|^2 = r^2 + x^2 + y^2.$$

Отложим на прямой SC отрезок SM длины $SM = SL$.

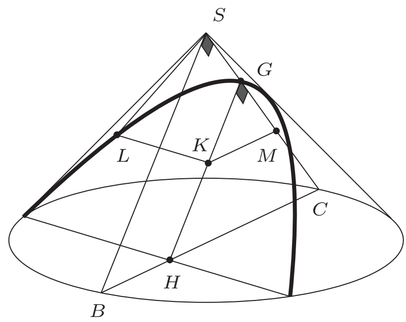


Рис. 9.17. Парабола как коническое сечение

9.55. В случае $2\alpha = 90^\circ$ имеем $|GM| = |GK| = x$, поэтому

$$|SM| = |SG| + |GM| = r + x \quad \Rightarrow \quad r^2 + x^2 + y^2 = (r + x)^2 = r^2 + 2rx + x^2,$$

т.е.

$$y^2 = 2rx.$$

Это каноническое уравнение параболы.

[1]Изложение следует книге [20]. Рисунки 9.20–9.23 взяты из книги [1].

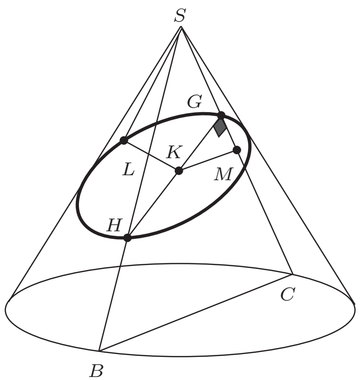


Рис. 9.18. Эллипс
как коническое сечение

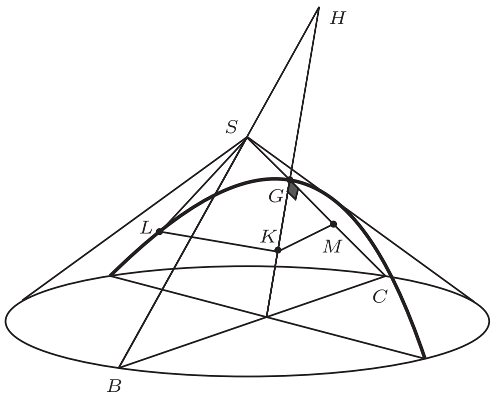


Рис. 9.19. Гипербола
как коническое сечение

9.56. Рассмотрим случай $2\alpha \neq 90^\circ$. Обозначим через H точку пересечения прямых GK и SB и через $2a$ длину отрезка GH. (Впоследствии станет ясно, что точки G и H являются вершинами эллипса или гиперболы.)

В $\triangle GKM$ имеем: $\angle KGM = 90^\circ$, $\angle GKM = \alpha$. Поэтому $|GM| = x \operatorname{tg} \alpha$. Получаем

$$r^2 + x^2 + y^2 = (r + x \operatorname{tg} \alpha)^2 = r^2 + 2xr \operatorname{tg} \alpha + x^2 \operatorname{tg}^2 \alpha,$$

т.е.

$$y^2 = 2xr \operatorname{tg} \alpha + x^2(\operatorname{tg}^2 \alpha - 1).$$

Введя обозначения $\varepsilon = \operatorname{tg} \alpha$, $p = r \operatorname{tg} \alpha$, перепишем уравнение в виде

$$y^2 = 2px + x^2(\varepsilon^2 - 1),$$

т.е. получили уравнение эллипса и гиперболы, отнесённое к вершине.

9.57. Аналогичные результаты получаются, если плоскость сечения проводить не перпендикулярно образующей конуса, а под произвольным углом. Чтобы продемонстрировать это, воспользуемся конструкцией, предложенной Ж. Данделеном[1].

Рассмотрим сечение конуса с вершиной S плоскостью π, пересекающей только одну полу конуса и не перпендикулярной оси m. Впишем в конус две сферы, касающиеся плоскости π в точках F_1 и F_2 (см. рис. 9.20).

Пусть M — произвольная точка на линии пересечения конуса с плоскостью π. Проведём через точку M образующую SM и обозначим точки A_1 и A_1 её пересечения с вписанными сферами. Тогда

$$|MF_1| = |MA_1|, \quad |MF_2| = |MA_2|$$

согласно свойству касательных к сфере, проведённых из одной точки. Следовательно,

$$|MF_1| + |MF_2| = |MA_1| + |MA_2| = |A_1A_2| = \text{const},$$

[1]Ж. Данделен (G. Dandelin) — бельгийский математик (1794–1847).

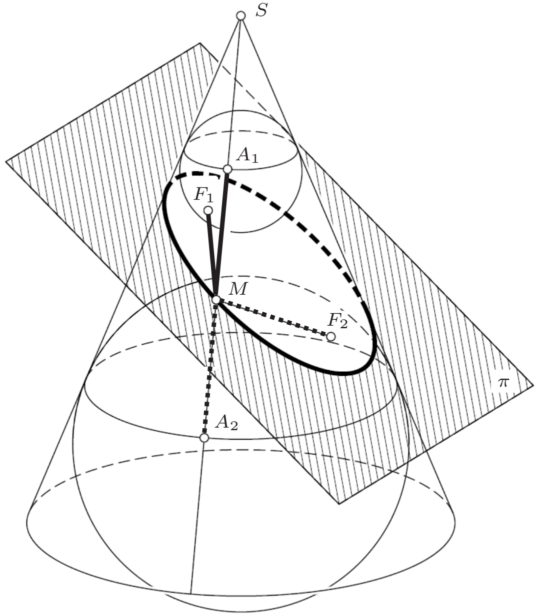


Рис. 9.20. Конструкция Данделена для эллипса

поскольку отрезок A_1A_2 — это отрезок образующей, заключённый между двумя параллельными плоскостями, длина которого не зависит от выбора образующей (т.е. от выбора точки M). Итак, линия пересечения конуса с плоскостью π является эллипсом.

9.58. В точности такими же рассуждениями можно показать, что сечениями кругового цилиндра будут эллипсы, так что эллипс может быть получен как параллельная проекция окружности на плоскость, не параллельную плоскости окружности (см. рис. 9.21).

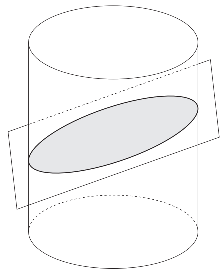


Рис. 9.21. Эллипс как сечение цилиндра

9.59. Аналогично доказывается, что если секущая плоскость параллельна двум различным образующим конуса (т.е. пересекает обе его полы), то в сечении получается гипербола (см. рис. 9.22).

9.60. Рассмотрим случай, когда секущая плоскость π параллельна одной из образующей (см. рис. 9.23). Впишем в конус сферу, касающуюся этой плоскости в точке F. Эта сфера касается конуса по окружности, лежащей в плоскости, которую мы обозначим σ. Пусть l — точка пересечения плоскостей π и σ.

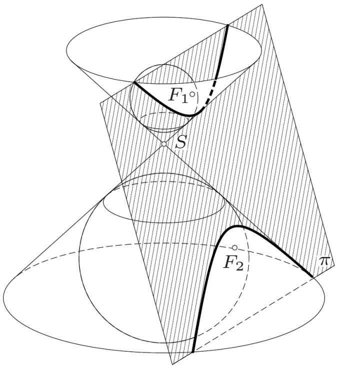


Рис. 9.22. Конструкция Данделена для гиперболы

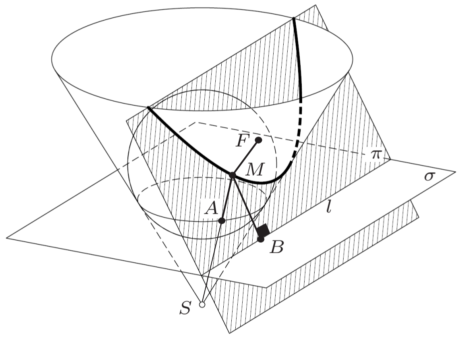


Рис. 9.23. Конструкция Данделена для параболы

Возьмём произвольную точку M сечения конуса плоскостью π и найдём точку A пересечения образующей SM с плоскостью σ и проекцию B точки M на прямую l. Тогда $|MF| = |MA|$ как касательные к сфере. С другой стороны, точки A и B лежат в плоскости σ, угол между прямой MA и плоскостью σ (т.е. угол между образующей конуса и плоскостью, перпендикулярной его оси) равен углу между прямой MB и плоскостью σ (т.е. углу между плоскостями π и σ). Поэтому $|MA| = |MB|$ как наклонные, образующие равные углы с плоскостью σ и, следовательно, $|MF| = |MB|$, так что точка M лежит на параболе с фокусом F и директрисой l.

9.61. Эксцентриситет и директрисы эллипса в конструкции Данделена получаются следующим образом. Пусть плоскость π пересекает все образующие конуса с вершиной S, так что в сечении образуется эллипс. Впишем в конус сферу, касающуюся плоскости π в точке F (как было показано выше, эта точка является одним из фокусов эллипса). Построим плоскость σ, содержащую окружность, по которой касаются конус и сфера, и обозначим через l прямую, по которой пересекаются плоскости π и σ. Пусть M — произвольная точка эллипса. Обозначим через A точку пересечения прямой SM с плоскостью σ, а через B — проекцию точки M на прямую l. Покажем, что отношение $|MA|/|MB|$ не зависит от выбора точки M, т.е. прямая l является директрисой, а указанное отношение — эксцентриситетом эллипса.

9.62. Обозначим через C проекцию точки M на плоскость σ. Имеем

$$\frac{|MC|}{|MA|} = \cos\alpha, \quad \frac{|MC|}{|MB|} = \cos\beta,$$

где α — угол между образующей конуса и его осью, β — угол между плоскостью π и осью конуса. Поэтому

$$\frac{|MA|}{|MB|} = \frac{|MA|}{|MC|} \cdot \frac{|MC|}{|MB|} = \frac{\cos\beta}{\cos\alpha}.$$

Аналогичные построения и рассуждения можно провести и для гиперболы.

ГЛАВА 10

Преобразования плоскости

ГЕОМЕТРИЧЕСКИЕ преобразования играют в геометрии важнейшую роль. Из школьного курса читателю хорошо известно понятие движения плоскости — преобразования евклидовой плоскости, при котором «ничего существенного не меняется»: фигуры, переводимые друг в друга некоторым движением, считаются равными. Это определение равенства фигур характеризует евклидову геометрию. Таким образом, каждое движение представляет собой изоморфизм евклидовой плоскости на себя (автоморфизм). Аналогичным образом можно определить и автоморфизмы аффинных пространств — аффинные преобразования. Мы ограничимся рассмотрением только планиметрического случая.

1. Преобразование базисов и координат на плоскости

В этом разделе мы изучим преобразования ортогональных базисов и координат на евклидовой плоскости, пользуясь наглядными представлениями. Аналогичный анализ в рамках аксиоматической модели проведён в разделах 2 и 3.

А. Поворот системы координат.

10.1. Пусть на евклидовой плоскости $\mathcal{E}_2$ заданы две ортогональные системы координат с общим началом: $O\boldsymbol{I}$ («старая») и $O\boldsymbol{I}'$ («новая»), где $\boldsymbol{I} = \big(\boldsymbol{i}, \boldsymbol{j}\big)$ и $\boldsymbol{I}' = \big(\boldsymbol{i}', \boldsymbol{j}'\big)$ — «старый» и «новый» базисы в ассоциированном векторном пространстве $\mathcal{V}$; оба эти базиса ортонормированные. Новая система координат получена из старой поворотом осей координат на угол α (см. рис. 10.1).

Разложим векторы нового базиса по старому базису (ср. п. 6.86, с. 153):

$$\begin{cases} \boldsymbol{i}' = r_1^1\boldsymbol{i} + r_1^2\boldsymbol{j}, \\ \boldsymbol{j}' = r_2^1\boldsymbol{i} + r_2^2\boldsymbol{j} \end{cases} \iff \boldsymbol{I}' = \boldsymbol{I}R, \quad R = \begin{pmatrix} r_1^1 & r_2^1 \\ r_1^2 & r_2^2 \end{pmatrix}. \qquad (10.1)$$

Матрица R, составленная из коэффициентов этих разложений, называется *матрицей перехода* от базиса $\boldsymbol{I}$ к базису $\boldsymbol{I}'$. Обратите внимание, что коэффициенты разложения вектора $\boldsymbol{i}'$ записываются в первый столбец (а не в первую строку!) матрицы R.

Для нахождения коэффициентов разложения r_k^j воспользуемся формулами Гиббса (8.9) (см. с. 191):

$$r_1^1 = (\boldsymbol{i}', \boldsymbol{i}) = \cos\alpha, \qquad r_2^1 = (\boldsymbol{j}', \boldsymbol{i}) = -\sin\alpha,$$
$$r_1^2 = (\boldsymbol{i}', \boldsymbol{j}) = \sin\alpha, \qquad r_2^2 = (\boldsymbol{j}', \boldsymbol{j}) = \cos\alpha,$$

т.е.

$$R = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}.$$

Матрица перехода в рассматриваемом случае называется *матрицей поворота* (или вращения). Она обладает легко проверяемым (см. формулу (3.3), с. 57) замечательным свойством

$$R^{-1} = R^T,$$

которое называется *свойством ортогональности.*

10.2. Пусть A — произвольная точка евклидовой плоскости $\mathcal{E}_2$, X и X' — столбцы координат радиус-вектора $\overrightarrow{OA}$ относительно базисов $\boldsymbol{I}$ и $\boldsymbol{I}'$ соответственно. Имеем

$$\overrightarrow{OA} = \boldsymbol{I}X = \boldsymbol{I}'X';$$

поскольку $\boldsymbol{I}' = \boldsymbol{I}R$, отсюда находим $\boldsymbol{I}X = (\boldsymbol{I}R)X' = \boldsymbol{I}(RX')$. В силу единственности разложения по базису получаем $X = RX'$, откуда $X' = R^{-1}X$.

10.3. Запишем все полученные формулы ещё раз:

$$\text{переход } O\boldsymbol{ij} \to O\boldsymbol{i}'\boldsymbol{j}': \quad \begin{cases} \boldsymbol{I}' = \boldsymbol{I}R, \\ X' = R^{-1}X = R^T X; \end{cases} \tag{10.2}$$

$$\text{переход } O\boldsymbol{i}'\boldsymbol{j}' \to O\boldsymbol{ij}: \quad \begin{cases} \boldsymbol{I} = \boldsymbol{I}'R^{-1} = \boldsymbol{I}'R^T, \\ X = RX'. \end{cases} \tag{10.3}$$

Преобразование строк, составленных из векторов базиса, и преобразование столбцов координат производятся взаимно обратными матрицами. Этот факт незаметен (более того, создаётся противоположное впечатление!), если формулы записать в координатном виде. Для перехода $O\boldsymbol{ij} \to O\boldsymbol{i}'\boldsymbol{j}'$ имеем

$$\begin{cases} \boldsymbol{i}' = \boldsymbol{i}\cos\alpha + \boldsymbol{j}\sin\alpha, \\ \boldsymbol{j}' = -\boldsymbol{i}\sin\alpha + \boldsymbol{j}\cos\alpha, \end{cases} \qquad \begin{cases} x' = x\cos\alpha + y\sin\alpha, \\ y' = -x\sin\alpha + y\cos\alpha \end{cases} \tag{10.4}$$

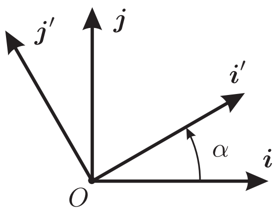


Рис. 10.1. Поворот ортогональной системы координат

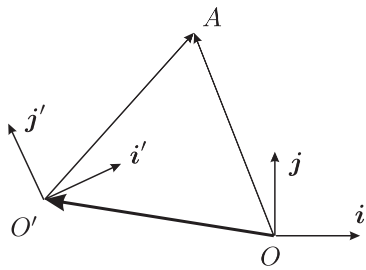


Рис. 10.2. Ортогональное преобразование

а для для обратного перехода $O\boldsymbol{i}'\boldsymbol{j}' \to O\boldsymbol{i}\boldsymbol{j}$ —

$$\begin{cases} \boldsymbol{i} = \boldsymbol{i}'\cos\alpha - \boldsymbol{j}'\sin\alpha, \\ \boldsymbol{j} = \boldsymbol{i}'\sin\alpha + \boldsymbol{j}'\cos\alpha, \end{cases} \qquad \begin{cases} x = x'\cos\alpha - y'\sin\alpha, \\ y = x'\sin\alpha + y'\cos\alpha. \end{cases} \tag{10.5}$$

Конечно, всё дело здесь в порядке матричных сомножителей и ортогональности матрицы перехода.

Б. Композиция поворота и параллельного переноса.

10.4. Рассмотрим теперь общий случай, когда новая ортогональная система координат $O'\boldsymbol{i}'\boldsymbol{j}'$ получается из старой параллельным переносом начала на вектор $\overrightarrow{OO'}$ и поворотом базисных векторов на угол α (см. рис. 10.2). Поворот координатных осей ортогональной декартовой системы и перенос её начала являются примерами *ортогональных преобразований* евклидовой плоскости; общее определение будет дано в разделе 3.

Пусть A — произвольная точка плоскости. Очевидно,

$$\overrightarrow{OA} = \overrightarrow{OO'} + \overrightarrow{O'A}.$$

Радиус-векторы $\overrightarrow{OA}$ и $\overrightarrow{OO'}$ точек A и O' разложим по старому базису $\boldsymbol{I}$, а радиус-вектор $\overrightarrow{O'A}$ — по новому базису $\boldsymbol{I}'$. Введём столбцы координат этих радиус-векторов:

$$\overrightarrow{OA}:\ X = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \overrightarrow{OO'}:\ X_0 = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}, \quad \overrightarrow{OA'}:\ X' = \begin{pmatrix} x' \\ y' \end{pmatrix}.$$

Тогда

$$\boldsymbol{I}X = \boldsymbol{I}X_0 + \boldsymbol{I}'X'.$$

Учитывая формулу $\boldsymbol{I}' = \boldsymbol{I}R$ (см. (10.1)) и пользуясь единственностью разложения по базису, получаем

$$X = X_0 + RX'. \tag{10.6}$$

Отсюда легко получить и формулу обратного преобразования:

$$X' = R^{-1}(X - X_0) = R^T(X - X_0). \tag{10.7}$$

Запишем эти формулы в развёрнутом виде:

$$\begin{cases} x = x' \cos\alpha - y' \sin\alpha + x_0, \\ y = x' \sin\alpha + y' \cos\alpha + y_0 \end{cases} \tag{10.8}$$

и

$$\begin{cases} x' = (x - x_0)\cos\alpha + (y - y_0)\sin\alpha, \\ y' = -(x - x_0)\sin\alpha + (y - y_0)\cos\alpha. \end{cases} \tag{10.9}$$

10.5. Часто для удобства вычислений (особенно в компьютерной графике) используются так называемые *однородные координаты*[1]: столбцы координат векторов и точек дополняются ещё одним элементом, равным 1:

$$Z = \begin{pmatrix} x \\ y \\ \hline 1 \end{pmatrix} = \left\| \begin{array}{c} X \\ \hline 1 \end{array} \right\|, \quad Z' = \begin{pmatrix} x' \\ y' \\ \hline 1 \end{pmatrix} = \left\| \begin{array}{c} X' \\ \hline 1 \end{array} \right\| \tag{10.10}$$

(напомним, что квадратные скобки обозначают блочную матрицу). Вводя *расширенную матрицу преобразования*

$$P = \left(\begin{array}{cc|c} \cos\alpha & -\sin\alpha & x_0 \\ \sin\alpha & \cos\alpha & y_0 \\ \hline 0 & 0 & 1 \end{array} \right) = \left\| \begin{array}{c|c} R & X_0 \\ \hline O & 1 \end{array} \right\|, \tag{10.11}$$

можно записать преобразования координат (10.6) и (10.7) в виде

$$Z = PZ', \quad Z' = P^{-1}Z, \tag{10.12}$$

причём, как легко проверить (сделайте это!), обратная матрица P^{-1} имеет вид

$$P^{-1} = \left\| \begin{array}{c|c} R^T & -R^T X_0 \\ \hline O & 1 \end{array} \right\| = \left(\begin{array}{cc|c} \cos\alpha & \sin\alpha & x_0' \\ -\sin\alpha & \cos\alpha & y_0' \\ \hline 0 & 0 & 1 \end{array} \right), \tag{10.13}$$

[1]Термин обязан своему происхождению проективной геометрии.

где

$$\begin{cases} x_0' = -x_0\cos\alpha - y_0\sin\alpha, \\ y_0' = x_0\sin\alpha - y_0\cos\alpha \end{cases} \quad\Leftrightarrow\quad X_0' = -R^T X_0$$

— координаты точки O (начала системы координат $O\boldsymbol{ij}$) в системе координат $O'\boldsymbol{i}'\boldsymbol{j}'$.

B. Ортогональные преобразования уравнения прямой на плоскости.

10.6. Рассмотрим общее уравнение прямой

$$Ax + By + d = 0$$

в прямоугольной декартовой системе координат $O\boldsymbol{ij}$. Коэффициенты A и B являются координатами вектора нормали $\boldsymbol{n}$ этой прямой[1]. В матричных обозначениях это уравнение имеет вид

$$N^T X + d = 0, \tag{10.14}$$

где $N = \big(A, B\big)^T$ — столбец координат вектора нормали $\boldsymbol{n}$ прямой. При переходе к новой системе координат $O'\boldsymbol{i}'\boldsymbol{j}'$ по формулам (10.6) уравнение преобразуется следующим образом:

$$N^T\big(RX' + X_0\big) + d = 0 \;\Leftrightarrow\; (N^T R)X' + N^T X_0 + d = 0,$$

т.е. принимает вид, совпадающий с (10.14):

$$N'^T X' + d' = 0,$$

где

$$N' = (N^T R)^T = R^T N = R^{-1}N, \quad d' = N^T X_0 + d.$$

(Обратите внимание на соответствие соотношения $N' = R^{-1}N$ формуле (10.2), по которой преобразуются координаты вектора при повороте системы координат!)

10.7. Если записать уравнение прямой в однородных координатах, т.е. в виде

$$MZ = 0,$$

где $M = \big(A, B, d\big)$, $Z = \big(x, y, 1\big)^T$, а преобразование координат осуществить по формулам (10.12), получим

$$M(PZ') = 0 \;\Leftrightarrow\; (MP)Z' = 0,$$

т.е. матричный коэффициент M уравнения прямой преобразуется по формуле

$$M' = MP.$$

[1]Напомним, что это утверждение верно только для прямоугольной системы координат (см. замечание 8.53, с. 203).

2. Аффинные преобразования

10.8. Формулу $X = X_0 + RX'$ (см. (10.6)) можно трактовать не только как соотношение, связывающее координаты одной и той же точки в разных реперах, но и как формулу, выражающую координаты точки X через координаты точки X' в одном и том же репере, т.е. трактовать её как отображение евклидовой плоскости $\mathcal{E}_2$ на себя,

$$f : \mathcal{E}_2 \to \mathcal{E}_2, \quad X' \mapsto X = X_0 + RX', \tag{10.15}$$

где точка X' считается исходной (прообразом), а X — результатом отображения (образом). Эти две точки зрения называются соответственно «пассивной» и «активной»: в первом случае мы описываем одну и ту же точку с точки зрения разных наблюдателей (в разных реперах). Во втором случае наблюдатель (репер) один, а точка подвергается преобразованию с помощью отображения f.

10.9. Кроме того, можно рассматривать преобразования, аналогичные (10.15), не только евклидовой плоскости, но и аффинной; в этом случае естественным будет отказаться от свойства ортогональности матрицы преобразования R.

A. Основные свойства аффинных преобразований. Рассмотрим аффинное пространство $(\mathcal{A}, \mathcal{V}, \mathbb{K})$, имеющее ассоциированное векторное пространство $\mathcal{V}$ над полем $\mathbb{K}$. Зафиксируем в $\mathcal{A}$ произвольный аффинный репер $O\boldsymbol{E}$ и будем в дальнейшем координаты всех точек задавать относительно этого репера, а координаты всех векторов — относительно базиса $\boldsymbol{E}$ ассоциированного векторного пространства $\mathcal{V}$.

10.10. Определение. Отображение

$$f : \mathcal{A} \to \mathcal{A}, \quad X \mapsto X' = CX + X_0, \tag{10.16}$$

где C — невырожденная матрица, называется *аффинным преобразованием*[1] пространства $\mathcal{A}$. Аффинные преобразования можно записывать и с помощью однородных координат:

$$Z' = PZ, \quad Z = \left\| \begin{array}{c} X \\ \hline 1 \end{array} \right\|, \quad Z' = \left\| \begin{array}{c} X' \\ \hline 1 \end{array} \right\|, \quad P = \left\| \begin{array}{c|c} C & X_0 \\ \hline O & 1 \end{array} \right\|.$$

Матрицы C и P называются соответственно *основной* и *расширенной матрицами* аффинного преобразования f.

10.11. Множество всех аффинных преобразований аффинного пространства $\mathcal{A}$ обозначается $\operatorname{Aff} \mathcal{A}$. Легко видеть, что оно является группой относительно операции композиции преобразований (проверьте выполнение групповых аксиом **Г1–Г3**; см. с. 338).

10.12. Теорема.

1. *Каждое аффинное преобразование взаимно однозначно.*
2. *При аффинном преобразовании p-плоскость переходит в p-плоскость.*
3. *При аффинном преобразовании пересекающиеся плоскости переходят в пересекающиеся, параллельные — в параллельные, скрещивающиеся — в скрещивающиеся.*

---

[1] Обратите внимание на различие в записи формул (10.16) и (10.6): сейчас точка X — прообраз, точка X' — образ, т.е. результат действия преобразования.

4. *Пусть X_1, X_2, X_3 — три коллинеарные*[1] *точки,* $\overrightarrow{X_1X_3} = k\overrightarrow{X_1X_2}$. *Тогда для образов X_1', X_2', X_3' этих точек выполняется соотношение* $\overrightarrow{X_1'X_3'} = k\overrightarrow{X_1'X_2'}$. *Иными словами, аффинные преобразования сохраняют отношения коллинеарных отрезков*[2].

Доказательство. 1. Очевидно, каждой точке X соответствует единственная точка $X' = CX + X_0$. В силу невырожденности матрицы C каждой точке X' соответствует единственная точка $X = C^{-1}(X - X_0)$.

2. Пусть p-плоскость в n-мерном аффинном пространстве задана параметрическим уравнением

$$X = A_0 + t^1 A_1 + \dots + t^p A_p,$$

где A_0 — опорная точка, $A_1, \dots, A_p$ — линейно независимые направляющие векторы. После преобразования (10.16) получим

$$X' = C\Big(A_0 + t^1 A_1 + \dots + t^p A_p\Big) + X_0 \;\Leftrightarrow$$
$$\Leftrightarrow\; X' = (CA_0 + X_0) + t^1(CA_1) + \dots + t^p(CA_p).$$

Поскольку векторы $CA_1, \dots, CA_p$ линейно независимы (объясните!), полученное уравнение определяет p-плоскость с опорной точкой $CA_0 + X_0$ и направляющими векторами $CA_1, \dots, CA_p$.

Утверждение 3 читатель легко докажет самостоятельно, основываясь на использованных идеях.

Доказательство утверждения 4 сводится к лёгкому вычислению. Если $X_j' = CX_j + X_0$ $(j = 1, 2, 3)$, то

$$\begin{aligned}\overrightarrow{X_1'X_3'} &= X_3' - X_1' = (CX_3 + X_0) - (CX_1 + X_0) = \\ &= C(X_3 - X_1) = C\overrightarrow{X_1X_3} = Ck\overrightarrow{X_1X_2} = kC(X_2 - X_1) = \\ &= k\Big[(CX_2 + X_0) - (CX_1 + X_0)\Big] = k(X_2' - X_1') = k\overrightarrow{X_1'X_2'}. \qquad \square\end{aligned}$$

Б. Примеры аффинных преобразований плоскости.

1. *Параллельный перенос*, или *трансляция* на вектор X_0: $C = \mathbb{1}$, $X' = X + X_0$.
2. *Гомотетия*, или *равномерное растяжение* (*сжатие*) с центром в начале координат[3]: $X' = kX$, где $k \neq 0$. Матрицей гомотетии является матрица вида $k\mathbb{1}$ (такие матрицы называются *скалярными*).
3. *Растяжение* (*сжатие*) *вдоль оси абсцисс*:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} kx \\ y \end{pmatrix}; \quad C = \begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix}.$$

[1] Т.е. лежащие на одной прямой.

[2] Для неколлинеарных отрезков это утверждение не имеет места; приведите пример самостоятельно.

[3] Конечно, можно рассматривать гомотетию с центром, отличным от начала координат.

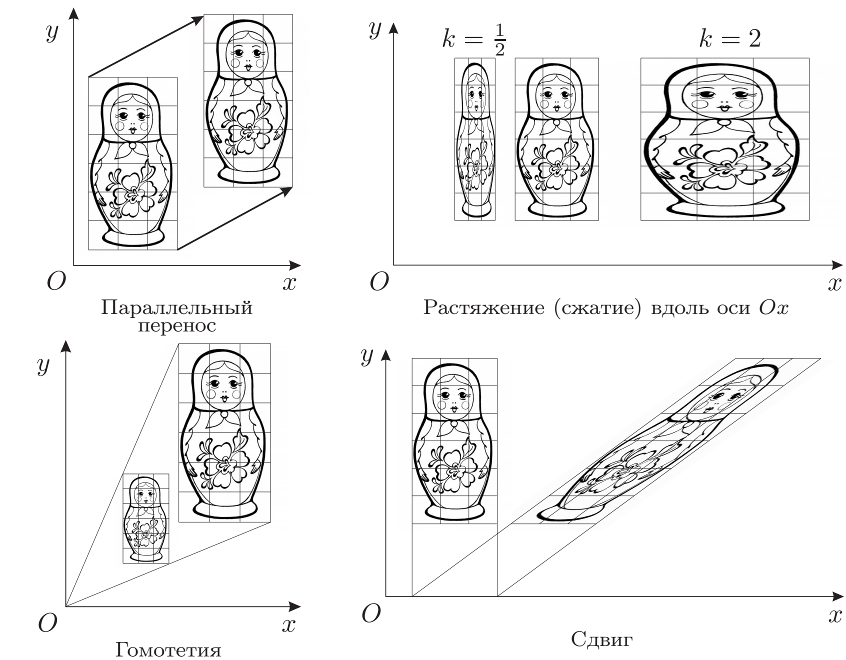


Рис. 10.3. Аффинные преобразования плоскости

Аналогично определяется растяжение (сжатие) вдоль любой другой оси (прямой). Комбинация двух растяжений (сжатий) вдоль обеих координатных осей задаётся формулами

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} kx \\ ly \end{pmatrix}; \quad C = \begin{pmatrix} k & 0 \\ 0 & l \end{pmatrix}.$$

4. *Сдвиг*

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x + py \\ y \end{pmatrix}; \quad C = \begin{pmatrix} 1 & p \\ 0 & 1 \end{pmatrix}.$$

3. Ортогональные преобразования

В этом разделе мы рассматриваем преобразования евклидовых пространств.

А. Основное свойство ортогональных преобразований.

10.13. Напомним (см. п. 10.1), что матрица C называется *ортогональной*, если $C^{-1} = C^T$.

Ортогональное преобразование, или *движение*, евклидова пространства $\mathcal{E}$ — это аффинное преобразование, основная матрица C которого ортогональна:

$$X' = CX + X_0.$$

10.14. Предложение. *Движения сохраняют расстояния между точками и величины углов.*

Доказательство. Пусть X, Y — произвольные точки евклидова точечного пространства $\mathcal{E}$; столбцы координат этих точек (относительно произвольной прямоугольной декартовой системы) будем обозначать теми же символами. Квадрат расстояния между этими точками равен $(Y-X)^T(Y-X)$, а квадрат расстояния между их образами $X' = CX + X_0$ и $Y' = CY + X_0$ —

$$\begin{aligned}(Y'-X')^T(Y'-X') &= (CY-CX)^T(CY-CX) = \\ &= \big(C(Y-X)\big)^T C(Y-X) = (Y-X)^T \underbrace{C^T C}_{=\mathbb{1}}(Y-X) = \\ &= (Y-X)^T(Y-X).\end{aligned}$$

Сохранение углов (т.е. скалярных произведений) при ортогональных преобразованиях доказывается аналогично (читателю предлагается сделать это самостоятельно). □

Б. Ортогональные матрицы и их свойства.

10.15. Применяя теоремы об определителе произведения матриц и об определителе транспонированной матрицы, из равенства $C^T C = \mathbb{1}$ получаем $\det C^T \cdot \det C = 1$, откуда

$$|\det C| = 1 \;\Leftrightarrow\; \det C = \pm 1.$$

Ортогональная матрица с определителем 1 называется *собственной*, а с определителем -1 — *несобственной*.

10.16. Предложение. *Множество всех ортогональных матриц порядка n образует группу относительно операции умножения матриц, которая называется* ортогональной группой *и обозначается $O(n)$. Собственные ортогональные матрицы образуют группу, которая называется* специальной ортогональной группой *и обозначается $SO(n)$.*

Доказательство. Проверим, что множество ортогональных матриц замкнуто относительно операции умножения матриц. Действительно, произведение ортогональных матриц C_1 и C_2 является ортогональной матрицей:

$$(C_1C_2)^T(C_1C_2) = C_2^T \underbrace{C_1^T C_1}_{=\mathbb{1}} C_2 = C_2^T C_2 = \mathbb{1}.$$

Если при этом матрицы C_1 и C_2 собственные, т.е. $\det C_1 = \det C_2 = 1$, то C_1C_2 — тоже собственная:

$$\det(C_1C_2) = \det C_1 \cdot \dot{C}_2 = 1.$$

Проверим аксиомы группы **Г1–Г3** (см. с. 338). Очевидно, единичная матрица является ортогональной: $\mathbb{1}^T\mathbb{1} = \mathbb{1}$. Ассоциативность умножения матрицы имеет место для всех матриц, в том числе для ортогональных. Матрица, обратная ортогональной, также ортогональна:

$$(C^{-1})^T C^{-1} = (C^T)^T C^{-1} = CC^{-1} = \mathbb{1};$$

если при этом $\det C = 1$, то $\det C^{-1} = 1$. □

10.17. Обратите внимание, что несобственные ортогональные матрицы не образуют группу: множество таких матриц не содержит единичной матрицы, а произведение двух несобственных матриц является матрицей собственной.

10.18. При $n = 1$ ортогональная матрица имеет единственный элемент $c^1_{1'}$, удовлетворяющий условию $(c^1_{1'})^2 = 1$. Таким образом,

$$O(1) = \{1,\ -1\}, \quad SO(1) = \{1\}.$$

В. Группы $O(2)$ и $SO(2)$. Найдём общий вид ортогональной матрицы порядка 2. Записав в развёрнутом виде условие $C^T C = \mathbb{1}$ для матрицы

$$C = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

получим уравнения

$$a^2 + c^2 = 1, \quad ab + cd = 0, \quad b^2 + d^2 = 1.$$

Первое и последнее уравнения означают, что найдутся такие углы α и β, что

$$a = \cos\alpha,\ c = \sin\alpha, \quad b = \cos\beta,\ d = \sin\beta,$$

а второе — что эти углы связаны соотношением

$$\cos\alpha\cos\beta + \sin\alpha\sin\beta = 0 \ \Leftrightarrow\ \cos(\alpha - \beta) = 0.$$

Таким образом, либо $\beta = \alpha + \pi/2$, и матрица C имеет вид

$$R = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}, \tag{10.17}$$

либо $\beta = \alpha - \pi/2$, и матрица C имеет вид

$$S = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}, \tag{10.18}$$

В первом случае матрица собственная, во втором — несобственная.

Г. Движения евклидовой плоскости.

10.19. Каждое движение евклидовой плоскости $\mathcal{E}_2$ может быть записано в виде $X' = CX + X_0$, где $C \in O(2)$ — ортогональная матрица порядка 2, т.е. матрица одного из следующих типов:

$$R = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \quad \text{либо} \quad S = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}.$$

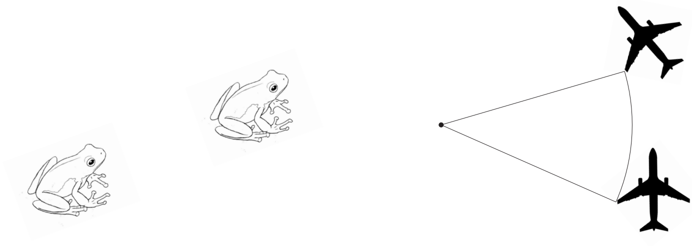
В первом случае $\det R = 1$, соответствующее движение *сохраняет* ориентацию плоскости и называется *движением первого рода*, или *собственным движением*:

$$\begin{cases} x' = x\cos\alpha - y\sin\alpha + x_0, \\ y' = x\sin\alpha + y\cos\alpha + y_0, \end{cases} \tag{10.19}$$

Во втором случае $\det S = -1$; такое движение *изменяет* ориентацию плоскости и называется *движением второго рода*, или *несобственным движением*:

$$\begin{cases} x' = x\cos\alpha + y\sin\alpha + x_0, \\ y' = x\sin\alpha - y\cos\alpha + y_0. \end{cases} \tag{10.20}$$

10.20. Классификацию движений удобно провести, используя понятие *неподвижной точки*. Пусть $f : \mathcal{E}_2 \to \mathcal{E}_2$ — какое-либо преобразование плоскости. Точка M называется *неподвижной точкой* преобразования f, если $f(M) = M$.



(а) (б)

Рис. 10.4. Собственные движения плоскости:
(а) параллельный перенос; (б) вращение

Д. Собственные движения плоскости.

10.21. При $\alpha = 0$ (в этом случае $R = \mathbb{1}$) собственное движение (10.19) записывается формулами

$$\begin{cases} x' = x + x_0, \\ y' = y + y_0. \end{cases} \tag{10.21}$$

Если $x_0 = y_0 = 0$, то это тождественное преобразование; в противном случае указанное преобразование неподвижных точек не имеет.

10.22. Собственное движение плоскости, не имеющее неподвижных точек, называется *параллельным переносом*, или *трансляцией* (см. рис. 10.4(а)).

Ясно, что каждая трансляция $\boldsymbol{T}$ однозначно задаётся вектором (x_0, y_0) и может быть отождествлена с ним. Это объясняет ещё одну возможную точку зрения на свободный вектор как на параллельный перенос.

10.23. Условие того, что точка M является неподвижной точкой собственного движения (10.19), записывается в виде системы уравнений

$$\begin{cases} x = x\cos\alpha - y\sin\alpha + x_0, \\ y = x\sin\alpha + y\cos\alpha + y_0 \end{cases}$$

или, эквивалентно,

$$\begin{cases} x(1-\cos\alpha) + \quad y\sin\alpha \quad = x_0, \\ \quad -x\sin\alpha \quad + y(1-\cos\alpha) = y_0. \end{cases} \tag{10.22}$$

Определитель этой системы

$$\begin{vmatrix} 1-\cos\alpha & \sin\alpha \\ -\sin\alpha & 1-\cos\alpha \end{vmatrix} = (1-\cos\alpha)^2 + \sin^2\alpha = 2(1-\cos\alpha)$$

при $\alpha \neq 0$ отличен от нуля, так что система имеет единственное решение, а соответствующее собственное движение — единственную неподвижную точку.

10.24. Собственное движение плоскости, имеющее одну неподвижную точку, называется *вращением* или *поворотом*, а сама неподвижная точка — центром вращения.

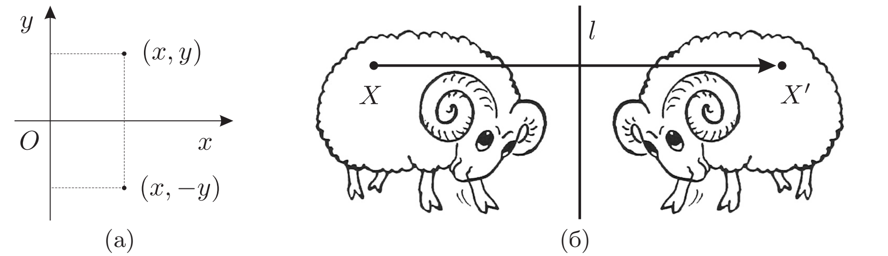


Рис. 10.5. Осевая симметрия

10.25. Геометрический смысл вращения достаточно очевиден: поскольку вращение, будучи движением, сохраняет расстояния, то любая точка плоскости M и её образ равноудалены от центра вращения, т.е. лежат на окружности с центром в неподвижной точке (см. рис. 10.4(б)).

Е. Несобственные движения плоскости.

10.26. Обсудим теперь неподвижные точки несобственных движений плоскости (10.20); эти точки определяются системой уравнений

$$\begin{cases} x = x\cos\alpha + y\sin\alpha + x_0, \\ y = x\sin\alpha - y\cos\alpha + y_0 \end{cases}$$

или, эквивалентно,

$$\begin{cases} x(1-\cos\alpha) - \quad y\sin\alpha \quad = x_0, \\ \quad -x\sin\alpha \quad + y(1+\cos\alpha) = y_0. \end{cases} \tag{10.23}$$

Определитель системы равен нулю:

$$\begin{vmatrix} 1-\cos\alpha & -\sin\alpha \\ -\sin\alpha & 1+\cos\alpha \end{vmatrix} = 1 - \cos^2\alpha - \sin^2\alpha = 0,$$

так что система либо не имеет решений (а соответствующее несобственное движение плоскости не имеет неподвижных точек), либо имеет бесконечно много решений (а у движения имеется бесконечно много неподвижных точек, причём эти точки заполняют целую прямую: вспомните устройство множества решений неоднородной системы).

Ж. Осевая симметрия.

10.27. Выясним геометрический смысл несобственного движения, неподвижные точки которого заполняют прямую. Например, неподвижные точки несобственного движения вида (10.20) с параметрами $\alpha = 0$, $x_0 = y_0 = 0$, т.е.

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ -y \end{pmatrix}$$

заполняют ось Ox, а само это движение является, очевидно, отражением в оси Ox (см. рис. 10.5(а)).

Докажем, что любое несобственное движение, имеющее прямую неподвижных точек, имеет аналогичный геометрический смысл.

10.28. Предложение. *Пусть $\boldsymbol{S}$ — несобственное движение, для которого прямая l состоит из неподвижных точек. Справедливы следующие утверждения:*

(i) *для любой точки X и её образа $X' = \boldsymbol{S}(X)$ вектор $\overrightarrow{XX'}$ перпендикулярен прямой l;*
(ii) *середина отрезка XX' лежит на прямой l;*
(iii) *движение $\boldsymbol{S}$ инволютивно, т.е. его квадрат равен тождественному отображению:*

$$\boldsymbol{S} \circ \boldsymbol{S} = \mathrm{id} \;\Leftrightarrow\; \boldsymbol{S}^{-1} = \boldsymbol{S}.$$

Доказательство. В рассматриваемом случае система (10.23), определяющая неподвижные точки, состоит из пропорциональных уравнений:

$$\frac{1-\cos\alpha}{-\sin\alpha} = \frac{-\sin\alpha}{1+\cos\alpha} = \frac{x_0}{y_0}.$$

Первое из этих равенств выполняется тождественно, в то время как второе является условием совместности системы (10.23). Итак, для того чтобы рассматриваемое несобственное движение обладало прямой неподвижных точек, необходимо и достаточно выполнение соотношений

$$x_0 \sin\alpha + y_0(1-\cos\alpha) = 0 \;\Leftrightarrow\; x_0(1+\cos\alpha) + y_0 \sin\alpha = 0. \tag{10.24}$$

Каждое из уравнение системы (10.23) при этом является уравнением прямой неподвижных точек. Таким образом, вектор нормали этой прямой можно взять, например, в виде $\boldsymbol{n} = (1-\cos\alpha, -\sin\alpha)^T$; тогда направляющий вектор этой прямой есть $\boldsymbol{a} = (\sin\alpha, 1-\cos\alpha)^T$.

1. Рассмотрим вектор $\overrightarrow{XX'}$, где $X = (x, y)^T$ — произвольная точка плоскости, а $X' = \boldsymbol{S}(X) = (x', y')^T$ — её образ, вычисляемый по формуле (10.20):

$$\begin{aligned}\overrightarrow{XX'} &= X' - X = (SX + X_0) - X = \\ &= \begin{pmatrix} x\cos\alpha + y\sin\alpha + x_0 \\ x\sin\alpha - y\cos\alpha + y_0 \end{pmatrix} - \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x(\cos\alpha - 1) + y\sin\alpha + x_0 \\ x\sin\alpha - y(\cos\alpha + 1) + y_0 \end{pmatrix}.\end{aligned}$$

Найдём скалярное произведение вектора $\overrightarrow{XX'}$ и направляющего вектора $\boldsymbol{a}$ неподвижной прямой:

$$\left(\overrightarrow{XX'}, \boldsymbol{a}\right) = \begin{pmatrix} x(\cos\alpha - 1) + y\sin\alpha + x_0 \\ x\sin\alpha - y(\cos\alpha + 1) + y_0 \end{pmatrix}^T \cdot \begin{pmatrix} \sin\alpha \\ 1-\cos\alpha \end{pmatrix} = 0,$$

где приняты во внимание соотношения (10.24). Утверждение 1 доказано.

2. Доказательство того факта, что точка Y, являющаяся серединой отрезка XX', лежит на прямой неподвижных точек l, чисто техническое. Найдём координаты точки Y:

$$Y = \frac{1}{2}(X + X') = \frac{1}{2}\begin{pmatrix} x(1+\cos\alpha) + y\sin\alpha + x_0 \\ x\sin\alpha + y(1-\cos\alpha) + y_0 \end{pmatrix}$$

и подставим их в уравнение прямой l, т.е. в любое из уравнений (10.23). Принимая во внимание соотношения (10.24), обнаруживаем, что точка Y действительно лежит на прямой l:

$$\frac{1}{2}\Big(x(1+\cos\alpha)+y\sin\alpha+x_0\Big)(1-\cos\alpha)- \\ -\frac{1}{2}\Big(x\sin\alpha+y(1-\cos\alpha)+y_0\Big)\sin\alpha=x_0.$$

3. Найдём результат двукратного действия движения $\boldsymbol{S}$ на произвольную точку X, т.е. $\boldsymbol{S}\big(\boldsymbol{S}(X)\big)$. Если $X=(x,y)^T$, то

$$X'=\boldsymbol{S}(X)=SX+X_0,$$
$$X''=\boldsymbol{S}(X')=S\big(SX+X_0\big)+X_0=S^2X+SX_0+X_0.$$

Учитывая соотношения (10.24), находим

$$S^2=\begin{pmatrix}\cos\alpha & \sin\alpha\\ \sin\alpha & -\cos\alpha\end{pmatrix}^2=\mathbb{1},$$

$$SX_0+X_0=\begin{pmatrix}x_0\cos\alpha+y_0\sin\alpha+x_0\\ x_0\sin\alpha-y_0\cos\alpha+y_0\end{pmatrix}=\begin{pmatrix}x_0(1+\cos\alpha)+y_0\sin\alpha\\ x_0\sin\alpha+y_0(1-\cos\alpha)\end{pmatrix}=O,$$

т.е. $\boldsymbol{S}\big(\boldsymbol{S}(X)\big)=X$, т.е. $\boldsymbol{S}^2=\mathrm{id}$. □

10.29. Определение. Несобственное движение плоскости, у которого имеется бесконечно много неподвижных точек, заполняющих прямую, называется (осевой) *симметрией* или *отражением* относительно этой прямой, а сама неподвижная прямая — *осью симметрии*.

10.30. Из предложения 10.28(i) вытекает, что при осевой симметрии любая прямая, перпендикулярная оси симметрии, переходит в себя, т.е. является неподвижной, или *инвариантной*. Не следует смешивать понятия «неподвижная прямая» и «прямая, состоящая из неподвижных точек»: точки «неподвижной прямой» могут переходить друг в друга, в то время как каждая точка «прямой, состоящей из неподвижных точек», переходит в точности в себя.

3. Скользящая симметрия.

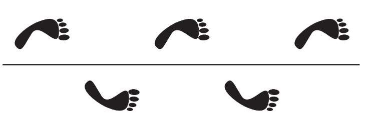


Рис. 10.6. Скользящая симметрия

10.31. Определение. Несобственное движение плоскости, не имеющее неподвижных точек, называется *скользящей симметрией*.

10.32. Предложение. *Любая скользящая симметрия $\boldsymbol{G}$ является композицией осевой симметрии $\boldsymbol{S}$ и трансляции $\boldsymbol{T}$ вдоль её оси: $\boldsymbol{G}=\boldsymbol{T}\circ\boldsymbol{S}$.*

Доказательство. Пусть $\boldsymbol{G}$ — скользящая симметрия, т.е. движение плоскости вида (10.20), не имеющее неподвижных точек; в этом случае система (10.23) несовместна. Покажем, что существует такой параллельный перенос, т.е. вектор $\boldsymbol{T}=(p,q)$, что преобразование $\boldsymbol{T}^{-1}\circ\boldsymbol{G}$ является осевой симметрией $\boldsymbol{S}$ с осью, параллельной вектору $\boldsymbol{T}$: этим предложение будет доказано, поскольку из равенства $\boldsymbol{T}^{-1}\circ\boldsymbol{G}=\boldsymbol{S}$ вытекает, что $\boldsymbol{G}=\boldsymbol{T}\circ\boldsymbol{S}$.

Очевидно, преобразование $\boldsymbol{S}=\boldsymbol{T}^{-1}\circ\boldsymbol{G}$ задаётся формулами

$$\begin{cases} x'=x\cos\alpha+y\sin\alpha+x_0-p,\\ y'=x\sin\alpha-y\cos\alpha+y_0-q.\end{cases}$$

Его неподвижные точки определяются из системы уравнений

$$\begin{cases} x(1-\cos\alpha)-\quad y\sin\alpha \quad =x_0-p,\\ \quad -x\sin\alpha \quad +y(1+\cos\alpha)=y_0-q.\end{cases} \tag{10.25}$$

Задача состоит в том, чтобы найти такие p и q, чтобы

(а) эта система определяла прямую (тогда преобразование $\boldsymbol{S}$ будет симметрией),

(б) полученная прямая была параллельна вектору $\boldsymbol{T}$.

Условие (а) будет выполнено, если уравнения системы (10.25) пропорциональны:

$$\frac{1-\cos\alpha}{-\sin\alpha}=\frac{-\sin\alpha}{1+\cos\alpha}=\frac{x_0-p}{y_0-q}.$$

Как и в доказательстве предложения 10.28, первое из этих равенств выполняется тождественно, а второе даёт уравнение для p и q:

$$\frac{-\sin\alpha}{1+\cos\alpha}=\frac{x_0-p}{y_0-q} \;\Leftrightarrow\; (x_0-p)\cos\frac{\alpha}{2}+(y_0-q)\sin\frac{\alpha}{2}=0$$

или, эквивалентно,

$$p\cos\frac{\alpha}{2}+q\sin\frac{\alpha}{2}=x_0\cos\frac{\alpha}{2}+y_0\sin\frac{\alpha}{2}.$$

Условие (б) означает ортогональность вектора трансляции $\boldsymbol{T}$ и нормального вектора $\boldsymbol{n}$ прямой (10.25):

$$(\boldsymbol{n},\boldsymbol{T})=\begin{pmatrix}1-\cos\alpha\\ -\sin\alpha\end{pmatrix}^T\cdot\begin{pmatrix}p\\ q\end{pmatrix}=0 \;\Leftrightarrow\; p\sin\frac{\alpha}{2}-q\cos\frac{\alpha}{2}=0.$$

Таким образом, для p и q получены два уравнения:

$$\begin{cases} p\sin\dfrac{\alpha}{2}-q\cos\dfrac{\alpha}{2}=0,\\ p\cos\dfrac{\alpha}{2}+q\sin\dfrac{\alpha}{2}=x_0\cos\dfrac{\alpha}{2}+y_0\sin\dfrac{\alpha}{2}.\end{cases}$$

Определитель этой системы равен единице, так что она имеет единственное решение, которое при необходимости легко найти, например, по формулам Крамера (нам это решение не понадобится, поэтому находить его не будем). Предложение доказано. □

И. Композиции движений плоскости.

10.33. Теорема (теорема Шаля[1]). *Всякое собственное (сохраняющее ориентацию) движение плоскости представляет собой либо вращение* $\boldsymbol{R}$, *либо трансляцию* $\boldsymbol{T}$. *Всякое несобственное (изменяющее ориентацию) движение плоскости является осевой симметрией* $\boldsymbol{S}$ *или скользящей симметрией* $\boldsymbol{G}$:

---

[1] М. Шаль (M. Chasles) — французский математик (геометр) и историк математики (1793–1880).

	Количество неподвижных точек		
	0	1	∞
Движение I рода	трансляция $\boldsymbol{T}$	вращение $\boldsymbol{R}$	тождественное преобразование id
Движение II рода	скользящая симметрия $\boldsymbol{G}$		осевая симметрия $\boldsymbol{S}$

10.34. Очевидно, что композиция двух собственных или двух несобственных движений является собственным движением, а композиция собственного и несобственного — несобственным движением.

10.35. Предложение. *Каждое несобственное движение $\boldsymbol{F}_2$ плоскости является композицией*

$$\boldsymbol{F}_2 = \boldsymbol{S} \circ \boldsymbol{F}_1$$

некоторого собственного движения $\boldsymbol{F}_1$ и симметрии $\boldsymbol{S}$ относительно произвольно выбранной (но фиксированной) прямой.

Доказательство. Для данного несобственного движения $\boldsymbol{F}_2$ и произвольной симметрии $\boldsymbol{S}$ преобразование $\boldsymbol{F}_1 = \boldsymbol{S} \circ \boldsymbol{F}_2$ является собственным движением (см. замечание 10.34). Поскольку $\boldsymbol{S} \circ \boldsymbol{S} = \mathrm{id}$, имеем

$$\boldsymbol{F}_2 = (\boldsymbol{S} \circ \boldsymbol{S}) \circ \boldsymbol{F}_2 = \boldsymbol{S} \circ (\boldsymbol{S} \circ \boldsymbol{F}_2) = \boldsymbol{S} \circ \boldsymbol{F}_1. \qquad \square$$

10.36. Вообще говоря, $\boldsymbol{F}_1 \circ \boldsymbol{S} \neq \boldsymbol{S} \circ \boldsymbol{F}_1$. Тем не менее несобственное движение $\boldsymbol{F}_2$ допускает также разложение $\boldsymbol{F}_2 = \boldsymbol{F}'_1 \circ \boldsymbol{S}$, где $\boldsymbol{S}$ — та же симметрия, что и выше, а $\boldsymbol{F}'_1$ — некоторое собственное движение (вообще говоря, отличное от $\boldsymbol{F}_1$). Докажите это утверждение самостоятельно.

10.37. Предложение. *Композиция $\boldsymbol{R} \circ \boldsymbol{S}$ осевой симметрии $\boldsymbol{S}$ и вращения $\boldsymbol{R}$ относительно некоторой точки O, лежащей на оси симметрии, является осевой симметрией относительно некоторой оси, проходящей через точку O.*

Доказательство. Указанная композиция является несобственным движением (см. п. 10.34), имеющим неподвижную точку O, так что по теореме Шаля представляет собой осевую симметрию. $\square$

10.38. Предложение.

1. *Разложение параллельного переноса:*
   (i) *композиция двух осевых симметрий с параллельными (не совпадающими) осями является параллельным переносом;*
   (ii) *обратно, каждый параллельный перенос может быть разложен в композицию двух осевых симметрий с параллельными осями.*
2. *Разложение вращения:*
   (i) *композиция двух осевых симметрий с пересекающимися осями является вращением;*
   (ii) *обратно, каждое вращение может быть разложено в композицию двух осевых симметрий.*

Доказательство. Докажем утверждения (i). Прежде всего заметим, что композиция двух осевых симметрий (несобственных движений) является собственным движением, т.е. может быть либо тождественным преобразованием (это будет в случае двукратного применения осевой симметрии; см. предложение 10.30(iii)), либо трансляцией, либо вращением. Композиция двух симметрий с параллельными несовпадающими осями не имеет неподвижных точек, а потому является трансляцией. Если же оси двух симметрий пересекаются, то точка пересечения осей является неподвижной точкой, так что композиция представляет собой вращение.

Докажем утверждение 1(ii). Введём прямоугольную декартову систему координат, ось Ox которой параллельна вектору трансляции $\boldsymbol{T}$. Тогда перенос задаётся формулами

$$\boldsymbol{T}\colon \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x+a \\ y \end{pmatrix}.$$

Рассмотрим симметрию относительно оси Oy, перпендикулярной направлению трансляции:

$$\boldsymbol{S}_1\colon \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} -x \\ y \end{pmatrix}.$$

Их композиция

$$\boldsymbol{T}\circ\boldsymbol{S}_1\colon \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} -x+a \\ y \end{pmatrix}$$

имеет неподвижную прямую $x = a/2$, т.е. является осевой симметрией $\boldsymbol{S}_2$ с осью, параллельной оси симметрии $\boldsymbol{S}_1$. Итак, $\boldsymbol{T}\circ\boldsymbol{S}_1 = \boldsymbol{S}_2$, откуда $\boldsymbol{T} = \boldsymbol{S}_2\circ\boldsymbol{S}_1$.

Докажем 2(ii). Заметим, что $\boldsymbol{R} = (\boldsymbol{R}\circ\boldsymbol{S})\circ\boldsymbol{S}$, а преобразование $\boldsymbol{R}\circ\boldsymbol{S}$ в силу предложения 10.37 является осевой симметрией. □

Из предложений 10.35 и 10.38 вытекает следующее утверждение.

10.39. Теорема. *Любое движение плоскости может быть представлено как композиция не более чем трёх осевых симметрий.*

4. Групповой подход в геометрии

A. Основные определения.

10.40. В этом разделе мы будем рассматривать множества различной природы, и для того, чтобы избежать путаницы, договоримся о следующей терминологии. Множества, являющиеся объектом геометрических исследований, будем называть *пространствами* (наиболее интересными для нас будут аффинные и евклидовы пространства), подмножества этих пространств — *геометрическими фигурами*, а слово *множество* будем использовать в остальных случаях.

10.41. В геометрии *преобразованием* пространства X принято называть любое взаимно однозначное отображение $f\colon X \to X$ этого пространства на себя. Множество всех преобразований пространства X будем обозначать $S(X)$.

10.42. На множестве $S(X)$ естественным образом определяется операция композиции преобразований: если $f, g \in S(X)$ (т.е. $f\colon X \to X$ и $g\colon X \to X$ — преобразование пространства X), то их композиция $g \circ f$ определяется правилом

$$g\circ f\colon X \to X, \quad x \mapsto g\bigl(f(x)\bigr).$$

Ясно, что $g \circ f \in S(X)$, поскольку композиция взаимно однозначных отображений сама является взаимно однозначным отображением. Тождественное отображение $\mathrm{id}\colon X \to X$ играет роль нейтрального элемента при выполнении композиции, т.е.

$$\forall f \in S(X): \quad f \circ \mathrm{id} = \mathrm{id} \circ f = f.$$

Будучи взаимно однозначным, каждое преобразование $f \in S(X)$ имеет обратное f^{-1}, которое также взаимно однозначно (см. предложение II.57, с. 313).

10.43. Таким образом, множество $S(X)$ всех преобразований пространства X образует группу относительно операции композиции преобразований, которая называется *группой преобразований пространства* X.

10.44. Пусть G — некоторая подгруппа группы преобразований $S(X)$ пространства X. Будем говорить, что геометрическая фигура $\Psi \subset X$ в пространстве X *эквивалентна* фигуре $\Phi \subset X$ относительно группы G, если существует такое преобразование $g \in G$, что $\Psi = g(\Phi)$; обозначение $\Psi \overset{G}{\sim} \Phi$. Легко проверить, опираясь на свойство взаимной однозначности каждого преобразования из G, что указанное отношение действительно является отношением эквивалентности (см. определение II.33, с. 308), т.е. обладает следующими свойствами:

(1) рефлексивность: $\Phi \overset{G}{\sim} \Phi$;

(2) симметричность: если $\Psi \overset{G}{\sim} \Phi$, то $\Phi \overset{G}{\sim} \Psi$;

(3) транзитивность: если $\Psi \overset{G}{\sim} \Phi$ и $\Xi \overset{G}{\sim} \Psi$, то $\Xi \overset{G}{\sim} \Phi$.

Б. Эрлангенская программа Ф. Клейна.

10.45. По существу, геометрия занимается изучением тех свойств геометрических фигур в пространстве X, которые остаются инвариантными (т.е. неизменными) при преобразованиях из некоторой группы G, являющейся подгруппой в группе $S(X)$. При различном выборе группы G получаются различные «геометрии».

Такой подход к геометрии был впервые предложен Ф. Клейном[1] в 1872 г. в его работе, известной в истории математики как «Эрлангенская программа».

10.46. Среди различных «геометрий» наиболее известными являются аффинная и евклидова, изучать которые читатель начал ещё в средней школе; каждая из них является вполне самостоятельным содержательным разделом математики. В аффинной геометрии в качестве группы G принимается группа аффинных преобразований (см. раздел 2), которые переводят плоскости в плоскости, сохраняют отношения коллинеарных направленных отрезков и т. п. Типичной теоремой аффинной геометрии является теорема о том, что медианы треугольника пересекаются в одной точке и делятся в ней в отношении 2 : 1. Евклидова геометрия изучает свойства фигур, инвариантные относительно группы движений (см. раздел 3).

В. Группы симметрий. Группы являются также наиболее удобным и адекватным средством описания симметрий геометрических фигур.

10.47. Пусть Φ — некоторая геометрическая фигура (т.е. множество точек) на евклидовой плоскости $\mathcal{E}_2$. Преобразование плоскости $f : \mathcal{E}_2 \to \mathcal{E}_2$ называется *симметрией* фигуры Φ, если $f(\Phi) = \Phi$. Например, отражение относительно прямой, содержащей диагональ ромба, является симметрией этого ромба.

[1] Ф. Клейн (F. Klein) — немецкий математик (1849–1925).

10.48. Легко видеть, что множество всех симметрий произвольной фигуры Φ представляет собой группу относительно операции композиции преобразований. Действительно, композиция отображений ассоциативна, нейтральным элементом является тождественное преобразование, а обратимость любой симметрии следует из её биективности. Группа симметрий фигуры Φ, обозначаемая символом $\operatorname{Sym}\Phi$, позволяет «измерить» степень симметричности фигуры: чем обширнее эта группа, тем фигура «более симметрична».

10.49. Примеры. 1. Отрезок AB прямой обладает, помимо тождественного преобразования, единственной симметрией: симметрией отражения относительно собственного центра.

2. Группа симметрий правильного треугольника $\triangle$ содержит кроме тождественного преобразования ещё пять преобразований плоскости: три отражения относительно высот треугольника и два поворота на углы $\pm 2\pi/3$ вокруг его центра.

ГЛАВА 11

Квадрики на евклидовой плоскости. Канонизация уравнения квадрики

В ГЛАВЕ 9 мы рассмотрели три важные линии на евклидовой плоскости — эллипс, гиперболу и параболу, и установили, что каждая из них задаётся в некоторой ортогональной системе координат уравнением вида $F(x, y) = 0$, где $F(x, y)$ — многочлен второй степени от двух переменных:

(а) для эллипса $F(x, y) = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - 1$,

(б) для гиперболы $F(x, y) = \dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} - 1$,

(в) для параболы $F(x, y) = y^2 - 2px$.

Теперь нам предстоит выяснить, какие ещё линии на плоскости можно задать уравнением вида $F(x, y) = 0$, где $F(x, y)$ — произвольный многочлен второй степени от двух переменных, и разработать алгоритм упрощения уравнения посредством специального выбора системы координат. Отметим, что в следующем семестре будет построена общая теория, позволяющая с лёгкостью выполнять такое упрощение в случае пространства любой размерности; сейчас мы получим необходимые результаты для квадрик на плоскости с помощью эвристических рассуждений.

1. Уравнения квадрики на плоскости

А. Различные формы записи уравнения квадрики.

11.1. Рассмотрим уравнение второй степени от двух переменных

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2b_1x + 2b_2y + c = 0; \qquad (11.1)$$

многочлен в левой части этого уравнения обозначим $F(x, y)$. Пусть на евклидовой плоскости $\mathcal{E}_2$ задана некоторая прямоугольная декартова система координат Oxy. Линия, точки которой (точнее, координаты точек) и только они являются решениями уравнения (11.1), называется *квадрикой* (на плоскости). При преобразовании системы координат уравнение квадрики, разумеется, изменяется; нашей окончательной целью является нахождение такой системы координат, в которой уравнение квадрики имеет наиболее простой вид.

11.2. Для сокращения записи введём матрицы

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \text{ где } a_{12} = a_{21},\ B = \begin{pmatrix} b_1 & b_2 \end{pmatrix},\ X = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Тогда уравнение (11.1) можно записать в виде

$$X^T AX + 2BX + c = 0. \tag{11.2}$$

11.3. Вид уравнения можно ещё упростить, используя однородные координаты (см. п. 10.5, с. 240). Введя матрицы

$$D = \left(\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ \hline b_1 & b_2 & c \end{array}\right) = \left\|\begin{array}{c|c} A & B^T \\ \hline B & c \end{array}\right\|, \quad Z = \left(\begin{array}{c} x \\ y \\ \hline 1 \end{array}\right) = \left\|\begin{array}{c} X \\ \hline 1 \end{array}\right\|, \tag{11.3}$$

можно записать уравнение квадрики в виде

$$Z^T DZ = 0. \tag{11.4}$$

11.4. *Конической поверхностью* (*конусом*) с вершиной O и образующей G называется множество всех прямых, проходящих через точку O и пересекающихся с линией G. Уравнение

$$W^T DW = 0, \quad \text{где} \quad W = \begin{pmatrix} x, y, z \end{pmatrix}^T, \tag{11.5}$$

определяет в трёхмерном евклидовом пространстве некоторую коническую поверхность. Действительно, начало координат удовлетворяет уравнению (11.5), а вместе с точкой, имеющей столбец координат W, уравнению удовлетворяют и все точки, имеющие координаты tW, где t — произвольное вещественное число. Таким образом, прямая, проходящая через начало координат и точку W, целиком лежит на поверхности, определяемой уравнением (11.5), а это и означает, что поверхность является конусом. В таком случае квадрику (11.4) можно рассматривать как сечение указанного конуса плоскостью $z = 1$.

Б. Ортогональные преобразования уравнения квадрики.

11.5. Рассмотрим на евклидовой плоскости $\mathcal{E}_2$ другую прямоугольную декартову систему координат $O'x'y'$, полученную из исходной Oxy поворотом осей координат и параллельным переносом начала координат; такое преобразование системы координат описывается формулой (10.12) (см. с. 240).

При указанном преобразовании координат $Z = PZ'$ уравнение квадрики $Z^T DZ = 0$ (см. (11.4)) изменится следующим образом:

$$Z^T DZ = (PZ')^T D(PZ') = Z'^T P^T DPZ' = Z'^T D'Z',$$

т.е. примет вид

$$Z'^T D'Z' = 0,$$

где

$$D' = P^T DP, \tag{11.6}$$

формально не отличающийся от (11.4), но с другими коэффициентами:

$$D' = \left\| \begin{array}{c|c} A' & B'^T \\ \hline B' & c' \end{array} \right\| = P^T DP =$$

$$= \left\| \begin{array}{c|c} R & X_0 \\ \hline O & 1 \end{array} \right\|^T \left\| \begin{array}{c|c} A & B^T \\ \hline B & c \end{array} \right\| \left\| \begin{array}{c|c} R & X_0 \\ \hline O & 1 \end{array} \right\| =$$

$$= \left\| \begin{array}{c|c} R^T AR & R^T(AX_0 + B^T) \\ \hline (X_0^T A + B)R & X_0^T AX_0 + 2BX_0 + c \end{array} \right\|,$$

где учтено, что $X_0^T B^T = BX_0$, поскольку это матрицы порядка 1, т.е. числа.

11.6. Запишем формулы преобразования матричных коэффициентов уравнения квадрики (11.2) без использования однородных координат:

$$\begin{gathered} A' = R^T AR, \quad B' = (X_0^T A + B)R, \\ c' = X_0^T AX_0 + 2BX_0 + c. \end{gathered} \tag{11.7}$$

Обратите внимание, что матрица A изменяется только при повороте, а свободный член c уравнения — только при переносе начала координат. Матрица B, содержащая коэффициенты линейных слагаемых уравнения квадрики, меняется и при повороте, и при переносе.

B. Ортогональные инварианты уравнения квадрики.

11.7. Теорема. *При ортогональных преобразованиях системы координат величины*

$$S = \operatorname{tr} A, \quad \delta = \det A, \quad \Delta = \det D$$

не изменяются. Эти величины называют ортогональными инвариантами уравнения квадрики.

Доказательство. Используя формулу $\mathrm{tr}(UV) = \mathrm{tr}(VU)$, где U, V — произвольные матрицы, для которых определены оба произведения UV и VU (см. предложение 3.21(5), с. 53), имеем

$$\mathrm{tr}\, A' = \mathrm{tr}\left(R^T A R\right) = \mathrm{tr}\left((R^T A)R\right) = \\ = \mathrm{tr}\left(R(R^T A)\right) = \mathrm{tr}\left((R^T R)A\right) = \mathrm{tr}\, A.$$

По теореме об определителе произведения матриц

$$\det A' = \det(R^T A R) = \det R^T \cdot \det A \cdot \det R = \det A,$$

поскольку $\det R = \det R^T = 1$. По той же теореме

$$\det D' = \det(P^T D P) = \det P^T \cdot \det D \cdot \det P = \det D,$$

поскольку $\det P = \det P^T = 1$, в чём легко убедиться, разложив определитель матрицы P (см. (10.11)) по последней строке. □

2. Упрощение уравнения квадрики: уничтожение слагаемого $2a_{12}xy$ при помощи поворота

11.8. Попытаемся выбрать угол поворота координатных осей таким образом, чтобы слагаемое вида $2a_{12}xy$ в преобразованном уравнении исчезло, т.е. чтобы матрица A' уравнения, отнесённого к новой системе координат, стала диагональной:

$$A' = R^T A R = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

11.9. Отметим, что при наличии в уравнении квадрики слагаемого $2a_{12}xy$ необходимо выполняется условие $\lambda_1 \neq \lambda_2$. Действительно, если предположить, что $\lambda_1 = \lambda_2$, то $A' = \lambda_1 \mathbb{1}$, так что

$$A = (R^T)^{-1} A' R^{-1} = R(\lambda_1 \mathbb{1})R^{-1} = \lambda_1 R R^{-1} = \lambda_1 \mathbb{1},$$

т.е. матрица A также диагональна, т.е. $a_{12} = a_{21} = 0$, причём в любой системе координат.

11.10. Числа λ_1 и λ_2 могут быть найдены с помощью ортогональных инвариантов уравнения квадрики: поскольку

$$\mathrm{tr}\, A' = \lambda_1 + \lambda_2 = \mathrm{tr}\, A = S, \quad \det A' = \lambda_1 \cdot \lambda_2 = \det A = \delta,$$

согласно теореме Виета числа λ_1 и λ_2 являются корнями квадратного уравнения

$$\lambda^2 - S\lambda + \delta = 0.$$

11.11. Умножая обе части соотношения $A' = R^T AR$ слева на матрицу R, запишем его в виде $AR = RA'$. Обозначим столбцы координат векторов $\boldsymbol{i}', \boldsymbol{j}'$ нового (повернутого) базиса относительно старого базиса через R_1, R_2; эти столбцы являются столбцами матрицы поворота R: $R = \big\|R_1,\ R_2\big\|$. Соотношение $AR = RA'$ даёт

$$AR = RA' \ \Leftrightarrow$$
$$\Leftrightarrow\ A\big\|R_1,\ R_2\big\| = \big\|R_1,\ R_2\big\| \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \big\|\lambda_1 R_1,\ \lambda_2 R_2\big\|,$$

так что

$$AR_1 = \lambda_1 R_1, \quad AR_2 = \lambda_2 R_2.$$

Эти уравнения можно переписать в виде

$$(A - \lambda_1 \mathbb{1})R_1 = 0, \quad (A - \lambda_2 \mathbb{1})R_2 = 0,$$

т.е. столбцы R_1 и R_2 представляют собой решения однородных систем линейных уравнений. Нетривиальные решения у однородной системы имеются лишь в том случае, когда определитель основной матрицы системы равен нулю, так что числа λ_1 и λ_2 должны являться корнями уравнения

$$\det(A - \lambda \mathbb{1}) = 0,$$

называемого *характеристическим уравнением* рассматриваемой квадрики. Вычисляя определитель, получаем:

$$\det(A - \lambda \mathbb{1}) = \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} =$$
$$= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} = \lambda^2 - S\lambda + \delta,$$

что в точности совпадает с найденным выше многочленом для определения λ_1 и λ_2. Этот многочлен называется *характеристическим многочленом* матрицы A

11.12. Корни λ_1 и λ_2 характеристического многочлена матрицы A называют *собственными значениями*, а любые ненулевые решения системы уравнений $(A - \lambda \mathbb{1})X = O$, существующие лишь при $\lambda = \lambda_1$ или $\lambda = \lambda_2$ — *собственными векторами* этой матрицы, отвечающими соответствующему собственному значению.

11.13. Поскольку коэффициенты характеристического многочлена являются ортогональными инвариантами уравнения квадрики (см. теорему 11.7), то и сам характеристический многочлен инвариантен, т.е. не зависит от выбора ортогональной системы координат.

11.14. Дискриминант характеристического многочлена неотрицателен:

$$S^2 - 4\delta = (a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21}) = \\ = (a_{11} - a_{22})^2 + 4a_{12}^2 \geqslant 0,$$

так что он всегда имеет вещественные корни.

11.15. Найдя корни λ_1 и λ_2 характеристического многочлена и решив однородные системы уравнений

$$(A - \lambda_1 \mathbb{1})R_1 = 0, \quad (A - \lambda_2 \mathbb{1})R_2 = 0,$$

мы получим столбцы координат векторов $\boldsymbol{i}'$ и $\boldsymbol{j}'$ новой (повёрнутой) системы координат. Решение каждой из этих систем не единственно, но определено с точностью до произвольного множителя; будем выбирать эти множители (их называют нормировочными) так, чтобы векторы $\boldsymbol{i}'$ и $\boldsymbol{j}'$ имели единичную длину и образовывали базис, одноимённый с исходным базисом $\boldsymbol{i}, \boldsymbol{j}$.

11.16. Нетрудно найти выражение для угла поворота α новых координатных осей относительно старых:

$$AR_2 = \lambda_2 R_2 \;\Leftrightarrow\; \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix} = \begin{pmatrix} -\lambda_2 \sin\alpha \\ \lambda_2 \cos\alpha \end{pmatrix} \;\Leftrightarrow \\ \Leftrightarrow\; \begin{cases} -a_{11}\sin\alpha + a_{12}\cos\alpha = -\lambda_2 \sin\alpha, \\ -a_{21}\sin\alpha + a_{22}\cos\alpha = \lambda_2 \cos\alpha. \end{cases}$$

Умножая первое уравнение на $\cos\alpha$, а второе — на $\sin\alpha$ и складывая полученные соотношения, получаем

$$-a_{11}\sin\alpha\cos\alpha + a_{12}\cos^2\alpha - a_{21}\sin^2\alpha + a_{22}\cos\alpha\sin\alpha = 0,$$

откуда

$$\operatorname{ctg} 2\alpha = \frac{a_{11} - a_{22}}{2a_{12}}. \tag{11.8}$$

Итак, доказана следующая теорема.

11.17. Теорема. *Для уничтожения слагаемого* $2a_{12}xy$ *в уравнении квадрики* (11.1) *нужно перейти к новой системе координат, оси которой повёрнуты относительно осей исходной системы на угол* α, *определяемый соотношением* (11.8). *В этом случае столбцы координат* R_1, R_2 *ортов* $\boldsymbol{i}', \boldsymbol{j}'$ *новой системы координат являются нормированными решениями однородных систем линейных уравнений*

$$(A - \lambda_k \mathbb{1})R_k = 0, \quad k = 1, 2,$$

где λ_1, λ_2 — *корни характеристического уравнения*

$$\lambda^2 - S\lambda + \delta = 0 \;\Leftrightarrow\; \det(A - \lambda\mathbb{1}) = 0. \tag{11.9}$$

3. Упрощение уравнения квадрики: уничтожение линейных слагаемых $2b_1x + 2b_2y$ при помощи переноса начала координат

11.18. Попытаемся теперь выбрать вектор переноса X_0 начала системы координат так, чтобы уравнение квадрики в новых координатах не содержало бы линейных слагаемых, т.е. $B' = O$ (O — нулевая матрица).

11.19. Поскольку коэффициенты линейных членов уравнения квадрики преобразуются по формуле $B' = (X_0^T A + B)R$, для определения X_0 получаем уравнение

$$(X_0^T A + B)R = O \;\Leftrightarrow\; X_0^T A + B = O \;\Leftrightarrow\; (X_0^T A + B)^T = O$$

или, окончательно,

$$AX_0 = -B^T \tag{11.10}$$

(здесь использованы факты симметричности матрицы A и обратимости матрицы R). Полученное уравнение не всегда разрешимо, так что уничтожение линейных слагаемых в уравнении квадрики возможно не во всех случаях.

A. Случай $\delta = \det A \neq 0$: центральные квадрики.

11.20. Если $\det A \neq 0$, то уравнение $AX_0 = -B^T$ имеет единственное решение $X_0 = -A^{-1}B^T$, и матрица преобразования координат (10.11) определяется однозначно:

$$P = \left\| \begin{array}{c|c} R & -A^{-1}B^T \\ \hline O & 1 \end{array} \right\|,$$

где R — матрица поворота координатных осей, найденная ранее. После такого преобразования координат в уравнении квадрики исчезают слагаемые вида $2a_{12}xy$ и $2b_1x + 2b_2y$, т.е. оно приводится к так называемому «полуканоническому» виду

$$\lambda_1 x'^2 + \lambda_2 y'^2 + c' = 0, \tag{11.11}$$

где λ_1, λ_2 — корни характеристического уравнения (11.9), а c' — свободный член преобразованного уравнения, определяемый по формуле (11.7), в которой X_0 — решение уравнения (11.10).

11.21. Матрица D' коэффициентов уравнения квадрики в полуканонической системе координат имеет вид

$$D' = \left(\begin{array}{cc|c} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ \hline 0 & 0 & c' \end{array} \right). \tag{11.12}$$

Коэффициент c' можно найти по формуле (11.7), однако если принять во внимание соотношение $AX_0 = -B^T$, вычисление можно упростить:

$$c' = X_0^T AX_0 + 2BX_0 + c = \underbrace{X_0^T(-B^T)}_{=-BX_0} + 2BX_0 + c = BX_0 + c. \quad (11.13)$$

Кроме того, поскольку определитель матрицы (11.12) равен

$$\det D' = \lambda_1\lambda_2 c' = \Delta,$$

легко получаем выражение для c' через ортогональные инварианты уравнения:

$$c' = \frac{\Delta}{\delta}. \quad (11.14)$$

11.22. Начало канонической системы координат $O'x'y'$ является центром симметрии линии, определяемой уравнением (11.11), а координатные оси $O'x'$ и $O'y'$ — её осями симметрии. Действительно, если точка с координатами (x', y') является решением уравнения (11.11), то решениями являются также и точки $(-x', y')$, $(x', -y')$, $(-x', -y')$. По этой причине квадрики данного типа называют *центральными* (даже в случае, когда уравнение квадрики не имеет ни одного вещественного решения).

11.23. С помощью дальнейших алгебраических преобразований полученное полуканоническое уравнение можно привести к каноническому виду.

В зависимости от знака $\det A$ выделяются следующие два типа квадрик.

11.24. ***Эллиптический тип:*** $\delta = \det A > 0$**.** В этом случае числа λ_1 и λ_2 одного знака, и в зависимости от знака c' возможны следующие ситуации.

11.25. Пусть c' и $\lambda_{1,2}$ имеют разные знаки; тогда

$$S \cdot \Delta < 0$$

(напомним, что $c' = \Delta/\delta$, а $\delta > 0$). Тогда из (11.11) получаем

$$\frac{x'^2}{-c'/\lambda_1} + \frac{y'^2}{-c'/\lambda_2} = 1;$$

знаменатели обеих дробей положительны, так что имеем уравнение эллипса с полуосями $a = \sqrt{-c'/\lambda_1}$ и $b = \sqrt{-c'/\lambda_2}$. Если $|\lambda_1| < |\lambda_2|$, то $a > b$, и получится каноническое уравнение эллипса. Кроме того, полуоси эллипса можно выразить через ортогональные инварианты уравнения:

$$a = \sqrt{-\frac{\Delta}{\delta\lambda_1}}, \quad b = \sqrt{-\frac{\Delta}{\delta\lambda_2}}, \quad \lambda_1 < \lambda_2.$$

11.26. Пусть c' и $\lambda_{1,2}$ имеют один и тот же знак, т.е.

$$S \cdot \Delta > 0.$$

Тогда из (11.11) получаем

$$\frac{x'^2}{c'/\lambda_1} + \frac{y'^2}{c'/\lambda_2} = -1;$$

здесь знаменатели обеих дробей также положительны, так что получаем каноническое уравнение

$$\frac{x'^2}{a^2} + \frac{y'^2}{b^2} = -1, \quad \text{где} \quad a^2 = \frac{c'}{\lambda_1} > 0,\ b^2 = \frac{c'}{\lambda_2} > 0,$$

которое не имеет вещественных решений; благодаря сходству с уравнением эллипса оно называется *уравнением мнимого эллипса.* «Полуоси» мнимого эллипса также можно выразить через ортогональные инварианты уравнения:

$$a = \sqrt{\frac{\Delta}{\delta\lambda_1}}, \quad b = \sqrt{\frac{\Delta}{\delta\lambda_2}}, \quad \lambda_1 < \lambda_2. \tag{11.15}$$

11.27. Наконец, в случае $c' = 0$, т.е. $\Delta = 0$, из (11.11) получаем уравнение

$$\lambda_1 x'^2 + \lambda_2 y'^2 = 0 \ \Leftrightarrow \ a^2 x'^2 + b^2 y'^2 = 0,$$

где $a = \sqrt{|\lambda_1|}$, $b = \sqrt{|\lambda_2|}$. Это уравнение имеет единственное вещественное решение $x' = y' = 0$, а геометрическая фигура, определяемая уравнением, состоит из единственной точки — начала координат канонической системы. Однако если рассматривать это уравнение над множеством комплексных чисел, то его левую часть можно разложить на множители:

$$(ax' + iby')(ax' - iby') = 0 \ \Leftrightarrow \ \left[\begin{array}{l} (ax' + iby') = 0, \\ (ax' - iby') = 0, \end{array}\right.$$

что объясняет наименование рассматриваемой квадрики: *пара мнимых прямых, пересекающихся в вещественной точке.*

11.28. ***Гиперболический тип:*** $\boldsymbol{\delta = \det A < 0}$**.** В этом случае числа λ_1 и λ_2 имеют разные знаки, и возможны следующие ситуации.

11.29. Если $c' \neq 0$ (т.е. $\Delta \neq 0$), то из полуканонического уравнения (11.11) получаем каноническое уравнение гиперболы
$$\frac{x'^2}{a^2} - \frac{y'^2}{b^2} = 1,$$
где
$$a^2 = -\frac{c'}{\lambda_1} = -\frac{\Delta}{\delta\lambda_1} > 0, \quad b^2 = \frac{c'}{\lambda_2} = \frac{\Delta}{\delta\lambda_2} > 0.$$
Ясно, что корни характеристического уравнения квадрики нужно нумеровать так, чтобы числа λ_1 и c' имели противоположные знаки, а числа λ_2 и c' — одинаковые, т.е. чтобы числа λ_1 и Δ имели одинаковые знаки, а λ_2 и Δ — противоположные.

11.30. Если $c' = 0$ (т.е. $\Delta = 0$), то полуканоническое уравнение (11.11) приводится к виду
$$a^2x'^2 - b^2y'^2 = 0, \quad \text{где} \quad a^2 = |\lambda_1|,\ b^2 = |\lambda_2|,$$
или, эквивалентно,
$$(ax' + by')(ax' - by') = 0 \ \Leftrightarrow\ \left[\begin{aligned} &(ax' + by') = 0,\\ &(ax' - by') = 0.\end{aligned}\right.$$
Последняя совокупность уравнений определяет пару пересекающихся прямых (очевидно, точкой пересечения является начало канонической системы координат).

11.31. Отметим, что полуканоническое уравнение получается из исходного уравнения при помощи ортогонального преобразования координат, и величины $S = \operatorname{tr} A$, $\delta = \det A$, $\Delta = \det D$ являются инвариантными относительно этих преобразований. Каноническое же уравнение получается из полуканонического при помощи алгебраических преобразований (например, деления обеих частей уравнения на свободный член уравнения и т. п.); относительно таких преобразований величины S, δ и Δ инвариантами не являются.

Б. Случай $\delta = \det A = 0$: нецентральные квадрики.

11.32. Условие $\det A = 0$ эквивалентно тому, что один из корней характеристического уравнения (11.9) равен нулю; будем считать, что $\lambda_1 = 0$; в этом случае $\lambda_2 = S = \operatorname{tr} A$, а матрица D' коэффициентов уравнения квадрики после поворота осей координат имеет следующий вид (обратите внимание, что свободный член c при повороте не изменяется, см. формулы (11.7), с. 258):
$$D' = \left(\begin{array}{cc|c} 0 & 0 & b_1' \\ 0 & S & b_2' \\ \hline b_1' & b_2' & c \end{array}\right), \quad \Delta = \det D' = -b_1'^2 S. \tag{11.16}$$

Если $\det A = 0$, то система уравнений (11.10) может либо оказаться несовместной, либо иметь бесконечно много решений.

11.33. Отметим следующий важный факт: любые переносы начала координат, выполняемые *после* поворота, приводящего матрицу коэффициентов квадрики к виду (11.16), не изменяют коэффициента b'_1. Действительно, такие переносы сохраняют вид матрицы (11.16), а указанный коэффициент может быть выражен через инварианты, $b'_1 = \sqrt{-\Delta/S}$, и потому сам является инвариантом, но только относительно сдвигов.

Таким образом, вектор переноса начала координат нужно выбирать так, чтобы добиться, по возможности, исчезновения в матрице (11.16) элементов b'_2 и c.

11.34. ***Невырожденный параболический тип:*** $\Delta \neq 0$**.** Согласно теореме Кронекера—Капелли система (11.10) несовместна, если ранг матрицы A меньше ранга матрицы $\left\| A \mid -B^T \right\|$. Поскольку $\operatorname{rk} A = \operatorname{rk} A'$ и $\operatorname{rk} \left\| A \mid -B^T \right\| = \operatorname{rk} \left\| A' \mid -R^T B^T \right\|$, ибо R — невырожденная матрица, это эквивалентно неравенствам

$$\operatorname{rk} A' < \operatorname{rk} \left\| A' \mid -R^T B^T \right\| \;\Leftrightarrow\; \operatorname{rk} \begin{pmatrix} 0 & 0 \\ 0 & S \end{pmatrix} < \operatorname{rk} \begin{pmatrix} 0 & 0 & -b'_1 \\ 0 & S & -b'_2 \end{pmatrix},$$

что имеет место при $b'_1 \neq 0$. В этом случае после поворота координатных осей уравнение квадрики примет вид

$$Sy'^2 + 2b'_1 x' + 2b'_2 y' + c = 0,$$

где $b'_1 \neq 0$, и его можно привести к каноническому уравнению параболы, выделяя полные квадраты:

$$S\left(y'^2 + 2\frac{b'_2}{S}y' + \frac{b'^2_2}{S^2}\right) - \frac{b'^2_2}{S} + 2b'_1\left(x' + \frac{c}{2b'_1}\right) = 0,$$

так что

$$S\left(y' + \frac{b'_2}{S}\right)^2 + 2b'_1\left(x' + \frac{c}{2b'_1} - \frac{b'^2_2}{2b'_1 S}\right) = 0.$$

Вводя новые переменные

$$\begin{cases} x'' = \pm\left(x' + \dfrac{c}{2b'_1} - \dfrac{b'^2_2}{2b'_1 S}\right), \\[2ex] y'' = y' + \dfrac{b'_2}{S}, \end{cases} \tag{11.17}$$

получим полуканоническое уравнение

$$Sy''^2 + 2b'_1 x'' = 0, \tag{11.18}$$

а затем, выбирая надлежащим образом знак + или − в первой из формул (11.17), и каноническое уравнение параболы

$$y''^2 = 2px'', \tag{11.19}$$

где введено обозначение $p = \left|\dfrac{b_1'}{S}\right|$. Матрица коэффициентов полуканонического уравнения в системе координат $O''x''y''$ имеет вид

$$D'' = \left(\begin{array}{cc|c} 0 & 0 & b_1' \\ 0 & S & 0 \\ \hline b_1' & 0 & 0 \end{array}\right), \quad \det D'' = -b_1'^2 S. \tag{11.20}$$

Ясно, что итоговое преобразование системы координат сводится к последовательно выполненным повороту (переход $Oxy \to Ox'y'$) и переносу начала (переход $Ox'y' \to O''x''y''$), описываемому формулами (11.17).

11.35. Из формул (11.17) следует, что координаты вершины параболы, т.е. начала канонической системы координат $O''x''y''$, отнесённые к повёрнутой системе $Ox'y'$, равны

$$\begin{cases} x_0' = \mp\left(\dfrac{c}{2b_1'} - \dfrac{b_2'^2}{2b_1'S}\right), \\ y_0' = -\dfrac{b_2'}{S}. \end{cases}$$

Для нахождения координат вершины в исходной системе Oxy нужно воспользоваться формулами (10.3):

$$X_0 = RX_0'.$$

11.36. Величина фокального параметра параболы в каноническом уравнении (11.19) также может быть выражена через инварианты:

$$p = \left|\frac{b_1'}{S}\right| = \left|\frac{\sqrt{-\Delta/S}}{S}\right| = \sqrt{-\frac{\Delta}{S^3}}.$$

11.37. ***Вырожденный параболический тип:* $\Delta = 0$.** В случае, когда ранг матрицы A равен рангу матрицы $\left\|A \mid -B^T\right\|$, система (11.10) имеет бесконечно много решений. Выбрав любое из них, получаем следующую матрицу коэффициентов уравнения квадрики в преобразованной системе координат

$$D' = \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & S & 0 \\ \hline 0 & 0 & c' \end{array}\right), \tag{11.21}$$

т.е. само уравнение принимает полуканонический вид

$$Sy'^2 + c' = 0. \tag{11.22}$$

В зависимости от знаков чисел S и c' возможны следующие случаи.

11.38. Числа S и c' имеют разные знаки; тогда уравнение (11.22) приводится к каноническому виду

$$y''^2 - a^2 = 0, \quad a^2 = \left|\frac{c'}{S}\right|;$$

это уравнение определяет *две параллельные прямые* $y'' = a$ и $y'' = -a$.

11.39. В случае, когда знаки чисел S и c' совпадают, получаем

$$y''^2 + a^2 = 0, \quad a^2 = \frac{c'}{S}.$$

Это уравнение не имеет вещественных решений, т.е. определяемая им геометрическая фигура есть пустое множество. Если допускать комплексные решения, то по аналогии с предыдущим случаем можно назвать соответствующую квадрику *парой мнимых параллельных прямых.*

11.40. Наконец, если $c' = 0$, уравнение имеет вид $y''^2 = 0$ и называется *уравнением пары совпадающих прямых.*

B. Полуинвариант K. Мы установили, что тип и каноническое уравнение квадрики может быть получено с помощью инвариантов во всех случаях, кроме вырожденного параболического случая, когда $\delta = \Delta = 0$. Вырожденные параболические квадрики можно различать (и составлять их канонические уравнения) при помощи дополнительной величины, являющейся ортогональным инвариантом уравнения квадрики в указанном случае.

11.41. Рассмотрим определитель[1]

$$f_D(\lambda) = \det(D - \lambda\mathbb{1}) = \det\left\|\begin{array}{c|c} A - \lambda\mathbb{1} & B^T \\ \hline B & c - \lambda \end{array}\right\|.$$

После раскрытия этого определителя получится кубический многочлен от переменной λ, называемый *характеристическим многочленом* матрицы D.

11.42. Докажем, что характеристический многочлен инвариантен относительно поворотов, т.е. ортогональных преобразований, описываемых матрицей

$$P = \left\|\begin{array}{c|c} R & O \\ \hline O & 1 \end{array}\right\|.$$

[1]Обратите внимание, что в этой формуле $\mathbb{1}$ в одном случае обозначает единичную матрицу порядка 3, а в другом — порядка 2.

Матрица P ортогональна, т.е. $P^{-1} = P^T$ (проверьте!), поэтому, принимая во внимание формулу (11.6) преобразования матрицы D коэффициентов уравнения квадрики, имеем

$$\begin{aligned} f_{D'}(\lambda) &= \det(D' - \lambda\mathbb{1}) = \det\left(P^T DP - \lambda\mathbb{1}\right) = \\ &= \det\left(P^{-1}DP - \lambda P^{-1}P\right) = \det\left(P^{-1}(D - \lambda\mathbb{1})P\right) = \\ &= \det P^{-1}\cdot\det(D-\lambda\mathbb{1})\cdot\det P = \det(D-\lambda\mathbb{1}) = f_D(\lambda). \end{aligned}$$

11.43. Получим развёрнутое выражение для многочлена $f_D(\lambda)$:

$$f_D(\lambda) = \begin{vmatrix} a_{11}-\lambda & a_{12} & b_1 \\ a_{21} & a_{22}-\lambda & b_2 \\ b_1 & b_2 & c-\lambda \end{vmatrix} =$$

воспользуемся свойством линейности определителя, представив его первый столбец в виде $(a_{11}, a_{21}, b_1)^T - (\lambda, 0, 0)^T$ (в дальнейшем эта процедура будет несколько раз повторена):

$$= \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22}-\lambda & b_2 \\ b_1 & b_2 & c-\lambda \end{vmatrix} - \begin{vmatrix} \lambda & a_{12} & b_1 \\ 0 & a_{22}-\lambda & b_2 \\ 0 & b_2 & c-\lambda \end{vmatrix} =$$

в первом определителе ещё раз используем линейность, второй раскладываем по первому столбцу:

$$= \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ b_1 & b_2 & c-\lambda \end{vmatrix} - \begin{vmatrix} a_{11} & 0 & b_1 \\ a_{21} & \lambda & b_2 \\ b_1 & 0 & c-\lambda \end{vmatrix} - \lambda\begin{vmatrix} a_{22}-\lambda & b_2 \\ b_2 & c-\lambda \end{vmatrix} =$$

в первом и третьем определителях снова применяем свойство линейности, второй раскладываем по второму столбцу:

$$\begin{aligned} &= \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ b_1 & b_2 & c \end{vmatrix} - \begin{vmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ b_1 & b_2 & \lambda \end{vmatrix} - \lambda\begin{vmatrix} a_{11} & b_1 \\ b_1 & c-\lambda \end{vmatrix} - \\ &\quad -\lambda\left(\begin{vmatrix} a_{22} & b_2 \\ b_2 & c-\lambda \end{vmatrix} - \begin{vmatrix} \lambda & b_2 \\ 0 & c-\lambda \end{vmatrix}\right) = \\ &= \Delta - \lambda\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} - \lambda\left(\begin{vmatrix} a_{11} & b_1 \\ b_1 & c \end{vmatrix} - \begin{vmatrix} a_{11} & 0 \\ b_1 & \lambda \end{vmatrix}\right) - \\ &\quad -\lambda\left(\begin{vmatrix} a_{22} & b_2 \\ b_2 & c \end{vmatrix} - \begin{vmatrix} a_{22} & 0 \\ b_2 & \lambda \end{vmatrix} - \begin{vmatrix} \lambda & b_2 \\ 0 & c \end{vmatrix} + \begin{vmatrix} \lambda & 0 \\ 0 & \lambda \end{vmatrix}\right) = \\ &= \Delta - \Bigg(\delta + \underbrace{\begin{vmatrix} a_{11} & b_1 \\ b_1 & c \end{vmatrix} + \begin{vmatrix} a_{22} & b_2 \\ b_2 & c \end{vmatrix}}_{=K}\Bigg)\lambda + S\lambda^2 - \lambda^3. \end{aligned}$$

Так как S, δ, Δ и $f_D(\lambda)$ инвариантны относительно поворотов (а первые три величины также и относительно сдвигов), заключаем, что

$$K = \begin{vmatrix} a_{11} & b_1 \\ b_1 & c \end{vmatrix} + \begin{vmatrix} a_{22} & b_2 \\ b_2 & c \end{vmatrix}$$

— инвариант относительно поворотов. Напомним, что имеется ещё одна величина, инвариантная относительно поворотов: свободный член c уравнения (см. (11.7)); c и K называются *полуинвариантами* (*семиинвариантами*).

11.44. Докажем, что в случае $\delta = \Delta = 0$ величина K является инвариантом также и относительно сдвигов.

Поскольку K — инвариант относительно поворотов, можем считать, что в уравнении квадрики мы с помощью поворота уже добились того, что $a_{12} = a_{21} = 0$, т.е. матрица коэффициентов имеет вид

$$D = \left(\begin{array}{cc|c} 0 & 0 & b_1 \\ 0 & S & b_2 \\ \hline b_1 & b_2 & c \end{array}\right) \quad \Rightarrow \quad \Delta = -b_1^2 S.$$

Поскольку по условию $\Delta = 0$, отсюда получаем $b_1 = 0$, так что

$$D = \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & S & b_2 \\ \hline 0 & b_2 & c \end{array}\right).$$

После сдвига начала координат, описываемого формулами

$$x = x' + x_0, \quad y = y' + y_0,$$

уравнение квадрики

$$Sy^2 + 2b_2 y + c = 0$$

принимает вид

$$S(y' + y_0) + 2b_2(y' + y_0) + c = 0 \;\Leftrightarrow\; Sy'^2 + 2(Sy_0 + b_2) + (Sy_0^2 + 2b_2 y_0 + c) = 0.$$

Этому уравнению соответствует матрица коэффициентов

$$D' = \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & S & Sy_0 + b_2 \\ \hline 0 & Sy_0 + b_2 & Sy_0^2 + 2b_2 y_0 + c \end{array}\right).$$

Сравним величины K и K' для матриц D и D':

$$K' = \begin{vmatrix} S & Sy_0 + b_2 \\ Sy_0 + b_2 & Sy_0^2 + 2b_2 y_0 + c \end{vmatrix} =$$
$$= S\bigl(Sy_0^2 + 2b_2 y_0 + c\bigr) - \bigl(Sy_0 + b_2\bigr)^2 = Sc - b_2^2 = K,$$

так что K — инвариант относительно сдвигов в случае $\delta = \Delta = 0$.

11.45. Завершим анализ вырожденного параболического случая. Уравнение квадрики допускает уничтожение слагаемого $b_2 y$ с помощью переноса, т.е. приводится к полуканоническому виду

$$Sy'^2 + c' = 0, \quad D' = \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & S & 0 \\ \hline 0 & 0 & c' \end{array}\right), \quad K = Sc',$$

Классификация квадрик на евклидовой плоскости

<table>
<tr><th></th><th>Невырожденные линии: $\Delta \neq 0$</th><th>Вырожденные линии: $\Delta = 0$</th></tr>
<tr><td rowspan="2">Эллиптический
тип: $\delta > 0$</td><td>$S\Delta < 0$: эллипс $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$
$a = \sqrt{-\dfrac{\Delta}{\delta\lambda_1}}$, $b = \sqrt{-\dfrac{\Delta}{\delta\lambda_2}}$, $|\lambda_1| < |\lambda_2|$</td><td rowspan="2">Пара мнимых пересекающихся прямых
$a^2x^2 + b^2y^2 = 0$
$a = \sqrt{|\lambda_1|}$, $b = \sqrt{|\lambda_2|}$</td></tr>
<tr><td>$S\Delta > 0$: мнимый эллипс $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = -1$
$a = \sqrt{-\dfrac{\Delta}{\delta\lambda_1}}$, $b = \sqrt{-\dfrac{\Delta}{\delta\lambda_2}}$</td></tr>
<tr><td>Гиперболический
тип: $\delta < 0$</td><td>Гипербола $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$
$a = \sqrt{-\dfrac{\Delta}{\delta\lambda_1}}$, $b = \sqrt{\dfrac{\Delta}{\delta\lambda_2}}$</td><td>Пара вещественных пересекающихся прямых
$a^2x^2 - b^2y^2 = 0$
$a = \sqrt{|\lambda_1|}$, $b = \sqrt{|\lambda_2|}$</td></tr>
<tr><td rowspan="3">Параболический
тип $\delta = 0$</td><td rowspan="3">Парабола $y^2 = 2px$
$p = \sqrt{-\dfrac{\Delta}{S^3}}$</td><td>$K < 0$: пара веществ. паралл. прямых $y^2 = a^2$, $a = \sqrt{-\dfrac{K}{S^2}}$</td></tr>
<tr><td>$K > 0$: пара мнимых паралл. прямых $y^2 = -a^2$, $a = \sqrt{\dfrac{K}{S^2}}$</td></tr>
<tr><td>$K = 0$: пара совпад. прямых $y^2 = 0$</td></tr>
</table>

откуда $c' = K/S$, а уравнение квадрики принимает канонический вид вид

$$y'^2 + \frac{K}{S^2} = 0 \;\Leftrightarrow\; \begin{cases} y'^2 = a^2, & \text{где } a^2 = -\dfrac{K}{S^2}, \text{ если } K < 0, \\ y'^2 = -a^2, & \text{где } a^2 = \dfrac{K}{S^2}, \text{ если } K > 0, \\ y'^2 = 0, & \text{если } K = 0. \end{cases}$$

Таким образом, доказана следующая теорема.

11.46. Теорема (классификация квадрик на евклидовой плоскости)**.** *При помощи собственных ортогональных преобразований координат (т.е. поворота осей координат и переноса начала координат) уравнение квадрики* (11.1) *может быть приведено к одному из девяти канонических типов, перечисленных в таблице на с.* 271.

ГЛАВА 12

Квадрики в евклидовом пространстве

ЗАКЛЮЧИТЕЛЬНАЯ глава посвящена краткому знакомству с квадриками в трёхмерном евклидовом пространстве — поверхностями второго порядка, многие из которых часто встречаются в прикладных задачах. Мы ограничимся перечислением типов таких поверхностей и кратким обсуждением некоторых их геометрических свойств.

1. Классификация квадрик. Канонические уравнения

Квадрика в трёхмерном евклидовом пространстве, т.е. множество точек, координаты которых удовлетворяют уравнению

$$X^T AX + 2BX + c = 0, \tag{12.1}$$

где

$$X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad B = \begin{pmatrix} b_1; & b_2; & b_3 \end{pmatrix},$$

представляет собой некоторую поверхность.

Аналогично[1] тому, как это было сделано для линий второго порядка (см. гл. 11), каждое уравнение вида (12.1) можно привести к одному из семнадцати канонических типов, перечисленных ниже (см. рис. 12.1).

А. Эллиптический тип.

12.1. ***Эллипсоид*** (см. рис. 12.1(a)):

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1, \quad a > 0,\ b > 0,\ c > 0.$$

[1] В процессе канонизации уравнения используются некоторые факты линейной алгебры, которые будут изложены в следующем семестре.

12.2. ***Мнимый эллипсоид:***

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = -1, \quad a > 0,\ b > 0,\ c > 0.$$

Это уравнение не имеет ни одного вещественного решения.

12.3. ***Мнимый конус:***

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 0, \quad a > 0,\ b > 0,\ c > 0.$$

Поверхность имеет единственную вещественную точку $O(0, 0, 0)$.

Б. Гиперболический тип.

12.4. ***Однополостный гиперболоид*** (см. рис. 12.1(б)):

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1, \quad a > 0,\ b > 0,\ c > 0.$$

12.5. ***Двуполостный гиперболоид*** (см. рис. 12.1(в)):

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1, \quad a > 0,\ b > 0,\ c > 0.$$

12.6. ***Конус*** (см. рис. 12.1(г)):

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0, \quad a > 0,\ b > 0,\ c > 0.$$

В. Параболический тип.

12.7. ***Эллиптический параболоид*** (см. рис. 12.1(д)):

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 2z, \quad a > 0,\ b > 0.$$

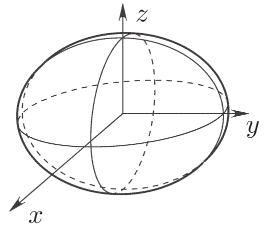
12.8. ***Гиперболический параболоид*** (см. рис. 12.1(е)):

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 2z, \quad a > 0,\ b > 0.$$

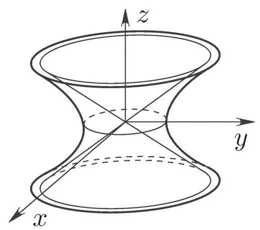
12.9. ***Эллиптический цилиндр*** (см. рис. 12.1(ж)):

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad a > 0,\ b > 0.$$

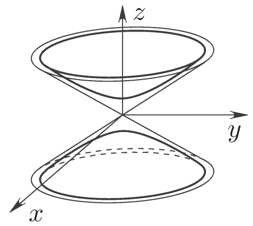
Направляющей цилиндра является эллипс, образующие параллельны оси Oz.



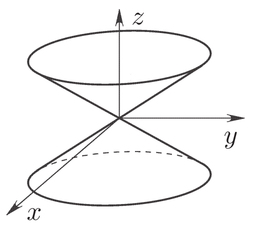
(а) эллипсоид



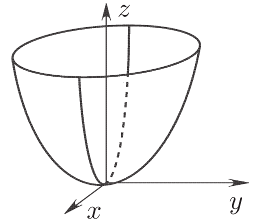
(б) однополостный гиперболоид



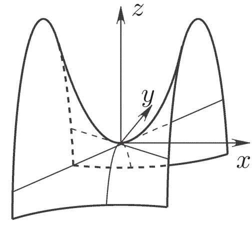
(в) двуполостный гиперболоид



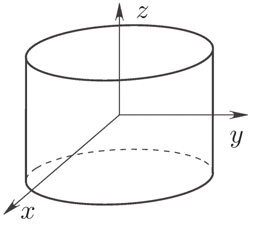
(г) конус



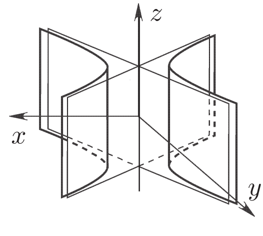
(д) эллиптический параболоид



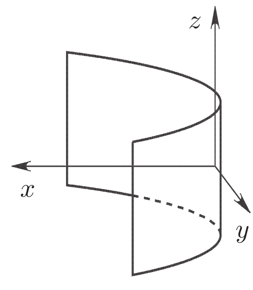
(е) гиперболический параболоид



(ж) эллиптический цилиндр



(з) гиперболический цилиндр



(и) параболический цилиндр

Рис. 12.1. Основные типы поверхностей второго порядка

12.10. ***Мнимый эллиптический цилиндр:***

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1, \quad a > 0,\ b > 0.$$

Это уравнение не имеет ни одного вещественного решения.

12.11. ***Пара мнимых пересекающихся плоскостей:***

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0, \quad a > 0,\ b > 0.$$

где $a > 0$, $b > 0$. Вещественные точки этой поверхности заполняют прямую (ось Oz).

12.12. ***Гиперболический цилиндр*** (см. рис. 12.1(з)):

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \quad a > 0,\ b > 0.$$

Направляющей является гипербола, образующие параллельны оси Oz.

12.13. ***Пара пересекающихся плоскостей:***

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0, \quad a > 0,\ b > 0.$$

Линией пересечения плоскостей является ось Oz.

12.14. ***Параболический цилиндр*** (см. рис. 12.1(и)):

$$y^2 = 2px, \quad p > 0.$$

Направляющей является парабола, образующие параллельны оси Oz.

12.15. ***Пара параллельных плоскостей:***

$$y^2 = a^2, \quad a > 0.$$

Плоскости параллельны плоскости Oxz.

12.16. ***Пара мнимых параллельных плоскостей:***

$$y^2 = -a^2, \quad a > 0.$$

Это уравнение не имеет ни одного вещественного решения.

12.17. ***Пара совпадающих плоскостей:***

$$y^2 = 0.$$

2. Исследование формы поверхностей

Получить представление о внешнем виде поверхностей можно при помощи метода сечений. Рассмотрим несколько примеров.

А. Эллипсоид.

12.18. В прямоугольной декартовой системе координат эллипсоид задается уравнением

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1,$$

где $a>0$, $b>0$, $c>0$. Ясно, что координатные плоскости являются плоскостями симметрии эллипсоида, начало координат — его центром симметрии. Эллипсоид целиком расположен в параллелепипеде с центром в точке $O(0,0,0)$, с гранями, параллельными координатным плоскостям, и со сторонами, равными $2a$, $2b$ и $2c$.

12.19. Сечением эллипсоида плоскостью $z=h$ является линия

$$\begin{cases} z=h, \\ \dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1-\dfrac{h^2}{c^2}, \end{cases}$$

или, эквивалентно,

$$\frac{x^2}{\left(a\sqrt{1-\dfrac{h^2}{c^2}}\right)^2}+\frac{y^2}{\left(b\sqrt{1-\dfrac{h^2}{c^2}}\right)^2}=1.$$

Следовательно, плоскость $z=h$ при $|h|>c$ не пересекает эллипсоид, при $|h|=c$ имеет единственную общую точку с эллипсоидом (это точка $(0,0,c)$ при $h=c$ и точка $(0,0,-c)$ при $h=-c$), а при $|h|<c$ пересекает эллипсоид по эллипсу с полуосями

$$a\sqrt{1-\frac{h^2}{c^2}},\quad b\sqrt{1-\frac{h^2}{c^2}},$$

которые максимальны (и равны a и b соответственно) при $h=0$ и монотонно уменьшаются до нуля, когда $|h|$ возрастает от нуля до c.

Аналогично анализируются сечения эллипсоида плоскостями $x=h$ и $y=h$; все такие сечения представляют собой эллипсы.

Б. Двуполостный гиперболоид.

12.20. Уравнение двуполостного гиперболоида имеет вид

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1 \;\Leftrightarrow\; -\frac{x^2}{a^2}-\frac{y^2}{b^2}+\frac{z^2}{c^2}=1,$$

где $a > 0$, $b > 0$ и $c > 0$. Ясно, что координатные плоскости являются плоскостями симметрии двуполостного гиперболоида, начало координат — его центром симметрии. Поверхность состоит из двух симметричных частей, расположенных в полупространствах $z \geqslant c$ и $z \leqslant -c$.

12.21. Плоскость $z = h$ при $|h| < c$ не пересекает гиперболоид, при $|h| = c$ имеет единственную общую точку с гиперболоидом (это точка $(0, 0, c)$ при $h = c$ или $(0, 0, -c)$ при $h = -c$) и при $|h| > c$ пересекает гиперболоид по эллипсу

$$\frac{x^2}{\left(a\sqrt{\frac{h^2}{c^2} - 1}\right)^2} + \frac{y^2}{\left(b\sqrt{\frac{h^2}{c^2} - 1}\right)^2} = 1,$$

полуоси которого монотонно возрастают от 0 до $+\infty$, когда $|h|$ возрастает от c до $+\infty$.

12.22. Плоскость $y = h$ пересекает гиперболоид по гиперболе

$$\frac{z^2}{\left(c\sqrt{1 + \frac{h^2}{b^2}}\right)^2} - \frac{x^2}{\left(a\sqrt{1 + \frac{h^2}{b^2}}\right)^2} = 1,$$

полуоси которой монотонно возрастают (от c и a соответственно) до $+\infty$, когда $|h|$ возрастает от 0 до $+\infty$. Аналогичные результаты можно получить для сечений плоскостями $x = h$.

B. Однополостный гиперболоид.

12.23. Уравнение однополостного гиперболоида

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1,$$

где $a > 0$, $b > 0$ и $c > 0$. Координатные плоскости являются плоскостями симметрии однополостного гиперболоида, начало координат — его центром симметрии.

12.24. Плоскость $z = h$ пересекает гиперболоид по эллипсу

$$\frac{x^2}{\left(a\sqrt{1 + \frac{h^2}{c^2}}\right)^2} + \frac{y^2}{\left(b\sqrt{1 + \frac{h^2}{c^2}}\right)^2} = 1,$$

полуоси которого монотонно возрастают от a и b соответственно до $+\infty$, когда $|h|$ возрастает от 0 до $+\infty$. Эллипс

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

получающийся при $h = 0$, называется *горловым эллипсом* гиперболоида.

12.25. Плоскость $y = h$ при $|h| < b$ пересекает гиперболоид по гиперболе

$$\frac{x^2}{\left(a\sqrt{1 - \frac{h^2}{b^2}}\right)^2} - \frac{z^2}{\left(c\sqrt{1 - \frac{h^2}{b^2}}\right)^2} = 1,$$

полуоси которой монотонно убывают от a и c соответственно до 0, когда $|h|$ возрастает от 0 до b. При $|h| = b$ сечением является пара пересекающихся прямых

$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = 0.$$

При $|h| > b$ сечение представляет собой гиперболу

$$\frac{z^2}{\left(c\sqrt{\frac{h^2}{b^2} - 1}\right)^2} - \frac{x^2}{\left(c\sqrt{\frac{h^2}{b^2} - 1}\right)^2} = 1,$$

полуоси которой возрастают от 0 до $+\infty$, когда $|h|$ возрастает от b до $+\infty$. Аналогичные результаты можно получить для сечений плоскостями $x = h$.

Г. Конус.

12.26. Уравнение конуса

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0,$$

где $a > 0$, $b > 0$, $c > 0$. Координатные плоскости являются плоскостями симметрии конуса, начало координат — его центром симметрии.

12.27. Коническая поверхность — это поверхность, образованная прямыми (прямолинейными образующими), проходящими через одну точку, называемую вершиной конуса. Направляющая конической поверхности — это произвольная расположенная на ней линия, обладающая тем свойством, что любая прямолинейная образующая пересекает её в единственной точке.

12.28. Сечение конуса плоскостью $z = h$, $h \neq 0$, представляет собой эллипс
$$\frac{x^2}{a^2h^2/c^2} + \frac{y^2}{b^2h^2/c^2} = 1,$$
полуоси которого пропорциональны $|h|$. Прямая, проходящая через центры этих эллипсов, называется *осью* конуса. Рассматривая сечения конуса плоскостями, не перпендикулярными оси, можем получить окружность.

Сечение конуса плоскостью $z = 0$ состоит из одной точки $O(0,0,0)$.

12.29. Сечение конуса плоскостью $y = h$, $h \neq 0$, является гиперболой
$$\frac{z^2}{c^2h^2/b^2} - \frac{x^2}{a^2h^2/b^2} = 1,$$
полуоси которой пропорциональны $|h|$. Аналогично для сечений плоскостями $x = h$. Таким образом, в качестве направляющей конуса может быть выбрана гипербола.

12.30. Сечение конуса плоскостью $y = 0$ представляет собой пару пересекающихся прямых
$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = 0.$$
Парабола также может быть получена как плоское сечение конуса.

Д. Эллиптический параболоид.

12.31. Уравнение эллиптического параболоида
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 2z,$$
где $a > 0$, $b > 0$. Координатные плоскости $x = 0$ и $y = 0$ являются плоскостями симметрии эллиптического параболоида, других плоскостей симметрии и центра симметрии у него нет.

12.32. Плоскость $z = h$ при $h < 0$ не пересекает параболоид, при $h = 0$ имеет с ним единственную общую точку $O(0,0,0)$, при $h > 0$ пересекает параболоид по эллипсу
$$\frac{x^2}{2ha^2} + \frac{y^2}{2hb^2} = 1,$$
полуоси которого монотонно возрастают вместе с h от 0 до $+\infty$.

12.33. Плоскости $y = h$ и $x = h$ пересекают параболоид по параболам с фокальными параметрами a^2 и b^2, с вершинами в точках $(0, h, h^2/2b^2)$ и $(h, 0, h^2/2a^2)$ и ветвями, направленными вверх.

Е. Гиперболический параболоид.

12.34. Уравнение гиперболического параболоида
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 2z,$$
где $a > 0$, $b > 0$. Плоскости $x = 0$ и $y = 0$ являются плоскостями симметрии гиперболического параболоида; других плоскостей симметрии нет.

12.35. Плоскость $z = h$ при $h < 0$ пересекает параболоид по гиперболе
$$\frac{y^2}{-2hb^2} - \frac{x^2}{-2ha^2} = 1;$$
действительная ось этой гиперболы параллельна оси Oy, а мнимая — оси Ox. Плоскость $z = h$ при $h > 0$ пересекает параболоид по гиперболе
$$\frac{x^2}{2ha^2} - \frac{y^2}{2hb^2} = 1;$$
действительная ось этой гиперболы параллельна оси Ox, а мнимая — оси Oy. Плоскость $z = 0$ пересекает параболоид по паре прямых
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0.$$

12.36. Плоскости $y = h$ и $x = h$ пересекают параболоид по параболам с фокальными параметрами a^2 и b^2 и с вершинами в точках $\left(0, h, -\dfrac{h^2}{2b^2}\right)$ и $\left(h, 0, \dfrac{h^2}{2b^2}\right)$; ветви первой параболы направлены вверх, второй — вниз. Вершины парабол, высекаемых плоскостями $y = h$, лежат на параболе, высекаемой плоскостью $x = 0$, а вершины парабол, высекаемых плоскостями $x = h$, — на параболе, высекаемой плоскостью $y = 0$.

3. Линейчатые поверхности

12.37. Поверхность называется *l-кратно линейчатой* поверхностью, если через каждую её точку проходит ровно l различных прямых, называемых *прямолинейными образующими.*

12.38. Все цилиндры являются 1-линейчатыми поверхностями. Конус также является 1-линейчатой поверхностью: все прямолинейные образующие которой проходят через одну точку — вершину конуса.

А. Однополостный гиперболоид.

12.39. Предложение. *Однополостный гиперболоид является дважды линейчатой поверхностью, т.е. через каждую его точку проходят две различные прямые, целиком лежащие на нём.*

Доказательство. Пусть (x_0, y_0, z_0) — точка, лежащая на однополостном гиперболоиде

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1. \tag{12.2}$$

Рассмотрим прямую, проходящую через эту точку:

$$x = x_0 + lt, \quad y = y_0 + mt, \quad z = z_0 + nt. \tag{12.3}$$

Для краткости введём обозначения[1]

$$\begin{array}{lll} X = \dfrac{x}{a}, & X_0 = \dfrac{x_0}{a}, & L = \dfrac{l}{a}, \\ Y = \dfrac{y}{b}, & Y_0 = \dfrac{y_0}{b}, & M = \dfrac{m}{b}, \\ Z = \dfrac{z}{c}, & Z_0 = \dfrac{z_0}{c}, & N = \dfrac{n}{c}. \end{array}$$

Уравнение гиперболоида (12.2) в новых обозначениях

$$X^2 + Y^2 - Z^2 = 1 \;\Leftrightarrow\; X^2 + Y^2 = 1 + Z^2, \tag{12.4}$$

а уравнение прямой (12.3) —

$$X = X_0 + Lt, \quad Y = Y_0 + Mt, \quad Z = Z_0 + Nt. \tag{12.5}$$

Подставляя (12.5) в (12.4), получим

$$(X_0 + Lt)^2 + (Y_0 + Mt)^2 - (Z_0 + Nt)^2 = 1 \;\Leftrightarrow$$
$$\Leftrightarrow\; \underbrace{(X_0^2+Y_0^2-Z_0^2)}_{=1} + 2t(X_0L+Y_0M-Z_0N) + t^2(L^2+M^2-N^2) = 1$$
$$\Leftrightarrow\; 2t(X_0L + Y_0M - Z_0N) + t^2(L^2 + M^2 - N^2) = 0.$$

Это уравнение выполняется тождественно (т.е. прямая (12.5) целиком лежит на гиперболоиде (12.4)) тогда и только тогда, когда

$$X_0L + Y_0M - Z_0N = 0, \quad L^2 + M^2 - N^2 = 0.$$

Поскольку направляющий вектор прямой может быть выбран с точностью до произвольного ненулевого множителя, положим $n = c$, тогда $N = 1$, и получим

$$X_0L + Y_0M = Z_0, \quad L^2 + M^2 = 1.$$

Второе уравнение допускает параметризацию

$$L = \cos\varphi, \quad M = \sin\varphi,$$

после чего первое уравнение примет вид

$$X_0\cos\varphi + Y_0\sin\varphi = Z_0 \;\Leftrightarrow$$

[1] Переход к переменным X, Y, Z можно рассматривать как растяжение или сжатие вдоль координатных осей.

$$\Leftrightarrow \sqrt{X_0^2+Y_0^2}\left(\underbrace{\frac{X_0}{\sqrt{X_0^2+Y_0^2}}}_{=\cos\theta}\cos\varphi+\underbrace{\frac{Y_0}{\sqrt{X_0^2+Y_0^2}}}_{=\sin\theta}\sin\varphi\right)=Z_0 \Leftrightarrow$$

$$\Leftrightarrow \cos(\varphi-\theta)=\frac{Z_0}{\sqrt{X_0^2+Y_0^2}}=\frac{Z_0}{\sqrt{1+Z_0^2}};$$

последняя дробь строго меньше единицы по модулю, поэтому тригонометрическое уравнение имеет ровно 2 решения на промежутке $[0,2\pi)$:

$$\varphi-\theta=\pm\arccos\frac{Z_0}{\sqrt{1+Z_0^2}} \Leftrightarrow \varphi=\theta\pm\arccos\frac{Z_0}{\sqrt{1+Z_0^2}}.$$

Пусть

$$\beta=\arccos\frac{Z_0}{\sqrt{1+Z_0^2}}\in[0,\pi];$$

тогда

$$\cos\beta=\frac{Z_0}{\sqrt{1+Z_0^2}},\quad \sin\beta=\frac{1}{\sqrt{1+Z_0^2}}$$

и далее

$$\begin{aligned}\cos\varphi&=\cos(\theta\pm\beta)=\cos\theta\cos\beta\mp\sin\theta\sin\beta\\&=\frac{X}{\sqrt{X_0^2+Y_0^2}}\frac{Z_0}{\sqrt{1+Z_0^2}}\mp\frac{Y_0}{\sqrt{X_0^2+Y_0^2}}\frac{1}{\sqrt{1+Z_0^2}}=\frac{X_0Z_0\mp Y_0}{1+Z_0^2},\\ \sin\varphi&=\sin(\theta\pm\beta)=\sin\theta\cos\beta\pm\cos\theta\sin\beta\\&=\frac{Y_0}{\sqrt{X_0^2+Y_0^2}}\frac{Z_0}{\sqrt{1+Z_0^2}}\pm\frac{X_0}{\sqrt{X_0^2+Y_0^2}}\frac{1}{\sqrt{1+Z_0^2}}=\frac{Y_0Z_0\pm X_0}{1+Z_0^2}.\end{aligned}$$

Итак, мы получили два решения

$$L_\varepsilon=\cos\varphi=\frac{X_0Z_0-\varepsilon Y_0}{1+Z_0^2},\quad M_\varepsilon=\sin\varphi=\frac{Y_0Z_0+\varepsilon X_0}{1+Z_0^2},\quad \varepsilon=\pm1.$$

Каждое из возможных значений ε определяет направляющий вектор прямолинейной образующей.

Таким образом, через точку (X_0,Y_0,Z_0) гиперболоида (12.4) проходят ровно две прямые, целиком лежащие на гиперболоиде:

$$X=X_0+\frac{X_0Z_0-\varepsilon Y_0}{1+Z_0^2}t,\quad Y=Y_0+\frac{Y_0Z_0+\varepsilon X_0}{1+Z_0^2}t,\quad Z=Z_0+t,$$

где $\varepsilon=\pm1$. Эти прямые пересекаются с плоскостью $z=0$ ($Z=0$) соответственно в точках (X_1,Y_1,Z_1), (X_{-1},Y_{-1},Z_{-1}), отвечающих значению параметра $t_0=-Z_0$:

$$\begin{aligned}X_\varepsilon&=X_0-\frac{X_0Z_0-\varepsilon Y_0}{1+Z_0^2}Z_0=\frac{X_0+\varepsilon Y_0Z_0}{1+Z_0^2}=\varepsilon\frac{Y_0Z_0+\varepsilon X_0}{1+Z_0^2}=\varepsilon M_\varepsilon,\\ Y_\varepsilon&=Y_0-\frac{Y_0Z_0+\varepsilon X_0}{1+Z_0^2}Z_0=\frac{Y_0-\varepsilon X_0Z_0}{1+Z_0^2}=-\varepsilon\frac{X_0Z_0-\varepsilon Y_0}{1+Z_0^2}=-\varepsilon L_\varepsilon,\\ Z_\varepsilon&=0,\qquad \varepsilon=\pm1.\end{aligned}$$

Ясно, что эти две точки лежат на горловом эллипсе гиперболоида. Теперь можно записать уравнения прямолинейных образующих в виде

$$\begin{cases} X = X_\varepsilon - \varepsilon Y_\varepsilon t, \\ Y = Y_\varepsilon + \varepsilon X_\varepsilon t, \\ Z = t, \end{cases} \tag{12.6}$$

где $(X_\varepsilon, Y_\varepsilon)$ — точка горлового эллипса, $\varepsilon = \pm 1$.

Обратно, пусть (X_*, Y_*) — точка горлового эллипса однополостного гиперболоида (12.4); её координаты удовлетворяют уравнению

$$X_*^2 + Y_*^2 = 1.$$

Рассмотрим две прямые

$$X = X_* - \varepsilon Y_* t, \quad Y = Y_* + \varepsilon X_* t, \quad Z = t, \qquad \varepsilon = \pm 1.$$

Поскольку

$$X^2 + Y^2 - Z^2 = (X_* - \varepsilon Y_* t)^2 + (Y_* + \varepsilon X_* t)^2 - t^2 = \\ = \underbrace{(X_*^2 + Y_*^2)}_{=1} + 2t(X_\varepsilon Y_\varepsilon - X_\varepsilon Y_\varepsilon) + t^2 \underbrace{(X_*^2 + Y_*^2 - 1)}_{=0} = 1,$$

обе эти прямые целиком лежат на гиперболоиде.

Итак, через любую точку гиперболоида проходит ровно две прямолинейные образующие, одна из которых отвечает значению $\varepsilon = 1$, а другая — значению $\varepsilon = -1$. Все прямолинейные образующие разбиваются на два семейства, отвечающих значениям $\varepsilon = 1$ и $\varepsilon = -1$ соответственно. □

12.40. При решении задач удобнее пользоваться другим методом нахождения прямолинейных образующих. Запишем уравнение однополостного гиперболоида в виде

$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1 - \frac{y^2}{b^2} \;\Leftrightarrow\; \left(\frac{x}{a} - \frac{z}{c}\right)\left(\frac{x}{a} + \frac{z}{c}\right) = \left(1 - \frac{y}{b}\right)\left(1 + \frac{y}{b}\right).$$

Пусть точка (x_0, y_0, z_0) лежит на гиперболоиде. Рассмотрим две системы однородных линейных уравнений

$$\begin{cases} \alpha\left(\dfrac{x_0}{a} - \dfrac{z_0}{c}\right) = \beta\left(1 - \dfrac{y_0}{b}\right), \\ \beta\left(\dfrac{x_0}{a} + \dfrac{z_0}{c}\right) = \alpha\left(1 + \dfrac{y_0}{b}\right), \end{cases} \qquad \begin{cases} \gamma\left(\dfrac{x_0}{a} - \dfrac{z_0}{c}\right) = \delta\left(1 + \dfrac{y_0}{b}\right), \\ \delta\left(\dfrac{x_0}{a} + \dfrac{z_0}{c}\right) = \gamma\left(1 - \dfrac{y_0}{b}\right) \end{cases}$$

относительно неизвестных (α, β) для первой системы и (γ, δ) для второй. Легко убедиться, что определители этих систем равны нулю; например, для первой системы

$$\begin{vmatrix} \dfrac{x_0}{a} - \dfrac{z_0}{c} & -\left(1 - \dfrac{y_0}{b}\right) \\ -\left(1 + \dfrac{y_0}{b}\right) & \dfrac{x_0}{a} + \dfrac{z_0}{c} \end{vmatrix} = \left(\frac{x_0^2}{a^2} - \frac{z_0^2}{c^2}\right) - \left(1 - \frac{y_0^2}{b^2}\right) = 0.$$

Таким образом, каждая из систем обладает нетривиальным решением; обозначим эти решения через (α_0, β_0) и (γ_0, δ_0) соответственно. Рассмотрим теперь системы уравнений

$$\begin{cases} \alpha_0\left(\dfrac{x}{a}-\dfrac{z}{c}\right)=\beta_0\left(1-\dfrac{y}{b}\right), \\ \beta_0\left(\dfrac{x}{a}+\dfrac{z}{c}\right)=\alpha_0\left(1+\dfrac{y}{b}\right), \end{cases} \qquad \begin{cases} \gamma_0\left(\dfrac{x}{a}-\dfrac{z}{c}\right)=\delta_0\left(1+\dfrac{y}{b}\right), \\ \delta_0\left(\dfrac{x}{a}+\dfrac{z}{c}\right)=\gamma_0\left(1-\dfrac{y}{b}\right) \end{cases}$$

относительно неизвестных (x, y, z). Точка (x_0, y_0, z_0) является решением каждой из систем, и при этом каждая из систем определяет прямую, проходящую через указанную точку. Поскольку при перемножении уравнений каждой из систем получается уравнение гиперболоида, любое решение (x, y, z) каждой из систем представляет точку, лежащую на гиперболоиде. Таким образом, обе прямых, представляемых данными системами, целиком лежат на гиперболоиде.

12.41. Теорема. *Прямолинейные образующие однополостного гиперболоида обладают следующими свойствами:*

(1) *через каждую точку гиперболоида проходит одна и только одна образующая каждого семейства;*
(2) *каждая образующая пересекает горловой эллипс гиперболоида;*
(3) *любые две образующие, принадлежащие к одному семейству, скрещиваются;*
(4) *любые две образующие, принадлежащие к разным семействам, лежат в одной плоскости, причём они*
    (а) *параллельны, если проходят через диаметрально противоположные точки горлового эллипса,*
    (б) *и пересекаются в противном случае.*

Доказательство. Первые два утверждения были доказаны выше. Выясним вопрос о взаимном расположении прямолинейных образующих.

Рассмотрим две прямолинейные образующие, проходящие через точки (X_1, Y_1) и (X_2, Y_2) горлового эллипса:

$$\begin{cases} X = X_1 - Y_1 t, \\ Y = Y_1 + X_1 t, \\ Z = t, \end{cases} \qquad \begin{cases} X = X_2 - \varepsilon Y_2 t, \\ Y = Y_2 + \varepsilon X_2 t, \\ Z = t; \end{cases} \tag{12.7}$$

при $\varepsilon = 1$ эти образующие принадлежат одному семейству, а при $\varepsilon = -1$ — разным.

Выясним вопрос о взаимном расположении этих прямых, используя метод, изложенный в п. 7.78 (см. с. 180). Вопрос сводится к исследованию рангов матриц

$$\begin{pmatrix} -Y_1 & -\varepsilon Y_2 \\ X_1 & \varepsilon X_2 \\ 1 & 1 \end{pmatrix}, \qquad \begin{pmatrix} -Y_1 & -\varepsilon Y_2 & X_2 - X_1 \\ X_1 & \varepsilon X_2 & Y_2 - Y_1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Ранг r первой матрицы равен 1, если $Y_1 = \varepsilon Y_2$ и $X_1 = \varepsilon X_2$, т.е. если точки (X_1, Y_1) и (X_2, Y_2) либо совпадают (при $\varepsilon = 1$: этот случай бессодержателен), либо (при $\varepsilon = -1$) расположены на горловом эллипсе симметрично относительно начала координат, причём рассматриваемые прямолинейные образующие принадлежат к разным семействам. В остальных случаях $r = 2$.

Для нахождения ранга R второй матрицы вычислим её определитель:

$$\begin{array}{c} \begin{array}{ccc} \cdot(-1) & \qquad + & \qquad\qquad \end{array} \\ \begin{vmatrix} -Y_1 & -\varepsilon Y_2 & X_2 - X_1 \\ X_1 & \varepsilon X_2 & Y_2 - Y_1 \\ 1 & 1 & 0 \end{vmatrix} \end{array} = \begin{vmatrix} -Y_1 & Y_1 - \varepsilon Y_2 & X_2 - X_1 \\ X_1 & \varepsilon X_2 - X_1 & Y_2 - Y_1 \\ 1 & 0 & 0 \end{vmatrix} =$$

$$= \begin{vmatrix} Y_1 - \varepsilon Y_2 & X_2 - X_1 \\ \varepsilon X_2 - X_1 & Y_2 - Y_1 \end{vmatrix} = (Y_1 - \varepsilon Y_2)(Y_2 - Y_1) - (\varepsilon X_2 - X_1)(X_2 - X_1).$$

Если $\varepsilon = 1$, т.е. образующие принадлежат к одному семейству, то это выражение равно

$$-\Big[(X_2 - X_1)^2 + (Y_2 - Y_1)^2\Big] \neq 0,$$

т.е. $R = 3$. Кроме того, в этом случае $r = 2$, и согласно результатам, полученным в п. 7.78, рассматриваемые две прямые скрещиваются.

Если $\varepsilon = -1$, т.е. образующие принадлежат к разным семействам, то имеем

$$(Y_2 + Y_1)(Y_2 - Y_1) - (-X_2 - X_1)(X_2 - X_1) = Y_2^2 - Y_1^2 + X_2^2 - X_1^2 = 0,$$

т.е. $R = 2$, рассматриваемые две прямые лежат в одной плоскости. Если при этом $r = 2$, то эти прямые пересекаются в единственной точке. Если же $r = 1$, т.е. рассматриваемые прямолинейные образующие разных семейств проходят через диаметрально противоположные точки горлового эллипса, то они параллельны.

Рассмотрим три попарно различные одноименные прямолинейные образующие, проходящие через точки (X_1, Y_1), (X_2, Y_2), (X_3, Y_3) горлового эллипса, и определитель, составленный из координат направляющих векторов указанных прямых:

$$\begin{vmatrix} -Y_1 & -Y_2 & -Y_3 \\ X_1 & X_2 & X_3 \\ 1 & 1 & 1 \end{vmatrix} \neq 0,$$

поскольку точки (X_1, Y_1), (X_2, Y_2), (X_3, Y_3) не лежат на одной прямой. Таким образом, эти три прямолинейные образующие не компланарны. □

Б. Гиперболический параболоид.

12.42. Предложение. *Гиперболический параболоид является дважды линейчатой поверхностью.*

Доказательство. Пусть точка (x_0, y_0, z_0) лежит на гиперболическом параболоиде

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 2z. \tag{12.8}$$

Рассмотрим прямую, проходящую через эту точку:

$$x = x_0 + lt, \quad y = y_0 + mt, \quad z = z_0 + nt. \tag{12.9}$$

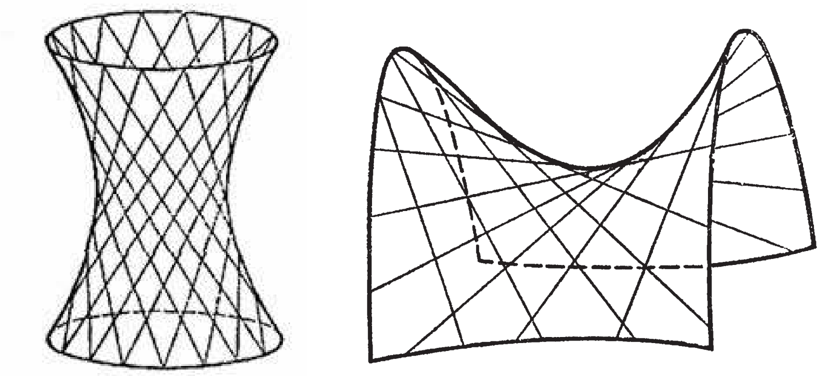


Рис. 12.2. Прямолинейные образующие однополостного гиперболоида и гиперболического параболоида

Для краткости введём обозначения[1]

$$X = \frac{x}{a}, \qquad X_0 = \frac{x_0}{a}, \qquad L = \frac{l}{a},$$
$$Y = \frac{y}{b}, \qquad Y_0 = \frac{y_0}{b}, \qquad M = \frac{m}{b},$$
$$Z = z, \qquad Z_0 = z_0, \qquad N = n.$$

Уравнения параболоида (12.8) и прямой (12.9) в новых обозначениях имеют вид

$$X^2 - Y^2 = 2Z, \qquad \begin{cases} X = X_0 + Lt, \\ Y = Y_0 + Mt, \\ Z = Z_0 + Nt. \end{cases}$$

Подставляя параметрические уравнения прямой в уравнение параболоида,получим

$$(X_0 + Lt)^2 - (Y_0 + Mt)^2 = 2(Z_0 + Nt) \;\Leftrightarrow$$
$$\Leftrightarrow\; \underbrace{(X_0^2 - Y_0^2)}_{=2Z_0} + 2t(X_0 L - Y_0 M) + t^2(L^2 - M^2) = 2Z_0 + 2Nt \;\Leftrightarrow$$
$$\Leftrightarrow\; 2t(X_0 L - Y_0 M) + t^2(L^2 - M^2) = 2Nt.$$

Это уравнение выполняется тождественно, т.е. прямая целиком лежит на параболоиде, тогда и только тогда, когда

$$X_0 L - Y_0 M = N, \quad L^2 - M^2 = 0.$$

Поскольку направляющий вектор прямой определен лишь с точностью до ненулевого множителя, положим $L = 1$; тогда $M = \varepsilon$, где $\varepsilon = \pm 1$, $N = X_0 - \varepsilon Y_0$.

[1]Как и в случае однополостного гиперболоида, это можно рассматривать как растяжение или сжатие вдоль координатных осей.

Итак, через точку (X_0, Y_0, Z_0) параболоида проходят ровно две прямых

$$\begin{cases} X = X_0 + t, \\ Y = Y_0 + \varepsilon t, \\ Z = Z_0 + (X_0 - \varepsilon Y_0)t, \end{cases} \qquad \varepsilon = \pm 1. \qquad \square$$

12.43. Для практического нахождения уравнений прямолинейных образующих гиперболического параболоида используется следующий приём. Запишем уравнение параболоида (12.8) в виде

$$\left(\frac{x}{a} - \frac{y}{b}\right)\left(\frac{x}{a} + \frac{y}{b}\right) = 2z. \tag{12.10}$$

Пусть точка (x_0, y_0, z_0) лежит на параболоиде. Рассмотрим две системы однородных линейных уравнений относительно неизвестных (α, β) и (γ, δ):

$$\begin{cases} \alpha\left(\dfrac{x_0}{a} - \dfrac{y_0}{b}\right) = \beta, \\ \beta\left(\dfrac{x_0}{a} + \dfrac{y_0}{b}\right) = 2\alpha z_0, \end{cases} \qquad \begin{cases} \gamma\left(\dfrac{x_0}{a} + \dfrac{y_0}{b}\right) = \delta, \\ \delta\left(\dfrac{x_0}{a} - \dfrac{y_0}{b}\right) = 2\gamma z_0. \end{cases}$$

Определители этих систем равны нулю (проверьте!), поэтому каждая из систем нетривиально разрешима; пусть (α_0, β_0) и (γ_0, δ_0) — какие-либо их решения. Рассмотрим теперь системы

$$\begin{cases} \alpha_0\left(\dfrac{x}{a} - \dfrac{y}{b}\right) = \beta_0, \\ \beta_0\left(\dfrac{x}{a} + \dfrac{y}{b}\right) = 2\alpha z, \end{cases} \qquad \begin{cases} \gamma_0\left(\dfrac{x}{a} + \dfrac{y}{b}\right) = \delta_0, \\ \delta_0\left(\dfrac{x}{a} - \dfrac{y}{b}\right) = 2\gamma_0 z; \end{cases}$$

каждая из них определяет прямую, проходящую через точку (x_0, y_0, z_0) параболоида. Перемножая уравнения каждой из систем, обнаруживаем, что любое решение системы является также и решением уравнения (12.8), т.е. прямая целиком лежит на параболоиде.

12.44. Теорема. *Прямолинейные образующие гиперболического параболоида обладают следующими свойствами:*

(1) *через каждую точку гиперболоида проходит одна и только одна образующая каждого семейства;*
(2) *любые две образующие, принадлежащие к одному семейству, скрещиваются;*
(3) *любые две образующие, принадлежащие к разным семействам, пересекаются;*
(4) *все образующие одного семейства параллельны одной плоскости.*

Доказательство. Первое утверждение было доказано выше. Выясним взаимное расположение двух одноименных прямолинейных образующих, проходящих через две различные точки (X_1, Y_1, Z_1), (X_2, Y_2, Z_2) параболоида:

$$\begin{cases} X = X_1 + t, \\ Y = Y_1 + t, \\ Z = Z_1 + (X_1 - Y_1)t, \end{cases} \qquad \begin{cases} X = X_2 + t, \\ Y = Y_2 + t, \\ Z = Z_2 + (X_2 - Y_2)t, \end{cases}$$

для чего найдём ранги матриц

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ X_1 - Y_1 & X_2 - Y_2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & X_2 - X_1 \\ 1 & 1 & Y_2 - Y_1 \\ X_1 - Y_1 & X_2 - Y_2 & Z_2 - Z_1 \end{pmatrix}.$$

Ранг первой из этих матриц равен

$$r = \begin{cases} 1, & \text{если } X_1 - Y_1 = X_2 - Y_2 \ \Leftrightarrow\ X_1 - X_2 - Y_1 + Y_2 = 0, \\ 2 & \text{в противном случае.} \end{cases}$$

Для вычисления ранга второй матрицы найдём её определитель:

$$\begin{vmatrix} 1 & 1 & X_2 - X_1 \\ 1 & 1 & Y_2 - Y_1 \\ X_1 - Y_1 & X_2 - Y_2 & Z_2 - Z_1 \end{vmatrix} \begin{matrix} \cdot(-1) \\ + \\ \ \end{matrix} =$$

$$= \begin{vmatrix} 1 & 1 & X_2 - X_1 \\ 0 & 0 & Y_2 - Y_1 - X_2 + X_1 \\ X_1 - Y_1 & X_2 - Y_2 & Z_2 - Z_1 \end{vmatrix} =$$

$$= \begin{vmatrix} 1 & 0 & X_2 - X_1 \\ 0 & 0 & Y_2 - Y_1 - X_2 + X_1 \\ x_1 - Y_1 & X_2 - Y_2 - X_1 + Y_1 & Z_2 - Z_1 \end{vmatrix} = (Y_2 - Y_1 - X_2 + X_1)^2.$$

Таким образом, в случае, когда $Y_2 - Y_1 - X_2 + X_1 \neq 0$, имеем

$$r = 2, \quad R = 3,$$

т.е. прямолинейные образующие скрещиваются.

Если же $Y_2 - Y_1 - X_2 + X_1 = 0$, то $r = 1$, но и $R = 1$, поскольку в этом случае $Y_2 - Y_1 = X_2 - X_1$, $X_2 - Y_2 = X_1 - Y_1$,

$$\operatorname{rk} \begin{pmatrix} 1 & 1 & X_2 - X_1 \\ 1 & 1 & Y_2 - Y_1 \\ X_1 - Y_1 & X_2 - Y_2 & Z_2 - Z_1 \end{pmatrix} = \operatorname{rk} \begin{pmatrix} 1 & 1 & X_2 - X_1 \\ 1 & 1 & X_2 - X_1 \\ X_1 - Y_1 & X_1 - Y_1 & Z_2 - Z_1 \end{pmatrix} =$$

$$= \operatorname{rk} \begin{pmatrix} 1 & X_2 - X_1 \\ X_1 - Y_1 & \frac{1}{2}(X_2^2 - Y_2^2) - \frac{1}{2}(X_1^2 - Y_1^2) \end{pmatrix},$$

а определитель последней матрицы равен нулю:

$$\begin{vmatrix} 1 & X_2 - X_1 \\ X_1 - Y_1 & \frac{1}{2}\underbrace{(X_2 - Y_2)}_{X_1 - Y_1}(X_2 + Y_2) - \frac{1}{2}(X_1 - Y_1)(X_1 + Y_1) \end{vmatrix} =$$

$$= (X_1 - Y_1)\begin{vmatrix} 1 & X_2 - X_1 \\ 1 & \frac{1}{2}(X_2 + Y_2 - X_1 - Y_1) \end{vmatrix} \begin{matrix} \cdot(-1) \\ + \end{matrix} =$$

$$= (X_1 - Y_1)\begin{vmatrix} 1 & X_2 - X_1 \\ 0 & \frac{1}{2}\underbrace{(-X_2 + Y_2 + X_1 - Y_1)}_{=0} \end{vmatrix} = 0.$$

Поэтому рассматриваемые прямолинейные образующие совпадают.

Теперь выясним взаимное расположение двух разноименных прямолинейных образующих, проходящих через две различные точки (X_1, Y_1, Z_1), (X_2, Y_2, Z_2) параболоида:

$$\begin{cases} X = X_1 + t, \\ Y = Y_1 + t, \\ Z = Z_1 + (X_1 - Y_1)t, \end{cases} \qquad \begin{cases} X = X_2 + t, \\ Y = Y_2 - t, \\ Z = Z_2 + (X_2 + Y_2)t; \end{cases}$$

для этого найдём ранги матриц

$$\begin{pmatrix} 1 & 1 \\ -1 & 1 \\ X_1 + Y_1 & X_2 - Y_2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & X_2 - X_1 \\ -1 & 1 & Y_2 - Y_1 \\ X_1 + Y_1 & X_2 - Y_2 & Z_2 - Z_1 \end{pmatrix}.$$

Ранг первой из этих матриц равен $r = 2$ (базисный минор — в левом верхнем углу). Для нахождения ранга второй вычислим её определитель:

$$\begin{vmatrix} 1 & 1 & X_2 - X_1 \\ -1 & 1 & Y_2 - Y_1 \\ X_1 + Y_1 & X_2 - Y_2 & Z_2 - Z_1 \end{vmatrix} = X_2^2 - Y_2^2 - 2Z_2 - \left(X_1^2 - Y_1^2 - 2Z_1\right) = 0$$

(проведите вычисление самостоятельно), так что $R = 2$. Таким образом, эти прямолинейные образующие пересекаются в единственной точке.

Наконец, направляющие векторы всех образующих одного семейства имеют вид $(1, \varepsilon, X_0 - \varepsilon Y_0)$ и, следовательно, параллельны плоскости $X - \varepsilon Y = 0$, что и требовалось доказать. □

ПРИЛОЖЕНИЕ I

Простейшие понятия математической логики

ПОЛУЧЕНИЕ новых математических фактов из уже известных осуществляется с помощью логического вывода. В этом приложении мы обсудим простейшие понятия логики, необходимые в повседневной математической работе.

1. Математический алфавит

A. Переменные.

I.1. Под *переменной* в математике понимается символ (обычно буква или буква с индексами), служащий для обозначения произвольного объекта из некоторой фиксированной совокупности объектов, называемой *областью возможных значений* рассматриваемой переменной. Часто область возможных значений переменной явно не указывается, но из контекста ясно, какова эта область.

I.2. Если переменная употребляется таким образом, что допускается подстановка вместо неё любых объектов из области её возможных значений, то переменная называется *свободной*. Однако в математике встречается и другое употребление переменных, при котором не предполагается и не допускается подстановка конкретных объектов вместо переменной. Например, в выражении $\int\limits_0^1 x^3\,dx$, где x — вещественная числовая переменная, нельзя подставлять вместо x какие-либо конкретные числа. В таких случаях, т.е. когда по смыслу выражения, содержащего переменную, подстановка конкретных объектов вместо переменной недопустима, эта переменная называется *связанной*. В сущности, связанная переменная не является переменной; её естественно было считать переменной на некотором (раннем) этапе образования рассматриваемого выражения из более простых выражений. Связанную переменную можно произвольным образом переименовать (т.е. обозначить другой буквой), при этом смысл выражения не изменится; так,

$$\int\limits_0^1 x^3\,dx = \int\limits_0^1 y^3\,dy = \int\limits_0^1 \xi^3\,d\xi = \frac{1}{4}.$$

I.3. В одном выражении могут встречаться одновременно и свободные, и связанные переменные; например, в выражении

$$\int\limits_0^1 (xy)^3\,dx$$

переменная x является связанной, а y — свободной. Различение свободных и связанных вхождений переменной в выражение важно для правильного понимания смысла выражения.

Б. Знак равенства. Символ $=$ используется в математике в разных смыслах.

I.4. Чаще всего символ $=$ употребляется в смысле, описываемом словами «два имени одного и того же объекта». Например, выражение $2 \cdot 6 = 3 \cdot 4$ означает, что $2 \cdot 6$ и $3 \cdot 4$ — два имени одного и того же числа (одно и то же число). Этот смысл символа $=$ будем считать его главным смыслом.

I.5. В другом смысле символ $=$ используется в уравнениях; например, когда рассматривается уравнение $x + 2 = 4$, то вовсе не утверждается, что $x + 2$ и 4 — два имени одного и того же, но требуется найти такие x, что значения выражений $x + 2$ и 4 совпадают.

I.6. Для указания тождественного равенства двух выражений (т.е. их совпадения для всех значений переменных из области возможных значений) вместо символа $=$ часто используется символ $\equiv$; например, $(x+y)^2 \equiv x^2+2xy+y^2$ для всех вещественных чисел x и y; $\dfrac{x^2-1}{x-1} \equiv x+1$ для всех вещественных $x \neq 1$ (здесь требуется указание о том, для каких значений переменной имеет место тождественное равенство).

I.7. В *определениях* символ $=$ имеет смысл знака присваивания имени вновь определяемому понятию, например, $n! = 1 \cdot 2 \cdot \ldots \cdot n$ при $n \geqslant 1$: здесь слева от знака $=$ стоит имя $n!$, которым обозначается произведение всех последовательных натуральных чисел от 1 до n включительно (*факториал* натурального числа n). В случаях, когда символ $=$ используется в указанном смысле, часто употребляется символ $\stackrel{\text{def}}{=\!=}$ («равенство по определению»); так, в рассмотренном примере лучше записать $n! \stackrel{\text{def}}{=\!=} 1 \cdot 2 \cdot \ldots \cdot n$.

В. Знак суммирования.

I.8. Типичный пример использования связанной переменной — *оператор суммирования* $\sum$. Пусть $A(k)$ — выражение, содержащее целочисленную переменную k, и пусть a, b — целые числа. Определим символ $\sum\limits_{k=a}^{b} A(k)$ следующим образом:

$$\sum_{k=a}^{b} A(k) \stackrel{\text{def}}{=\!=} \begin{cases} A(a), & \text{если } b = a, \\ A(a) + A(a+1) + \cdots + A(b), & \text{если } b > a, \\ 0, & \text{если } b < a. \end{cases}$$

Выражение $\sum\limits_{k=a}^{b} A(k)$ является числом, от k оно не зависит, а подставлять чи́сла вместо символа k бессмысленно (например, бессмысленна запись $\sum\limits_{5=a}^{b} A(5)$), и поэтому k уже не является переменной в данном выражении. В выражении $\sum\limits_{k=a}^{b} A(k)$ буква k является связанной переменной; в данном контексте её называют *индексом суммирования*. Как и любую связанную переменную, индекс суммирования можно произвольным образом переобозначить:

$$\sum_{k=a}^{b} A(k) = \sum_{n=a}^{b} A(n) = \sum_{\mu=a}^{b} A(\mu).$$

I.9. Подобно оператору суммирования $\sum$ определяется и оператор перемножения:

$$\prod_{k=a}^{b} A(k) \stackrel{\text{def}}{=\!=} \begin{cases} A(a), & \text{если } b = a, \\ A(a) \cdot A(a+1) \cdots A(b), & \text{если } b > a, \\ 1, & \text{если } b < a; \end{cases}$$

разумеется, индекс перемножения k является связанной переменной и допускает произвольное переобозначение.

I.10. Из школьного курса математики читателю хорошо известны следующие свойства операций сложения и умножения чисел:

(i) переместительное свойство, или *коммутативность*: для любых чисел a и b

$$a + b = b + a, \quad ab = ba;$$

(ii) сочетательное свойство, или *ассоциативность*: для любых чисел a, b и c

$$(a + b) + c = a + (b + c), \quad (ab)c = a(bc);$$

(iii) распределительное свойство, или *дистрибутивность*: для любых чисел a, b и c

$$a(b + c) = ab + ac.$$

Суммы и произведения, состоящие из бо́льшего числа слагаемых (сомножителей), также обладают подобными свойствами:

$$\sum_{k=a}^{b} c \cdot A(k) = c \cdot \sum_{k=a}^{b} A(k),$$

$$\sum_{k=a}^{b} A(k) + \sum_{k=a}^{b} B(k) = \sum_{k=a}^{b} \bigl(A(k) + B(k)\bigr),$$

$$\sum_{k=a}^{b} \left[\sum_{l=c}^{d} A(k,l)\right] = \sum_{l=c}^{d} \left[\sum_{k=a}^{b} A(k,l)\right],$$

$$\sum_{k=a}^{b} \left[\sum_{l=c}^{d} B(k) \cdot A(k,l)\right] = \sum_{k=a}^{b} \left[B(k) \cdot \sum_{l=c}^{d} A(k,l)\right],$$

$$\prod_{k=a}^{b} A(k) \cdot \prod_{k=a}^{b} B(k) = \prod_{k=a}^{b} \bigl(A(k) \cdot B(k)\bigr),$$

$$\prod_{k=a}^{b} \left[\prod_{l=c}^{d} A(k,l)\right] = \prod_{l=c}^{d} \left[\prod_{k=a}^{b} A(k,l)\right].$$

2. Логические операции

А. Логика высказываний.

I.11. *Высказывание* — это предложение, про которое разумно говорить, что оно истинно или ложно. Фразы «Есть ли жизнь на Марсе?» и «Добро пожаловать!» не являются высказываниями (как и любые вопросительные или восклицательные предложения). Определение «треугольник *называется* равносторонним, если все его стороны равны между собой» — *не* высказывание (как и вообще любое *определение*), но фраза «если все стороны треугольника равны, то он *является* равносторонним» — истинное (в силу только что приведённого определения) высказывание. Фразы «$2 + 2 = 4$» и «$5 > 7$» — высказывания (первое — истинное, второе — ложное), а $2 + 3$ — не высказывание.

Итак, *логическое высказывание* — это повествовательное предложение, которое формализует некоторое выражение мысли и представляет собой утверждение, которому можно поставить в соответствие одно из двух логических значений: ложь (0) или истина (1). Для обозначения высказываний будем использовать прописные латинские буквы.

I.12. Из высказываний можно образовывать новые высказывания при помощи основных *логических операций*, которые удобно определять с помощью таблиц истинности.

I.13. Определение. *Отрицание* — логическая операция, в результате которой из данного высказывания A получается новое высказывание $\neg A$ («не-A», «неверно, что A», «A не имеет места») по следующему правилу:

A	$\neg A$
0	1
1	0

(I.1)

В классической логике справедлив *закон исключенного третьего*, утверждающий, что из двух высказываний A и $\neg A$ одно и только одно обязательно является истинным.

I.14. Определение. *Конъюнкция* — логическая операция, которая заключается в соединении двух высказываний A и B в новое высказывание $A \wedge B$ («A и B») по следующему правилу:

A	B	$A \wedge B$
0	0	0
0	1	0
1	0	0
1	1	1

(I.2)

Из таблицы истинности видно, что высказывание $A \wedge B$ истинно только в том случае, когда оба высказывания A и B истинны. Например, высказывание «$2 < 3$ и 4 — чётное число» истинно, а высказывания «$2 < 3$ и 5 — чётное число», «$2 > 3$ и 5 — чётное число» ложны.

I.15. Определение. *Дизъюнкция* — логическая операция, которая заключается в соединении двух высказываний A и B в новое высказывание $A \vee B$ («A

или B») по следующему правилу:

A	B	$A \vee B$
0	0	0
0	1	1
1	0	1
1	1	1

(I.3)

Из таблицы истинности видно, что высказывание $A \vee B$ ложно только в том случае, когда оба высказывания A и B ложны. Например, высказывание «$2 > 3$ или 5 — чётное число» ложно, а высказывания «$2 < 3$ или 5 — чётное число», «$2 < 3$ или 4 — чётное число» истинны.

I.16. Определение. *Импликация* — логическая операция, которая заключается в соединении двух высказываний A и B в новое высказывание $A \Rightarrow B$ («если A, то B») по следующему правилу:

A	B	$A \Rightarrow B$
0	0	1
0	1	1
1	0	0
1	1	1

(I.4)

Высказывание A называется *посылкой* высказывания $A \Rightarrow B$, высказывание B — его *заключением*.

I.17. В обычной речи утверждение «если A, то B», как правило, предполагает наличие причинной связи между тем, что утверждается в высказывании A, и тем, что утверждается в высказывании B, и его истинность зависит от смысла этих высказываний. В логике учитывается лишь истинность или ложность высказываний, но не их смысл и тем более причинная связь.

Импликация двух высказываний ложна лишь в одном случае — когда посылка истинна, а заключение ложно. Импликация с ложной посылкой по определению истинна; например высказывания «если $2^2 = 5$, то $3^2 = 9$», «если $2^2 = 5$, то $3^2 = 10$» оба истинны. Такое определение на первый взгляд кажется не вполне естественным, однако математическая практика показывает, что оно удобно. В повседневной речи импликации с ложными посылками не употребляются, а в математическом языке принятое определение часто значительно сокращает речь, позволяя не оговаривать специально особые случаи рассуждений.

О высказываниях вида «если A, то B» с ложной посылкой A говорят, что они *истинны тривиальным образом*, т.е. истинны не в силу внутренней содержательной связи между высказываниями A и B, а вследствие соглашения о трактовке логического союза «если..., то...».

I.18. Приведем пример использования импликации с ложной посылкой в математике, приводящий к сокращению речи. Известен приём решения алгебраических иррациональных уравнений с квадратными корнями, основанный на возведении обеих частей уравнения в квадрат. При использовании этого приема могут появляться «посторонние корни», но «потеряться» корни не могут. Это означает, что *любой корень уравнения*

$$f(x) = g(x) \tag{I.5}$$

является также и корнем уравнения

$$[f(x)]^2 = [g(x)]^2. \tag{I.6}$$

Это высказывание мы считаем верным несмотря на то, что исходное уравнение может вообще не иметь корней. Если высказывание

«любой корень уравнения (I.5) является также корнем уравнения (I.6)»

представить в форме

«если число a является корнем уравнения (I.5), то оно является также корнем уравнения (I.6)»,

то становится ясным, что наша уверенность в истинности этого высказывания (независимо от того, имеет ли уравнение (I.5) корни) основана на бессознательном применении соглашения об истинности импликации с ложной посылкой. Если бы мы хотели сформулировать утверждение об уравнениях (I.5), (I.6) без учета этого соглашения, то пришлось бы формулировать так:

«если уравнение (I.5) имеет корни, то любой корень уравнения (I.5) является также корнем уравнения (I.6)».

Очевидно, что принятое соглашение сокращает речь.

I.19. Определение. *Эквивалентность* — логическая операция, которая заключается в соединении двух высказываний A и B в новое высказывание $A \Leftrightarrow B$ («A тогда и только тогда, когда B», «A эквивалентно B») по следующему правилу:

A	B	$A \Leftrightarrow B$
0	0	1
0	1	0
1	0	0
1	1	1

(I.7)

Из таблицы истинности видно, что высказывание $A \Leftrightarrow B$ истинно в том и только том случае, когда высказывания A и B оба истинны или оба ложны.

I.20. Предложение. *Справедливы следующие соотношения, в которых знак равенства означает совпадение логических значений («ложь» 0 или «истина» 1) левой и правой частей:*

(1) *коммутативность*:

$$A \vee B = B \vee A, \quad A \wedge B = B \wedge A;$$

(2) *ассоциативность*:

$$(A \vee B) \vee C = A \vee (B \vee C),$$
$$(A \wedge B) \wedge C = A \wedge (B \wedge C);$$

(3) *дистрибутивность*:

$$A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C),$$
$$A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C);$$

(4) *закон исключённого третьего*:

$$A \vee \neg A = 1, \quad A \wedge \neg A = 0;$$

(5) *закон двойного отрицания*:

$$\neg\neg A = A;$$

(6) *законы де Моргана*[1]:

$$\neg(A \wedge B) = \neg A \vee \neg B, \quad \neg(A \vee B) = \neg A \wedge \neg B;$$

(7) *закон контрапозиции*:

$$(A \Rightarrow B) = (\neg B \Rightarrow \neg A).$$

Доказательство. Каждое из приведенных соотношений может быть легко доказано при помощи таблиц истинности. Докажем, например, закон контрапозиции:

A	B	$A \Rightarrow B$	$\neg A$	$\neg B$	$\neg B \Rightarrow \neg A$
0	0	1	1	1	1
0	1	1	1	0	1
1	0	0	0	1	0
1	1	1	0	0	1

(здесь использованы таблицы истинности (I.1) и (I.4)). Сравнивая истинностные значения высказываний $A \Rightarrow B$ и $\neg B \Rightarrow \neg A$, видим, что они совпадают при любых истинностных значениях высказываний A и B. □

Б. Логика предикатов.

I.21. *Предикат* — это выражение, содержащее одну (одноместный предикат) или несколько переменных (s-местный предикат, если число переменных равно s), которое превращается в высказывание, если вместо этих переменных подставить объекты из области возможных значений.

Предикат может быть не определён при некоторых значениях входящих в него переменных; например, одноместный предикат $\sqrt{x} > 2$ с числовой переменной x, являющийся истинным при $x > 4$ и ложным при $0 \leqslant x \leqslant 4$, не определен при $x < 0$. В таких случаях, т.е. когда при некоторых значениях переменных рассматриваемый предикат не определён, он считается ложным при этих значениях переменных.

С помощью логических операций $\neg$, $\wedge$, $\vee$, $\Rightarrow$ из данных предикатов можно строить более сложные предикаты.

I.22. *Кванторы* — это логические операции, которые из предиката $P(x)$ строят высказывание, дающее количественную характеристику области истинности предиката $P(x)$. Наиболее употребительны квантор общности $\forall$ и квантор существования $\exists$.

I.23. Определение. Пусть P — одноместный предикат, X — область возможных значений переменной x. Обозначим через

$$(\forall x \in X) P(x) \tag{I.8}$$

следующее высказывание: «*для любого* значения $x \in X$ высказывание, полученное подстановкой этого значения в предикат P вместо переменной, истинно»[2] или, короче, «*для любого* $x \in X$ высказывание $P(x)$ истинно». Символ $\forall$ называется *квантором общности*, выражение $(\forall x \in X)$ — *квантором общности по переменной* x, а переход от предиката P к высказыванию (I.8) — *применением* к предикату P *квантора общности* по переменной x.

---

[1]О. де Морган (A. de Morgan) — шотландский математик (1806–1871).

[2]Разумеется, само высказывание, обозначенное выражением (I.8), может быть истинным или ложным.

I.24. Определение. Пусть P — одноместный предикат, X — область возможных значений переменной x. Обозначим через

$$(\exists x \in X)P(x) \tag{I.9}$$

следующее высказывание: «*существует* такое значение $x \in X$, что высказывание, полученное подстановкой этого значения в предикат P вместо переменной, истинно» или, короче, «*существует* такое $x \in X$, что высказывание $P(x)$ истинно». Символ $\exists$ называется *квантором существования*, выражение $(\exists x \in X)$ — *квантором существования по переменной* x, а переход от предиката P к высказыванию (I.9) — *применением* к предикату P *квантора существования* по переменной x.

I.25. Кванторы $\forall$ и $\exists$ называют взаимно двойственными.

I.26. Пример. $(\forall x \in \mathbb{R})(x^2 \geqslant 0)$ и $(\exists x \in \mathbb{R})(|x| \leqslant 0)$ — истинные высказывания; $(\forall x \in \mathbb{R})(x^2 > 0)$ и $(\exists x \in \mathbb{R})(|x| < 0)$ — ложные высказывания.

I.27. В выражениях (I.8), (I.9) буква x является *связанной переменной*; от x эти выражения не зависят. Таким образом, применение квантора связывает переменную.

I.28. Часто используется более вольная система обозначений. Так, в большинстве случаев опускаются скобки вокруг выражений $(\forall x \in X)$ и $(\exists x \in X)$. Указание области X возможных значений переменной также может быть задано не в виде $x \in X$, а в виде выражения иного типа, например, $\forall x > 0$. В ряде случаев указание области X вообще опускается; так, в случае, когда из контекста ясно, что x — вещественная числовая переменная, вместо $\forall x \in \mathbb{R}$ допустимо написать просто $\forall x$.

I.29. Предложение. *Справедливы следующие аналоги законов де Моргана*:

$$\neg(\forall x \in X)P(x) \quad \equiv \quad (\exists x \in X)\neg P(x), \tag{I.10}$$

$$\neg(\exists x \in X)P(x) \quad \equiv \quad (\forall x \in X)\neg P(x), \tag{I.11}$$

Доказательство. Докажем соотношение (I.10); для этого требуется установить истинность двух импликаций «левая часть $\Rightarrow$ правая часть» и «правая часть $\Rightarrow$ левая часть».

1. Пусть высказывание $\neg(\forall x)P(x)$ истинно, т.е., по определению отрицания, $(\forall x)P(x)$ ложно. Следовательно, не для любого значения переменной x высказывание $P(x)$ истинно. Значит, существует такое значение переменной x, для которого $P(x)$ ложно. Обозначим одно из таких значений через a. Итак, $a \in X$ и высказывание $P(a)$ ложно, так что $\neg P(a)$ истинно. Значит, существует такое x, что $\neg P(x)$ истинно, т.е. $(\exists x)\neg P(x)$.

2. Пусть теперь высказывание $(\exists x)\neg P(x)$ истинно. Тогда, по определению квантора существования, найдется такое значение переменной x, что $\neg P(x)$ истинно. Обозначим одно из таких значений через a. Итак, $a \in X$ и $\neg P(a)$ истинно, т.е. $P(a)$ ложно. Но тогда, по определению квантора общности, ложно и $(\forall x)P(x)$, т.е. $\neg(\forall x)P(x)$ истинно.

Соотношение (I.11) доказывается аналогично. □

I.30. Правило, выражаемое соотношениями (I.10), (I.11), можно сформулировать следующим образом: *чтобы получить отрицание высказывания, начинающегося с квантора, нужно квантор заменить на двойственный (т.е. квантор общности на квантор существования и наоборот) и перенести знак отрицания за квантор.*

I.31. Если P — двухместный предикат, то его можно превратить в высказывание, применяя кванторы по каждой переменной. В результате можно получить следующие восемь высказываний:

$$\begin{array}{ll}
(\forall y)(\forall x)P(x,y), & (\exists y)(\forall x)P(x,y),\\
(\forall x)(\forall y)P(x,y), & (\exists x)(\forall y)P(x,y),\\
(\forall y)(\exists x)P(x,y), & (\exists y)(\exists x)P(x,y),\\
(\forall x)(\exists y)P(x,y), & (\exists x)(\exists y)P(x,y),
\end{array}$$

Возникает вопрос: можно ли переставлять кванторы в этих выражениях?

Легко видеть, что

$$\begin{array}{lcl}
(\forall x)(\forall y)P(x,y) & = & (\forall y)(\forall x)P(x,y),\\
(\exists x)(\exists y)P(x,y) & = & (\exists y)(\exists x)P(x,y);
\end{array}$$

иными словами, *одноименные кванторы можно переставлять.* «Разноименные» же кванторы переставлять нельзя. Например, высказывание $(\forall y)(\exists x)(x > y)$ («для любого числа y существует большее число x») истинно, однако высказывание $(\exists x)(\forall y)(x > y)$ («существует число x, большее любого другого числа y») ложно.

I.32. При написании высказывания, имеющего вид предиката с применёнными к нему кванторами общности, эти кванторы общности (стоящие впереди) разрешается опускать. Так, истинное высказывание

$$(\forall x)(\forall y)\ (x+y)^2 = x^2 + 2xy + y^2 \tag{I.12}$$

можно написать короче:

$$(x+y)^2 = x^2 + 2xy + y^2;$$

эта запись по форме является двухместным предикатом, а по смыслу — высказыванием (I.12).

3. Теоремы и доказательства

I.33. В математике *доказательством* называется цепочка логических умозаключений, показывающая, что при некотором наборе аксиом и правил вывода верно некоторое утверждение. Доказанные утверждения называют *теоремами*; если ни утверждение, ни его отрицание ещё не доказаны, то такое утверждение называют *гипотезой.* В математических текстах теоремами обычно именуют только достаточно важные утверждения; менее важные и технически несложные утверждения-теоремы обычно называют леммами, предложениями, следствиями и прочими подобными терминами.

I.34. Основные виды формулировок теорем следующие:

(1) категорические: S есть P (пример: «сумма углов треугольника равна 180°»);
(2) импликативные: $A \Rightarrow B$ (пример: «если число делится на 10, то его последняя цифра — нуль»); высказывание A называют *условием* (*посылкой*) теоремы, B — *заключением* (*следствием*);
(3) разделительные (классификационные): S есть или A, или B, или C.

I.35. Определение. Следствие B импликативной теоремы «$A \Rightarrow B$» называется условием, *необходимым* для истинности посылки A. (Без выполнения B утверждение A не может быть истинным.) Теорема, выражающая необходимое условие, называется *свойством.*

I.36. Определение. Посылка A теоремы «$A \Rightarrow B$» называется условием, *достаточным* для выполнения следствия B. (Если A истинно, то утверждение B заведомо верно.) Теорема, выражающая достаточное условие, называется *признаком.*

I.37. Пример. Для того чтобы число делилось на 2, необходимо, чтобы последняя цифра в его десятичной записи не была семеркой:

$$\text{число делится на 2} \quad \Longrightarrow \quad \underbrace{\text{последняя цифра не 7.}}_{\text{необходимое условие делимости на 2}}$$

Это условие не является достаточным: число 13 не заканчивается на 7, но на 2 не делится; таким образом, необходимые условия содержат «лишние случаи». Утверждение «*если число делится на* 2*, то его последняя цифра не* 7» — свойство (одно из свойств) чётных чисел.

I.38. Пример. Для того чтобы число делилось на 2, достаточно, чтобы его последняя цифра была нулём:

$$\underbrace{\text{последняя цифра 0}}_{\text{достаточное условие делимости на 2}} \quad \Longrightarrow \quad \text{число делится на 2.}$$

Это условие не является необходимым: число 14 делится на 2, но его последняя цифра не 0; таким образом, достаточные условия содержат «не все случаи». Утверждение «*если последняя цифра числа — нуль, то число делится на* 2» является признаком (одним из признаков) делимости на 2.

I.39. Достаточные условия стараются сделать возможно более широкими, т.е. охватывающими возможно большее число случаев, в которых интересующий нас факт все ещё имеет место, а необходимые условия — возможно более узкими, т.е. охватывающими возможно меньше лишних случаев, в которых изучаемый факт уже не имеет места. Теорема, выражающая одновременно необходимое и достаточное условие, называется *критерием*: $A \Leftrightarrow B$. Критерии часто формулируются с помощью речевого оборота «*тогда и только тогда*».

I.40. Пример. Следующее утверждение может служить примером критерия: «*для того чтобы число делилось на* 3*, необходимо и достаточно, чтобы сумма его цифр делилась на* 3». В другой формулировке: «*число делится на* 3 *тогда и только тогда, когда сумма его цифр делится на* 3».

I.41. Определение. Рассмотрим следующие четыре импликативные теоремы, образованные из высказываний A и B:

$$A \Rightarrow B, \tag{I.13}$$

$$B \Rightarrow A, \tag{I.14}$$

$$\neg A \Rightarrow \neg B, \tag{I.15}$$

$$\neg B \Rightarrow \neg A. \tag{I.16}$$

Теоремы (I.13) и (I.14) (и соответственно, теоремы (I.15) и (I.16)) называются *взаимно обратными* теоремами: теорема (I.14) — это теорема, обратная теореме (I.13) и наоборот. Теоремы (I.13) и (I.15) (и соответственно, теоремы (I.14) и (I.16)) называются *взаимно противоположными теоремами.*

I.42. Как показывает следующая таблица истинности, взаимно обратные теоремы (I.13) и (I.14) логически почти не зависят друг от друга: истинность одной из них не влечет ни истинности, ни ложности другой, однако эти теоремы не могут быть одновременно ложными:

A	B	$A \Rightarrow B$	$B \Rightarrow A$
0	0	1	1
0	1	1	0
1	0	0	1
1	1	1	1

Взаимно противоположные теоремы (I.13) и (I.15) связаны аналогично (убедитесь в этом самостоятельно).

I.43. Если для некоторой теоремы $A \Rightarrow B$ верна и обратная теорема $B \Rightarrow A$, то они могут быть объединены в критерий $A \ \Leftrightarrow \ B$.

I.44. Пример. Исходная теорема: «*если в треугольнике один из углов прямой, то два других угла — острые*», — истинное утверждение. Обратная теорема: «*если два угла в треугольнике острые, то третий угол — прямой*», — ложное утверждение. Теорема Пифагора и обратная к ней теорема могут быть объединены в критерий: «*для того чтобы треугольник был прямоугольным, необходимо и достаточно, чтобы для его сторон a, b, c выполнялось соотношение $a^2 + b^2 = c^2$*».

I.45. В силу закона *контрапозиции*

$$(\neg B \Rightarrow \neg A) \quad == \quad (A \Rightarrow B)$$

теорема, противоположная к обратной, равносильна исходной. Этот прием может быть использован для доказательства в случае, когда теорема, противоположная к обратной, доказывается проще, чем исходная теорема. Такое доказательство часто называют «доказательством от противного».

ПРИЛОЖЕНИЕ II

Множества, соответствия, отображения

МНОЖЕСТВО — фундаментальное математическое понятие, появляющееся во всех её областях. Никакого определения понятию множества не даётся; мы считаем это понятие первичным. С элементами теории множеств читатель встречался в школьном курсе математики. В этом приложении, помимо множеств, обсуждаются важнейшие для математики понятия соответствий, отношений и отображений.

1. Множества и операции над ними

II.1. Общепризнано, что создателем теории множеств является Г. Кантор[1]. Он определил множество как «единое имя для совокупности всех объектов, обладающих данным свойством», а сами эти объекты назвал элементами множества. Теория, созданная Кантором, в настоящее время называется *наивной теорией множеств*. Эта теория не свободна от определённых недостатков; в частности, в 1903 г. Б. Расселл[2] обнаружил следующий парадокс, демонстрирующий противоречивость наивной теории множеств.

II.2. ***Парадокс Расселла.*** Пусть K — множество всех множеств, которые не содержат себя в качестве своего элемента. Содержит ли K само себя в качестве элемента? Если да, то, по определению K, оно не должно быть элементом K — противоречие. Если нет — то, по определению K, оно должно быть элементом K — вновь противоречие.

II.3. В наивной теории множеств Г. Кантора имеется два неопределяемых понятия — *множество* и *элемент*, и одно отношение между этими понятиями — *отношение принадлежности*, выражаемое словами «элемент принадлежит множеству» или «множество содержит элемент». Множества обычно обозначают прописными буквами A, B, ..., X, Y, ..., элементы — строчными буквами a, b, ..., x, y, ... Если элемент a принадлежит множеству A, пишем $a \in A$ ($A \ni a$), а если не принадлежит — то $a \notin A$ ($A \not\ni A$).

II.4. Запись $A = \{a, b, c, d\}$ означает, что множество A состоит из элементов a, b, c, d, а запись $A = \{a \mid P(a)\}$ — что множество A состоит из всех элементов, обладающих свойством $P(a)$ (характеристическим свойством); например, $[-2, 2] = \{x \mid x^2 \leqslant 4\}$.

II.5. Для удобства вводится понятие множества, не содержащего никаких элементов; оно называется *пустым множеством* и обозначается $\varnothing$.

[1] Г. Кантор (G. Cantor) — немецкий математик (1845–1918).

[2] Б. Расселл (B. Russell) — британский философ, общественный деятель и математик (1872–1970).

II.6. Следует чётко различать множества и элементы множеств. Например, число 2 и множество $\{2\}$ — разные объекты; так, о множестве $\{2\}$ нельзя задать вопрос, чётное оно или нечётное: множество не может быть чётным или нечётным, но может быть конечным или бесконечным, одно- или многоэлементным, пустым или непустым и т. п. Наоборот, о числе 2 нельзя спрашивать, одноэлементное оно или непустое и т. п. Итак, у объектов 2 и $\{2\}$ разные атрибуты: 2 — чётное простое положительное число, $\{2\}$ — непустое одноэлементное конечное множество. Очевидно, всегда $a \in \{a\}$. Очевидно также, что $b \in \{a\} \;\Leftrightarrow\; b = a$.

II.7. Определение. Множества A и B называются *равными* (запись $A = B$), если они состоят из одних и тех же элементов; таким образом, равенство $A = B$ можно понимать в том смысле, что одно и то же множество обозначено разными символами A и B.

II.8. Из определения равенства множеств вытекает, что порядок элементов при записи множеств несуществен; например, $\{a, b, c\}$ и $\{c, b, a\}$ — одно и то же множество. Мы считаем, что в множестве элементы расположены «в беспорядке», так что говорить о первом, втором и т. д. элементе данного множества бессмысленно (по крайней мере до тех пор, пока элементы множества не перенумерованы, что, кстати говоря, не всегда возможно).

II.9. Для того чтобы понятие «число элементов множества» было вполне определённым, нужно — при нашем понимании равенства множеств — условиться, что в множестве не бывает одинаковых (неразличимых) элементов. Таким образом, например, запись $\{a, a, b\}$ считается некорректной и должна быть заменена на $\{a, b\}$.

II.10. Элементами множеств могут быть множества. Например, рассмотрим множество $\{\{a, b\}, \{c, d, e\}\}$. Оно состоит из двух элементов, одним из которых является двухэлементное множество $\{a, b\}$, а другим — трёхэлементное множество $\{c, d, e\}$.

II.11. Определение. Множество B называется *подмножеством* множества A (запись $B \subset A$ или $A \supset B$), если каждый элемент B является также элементом A:

$$(B \subset A) \stackrel{\text{def}}{=\!=} (\forall x)(x \in B) \Rightarrow (x \in A).$$

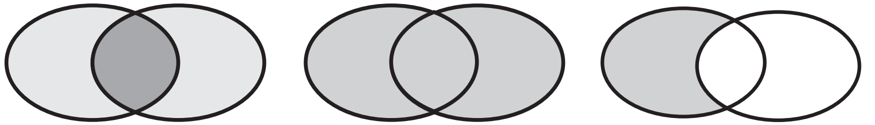
Отношение между множествами, выражаемое словами «B является подмножеством A», называется *отношением включения.* Очевидно, $A \subset A$.

II.12. Следует чётко различать знаки $\in$ и $\subset$ и обозначаемые ими понятия. Например, для множества $A = \{\{a, b\}, \{c, d, e\}\}$ запись $\{a, b\} \in A$ верна, а запись $\{a, b\} \subset A$ неверна, ибо в A нет ни элемента a, ни элемента b. Для множества $B = \{a, b, \{a, b\}\}$, состоящего из трёх элементов a, b и $\{a, b\}$, обе записи $\{a, b\} \in B$ и $\{a, b\} \subset B$ осмысленны, но смысл их различен.

II.13. Принято считать, по определению, что $\varnothing \subset A$ для любого множества A. Множества A и $\varnothing$ называются *несобственными подмножествами* множества A. Если же $B \subset A$ и существует такой элемент $a \in A$, что $a \notin B$, то множество B называется *собственным подмножеством* множества A. Запись $B \not\subset A$ означает, что B не является подмножеством множества A.

II.14. Определение. *Пересечением* $A \cap B$ множеств A и B называется множество, состоящее из всех элементов, которые одновременно принадлежат как множеству A, так и множеству B:

$$A \cap B \stackrel{\text{def}}{=\!=} \{x \mid x \in A \;\wedge\; x \in B\}.$$



(а) пересечение (б) объединение (в) разность

Рис. II.1. Диаграммы Эйлера, иллюстрирующие основные операции над множествами

II.15. Пустое множество является подмножеством самого себя, но при этом само с собой не пересекается: $\varnothing \subset \varnothing$, $\varnothing \cap \varnothing = \varnothing$.

II.16. Определение. *Объединением* $A \cup B$ множеств A и B называется множество, состоящее из всех элементов, которые принадлежат хотя бы одному из этих множеств:

$$A \cup B \overset{\text{def}}{=\!=} \{x \mid x \in A \ \lor \ x \in B\}.$$

Объединение непересекающихся множеств обозначают также знаком $\sqcup$ и называют *дизъюнктным объединением*.

II.17. Пусть $A_1, A_2, \ldots$ — некоторые множества (их число может быть конечным или бесконечным). Для обозначения пересечения и объединения этих множеств используют следующие символы:

$$\bigcap_{k=1}^{m} A_k, \quad \bigcup_{n=1}^{\infty} A_n, \quad \bigsqcup_{\alpha \in \mathfrak{A}} A_\alpha$$

и т. п. (в последней записи $\mathfrak{A}$ — множество возможных значений индекса α, нумерующего рассматриваемые множества).

II.18. Определение. *Разностью* $A \setminus B$ множеств A и B называется множество, состоящее из всех элементов, принадлежащих множеству A, но не принадлежащих множеству B:

$$A \setminus B \overset{\text{def}}{=\!=} \{x \mid x \in A \ \land \ x \notin B\}.$$

II.19. Для наглядного изображения множеств и отношений между ними часто используются диаграммы, называемые *диаграммами Эйлера* (диаграммами Эйлера—Венна). На рис. II.1 с помощью диаграмм Эйлера проиллюстрированы понятия пересечения, объединения и разности множеств.

II.20. Определение. *Дополнением* множества A называется множество A', состоящее из всех элементов, не принадлежащих A. Часто бывает удобно считать все множества, участвующие в некотором рассуждении, подмножествами некоторого множества U, называемого *универсальным множеством* (*универсумом*); в этом случае дополнение A' представляет собой разность $U \setminus A$. Дополнение множества A обозначается также $\overline{A}$, $\complement A$.

II.21. Предложение. *Операции над множествами обладают следующими свойствами*:

(1) *коммутативность*:

$$A \cap B = B \cap A, \quad A \cup B = B \cup A;$$

(2) *ассоциативность*:

$$(A \cap B) \cap C = A \cap (B \cap C),$$
$$(A \cup B) \cup C = A \cup (B \cup C);$$

(3) *взаимная дистрибутивность*:

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C),$$
$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C);$$

(4) $A \setminus B = A \cap B'$;
(5) $\varnothing' = U$; $U' = \varnothing$;
(6) *законы де Моргана*:

$$(A \cap B)' = A' \cup B', \quad (A \cup B)' = A' \cap B'. \tag{II.1}$$

Эти свойства легко проверить при помощи диаграмм Эйлера.

II.22. В математике часто встречается понятие упорядоченной пары (тройки и т. п.). Это понятие, интуитивно совершенно ясное и очевидное, тем не менее требует определения. Неформально говоря, упорядоченная пара — это объект (a, b), образованный из двух объектов a и b, причём выполняется следующее свойство:

$$(a_1, b_1) = (a_2, b_2) \;\Leftrightarrow\; (a_1 = a_2) \wedge (b_1 = b_2).$$

Приведём определение упорядоченной пары, предложенное К. Куратовским[1].

II.23. Предложение. *Множества* $\{\{a_1\}, \{a_1, b_1\}\}$ *и* $\{\{a_2\}, \{a_2, b_2\}\}$ *равны тогда и только тогда, когда* $a_1 = a_2$ *и* $b_1 = b_2$.

Доказательство. Докажем импликацию $\Rightarrow$. Пусть

$$\{\{a_1\}, \{a_1, b_1\}\} = \{\{a_2\}, \{a_2, b_2\}\}. \tag{II.2}$$

1. Предположим, что $a_1 \neq b_1$; тогда множество $\{a_1, b_1\}$ состоит из двух элементов. Так как оно принадлежит левой части равенства (II.2), то принадлежит и правой, так что либо $\{a_1, b_1\} = \{a_2, b_2\}$, либо $\{a_1, b_1\} = \{a_2\}$; второе невозможно, ибо двухэлементное множество не может быть равно одноэлементному. Итак, $\{a_1, b_1\} = \{a_2, b_2\}$. Далее, одноэлементное множество $\{a_1\}$ принадлежит левой части равенства (II.2), поэтому оно принадлежит также правой и следовательно равно $\{a_2\}$ (поскольку не может быть равным двухэлементному множеству $\{a_2, b_2\}$). Отсюда $a_1 = a_2$ и $b_1 = b_2$.

2. Аналогично рассматривается ситуация, когда $a_2 \neq b_2$.

3. Осталось рассмотреть случай $a_1 = b_1$. Левая часть (II.2) равна

$$\{\{a_1\}, \{a_1, b_1\}\} = \{\{a_1\}, \{a_1, a_1\}\} = \{\{a_1\}, \{a_1\}\} = \{\{a_1\}\},$$

т.е. представляет собой множество с единственным элементом $\{a_1\}$. Но тогда и правая часть должны быть одноэлементным множеством, т.е. $\{a_2\} = \{a_2, b_2\}$, что возможно лишь при $a_2 = b_2$; следовательно правая часть (II.2) равна $\{\{a_2\}\}$. Получаем, что $a_1 = a_2$, так что все четыре элемента a_1, a_2, b_1, b_2 совпадают.

Импликация $\Leftarrow$ очевидна. $\square$

Итак, корректно следующее определение.

II.24. Определение. Пусть A, B — множества. Упорядоченной парой (a, b) элементов $a \in A$ и $b \in B$ называется множество $\{\{a\}, \{a, b\}\}$.

[1]К. Куратовский (K. Kuratowski) — польский математик (1896–1980).

II.25. Определение. *Декартовым* (*прямым*) *произведением* $A \times B$ множеств A и B называется множество, состоящее из всех упорядоченных пар (a, b), где $a \in A$, $b \in B$:

$$A \times B = \{(a, b) \mid a \in A,\ b \in B\}.$$

II.26. Декартово произведение нескольких (конечного числа) множеств будет введено в п. II.69 (см. с. 315).

2. Соответствия и отношения

А. Соответствия.

II.27. Определение. *Соответствием* R, заданным на упорядоченной паре множеств A и B, называется любое подмножество R декартова произведения $A \times B$ этих множеств. Если $(a, b) \in R$, то говорят, что элементы $a \in A$ и $b \in B$ находятся в рассматриваемом соответствии R. Вместо записи $(a, b) \in R$ обычно используют *инфиксную* запись (т.е. запись, при которой знак соответствия ставится между элементами): $a\ R\ b$.

II.28. Для соответствия $R \subset A \times B$ введём следующие термины:

(а) *область отправления* — множество A;
(б) *область прибытия* — множество B;
(в) *область определения* $\operatorname{dom} R$ — множество всех *первых* элементов пар из R:

$$\operatorname{dom} R = \{a \mid \exists b : (a, b) \in R\};$$

(г) *область значений* $\operatorname{im} R$ — множество всех *вторых* элементов пар из R:

$$\operatorname{im} R = \{b \mid \exists a : (a, b) \in R\}.$$

II.29. В случае конечных множеств A и B соответствие $R \subset A \times B$ удобно иллюстрировать с помощью графика: элементы множеств A и B изображаются точками плоскости; точки $a \in A$ и $b \in B$ соединяются стрелкой, если $a\ R\ b$ (см. рис. II.2).

II.30. Определение. Соответствие $R \subset A \times B$ называется

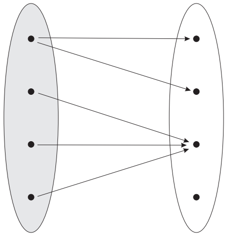
(а) *всюду определённым*, если область определения совпадает с областью отправления: $\operatorname{dom} R = A$; на графике из каждой точки области отправления исходит хотя бы одна стрелка (см. рис. II.2(а));
(б) *сюръективным*, если область значений совпадает с областью прибытия: $\operatorname{im} R = B$; на графике в каждую точку области прибытия приходит хотя бы одна стрелка (см. рис. II.2(б));
(в) *функциональным*, если в нём нет пар с одинаковыми первыми и различными вторыми компонентами, т.е.

$$\forall (a_1, b_1), (a_2, b_2) \in R \quad b_1 \neq b_2 \;\Rightarrow\; a_1 \neq a_2$$

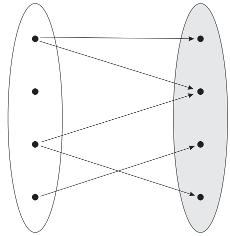
или, по контрапозиции,

$$\forall (a_1, b_1), (a_2, b_2) \in R \quad a_1 = a_2 \;\Rightarrow\; b_1 = b_2;$$

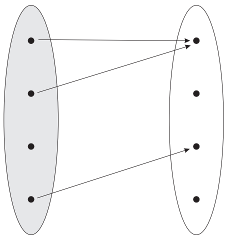
на графике из каждой точки области отправления исходит не более одной стрелки (см. рис. II.2(в)), т.е. запрещены ситуации вида $\bullet\!\begin{smallmatrix}\nearrow\\ \searrow\end{smallmatrix}$;



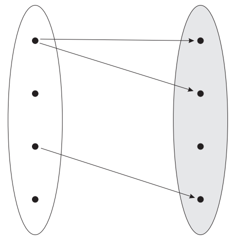
(а) всюду определённое



(б) сюръективное



(в) функциональное



(г) инъективное

Рис. II.2. Типы соответствий (к определению II.30)

(г) *инъективным*, если в нём нет пар с различными первыми и одинаковыми вторыми компонентами, т.е.

$$\forall (a_1, b_1), (a_2, b_2) \in R \quad a_1 \neq a_2 \;\Rightarrow\; b_1 \neq b_2$$

или, по контрапозиции,

$$\forall (a_1, b_1), (a_2, b_2) \in R \quad b_1 = b_2 \;\Rightarrow\; a_1 = a_2;$$

на графике в каждую точку области прибытия приходит не более одной стрелки (см. рис. II.2(г)), т.е. запрещены ситуации вида $\begin{smallmatrix}\searrow & \\ & \bullet \\ \nearrow & \end{smallmatrix}$;

(д) *взаимно однозначным соответствием*, или *биективным соответствием*, или *биекцией* между множествами A и B, если оно всюду определено, функционально, сюръективно и инъективно.

II.31. Неформально биекция между множествами A и B может быть описана следующим образом: *каждому* элементу множества A (определённость всюду) поставлен в соответствие *один* элемент множества B (функциональность), причём *разным* элементам множества A соответствуют *разные* элементы множества B (инъективность) и *каждый* элемент множества B соответствует *хотя бы* одному элементу множества A (сюръективность).

Б. Отношения эквивалентности.

II.32. Соответствия $R \subset A \times A$ (у которых область отправления совпадает с областью прибытия) называют *отношениями* на множестве A. Читателю хорошо знакомы следующие примеры отношений на множестве вещественных чисел: $=$, $>$, $\leqslant$.

Одним из наиболее важных типов отношений является отношение эквивалентности.

II.33. Определение. Отношение $R \subset A \times A$ называется *отношением эквивалентности*, если оно обладает следующими свойствами:

(i) *рефлексивность*: для любого $a \in A$ имеем $a\ R\ a$;

(ii) *симметричность*: для любых $a, b \in A$ из $a\ R\ b$ следует $b\ R\ a$;

(iii) *транзитивность*: для любых $a, b, c \in A$ из $a\ R\ b$ и $b\ R\ c$ следует $a\ R\ c$.

Типичные обозначения вместо $a\ R\ b$: $a \sim b$, $a \simeq b$, $a \cong b$, $a \approx b$.

II.34. Отношениями эквивалентности являются, например, отношения равенства чисел и множеств, подобия геометрических фигур, коллинеарности[1] прямых (компланарности плоскостей).

II.35. Определение. Пусть $\sim$ — отношение эквивалентности на множестве A. Для каждого элемента $a \in A$ его *классом эквивалентности относительно отношения* $\sim$ называется множество

$$[a] \stackrel{\text{def}}{=\!=} \{b \in A \mid b \sim a\}.$$

Например, если на множестве всех треугольников введено отношение подобия и $\triangle ABC$ — равносторонний, то классом эквивалентности $[\triangle ABC]$ является множество всех равносторонних треугольников.

II.36. Теорема.

1. *Если множество A представлено в виде объединения непересекающихся подмножеств, $A = \bigsqcup\limits_{\alpha \in \mathfrak{A}} A_\alpha$, то отношение «лежать в одном подмножестве» является отношением эквивалентности.*
2. *Любое отношение эквивалентности $\sim$ порождает разбиение множества A на непересекающиеся классы эквивалентности своих элементов:*

$$A = \bigsqcup [a].$$

Доказательство. Утверждение 1 очевидно.

Для доказательства утверждения 2 рассмотрим произвольное отношение эквивалентности $\sim$ на множестве A и построим для него соответствующее разбиение множества A в объединение непересекающихся подмножеств. Возьмём для каждого элемента $a \in A$ его класс эквивалентности. Докажем, что для двух различных $a_1, a_2 \in A$ классы $[a_1]$ и $[a_2]$ либо не пересекаются, либо совпадают. Предположим, что пересечение классов $[a_1]$ и $[a_2]$ непусто, т.е. $\exists c \in [a_1] \cap [a_2]$. Тогда $a_1 \sim c$ и $a_2 \sim c$, откуда $c \sim a_2$ (симметричность) и $a_1 \sim a_2$ (транзитивность), т.е. $a_1 \in [a_2]$. Аналогично получаем, что $a_2 \in [a_1]$. Итак, $[a_1] \subset [a_2]$ и $[a_2] \subset [a_1]$, поэтому $[a_1] = [a_2]$: пересекающиеся классы полностью совпадают. В силу рефлексивности каждый элемент $a \in A$ принадлежит задаваемому им классу, $a \in [a]$, т.е. действительно все множество A разбито на непересекающиеся классы: $A = \bigsqcup [a]$. □

II.37. Определение. Рассмотрим отношение эквивалентности $\sim$ на множестве A. Множество

$$A/\!\sim \stackrel{\text{def}}{=\!=} \{[a] \mid a \in A\},$$

состоящее из всех классов эквивалентности $[a]$ относительно $\sim$, называется *фактор-множеством* множества A по отношению $\sim$.

[1]Две прямые (плоскости) называются коллинеарными (компланарными), если они параллельны или совпадают.

II.38. В качестве простого примера рассмотрим множество A всех студентов некоторого учебного заведения, на котором введено отношение эквивалентности

$$(a \sim b) \overset{\text{def}}{=\!=} (\text{«}a \text{ и } b \text{ учатся в одной группе»})$$

(проверьте выполнение свойств (i)–(iii) определения II.33 самостоятельно). Классами эквивалентности являются в этом случае студенческие группы, а фактор-множеством — множество групп.

II.39. Каждый класс эквивалентности $[x]$ определяется любым своим элементом; составить представление об устройстве класса можно, взяв из него какой-либо любой элемент, называемый *представителем* этого класса.

II.40. На множестве $\mathbb{N}^2$ упорядоченных пар натуральных чисел введём отношение $\doteq$ следующим образом:

$$(a, b) \doteq (c, d) \;\Leftrightarrow\; ad = bc.$$

Проверим, что это отношение является отношением эквивалентности.

Рефлексивность: $(a, b) \doteq (a, b)$, поскольку $ab = ba$.

Симметричность: если $(a, b) \doteq (c, d)$, т.е. $ad = bc$, то $(c, d) = (a, b)$, поскольку $cb = da$.

Транзитивность: пусть $(a, b) \doteq (c, d)$ и $(c, d) \doteq (e, f)$. Это означает, что $ad = bc$ и $cf = de$. Умножая обе части первого равенства на f, а второго — на b, получим

$$adf = bcf,\ bcf = bde \;\Rightarrow\; adf = bde \;\Rightarrow\; af = be \;\Leftrightarrow\; (a, b) \doteq (e, f).$$

Записав упорядоченную пару натуральных чисел (a, b) в виде $\dfrac{a}{b}$, а закон эквивалентности — в виде

$$\frac{a}{b} \doteq \frac{c}{d} \;\Leftrightarrow\; ad = bc,$$

мы немедленно узнаём в этих парах положительные обыкновенные дроби. Отношение эквивалентности пар (дробей) позволяет сокращать дроби; классом эквивалентности каждой пары является *рациональное число*, т.е. множество всех равных между собой обыкновенных дробей, служащих «изображениями» этого рационального числа, например,

$$\left[\frac{2}{3}\right] = \left\{\frac{2}{3},\ \frac{4}{6},\ \frac{6}{9},\ \frac{8}{12},\ \ldots\right\} = \left[\frac{4}{6}\right] = \left[\frac{6}{9}\right] = \ldots.$$

В качестве представителя этого класса можно взять любую из дробей $2/3$, $4/6$ и т. д.

II.41. В математическом анализе рассматривается следующее отношение эквивалентности на множестве функций:

$$f(x) \sim g(x) \text{ при } x \to x_0, \quad \text{если } \lim_{x \to x_0} \frac{f(x)}{g(x)} = 1.$$

II.42. Два целых числа x и y называются *сравнимыми по модулю* m, обозначение $x \equiv y \pmod m$, если разность $x - y$ делится на m. Отношение сравнимости целых чисел по модулю является отношением эквивалентности; соответствующие классы эквивалентности обсуждаются в разделе 3 приложения IV (см. с. 342).

3. Отображения

Понятие отображения играет в математике центральную роль. Термины «отображение», «функция», «функционал», «оператор», «преобразование» являются синонимами; их употребление определяется сложившейся в разных разделах математики традицией.

А. Основные термины.

II.43. Определение. Пусть f — соответствие, заданное на множествах X и Y (т.е. подмножество декартова произведения $X \times Y$), *определённое всюду* на множестве X (см. определение II.30) и обладающее свойством *функциональности*: в нём нет пар с одинаковыми первыми и разными вторыми компонентами, т.е.

$$\forall (x_1, y_1), (x_2, y_2) \in f \quad y_1 \neq y_2 \implies x_1 \neq x_2.$$

Соответствие f называется *отображением*, определённым ***на*** множестве X и принимающим значения ***в*** множестве Y.[1]

II.44. Это определение уточняет знакомое читателю из школьного курса представление о функции как о правиле, которое *каждому* элементу одного множества (области определения) ставит в соответствие *единственный* элемент другого множества (области значений). В школе читатель имел дело в основном с числовыми функциями (область определения и множество значений — числовые множества, т.е. подмножества $\mathbb{R}$); в курсах высшей математики нам придётся иметь дело с отображениями множеств другой природы.

II.45. С понятием отображения связаны следующие термины и обозначения:

(а) если $(x, y) \in f$, то говорят, что y — *значение* отображения f на элементе x (в точке x), или y — *образ элемента* x при отображении f, и используют записи $y = f(x)$, $f\colon x \mapsto y$ и $x \overset{f}{\longmapsto} y$; говорят также, что отображение f переводит (преобразует, превращает; функция f отображает) элемент x в элемент y;[2]

(б) *образ множества* $C \subset X$ при отображении f — это множество

$$f(C) = \{f(c) \mid c \in C\} = \{y \in Y \mid \exists c \in C\colon (c, y) \in f\};$$

(в) X — *область определения* отображения f; используются обозначения $\operatorname{dom} f$ и $D(f)$;

(г) Y — *область прибытия* отображения f; говорят, что отображение f принимает значения ***в*** множестве Y;

(д) $\operatorname{im} f \overset{\text{def}}{=\!=} \{y \in Y \mid \exists x\colon y = f(x)\} \subset Y$ — *область значений* отображения f; говорят, что отображение f принимает значения ***на*** множестве $\operatorname{im} f$; используется также обозначение $R(f)$

(е) записи $f\colon X \to Y$, $X \xrightarrow{f} Y$ означают, что f — отображение с областью определения X и областью прибытия Y;

---

[1]Предлоги «на» и «в» в этом определении употребляются не произвольно, а играют самостоятельную роль, имеют терминологическое значение.

[2]Символ $f(x)$ обозначает *значение отображения на элементе* x; отображением является именно f, а не $f(x)$. Выражение «функция $y = f(x)$» некорректно: оно *задаёт* функцию, но *не является* ею. Однако во многих разделах математики часто обозначают через $f(x)$ как саму функцию, так и выражение, её задающее. Такое соглашение является в большинстве случаев удобным и вполне оправданным.

(ж) взаимно однозначное отображение $f\colon X \to X$ часто называют (особенно в геометрии) *преобразованием* множества X;

(з) *прообраз элемента* $y \in Y$ — это множество

$$f^{-1}(y) \stackrel{\text{def}}{=\!=} \{x \in X \mid f(x) = y\};$$

символ $f^{-1}(y)$ не следует ассоциировать с обратным отображением, которое может и не существовать (см. ниже п. II.56);

(и) *прообраз подмножества* $B \subset Y$ — это множество

$$f^{-1}(B) \stackrel{\text{def}}{=\!=} \{x \in X \mid f(x) \in B\} = \bigcup_{y \in B} f^{-1}(y).$$

Очевидно, если $y \in Y \setminus \operatorname{im} f$, то $f^{-1}(y)$ не существует или, эквивалентно, $f^{-1}(\{y\}) = \varnothing$.

II.46. Определение. Пусть $X \subset \tilde{X}$. Отображение $f\colon X \to Y$ называется *сужением* (или *ограничением*) отображения $g\colon \tilde{X} \to Y$ на подмножество X, если $f(x) = g(x)$ для всех $x \in X$; обозначение $f = g|_X$. В свою очередь, отображение $g\colon \tilde{X} \to Y$ называется *продолжением* отображения $f\colon X \to Y$ на множество $\tilde{X}$.

II.47. Определение. Отображение $f\colon X \to Y$ называется

(а) *сюръективным* (сюръекцией, или отображением «на»), если $\operatorname{im} f = Y$;

(б) *инъективным* (инъекцией), если из неравенства $x_1 \neq x_2$ вытекает $f(x_1) \neq f(x_2)$ или, по контрапозиции, из $f(x_1) = f(x_2)$ следует $x_1 = x_2$;

(в) *биективным* (взаимно однозначным, биекцией), если оно одновременно сюръективно и инъективно.

II.48. Два отображения f и g называются *равными*, если их соответствующие области совпадают: $X \xrightarrow{f} Y$, $X \xrightarrow{g} Y$, причём $f(x) = g(x)$ для любого $x \in X$.

II.49. Пример. Пусть $\mathbb{R}_{\geqslant 0} \stackrel{\text{def}}{=\!=} \{x \in \mathbb{R} \mid x \geqslant 0\}$. Три отображения $f\colon \mathbb{R} \to \mathbb{R}$, $g\colon \mathbb{R} \to \mathbb{R}_{\geqslant 0}$, $h\colon \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$, определённые одним и тем же правилом $x \mapsto x^2$, все различны:

(1) f не является ни сюръективным ($\operatorname{im} f = \mathbb{R}_{\geqslant 0}$), ни инъективным (числа x и $-x$ имеют один и тот же образ);

(2) g сюръективно, но не является инъективным;

(3) h биективно.

II.50. Отображение

$$\mathrm{id}_X\colon X \to X, \quad x \mapsto x,$$

которое каждому элементу x множества X ставит в соответствие сам этот элемент, называется *тождественным отображением* множества X. Очевидно, оно биективно.

II.51. Пусть $X \subseteq Y$. Отображение $f\colon X \to Y$, определённое на X правилом $f\colon x \mapsto x$, ставит каждому элементу $x \in X$ тот же самый элемент, но уже в множестве Y. Это отображение называется *вложением множества* X в множество Y и часто обозначается $f\colon X \hookrightarrow Y$.

Б. Композиция отображений. Обратное отображение.

II.52. Определение. *Композицией* (или *суперпозицией*) отображений $f\colon X \to Y$ и $g\colon Y \to Z$ называется отображение

$$g \circ f\colon X \to Z, \quad (g \circ f)(x) \stackrel{\text{def}}{=\!=} g\bigl(f(x)\bigr)$$

(правее пишется то отображение, которое применяется первым). То же самое наглядно изображается при помощи диаграммы

$$\begin{tikzcd} X \arrow[rr, "g\circ f"] \arrow[dr, "f"'] & & Z \\ & Y \arrow[ur, "g"'] & \end{tikzcd}$$

Про эту диаграмму говорят, что она *коммутативна*, т.е. результат перехода от X к Z не зависит от того, сделаем мы это непосредственно при помощи $g \circ f$ или воспользуемся промежуточным этапом Y.

II.53. Композиция отображений, вообще говоря, не коммутативна, т.е. $g \circ f \neq f \circ g$, хотя бы по той причине, что для $X \xrightarrow{f} Y$ и $Y \xrightarrow{g} Z$ отображение $X \xrightarrow{g\circ f} Z$ определено, а $f \circ g$ — не определено, если $Z \neq X$. Но даже в случае отображений $f, g\colon X \to X$, когда обе композиции $g \circ f$ и $f \circ g$ определены, коммутативность может не иметь места. Например, для числовых функций $f\colon x \mapsto x^2$ и $g\colon x \mapsto \sin x$ имеем

$$(g \circ f)(x) = g(f(x)) = \sin(x^2), \quad (f \circ g)(x) = f(g(x)) = (\sin x)^2.$$

II.54. Предложение. *Композиция отображений ассоциативна, т.е. для отображений $f\colon X \to Y$, $g\colon y \to Z$, $h\colon Z \to W$ имеет место соотношение*

$$(h \circ g) \circ f = h \circ (g \circ f),$$

что позволяет не ставить скобки и писать $h \circ g \circ f$.

Доказательство. Все необходимые рассуждения легко восстанавливаются по диаграмме

$$\begin{tikzcd} X \arrow[d, "f"'] \arrow[dr, dashed, "\alpha"] & W \\ Y \arrow[r, "g"'] \arrow[ur, dashed, "\beta"'] & Z \arrow[u, "h"'] \end{tikzcd}$$

где $\alpha = g \circ f$, $\beta = h \circ g$. В соответствии с определением равенства отображений нужно сравнить значения отображений $h \circ (g \circ f)$ и $h \circ (g \circ f)$ на элементе $x \in X$. Согласно определению композиции отображений имеем

$$\begin{aligned}(h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x))) = \\ = (h \circ g)(f(x)) = ((h \circ g) \circ f)(x).\end{aligned} \qquad \square$$

II.55. Определение. Пусть $f\colon X \to Y$ и $g\colon Y \to X$ — отображения. Если

$$g \circ f = \mathrm{id}_X \quad \text{и} \quad f \circ g = \mathrm{id}_Y, \tag{II.3}$$

то отображение g называется *обратным* к отображению f и обозначается f^{-1}. Таким образом,

$$f(x) = y \quad \Leftrightarrow \quad f^{-1}(y) = x.$$

Отображения f и f^{-1} называют *взаимно обратными.*

II.56. Обратное отображение обозначается тем же символом f^{-1}, что и операция взятия прообраза; следует избегать смешения этих понятий: прообраз элемента (или множества) существует даже в том случае, когда отображение не имеет обратного.

Сформулируем без доказательства следующие (интуитивно очевидные) утверждения.

II.57. Предложение.

1. *Отображение $f\colon X \to Y$ имеет обратное тогда и только тогда, когда оно биективно.*
2. *Если для отображения существует обратное, то оно единственно.*
3. *Если $f\colon X \to Y$, $h\colon Y \to Z$ — биективные отображения, то их композиция $h \circ f\colon X \to Z$ также биективна, причём*
$$(h \circ f)^{-1} = f^{-1} \circ h^{-1}. \tag{II.4}$$
4. *Если отображение $f\colon X \to Y$ биективно, то $f^{-1}\colon Y \to X$ также биективно, причём*
$$(f^{-1})^{-1} = f. \tag{II.5}$$

II.58. Пример. Отображения f и g из примера II.49 не имеют обратных, а биективное отображение $h : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$, $x \mapsto x^2$, имеет обратное $h^{-1}\colon x \mapsto \sqrt{x}$.

II.59. Пример. Отображение $\sin : \mathbb{R} \to [-1; 1]$ не имеет обратного, поскольку не является инъективным, однако его сужение

$$\sin\Big|_{[-\pi/2;\pi/2]}\colon \left[-\frac{\pi}{2}; \frac{\pi}{2}\right] \to [-1; 1],$$

будучи взаимно однозначным, имеет обратное отображение

$$\arcsin\colon [-1; 1] \to \left[-\frac{\pi}{2}; \frac{\pi}{2}\right].$$

B. Связь понятий эквивалентности и отображения.

II.60. Пусть $\sim$ — отношение эквивалентности на множестве X и $X/{\sim}$ — соответствующее фактор-множество. Определим отображение

$$f\colon X \to X/{\sim}, \quad x \mapsto [x],$$

которое каждому элементу x множества X ставит в соответствие его класс эквивалентности. Это отображение сюръективно, так как каждый элемент $x \in X$ принадлежит некоторому классу эквивалентности, т.е. для любого $[x] \in X/{\sim}$ справедливо $[x] = f(x)$.

II.61. Предложение. *Пусть* $f\colon X \to Y$ — *произвольное отображение. Отношение* $\overset{f}{\sim}$ *на* X, *определённое правилом*

$$x \overset{f}{\sim} y, \quad \text{если} \quad f(x) = f(y),$$

является отношением эквивалентности, причём существует биективное отображение $\varphi\colon X/\overset{f}{\sim} \to f(X)$.

Доказательство. Рефлексивность, симметричность и транзитивность отношения $\overset{f}{\sim}$ вытекают непосредственно из его определения.

Докажем, что отображение

$$\varphi\colon X/\overset{f}{\sim} \to f(X), \quad \varphi([x]) = f(x),$$

биективно, т.е. сюръективно и инъективно одновременно.

Инъективность. Пусть $[x] \neq [y]$; тогда $[x] \cap [y] = \varnothing$ и $x \overset{f}{\nsim} y$. Согласно определению отношения $\sim$ получаем $f(x) \neq f(y)$.

Сюръективность. Если $u \in f(X)$, то найдётся такой элемент $x \in X$, что $u = f(x) = \varphi([x])$, т.е. φ — сюръекция фактор-множества $X/\overset{f}{\sim}$ на $f(X)$. □

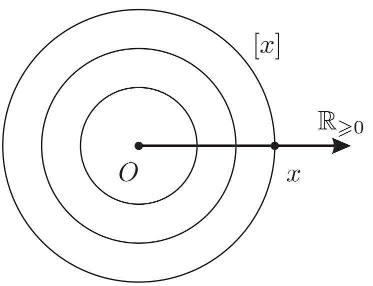


Рис. II.3.

II.62. Пример (см. рис. II.3)**.** Рассмотрим на евклидовой плоскости $\mathcal{E}_2$ функцию $f\colon \mathcal{E}_2 \to \mathbb{R}_{\geqslant 0}$, которая каждой точке $M \in \mathcal{E}_2$ ставит в соответствие расстояние от этой точки до некоторой фиксированной точки O: $f(M) = |OM|$. Классы эквивалентности относительно отношения эквивалентности $\overset{f}{\sim}$, порождённого этой функцией, представляют собой концентрические окружности с центром в точке O (а также сама точка O); фактор-множество $\mathcal{E}_2/\overset{f}{\sim}$ состоит из этих окружностей. Отображение $\varphi\colon \mathcal{E}_2/\overset{f}{\sim} \to \mathbb{R}_{\geqslant 0}$ ставит в соответствие каждой окружности её радиус (точке O соответствует число 0).

Г. Семейства элементов множества.

II.63. Определение. n-*Членным семейством* (*упорядоченным набором, конечной последовательностью*) элементов множества X будем называть отображение

$$f\colon \{1, 2, \ldots, n\} \to X, \quad k \mapsto x_k,$$

которое каждому натуральному числу $k \in \{1, 2, \ldots, n\}$ ставит в соответствие некоторый элемент $x_k \overset{\text{def}}{=\!=} f(k)$ множества X, называемый k-*м членом семейства.* Всё семейство обозначается при этом $(x_1, x_2, \ldots, x_n)$.

II.64. Множество всех n-членных семейств элементов множества X называется n-й *декартовой степенью* множества X и обозначается X^n:

$$X^n \overset{\text{def}}{=\!=} \Big\{(x_1, \ldots, x_n) \;\Big|\; x_k \in X,\ k = 1, \ldots, n\Big\}.$$

II.65. Если $\{j_1, j_2, \dots, j_s\} \subset \{1, 2, \dots, n\}$, то сужение отображения f

$$f\big|_{\{j_1, j_2, \dots, j_s\}} : \{j_1, j_2, \dots, j_s\} \to X$$

определяет *подсемейство* $(x_{j_1}, x_{j_2}, \dots, x_{j_s})$ семейства $(x_1, x_2, \dots, x_n)$.

II.66. Определение. Отображение

$$g\colon \mathbb{N} \to X, \quad k \mapsto x_k, \tag{II.6}$$

которое каждому натуральному числу $k \in \mathbb{N}$ ставит в соответствие некоторый элемент x_k множества X, называется (бесконечной) *последовательностью* элементов множества X.

II.67. Если M — бесконечное подмножество множества $\mathbb{N}$ натуральных чисел, $M = \{m_1, m_2, \dots\}$, то сужение отображения (II.6)

$$g\big|_M : M \to X, \quad m_k \mapsto x_{m_k}$$

называется *подпоследовательностью* рассмотренной последовательности.

II.68. Аналогично определению II.63 семейства элементов множества X можно сформулировать определения семейства, состоящего из элементов различных множеств, и декартова произведения множеств.

II.69. Определение. Пусть $X_1, \dots, X_n$ — множества. *Семейством* (упорядоченным набором) $(x_1, \dots, x_n)$, где $x_k \in X_k$, $k \in \{1, \dots, n\}$, будем называть отображение

$$f\colon \{1, 2, \dots, n\} \to X_1 \cup X_2 \cup \dots \cup X_n, k \mapsto x_k \in X_k,$$

которое каждому натуральному числу $k \in \{1, 2, \dots, n\}$ ставит в соответствие некоторый элемент $x_k \stackrel{\text{def}}{=\!=} f(k) \in X_k$ множества X_k, называемый k-м членом семейства. Множество всех семейств $(x_1, \dots, x_n)$ называется *декартовым произведением* множеств $X_1, \dots, X_n$ и обозначается $X_1 \times \dots \times X_n$.

ПРИЛОЖЕНИЕ III

Конечные и бесконечные множества

ПОНЯТИЕ «количества элементов множества», интуитивно ясное для множеств с конечным числом элементов (хотя пока мы и не обладаем определением «конечного числа»), допускает адекватное обобщение и на случай «бесконечных множеств» (определения «бесконечного множества» у нас тоже пока нет).

1. Основные определения

III.1. Определение. Два множества A и B называются *равномощными*, если между этими множествами существует взаимно однозначное соответствие (биекция) $f\colon A \to B$. Факт равномощности множеств A и B обозначается $|A| = |B|$, $\operatorname{card} A = \operatorname{card} B$ или $A \sim B$.

III.2. Предложение. *Отношение равномощности множеств обладает всеми свойствами отношения эквивалентности: оно рефлексивно, симметрично и транзитивно.*

III.3. Мы сознательно избегаем говорить, что отношение равномощности *является* отношением эквивалентности, чтобы не пришлось указывать, на каком множестве задано это отношение; как показывает парадокс Расселла (см. п. II.2, с. 302), о «множестве всех множеств» говорить нельзя, а конкретизировать класс всех множеств, которые допустимо рассматривать, в рамках наивной теории множеств невозможно.

Доказательство. Рефлексивность: каждое множество равномощно самому себе, поскольку в качестве взаимно однозначного отображения, о котором идёт речь в определении III.1, можно взять тождественное отображение $\mathrm{id}_A : A \to A$. *Симметричность* вытекает из того, что каждая биекция обратима, а *транзитивность* — из того факта, что композиция биекций также биективна (см. утверждения 3 и 4 теоремы II.57, с. 313). $\square$

III.4. Определение. Множество A называется *бесконечным*, если в нём существует собственное подмножество B (т.е. $B \neq A$), равномощное множеству A. Множество, не являющееся бесконечным, называется *конечным*.

III.5. Таким образом, в случае бесконечного множества имеется взаимно однозначное соответствие между ним самим и его частью, а в случае конечного множества такое соответствие невозможно.

III.6. Простейший пример бесконечного множества — множество натуральных чисел $\mathbb{N} = \{1, 2, 3, \dots\}$, равномощное своему собственному подмножеству

$2\mathbb{N} = \{2, 4, 6, \dots\}$, состоящему из чётных чисел. Биекция $f\colon \mathbb{N} \to 2\mathbb{N}$ задаётся формулой $n \mapsto 2n$, а обратное отображение $f^{-1}\colon 2\mathbb{N} \to \mathbb{N}$ — формулой $m \mapsto m/2$.

2. Натуральные числа и метод индукции

III.7. Натуральные числа возникают естественным образом при счёте. Последовательность всех натуральных чисел, расположенных в порядке их возрастания, называется *натуральным рядом*.

Натуральные числа служат фундаментом, на котором строятся все остальные числовые системы: множества целых, рациональных. вещественных и комплексных чисел. Каждое из перечисленных числовых множеств является расширением предыдущего, в результате чего становятся выполнимыми операции, невозможные в расширяемом множестве.

III.8. Существуют два подхода к определению натуральных чисел; один основан на использовании чисел для обозначения количества предметов (нет предметов, один предмет, два предмета и т. д.), а другой — для нумерации предметов (первый, второй, третий, …). В первом случае натуральный ряд начинается с нуля и обозначается $\mathbb{N}_0$, а во втором — с единицы и обозначается $\mathbb{N}$. Различие между этими двумя подходами по большей части терминологическое и на структуру теории натуральных чисел не влияет.

III.9. Натуральные числа можно рассматривать как мощности конечных множеств (см. ниже определение III.20, с. 319). Разумеется, чтобы избежать порочного круга в определениях[1] (т.е. ситуации, в которой понятие A определяется через понятие B, которое, в свою очередь, определяется снова через A), нужно определиться, какое из понятий — натуральное число или мощность множества — является первичным. Мы будем считать, что введение понятия натурального числа *логически предшествует* введению понятия мощности конечного множества; действительно, натуральные числа могут быть введены с помощью аксиоматики Пеано безотносительно к мощностям множеств, после чего понятие мощности конечного множества вводится определением III.20.

A. Аксиомы Пеано. Следующее определение вводит натуральные числа с помощью аксиом, предложенных Дж. Пеано[2] в конце XIX в. Отметим, что указанная аксиоматика базируется на наивной теории множеств.

III.10. Определение. Множество $\mathbb{N}_0$ будем называть *натуральным рядом*, (а его элементы — *натуральными числами*), если на нём задано отображение $S : \mathbb{N}_0 \to \mathbb{N}_0$, ставящее в соответствие каждому элементу $x \in \mathbb{N}_0$ *следующий* элемент $x' = S(x) \in \mathbb{N}_0$ (*отображение следования*) и удовлетворяющее следующим аксиомам:

$\mathbb{N}$**1:** нуль является натуральным числом: $0 \in \mathbb{N}_0$;

[1]Пример порочного круга в определениях, давно уже ставший хрестоматийным, привёл польский писатель-фантаст Станислав Лем в одном из рассказов цикла «Звёздные дневники Ийона Тихого» (путешествие четырнадцатое). Персонаж рассказа, изучая энциклопедию, последовательно читает следующие определения: «Сепульки — важный элемент цивилизации ардритов с планеты Энтеропия. См. *Сепулькарии*»; «Сепулькарии — устройства для сепуления. См. *Сепуление*»; «Сепуление — занятие ардритов с планеты Энтеропия. См. *Сепульки*».

[2]Дж. Пеано (G. Peano) — итальянский математик (1858–1932).

ℕ2: нуль не следует ни за каким натуральным числом:

$$\forall x \in \mathbb{N}_0\colon\ \ S(x) \neq 0;$$

ℕ3: если натуральное число z следует как за числом x, так и за числом y, то $x = y$:

$$(S(x) = z\ \wedge\ S(y) = z) \quad \Rightarrow \quad (x = y);$$

ℕ4: *аксиома индукции*: пусть A — подмножество натурального ряда, содержащее нуль и обладающее следующим свойством: если $x \in A$, то $S(x) \in A$; тогда A совпадает с натуральным рядом:

$$(0 \in A)\ \wedge\ (\forall x \in A\colon\ \ S(x) \in A) \quad \Rightarrow \quad (A = \mathbb{N}_0).$$

Элементы натурального ряда обозначаются привычными символами:

$$0,\ 1 \overset{\text{def}}{=\!=} S(0),\ 2 \overset{\text{def}}{=\!=} S(1),\ 3 \overset{\text{def}}{=\!=} S(2),\ \ldots .$$

Аксиомы ℕ**1**–ℕ**4** называются *аксиомами Пеано.*

III.11. Разумеется, заменив в этом определении 0 на 1, получим определение натурального ряда $\mathbb{N}$.

Б. Метод математической индукции. Аксиома ℕ**4** позволяет обосновать метод математической индукции, часто используемый для доказательств.

III.12. Теорема (метод математической индукции)**.** *Пусть $P(n)$ — некоторый одноместный предикат, зависящий от натурального параметра n. Если высказывание $P(0)$ истинно* (база индукции) *и доказано, что для любого $n \in \mathbb{N}_0$ из истинности высказывания $P(n)$ вытекает истинность высказывания $P(n')$, где $n' = S(n)$ — натуральное число, следующее за n* (шаг индукции), *то предикат P является тождественно истинным:*

$$P(0)\ \wedge\ \Big(\forall n\colon\ \ P(n) \Rightarrow P(n')\Big) \quad \Rightarrow \quad \Big(\forall n\colon\ \ P(n)\Big).$$

Доказательство. Пусть A — множество тех натуральных чисел, для которых предикат $P(n)$ является истинным:

$$A \overset{\text{def}}{=\!=} \{n \in \mathbb{N}_0 : P(n)\}.$$

Тогда $0 \in A$ (база индукции) и для любого $n \in A$ имеем $n' = S(n) \in A$ (шаг индукции). По аксиоме ℕ**4** имеем $A = \mathbb{N}_0$, что и требовалось доказать. □

В. Сложение и умножение натуральных чисел. Операции сложения и умножения натуральных чисел вводятся следующим образом.

III.13. Определение. *Сложение* на множестве $\mathbb{N}_0$ — это бинарная операция $(x, y) \mapsto x + y$ на $\mathbb{N}_0$, удовлетворяющая следующим аксиомам:
A1. $\forall x \in \mathbb{N}_0$: $x + 0 = x$;
A2. $\forall x, y \in \mathbb{N}_0$: $x + S(y) = S(x + y)$.

III.14. Согласно определению III.13 имеем

$$\begin{aligned}
&a + 0 = a,\\
&a + 1 = a + S(0) = S(a + 0) = S(a),\\
&a + 2 = a + S(1) = S(a + 1) = S(S(a)),\\
&a + 3 = a + S(2) = S(a + 2) = S(S(S(a))), \quad \ldots .
\end{aligned}$$

III.15. Определение. *Умножение* на множестве $\mathbb{N}_0$ — это бинарная операция $(x, y) \mapsto x \cdot y$ на bN_0, удовлетворяющая следующим аксиомам:
M1. $\forall x \in \mathbb{N}_0$: $x \cdot 0 = 0$;
M2. $\forall x, y \in \mathbb{N}_0$: $x \cdot S(y) = x \cdot y + x$.

III.16. Согласно определению III.15 имеем

$$\begin{aligned} &a \cdot 0 = 0, \\ &a \cdot 1 = a \cdot S(0) = a \cdot 0 + a = 0 + a = a, \\ &a \cdot 2 = a \cdot S(1) = a \cdot 1 + a = a + a, \\ &a \cdot 3 = a \cdot S(2) = a \cdot 2 + a = a + a + a, \quad \ldots . \end{aligned}$$

III.17. Можно доказать, что такие операции существуют, однозначно определены и обладают привычными свойствами, известными из школьного курса, например, свойствами коммутативности, ассоциативности, дистрибутивности умножения относительно сложения. Мы опускаем все эти доказательства; заинтересованный читатель может обратиться к литературе, посвящённой специально этим вопросам.

Г. Сравнение натуральных чисел.

III.18. Для двух натуральных чисел x и y выполняется одно и только одно из следующих соотношений:
(i) $x = y$;
(ii) $x = y + z$, где $z \neq 0$;
(iii) $y = x + z$, где $z \neq 0$.
Если $x = y + z$, где $z \neq 0$, то говорят, что x *больше* y и пишут $x > y$ или что y *меньше* x и пишут $y < x$. Если $x > y$ или $x = y$, то пишут $x \geqslant y$ или $y \leqslant x$.

III.19. Предложение. *Любое непустое подмножество $A \subset \mathbb{N}_0$ содержит наименьшее число, т.е. меньшее всех других чисел данного множества.*

3. Конечные множества. Комбинаторика

Комбинаторика — это раздел математики, в котором изучаются конечные множества и различные комбинации их элементов (комбинаторные конфигурации).

А. Мощность конечных множеств.

III.20. Определение. *Мощностью* конечного множества A называется натуральное число, равное количеству элементов этого множества, и обозначаемое $|A|$, $\operatorname{card} A$, $\#A$ (мы используем первое из этих обозначений). Конечное множество A мощности n для краткости будем называть *n-множеством.*

III.21. Предложение. *Мощность декартова произведения конечных множеств A и B равна произведению мощностей этих множеств*:

$$|A \times B| = |A| \cdot |B|.$$

Доказательство. Рассмотрим множества $A = \{a_1, \ldots, a_m\}$, $B = \{b_1, \ldots, b_n\}$. Предложение III.21 становится очевидным, если элементы множества $A \times B$ записать в таблицу вида

	b_1	b_2	$\dots$	b_n
a_1	(a_1, b_1)	(a_1, b_2)	$\dots$	(a_1, b_n)
a_2	(a_2, b_1)	(a_2, b_2)	$\dots$	(a_2, b_n)
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
a_m	(a_m, b_1)	(a_m, b_2)	$\dots$	(a_m, b_n)

□

III.22. Для декартова произведения нескольких множеств результат аналогичен:

$$\big|X_1 \times \dots \times X_s\big| = \big|X_1\big| \cdot \ldots \cdot \big|X_s\big|.$$

III.23. Предложение. *Пусть A — конечное множество, $|A| = n$. Мощность множества $\mathcal{P}(A)$ всех подмножеств множества A (т.е. количество всех его различных подмножеств) равна 2^n.*

Доказательство. Найдем число подмножеств множества $A = \{1, 2, \dots, n\}$. Закодируем каждое подмножество $X \subset A$ конечной n-членной последовательностью Y, состоящей из нулей и единиц, поставив в соответствие подмножеству X последовательность Y, на i-м месте которой стоит единица тогда и только тогда, когда $i \in X$. Например, при $n = 3$ имеем

X	$\{\varnothing\}$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
Y	000	100	010	001	110	101	011	111

Очевидно, построенное отображение $f\colon X \mapsto Y$ взаимно однозначно. Но каждая упорядоченная последовательность длины n, состоящая из нулей и единиц, является элементом декартовой степени $\{0,1\}^n$ двухэлементного множества $\{0,1\}$, так что $|\{0,1\}^n| = |\{0,1\}|^n = 2^n$. Таким образом, если A — конечное множество, то множество всех его подмножеств состоит из $2^{|A|}$ элементов; этот факт объясняет, почему множество $\mathcal{P}(A)$ всех подмножеств множества A обозначается также символом 2^A (даже в случае, когда A — бесконечное множество). □

III.24. Предложение III.21 позволяет сформулировать следующее комбинаторное правило (*правило произведения*): если объект α можно выбрать a способами, а после каждого выбора объекта α объект β можно выбрать b способами, то упорядоченную пару элементов (α, β) можно выбрать ab способами.

III.25. Предложение. *Мощность объединения двух множеств может быть найдена при помощи следующих формул:*

(1) *если множества A и B не пересекаются, то*

$$|A \sqcup B| = |A| + |B|;$$

(2) *если множества A и B пересекаются, то*

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Доказательство. Формула для непересекающихся множеств очевидна; более того, легко видеть (и доказать методом математической индукции), что для нескольких непересекающихся множеств справедлива формула

$$\left|\bigsqcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i|.$$

Для двух пересекающихся множеств в сумме $|A| + |B|$ элементы пересечения $A \cap B$ учтены дважды, и чтобы компенсировать это, нужно вычесть $|A \cap B|$ в правой части формулы. Более формальное рассуждение:

$$A \cup B = (A \setminus B) \sqcup (B \setminus A) \sqcup (A \cap B).$$

Обозначив $|A \cap B| = q$, $|A| = a$, $|B| = b$, находим $|A \setminus B| = a - q$, $|B \setminus A| = b - q$, так что

$$|A \cup B| = |A \setminus B| + |B \setminus A| + |A \cap B| = (a - q) + (b - q) + q = a + b - q. \quad \square$$

III.26. Предложение III.25 позволяет сформулировать следующее комбинаторное правило (*правило суммы*): если объект α можно выбрать a способами, а объект β — b способами, то выбор объекта α *или* β можно осуществить $a + b$ способами.

III.27. Определение. *Факториалом* $n!$ натурального числа n называется произведение $1 \cdot 2 \cdot \ldots \cdot n$. По определению полагают $0! = 1$. Факториал как функцию неотрицательного целого аргумента можно определить *рекурсивно*:

$$0! = 1, \quad (n + 1)! = n! \cdot (n + 1).$$

III.28. Значения факториала быстро растут при увеличении n:

n	0	1	2	3	4	5	6	7	8	9	10
$n!$	1	1	2	6	24	120	720	5 040	40 320	362 880	3 628 800

В курсе анализа доказывается *формула Стирлинга*[1], дающая приближённое значение $n!$ при больших значениях n:

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \tag{III.1}$$

где $\pi \approx 3{,}1415\ldots$ — отношение длины окружности к её диаметру, $e \approx 2{,}7182\ldots$ — основание натуральных логарифмов. Символ $\sim$ означает, что отношение левой и правой частей в формуле Стирлинга стремится к единице при $n \to \infty$. Хотя разность правой и левой частей в формуле Стирлинга неограниченно возрастает, относительная ошибка быстро убывает. Например, при $n = 10$ относительная ошибка составляет менее одного процента:

$$a = 10! = 3\,628\,800, \quad b = \sqrt{2\pi \cdot 10} \cdot \left(\frac{10}{e}\right) \approx 3\,598\,696, \quad \frac{a - b}{a} \approx 0{,}0083.$$

Б. Размещения.

III.29. В комбинаторике рассматриваются задачи о комбинаторных конфигурациях, т.е. задачи выбора и расположения элементов некоторого конечного множества в соответствии с заданными правилами. Основными типами комбинаторных конфигураций являются размещения, перестановки и сочетания.

III.30. Определение. *Размещением* объёма k из элементов данного n-множества A (k-размещением, размещением из n по k) называется произвольная упорядоченная последовательность $(a_1, \ldots, a_k)$, состоящая из k элементов множества A (также говорят, что k-размещение — это упорядоченная выборка объёма k из множества A). Если все элементы последовательности $(a_1, \ldots, a_k)$ различны, говорят о *бесповторной* упорядоченной выборке (размещении без повторений), в противном случае упорядоченная выборка называется *повторной* (размещение с повторениями).

[1] Дж. Стирлинг (J. Stirling) — шотландский математик (1692–1770).

III.31. Предложение.

1. *Число* $\overline{A}_n^k$ *повторных упорядоченных выборок объёма* k *из* n*-множества (т.е. размещений с повторениями из* n *по* k*) равно*
$$\overline{A}_n^k = n^k. \tag{III.2}$$
2. *Число* A_n^k *бесповторных упорядоченных выборок объёма* k *из* n*-множества (т.е. размещений без повторений из* n *по* k*) равно*
$$A_n^k = \frac{n!}{(n-k)!}. \tag{III.3}$$

Доказательство. 1. В повторной выборке $(a_1, \ldots, a_k)$ в качестве первого элемента a_1 может быть взят любой элемент исходного n-множества, т.е. имеется n способов выбора. То же относится к выбору каждого последующего элемента. Согласно правилу произведения получаем требуемый результат.

2. При образовании бесповторной выборки $(a_1, \ldots, a_k)$ в качестве первого элемента a_1 может быть взят любой элемент исходного n-множества, т.е. имеется n способов выбора; в качестве второго элемента a_2 может быть взят любой элемент исходного n-множества, кроме a_1, т.е. имеется $n-1$ способов выбора. При выборе каждого последующего элемента число способов выбора уменьшается на 1, и для выбора элемента a_k имеется $n-k+1$ возможностей. Согласно правилу произведения имеем

$$A_n^k = \underbrace{n(n-1)\cdots(n-k+1)}_{k \text{ сомножителей}} =$$
$$= n(n-1)\cdots(n-k+1) \cdot \overbrace{\frac{(n-k)(n-k-1)\cdots 2\cdot 1}{(n-k)(n-k-1)\cdots 2\cdot 1}}^{\text{дробь равна } 1} = \frac{n!}{(n-k)!}. \quad \square$$

В. Перестановки.

III.32. Размещения объёма n, составленные из элементов n-множества A, называются *перестановками* элементов множества A (см. определение 6.18, с. 126). Используя формулу (III.3) при $k=n$, находим число P_n перестановок элементов n-множества (см. предложение 6.21, с. 127):
$$P_n = n!. \tag{III.4}$$

Г. Сочетания и биномиальные коэффициенты.

III.33. Определение. *Сочетанием* объёма k из элементов данного n-множества A (k-сочетанием, сочетанием из n по k) называется произвольное k-подмножество множества A. Сочетания называют также *неупорядоченными бесповторными выборками.*

III.34. Предложение. *Число* C_n^k *неупорядоченных бесповторных выборок объёма* k *из* n*-множества, т.е. количество* k*-подмножеств* n*-множества (число сочетаний из* n *по* k*) равно*[1]
$$C_n^k = \frac{n!}{k!(n-k)!}. \tag{III.5}$$

[1] В зарубежной литературе вместо C_n^k используются обозначения C_k^n, ${}^nC_k$, $\binom{n}{k}$ (обратите внимание на положение индексов!).

Числа C_k^n называются *биномиальными коэффициентами*; причина такого названия вскоре выяснится.

Доказательство. Каждое k-сочетание $\{a_1, \ldots, a_k\}$ из элементов данного n-множества можно превратить в размещение $(a_{i_1}, \ldots, a_{i_k})$, по-разному упорядочивая его элементы, $k!$ способами. Таким образом, каждому сочетанию отвечает $k!$ различных размещений, поэтому

$$C_n^k = \frac{A_n^k}{k!} \overset{(\text{III.3})}{=\!=} \frac{n!}{k!(n-k)!}. \qquad \square$$

III.35. Теорема. *Биномиальные коэффициенты C_n^k обладают следующими свойствами*:

$$1. \quad \sum_{k=0}^{n} C_n^k = 2^n; \tag{III.6}$$

$$2. \quad C_n^0 = C_n^n = 1, \quad C_n^1 = C_n^{n-1} = n; \tag{III.7}$$

$$3. \quad C_n^k = \frac{\overbrace{n(n-1)\cdots(n-k+1)}^{k\ \text{сомножителей}}}{\underbrace{k(k-1)\cdots 1}_{k\ \text{сомножителей}}}; \tag{III.8}$$

$$4. \quad C_n^{n-k} = C_n^k; \tag{III.9}$$

$$5. \quad C_n^k + C_n^{k+1} = C_{n+1}^{k+1}. \tag{III.10}$$

Доказательство. Соотношение (III.6) легко доказывается на основе комбинаторных соображений: так как C_n^k — число k-подмножеств n-множества, то, суммируя эти числа, получим количество всех подмножеств данного n-множества, которое в силу теоремы III.23 равно 2^n. Алгебраическое доказательство будет приведено ниже (см. (III.12), с. 328).

Свойства (III.7), (III.8) вытекают непосредственно из формулы (III.5). Докажем (III.9):

$$C_n^{n-k} = \frac{n!}{(n-k)!(n-(n-k))!} = \frac{n!}{(n-k)!k!} = C_n^k.$$

Отметим, что соотношение (III.9) имеет простой комбинаторный смысл: каждая выборка k элементов из n-множества означает одновременно выборку $n-k$ элементов (конечно, это наблюдение можно положить в основу комбинаторного доказательства соотношения (III.9)).

Докажем (III.10) прямым вычислением:

$$\begin{aligned} C_n^k + C_n^{k+1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k+1)!(n-k-1)!} = \\ &= \frac{n!}{k!(n-k-1)!(n-k)} + \frac{n!}{k!(k+1)(n-k-1)!} = \\ &= \frac{n!(n-k) + n!(k+1)}{k!(n-k-1)!(k+1)(n-k)} = \frac{n!(n+1)}{(k+1)!(n-k)!} = \\ &= \frac{(n+1)!}{(k+1)!((n+1)-(k+1))!} = C_{n+1}^{k+1}. \end{aligned}$$

Можно доказать это равенство и с помощью комбинаторных рассуждений. Предположим, что из урны, содержащей $n+1$ шаров, один из которых чёрный, а остальные белые, извлекают $k+1$ шаров; это можно сделать C_{n+1}^{k+1} способами. Если среди извлечённых шаров имеется чёрный, то выбор может быть осуществлён C_n^k способами (выбираются лишь k шаров из n белых, поскольку один — чёрный — уже взят). Если же среди извлечённых шаров чёрного нет, то количество вариантов выбора равно C_n^{k+1}. По правилу суммы $C_{n+1}^{k+1} = C_n^k + C_n^{k+1}$. □

III.36. На свойстве (III.10) основан простой способ вычисления биномиальных коэффициентов, который можно проиллюстрировать следующей диаграммой:

$$
\begin{tikzcd}[column sep=small]
 & & & C_0^0 \arrow[dl] \arrow[dr] & & & \\
 & & C_1^0 \arrow[dl] \arrow[dr] & & C_1^1 \arrow[dl] \arrow[dr] & & \\
 & C_2^0 \arrow[dl] \arrow[dr] & & C_2^1 \arrow[dl] \arrow[dr] & & C_2^2 \arrow[dl] \arrow[dr] & \\
C_3^0 & & C_3^1 & & C_3^2 & & C_3^3
\end{tikzcd}
$$

Как указывает формула (III.10), каждое число C_{n+1}^{k+1} равно сумме чисел C_n^k и C_n^{k+1}, расположенных в диаграмме над C_{n+1}^{k+1}. Приведённая диаграмма носит название *треугольника Паскаля*[1]. Нумеровать строки треугольника принято сверху вниз, *начиная с нуля* (т.е. номер строки равен соответствующему значению n), а нумеровать числа C_n^k, стоящие в n-q строке, принято слева направо, также *начиная с нуля* (т.е. номер числа в строке равен соответствующему значению k). В соответствии с этим соглашением C_n^k — это k-е число в n-й строке треугольника Паскаля. Первые семь строк (с нулевой по шестую включительно) треугольника Паскаля имеют вид

$$
\begin{array}{c|ccccccccccccc}
n=0 & & & & & & & 1 & & & & & & \\
n=1 & & & & & & 1 & & 1 & & & & & \\
n=2 & & & & & 1 & & 2 & & 1 & & & & \\
n=3 & & & & 1 & & 3 & & 3 & & 1 & & & \\
n=4 & & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
n=5 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
n=6 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1
\end{array}
$$

Приведём без доказательства ещё ряд замечательных свойств, которыми обладают биномиальные коэффициенты C_n^k и треугольник Паскаля.

III.37. Боковые стороны треугольника состоят из единиц, а вдоль линий, параллельных боковым сторонам треугольника, стоят:

(i) натуральные числа;

(ii) *треугольные числа*, т.е. числа кругов, которые можно расположить в виде правильного треугольника;

---

[1]Б. Паскаль (B. Pascal) — французский математик, механик, физик, литератор и философ (1623–1662).

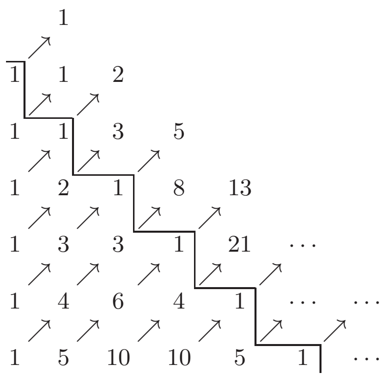
(iii) *тетраэдральные числа*, т.е. числа шаров, которые можно расположить в виде правильного тетраэдра;
(iv) числа, являющиеся обобщениями треугольных и тетраэдральных чисел на многомерный случай.

III.38. Если строки треугольника Паскаля выровнять по левому краю, то суммы чисел, расположенных вдоль наклонных линий, идущих слева направо и снизу вверх (см. диаграмму), равны числам Фибоначчи[1], т.е. элементам последовательности, определённой рекуррентно соотношениями

$$F_0 = 1, \quad F_1 = 1,$$
$$F_n = F_{n-1} + F_{n-2}.$$

Первые десять чисел Фибоначчи:

$$1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ 55.$$



III.39. Если в треугольнике Паскаля все места, занимаемые нечётными числами, окрасить в чёрный цвет, а чётными — в белый, то образуется орнамент, напоминающий один из известных фракталов — *треугольник Серпинского*[2], причём при увеличении числа строк треугольника Паскаля сходство становится все более отчетливым. В пределе, когда число строк треугольника Паскаля неограниченно возрастает, но периметр остается постоянным, и получается треугольник Серпинского.

III.40. Неформально говоря, фрактал — это сложная геометрическая фигура, обладающая свойством самоподобия, т.е. составленная из нескольких частей, каждая из которых подобна всей фигуре целиком. Для фрактала увеличение масштаба («рассматривание фигуры под микроскопом») не ведет к упрощению структуры: на всех шкалах мы увидим одинаково сложную картину.

III.41. Треугольник Серпинского строится следующим образом. Правильный треугольник T_0 делится прямыми, параллельными его сторонам, на четыре равных правильных треугольника и центральный треугольник удаляется. Получается множество T_1, состоящее из трёх оставшихся треугольников (назовем их треугольниками «первого ранга»). Поступая точно так же с каждым из треугольников первого ранга, получим множество T_2, состоящее из девяти равных правильных треугольников второго ранга.

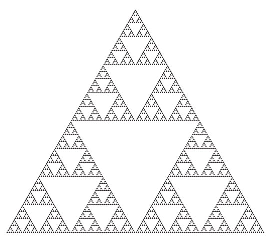


Рис. III.1. Треугольник Серпинского

[1]Леонардо Пизанский, известный под прозвищем Фибоначчи (Leonardo Pisano Fibonacci) (около 1170–1250) — первый крупный математик средневековой Европы.

[2]В. Серпинский (W. Sierpiński) — польский математик (1882–1969)

Продолжая этот процесс, получим бесконечную последовательность $T_0 \supset T_1 \supset \dots \supset T_n \supset \dots$, пересечение членов которой называется треугольником Серпинского.

Д. Бином Ньютона. Известные из школьного курса формулы сокращённого умножения

$$(a+b)^2 = a^2 + 2ab + b^2,$$
$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

допускают обобщение на случай произвольного натурального показателя. Числовые коэффициенты в последовательных слагаемых правых частей приведённых формул равны числам, стоящим во второй и третьей строках треугольника Паскаля. Это наблюдение подтверждается следующей теоремой.

III.42. Теорема. *Для любых чисел a и b и любого натурального n справедлива следующая формула, называемая формулой бинома Ньютона*[1]:

$$\begin{aligned}(a+b)^n = \sum_{k=0}^{n} C_n^k a^{n-k} b^k &= \\ = a^n + C_n^1 a^{n-1} b + C_n^2 a^{n-2} b^2 + \dots + C_n^k a^{n-k} b^k &+ \\ + \dots + C_n^{n-1} a b^{n-1} + b^n&. \qquad \text{(III.11)}\end{aligned}$$

Появление чисел C_n^k в формуле бинома Ньютона объясняет их наименование «биномиальные коэффициенты».

Приведем два доказательства формулы Ньютона.

Комбинаторное доказательство. Раскрывая скобки в произведении

$$(a+b)^n = \underbrace{(a+b) \cdot \dots \cdot (a+b)}_{n \text{ сомножителей}},$$

получим сумму 2^n слагаемых, каждое из которых имеет вид $d_1 \cdot d_2 \cdot \ldots \cdot d_n$, где каждый множитель d_i равен либо a, либо b. Разобьём все слагаемые на несколько подмножеств, относя к подмножеству B_k те слагаемые, в которых b встречается k раз, а a — $n-k$ раз; здесь, очевидно, k может принимать значения от 0 до n. Ясно, что $|B_k| = C_n^k$. Поскольку каждое слагаемое из B_k равно $a^{n-k}b^k$, получаем

$$(a+b)^n = \sum_{k=0}^{n} |B_k| a^{n-k} b^k = \sum_{k=0}^{n} C_n^k a^{n-k} b^k. \qquad \square$$

Индуктивное доказательство. Докажем формулу Ньютона методом математической индукции. Установим сначала базу индукции, т.е. справедливость формулы (III.11) при $n = 1$:

$$(a+b)^1 = a + b = C_1^0 a + C_1^1 b.$$

Предположение индукции состоит в том, что формула (III.11) справедлива при некотором значении n. Наша задача — пользуясь формулой (III.11), вывести её

[1] И. Нью́тон (I. Newton) — великий британский учёный (1643–1727).

справедливость для показателя степени $n+1$:

$$(a+b)^{n+1} = (a+b)\cdot(a+b)^n = (a+b)\left(\sum_{k=0}^{n} C_n^k\ a^{n-k}\ b^k\right) =$$

раскроем первую скобку:

$$= a\left(\sum_{k=0}^{n} C_n^k\ a^{n-k}\ b^k\right) + b\left(\sum_{k=0}^{n} C_n^k\ a^{n-k}\ b^k\right) =$$

внесём множитель «в скобки» (т.е. под знак оператора суммирования):

$$= \sum_{k=0}^{n} C_n^k\ a^{n-k+1}\ b^k + \sum_{k=0}^{n} C_n^k\ a^{n-k}\ b^{k+1} =$$

отделим первое слагаемое первой суммы и последнее слагаемое второй суммы и в полученных суммах произведём замену индекса суммирования:

$$= a^{n+1} + \underbrace{\sum_{k=1}^{n} C_n^k\ a^{n-k+1}\ b^k}_{\substack{\text{Замена } k=p+1,\\ k=1,\dots,n,\ p=0,\dots,n-1}} + \underbrace{\sum_{k=0}^{n-1} C_n^k\ a^{n-k}\ b^{k+1}}_{\text{Замена } k=p} + b^{n+1} =$$

перепишем суммы в новых обозначениях:

$$= a^{n+1} + \underbrace{\sum_{p=0}^{n-1} C_n^{p+1}\ \underline{a^{n-p}\ b^{p+1}} + \sum_{p=0}^{n-1} C_n^p\ \underline{a^{n-p}\ b^{p+1}}}_{\text{подобные слагаемые}} + b^{n+1} =$$

приведём подобные слагаемые:

$$= a^{n+1} + \sum_{p=0}^{n-1} \underbrace{\left(C_n^{p+1} + C_n^p\right)}_{=C_{n+1}^{p+1}}\ a^{n-p}\ b^{p+1} + b^{n+1} =$$

применим свойство (III.10) биномиальных коэффициентов, переобозначим индексы суммирования и напишем коэффициенты $C_{n+1}^0 = C_{n+1}^{n+1} = 1$ в первом и последнем слагаемом:

$$= C_{n+1}^0\ a^{n+1} + \underbrace{\sum_{p=0}^{n-1} C_{n+1}^{p+1}\ a^{(n+1)-(p+1)}\ b^{p+1}}_{\substack{\text{Замена } k=p+1,\\ p=0,\dots,n-1,\ k=1,\dots,n}} + C_{n+1}^{n+1}\ b^{n+1} =$$

$$= C_{n+1}^0\ a^{n+1} + \sum_{k=1}^{n} C_{n+1}^k\ a^{(n+1)-k}\ b^k + C_{n+1}^{n+1}\ b^{n+1} =$$

объединим всё в одну сумму:

$$= \sum_{k=0}^{n+1} C_{n+1}^k \; a^{(n+1)-k} \; b^k. \qquad \square$$

III.43. Формула бинома Ньютона позволяет установить ещё несколько интересных свойств биномиальных коэффициентов. Например, взяв в формуле (III.11) $a = b = 1$, получим

$$(1+1)^n = \sum_{k=0}^{n} C_n^k \; 1^{n-k} \; 1^k \quad \Rightarrow \quad \sum_{k=0}^{n} C_n^k = 2^n \tag{III.12}$$

(этот результат был получен комбинаторным способом в теореме III.35, формула (III.6), с. 323). Аналогично, взяв $a = 1$, $b = -1$, находим

$$\sum_{k=0}^{n} (-1)^k \, C_n^k = 0.$$

4. Бесконечные множества

Напомним (см. определение III.4, с. 316), что множество A называется *бесконечным*, если в нём существует собственное подмножество B, равномощное множеству A. Поскольку известно, что мощности конечных множеств могут быть различными, возникает вопрос, существуют ли бесконечные множества, не являющиеся равномощными.

A. Счётные множества.

III.44. Определение. Множество называется *счётным*, если оно равномощно множеству $\mathbb{N}$ натуральных чисел. Неформально говоря, множество счётно, если его элементы можно перенумеровать, т.е. представить множество в виде бесконечной последовательности $\{x_1, x_2, \dots\}$ (см. определение II.66, с. 315).

III.45. Теорема (свойства счётных множеств)**.**

1. *Любое подмножество счётного множества конечно или счётно.*
2. *Любое бесконечное множество содержит счётное подмножество* (в этом смысле счётные множества являются «наименьшими» бесконечными множествами).
3. *Объединение конечного или счётного числа конечных или счётных множеств является конечным или счётным множеством.*
4. *Прямое произведение конечного числа счётных множеств счётно.*

Доказательство. 1. Пусть множество B является подмножеством счётного множества $A = \{a_1, a_2, \dots\}$. Удалим из последовательности $a_1, a_2, \dots$ те элементы, которые не принадлежат B, сохраняя порядок оставшихся элементов. Тогда оставшиеся элементы образуют либо конечную последовательность (и тогда B конечно), либо бесконечную (и тогда B счётно).

2. Пусть A — произвольное бесконечное множество. Выберем какой-либо элемент b_1. Так как множество A бесконечно, то множество $A \setminus \{b_1\}$ также бесконечно; из него выберем элемент b_2 и отметим, что множество $A \setminus \{b_1, b_2\}$ бесконечно. Продолжая процесс, построим бесконечную последовательность $b_1, b_2, \dots$; построение не прервется ни на каком шаге, поскольку A бесконечно. Множество $B = \{b_1, b_2, \dots\}$ и является искомым счётным подмножеством.

3. Пусть имеется счётное число счётных множеств

$$A_1 = \{a_{11}, a_{12}, \dots\}, \quad A_2 = \{a_{21}, a_{22}, \dots\}, \quad \dots.$$

Расположив элементы каждого из них слева направо в последовательность и записав последовательности одну под другой, получим бесконечную таблицу

$$\begin{array}{cccccc}
a_{11} \rightarrow & a_{12} & a_{13} \rightarrow & a_{14} & a_{15} \rightarrow & \dots \\
a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & \dots \\
a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & \dots \\
a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & \dots \\
a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & \dots \\
\dots & \dots & \dots & \dots & \dots & \dots
\end{array}$$

элементы которой можно записать в виде последовательности

$$a_{11},\ a_{12},\ a_{21},\ a_{31},\ a_{22},\ a_{13},\ a_{14},\ a_{23},\ a_{32},\ a_{41},\ \dots.$$

Если множества A_i не пересекались, то искомое представление для их объединения получено, если пересекались — в построенной последовательности нужно удалить повторения, оставив только один из повторяющихся элементов. Если множеств конечное число или какие-либо из множеств конечны, то в описанной конструкции части членов не будет, и останется либо конечное, либо счётное множество.

4. Пусть $A = \{a_1, a_2, \dots\}$ — счётное множество; докажем по индукции, что A^n счётно. База индукции: $A^1 = A$ — счётное множество. Переход от $n = 1$ к $n = 2$: множество $A^2 = A \times A$ разбивается на счётное число непересекающихся счётных множеств $\{a_1\} \times A$, $\{a_2\} \times A$, … (элементами i-го множества являются пары, первый член которых равен a_i), поэтому A^2 — счётное множество. Аналогично выполняется индукционный переход в общем случае: если известно, что A^n счётно, то $A^{n+1} = A \times A^n$ разбивается на счётное число непересекающихся счётных множеств $\{a_1\} \times A^n$, $\{a_2\} \times A^n$ и т. д. и потому также счётно. $\square$

III.46. Примеры.

1. Множество $\mathbb{Z}$ целых чисел счётно, так как его элементы можно расположить в последовательность $0,\ 1,\ -1,\ 2,\ -2,\ \dots$.
2. Множество $\mathbb{Q}$ рациональных чисел счётно. Рациональные числа представляются несократимыми дробями с целыми числителем и знаменателем. Множество дробей с данным знаменателем счётно, поэтому $\mathbb{Q}$ представимо в виде объединения счётного числа счётных множеств.
3. Множества $\mathbb{N}^k$, $\mathbb{Z}^k$, $\mathbb{Q}^k$, состоящие из всех конечных последовательностей натуральных, целых или рациональных чисел соответственно, счётны.
4. Число называется *алгебраическим*, если оно является корнем ненулевого многочлена с целыми коэффициентами. Множество алгебраических чисел счётно, поскольку многочлен задаётся конечной последовательностью своих коэффициентов (целых чисел), а каждый многочлен имеет конечное число корней.

III.47. Предложение. *Если множество A бесконечно, а множество B конечно или счётно, то $A \cup B \sim A$.*

Доказательство. Без ограничения общности можно считать, что $B \cap A = \varnothing$, т.е. рассматривать дизъюнктное объединение $A \sqcup B$ (в противном случае вместо B рассмотрим множество $B \setminus A$, которое конечно либо счётно; ясно, что $A \cup B = A \sqcup (B \setminus A)$). Выделим в A счётное подмножество P и положим $Q = A \setminus P$; тогда $A = P \sqcup Q$. Нужно доказать, что

$$A \sqcup B = P \sqcup Q \sqcup B \sim P \sqcup Q = A.$$

Поскольку $P \sqcup B$ и P оба счётны, между ними существует биекция $f\colon P \sqcup B \to P$. Рассмотрим отображение

$$g\colon P \sqcup Q \sqcup B \to P \sqcup Q, \quad x \mapsto \begin{cases} f(x) \in P, & \text{если } x \in P \sqcup B, \\ x \in Q, & \text{если } x \in Q. \end{cases}$$

Очевидно, это отображение биективно. □

Б. Несчётные множества.

III.48. Предложение. *Множество*[1] $2^{\mathbb{N}}$*, элементами которого являются бесконечные последовательности, состоящие из нулей и единиц, несчётно.*

Доказательство. Предположим, что множество $2^{\mathbb{N}}$ счётно; тогда все последовательности нулей и единиц можно перенумеровать: $\alpha_1, \alpha_2, \ldots$. Составим бесконечную таблицу, строками которой являются наши последовательности:

$$\begin{array}{l} \alpha_1 = \underline{\alpha_{11}}\ \alpha_{12}\ \alpha_{13}\ \ldots \\ \alpha_2 = \alpha_{21}\ \underline{\alpha_{22}}\ \alpha_{23}\ \ldots \\ \alpha_3 = \alpha_{31}\ \alpha_{32}\ \underline{\alpha_{33}}\ \ldots \\ \qquad \vdots \end{array}$$

(через α_{ij} обозначен j-й член i-й последовательности). Рассмотрим теперь последовательность, образованную стоящими на диагонали членами $\alpha_{11}, \alpha_{22}, \ldots$; её i-й член равен α_{ii} и совпадает с i-м членом i-й последовательности. Заменив все члены этой последовательности на противоположные (т.е. нули на единицы и наоборот), получим последовательность β, у которой $\beta_i = 1 - \alpha_{ii}$, так что последовательность β отличается от любой из последовательностей α_i (в позиции i) и потому отсутствует в таблице. Но мы предположили, что таблица включает в себя все последовательности; полученное противоречие доказывает наше утверждение. □

III.49. Предложение. *Числовой отрезок* $[0,1]$*, множество* $2^{\mathbb{N}}$ *и множество* $\mathcal{P}(\mathbb{N})$ *всех подмножеств натурального ряда равномощны и несчётны:*

$$[0,1] \sim 2^{\mathbb{N}} \sim \mathcal{P}(\mathbb{N}).$$

[1]Обозначение этого множества объясняется следующим образом. Символом Y^X принято обозначать множество всех отображений $f\colon X \to Y$. Согласно этой нотации $\{0,1\}^{\mathbb{N}}$ — это множество отображений $f\colon \mathbb{N} \to \{0,1\}$, ставящих в соответствие каждому натуральному числу нуль или единицу, т.е. бесконечных последовательностей, состоящих из нулей и единиц. Для упрощения записи вместо $\{0,1\}^{\mathbb{N}}$ пишем $2^{\mathbb{N}}$ (число 2 — это мощность множества $\{0,1\}$).

Доказательство. Докажем, что $[0,1] \sim 2^{\mathbb{N}}$. Каждое число $x \in [0,1]$ записывается в виде последовательности нулей и единиц, построенной следующим образом: первый элемент равен 0 или 1 в зависимости от того, попадает ли число x в левую или правую половину отрезка; чтобы определить следующий знак, нужно выбранную половину снова поделить пополам и посмотреть, куда попадёт x и т. д. Обратное соответствие устроено так: последовательности $x_0x_1x_2\ldots$ соответствует число, являющееся суммой бесконечной геометрической прогрессии

$$\frac{x_0}{2} + \frac{x_1}{2^2} + \frac{x_2}{2^3} + \ldots.$$

Описанное соответствие пока не является биективным: двоично-рациональные числа, т.е. дроби вида $m/2^n$, имеют два представления. Например, числу $3/8$ отвечают последовательности $011000\ldots$ и $010111\ldots$. Соответствие станет взаимно однозначным, если отбросить все последовательности «с единицей в периоде», кроме последовательности $1111\ldots$, изображающей единицу. Но таких последовательностей счётное число, поэтому на мощность это не повлияет. Итак, $[0,1] \sim 2^{\mathbb{N}}$.

Докажем, что $2^{\mathbb{N}} \sim \mathcal{P}(\mathbb{N})$. Поставим в соответствие каждой последовательности из $2^{\mathbb{N}}$ (т.е. каждому отображению $f\colon \mathbb{N} \to \{0,1\}$) множество номеров мест, на которых стоят единицы (фактически это отображение является характеристической функцией). Например, последовательности, состоящей из одних нулей, соответствует пустое множество, последовательности, состоящей из одних единиц — весь натуральный ряд, а последовательности $101010\ldots$ — множество нечётных чисел. Полученное соответствие является взаимно однозначным. □

III.50. Несчётное множество вещественных чисел не может совпадать со счётным множеством алгебраических чисел; следовательно, существует вещественное число, не являющееся алгебраическим (т.е. не являющееся корнем никакого ненулевого многочлена с целыми коэффициентами). Такие числа называются *трансцендентными.* Отметим, что широко используемые в математике числа π и e являются трансцендентными.

B. Теорема Кантора и сравнение мощностей.

III.51. Теорема (теорема Кантора)**.** *Никакое множество X не равномощно множеству $\mathcal{P}(X) \equiv 2^X$ всех своих подмножеств.*

Доказательство. Предположим обратное, т.е. пусть $f\colon X \to 2^X$ — биекция. Рассмотрим множество

$$Z = \{x \in X \mid x \notin f(x)\} \subset X,$$

т.е. множество тех элементов $x \in X$, которые не принадлежат соответствующему им подмножеству. Докажем, что подмножеству Z не соответствует ни один элемент множества X. Пусть это не так, т.е. существует такой $z \in X$, что $Z = f(z)$. Тогда $z \in Z \;\Leftrightarrow\; z \notin f(z)$ по построению множества Z и $z \notin f(z)$ $\Leftrightarrow\; z \notin Z$ по предположению $f(z) = Z$. Полученное противоречие показывает, что Z не соответствует ни одному элементу множества X, т.е. отображение f не биективно. □

Определение равномощности уточняет интуитивную идею о множествах «одинакового размера». Дадим формальное определение ситуации, когда одно множество «больше» другого.

III.52. Определение. Будем говорить, что мощность множества A не больше мощности множества B (обозначение $A \preccurlyeq B$), если A равномощно некоторому подмножеству множества B (возможно, самому B).[1]

III.53. Теорема (теорема Кантора—Бернштейна[2])**.** *Если множество A равномощно некоторому подмножеству множества B, а B равномощно некоторому подмножеству множества A, то множества A и B равномощны.*

Заинтересованный читатель может найти доказательство этой теоремы в специальной литературе, посвящённой теории множеств.

III.54. Теорема Кантора—Бернштейна значительно упрощает доказательства равномощности. Например, для доказательства равномощности куба и шара в пространстве достаточно заметить, что из куба можно вырезать маленький шар (гомотетичный, а следовательно, равномощный большому), а из шара — маленький кубик.

III.55. Вернёмся к вопросу о сравнении мощностей. Для данных множеств A и B теоретически имеются четыре возможности:

(i) A равномощно некоторой части B, а B равномощно некоторой части A; в этом случае множества равномощны;

(ii) A равномощно некоторой части B, но B не равномощно никакой части A; в этом случае говорят, что A *имеет меньшую мощность, чем* B;

(iii) B равномощно некоторой части A, но A не равномощно никакой части B; в этом случае говорят, что A *имеет большую мощность, чем* B;

(iv) ни A не равномощно никакой части B, ни B не равномощно никакой части A. Этот случай на самом деле невозможен, но доказательство этого факта выходит за рамки данной книги.

Г. Континуум-гипотеза.

III.56. Мощность счётного множества обозначается символом $\aleph_0$ ($\aleph$ — первая буква еврейского алфавита; читается «áлеф»). Мощность континуума $\mathfrak{c}$ — это мощность множества вещественных чисел $\mathbb{R}$ (а также мощность равномощных ему множеств, например, $2^{\mathbb{N}}$). В силу того, что $\mathbb{R} \sim 2^{\mathbb{N}}$, мощность континуума обозначают также $2^{\aleph_0}$.

III.57. Возникает естественный вопрос: каков смысл индекса 0 в $\aleph_0$? Что такое, например, $\aleph_1$? Обычно $\aleph_1$ обозначает наименьшую несчётную мощность (можно доказать, что таковая существует). *Континуум-гипотеза* утверждает, что $\mathfrak{c} = \aleph_1$.

III.58. Континуум-гипотеза была высказана Г. Кантором в начале 80-х гг. XIX в. Многочисленные попытки самого Кантора и многих математиков конца XIX — начала XX вв. доказать континуум-гипотезу оказались безуспешными. Сложившаяся ситуация привела многих математиков к убеждению, что континуум-гипотеза не может быть решена традиционными средствами теории множеств. Это убеждение было подтверждено методами математической логики и аксиоматической теории множеств. В 1936 К. Гёдель[3] доказал, что

---

[1] Мы сознательно отказываемся от записи $|A| \leqslant |B|$, чтобы избежать ненужной аллюзии со сравнением чисел.

[2] Ф. Бернштейн (F. Bernstein) — немецкий математик (1878–1956).

[3] К. Гёдель (K. Gödel) — австрийский логик, математик и философ математики (1906–1978).

континуум-гипотеза не может быть опровергнута в рамках аксиоматической теории Цермело[1]—Френкеля[2], а в 1963 г. П. Коэн[3] доказал, что и отрицание континуум-гипотезы совместно с этой теорией, так что континуум-гипотезу невозможно доказать с помощью обычных методов теории множеств. Полученные результаты свидетельствуют о том, что на современном этапе развития теории множеств возможны различные подходы к основаниям этой науки, существенно различным образом отвечающие на естественные проблемы, возникающие в теории множеств.

[1]Э. Цермело (E. Zermelo) — немецкий математик (1871–1953).
[2]А. Френкель (A. Fraenkel) — израильский математик (1891–1965).
[3]П. Коэн (P. Cohen) — — американский математик (1934–2007).

ПРИЛОЖЕНИЕ IV

Алгебраические системы

КАЖДОЕ множество может быть снабжено некоторыми отношениями и операциями; например, на множестве вещественных чисел имеется отношение строгого порядка $<$, отношение равенства $=$, операции сложения $+$ и умножения $\cdot$ чисел и т. д. Множества с введёнными на них дополнительными математическими структурами называют алгебраическими системами. В этом разделе кратко представлены простейшие и в то же время наиболее важные типы алгебраических систем.

1. Операции на множестве

В математике под *операцией* понимается отображение, которое одному или нескольким элементам множества (аргументам, или операндам) ставит в соответствие другой элемент (значение).

А. Операции и их атрибуты.

IV.1. Определение. *Унарная операция* на множестве X — это отображение $f\colon X \to X$. Для обозначения унарной операции часто используется *префиксная* форма записи, при которой знак операции ставится перед элементом, к которому она применяется. Примерами унарных операций могут служить операция изменения знака числа $x \mapsto (-x)$, переход от множества к его дополнению $X \mapsto \complement X$ или $X \mapsto \overline{X}$, логическое отрицание высказывания $A \mapsto \neg A$, преобразование подобия в геометрии и т. п.

IV.2. Определение. *Бинарной операцией* f на множестве X называется отображение $f\colon X \times X \to X$, которое ставит в соответствие упорядоченной паре $(x, y) \in X \times X$ некоторый элемент $z = f(x, y)$ множества X. Для обозначения бинарной операции вместо записи $f(x, y)$ используется, как правило, *инфиксная* запись с помощью специальных знаков операций, например, $x + y$ для сложения чисел или $x \cdot y$ для умножения; при такой форме записи знак операции ставится *между* элементами, к которым применяется эта операция.

IV.3. Простейшими примерами бинарных операций являются операции сложения и умножения чисел или числовых функций.

IV.4. Каждая операция обладает определёнными свойствами (*атрибутами*), характеризующими её и определяющими взаимоотношения этой операции с другими отображениями рассматриваемого множества.

IV.5. Определение. Бинарная операция $*\colon X \times X \to X$ на множестве X называется *коммутативной*, если

$$\forall x, y \in X \quad x * y = y * x.$$

IV.6. Примеры.

1. Сложение и умножение вещественных чисел коммутативны:
$$a + b = b + a, \quad a \cdot b = b \cdot a.$$
2. Логические операции конъюнкция и дизъюнкция на множестве высказываний коммутативны:
$$A \wedge B \equiv B \wedge A, \quad A \vee B \equiv B \vee A.$$
3. Объединение и пересечение множеств коммутативны:
$$A \cup B = B \cup A, \quad A \cap B = B \cap A.$$
4. Вычитание, деление и возведение в степень вещественных чисел коммутативными не являются:
$$a - b \neq b - a, \quad \frac{a}{b} \neq \frac{b}{a}, \quad a^b \neq b^a.$$

IV.7. Определение. Бинарная операция $*\colon X \times X \to X$ на множестве X называется *ассоциативной*, если
$$\forall x, y, z \in X \quad (x * y) * z = x * (y * z).$$

IV.8. Примеры.

1. Сложение и умножение вещественных чисел ассоциативны:
$$(a + b) + c = a + (b + c), \quad (a \cdot b) \cdot c = a \cdot (b \cdot c).$$
2. Логические операции конъюнкция и дизъюнкция на множестве высказываний ассоциативны:
$$(A \wedge B) \wedge C \equiv A \wedge (B \wedge C), \quad (A \vee B) \vee C \equiv A \vee (B \vee C).$$
3. Объединение и пересечение множеств ассоциативны:
$$(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C).$$
4. Операция композиции отображений ассоциативна (см. теорему II.54, с. 312):
$$(h \circ g) \circ f = h \circ (g \circ f)$$
для любых отображений $f\colon X \to Y$, $g\colon y \to Z$, $h\colon Z \to W$.
5. Вычитание вещественных чисел не обладает свойством ассоциативности: $(a - b) - c \neq a - (b - c)$.

IV.9. Ассоциативность и коммутативность бинарной операции — независимые атрибуты. Примеры операций, одновременно ассоциативных и коммутативных, читателю хорошо известны. Операция $*$ на $\mathbb{Z}$, заданная правилом $n * m \overset{\text{def}}{=\!=} -n - m$, очевидно, коммутативна, но ассоциативной не является:
$$(1 * 2) * 3 = (-1 - 2) * 3 = -(-1 - 2) - 3 = 0,$$
$$1 * (2 * 3) = -1 - (2 * 3) = -1 - (-2 - 3) = 4.$$

Примером ассоциативной, но не коммутативной операции может служить композиция отображений.

IV.10. Пусть X — множество с заданной на нём бинарной операцией $*$, которую будем для определённости называть умножением и для краткости опускать знак операции, т.е. писать xy вместо $x * y$. Можно различными способами составлять различные произведения, состоящие из n сомножителей, не меняя их порядка:

(i) при $n = 2$ такое произведение лишь одно: $x_1 x_2$;

(ii) при $n = 3$ имеется два произведения: $(x_1x_2)x_3$ и $x_1(x_2x_3)$;
(iii) при $n = 4$ таких произведений уже пять: $((x_1x_2)x_3)x_4$, $(x_1(x_2x_3))x_4$, $x_1((x_2x_3)x_4)$, $x_1(x_2(x_3x_4))$, $(x_1x_2)(x_3x_4)$.

Докажем, что в случае ассоциативной операции все эти произведения совпадают.

IV.11. Предложение. *Если бинарная операция на множестве X ассоциативна, то результат её применения к n элементам множества не зависит от расстановки скобок.*

Доказательство. При $n = 1$ и $n = 2$ доказывать нечего, при $n = 3$ утверждение теоремы представляет собой определение ассоциативности. Применим индукцию по n. Предположим, что утверждение доказано для числа перемножаемых элементов $<n$ и докажем его для n сомножителей, т.е. проверим, что

$$(x_1 \dots x_k)(x_{k+1} \dots x_n) = (x_1 \dots x_l)(x_{l+1} \dots x_n) \tag{IV.1}$$

для любых k и l, $1 \leqslant k, l \leqslant n-1$. Мы выписали в последнем равенстве лишь внешние пары скобок, поскольку по предположению индукции расстановка внутренних скобок несущественна. Назовём расстановку скобок

$$\Big(\dots \big((x_1x_2)x_3\big) \dots x_{k-1}\Big)x_k$$

левонормированной. Если $k = n-1$, то произведение n сомножителей приводится к левонормированному виду:

$$(x_1 \dots x_{n-1})x_n = \Big(\dots \big((x_1x_2)x_3\big) \dots x_{n-1}\Big)x_n.$$

В случае $k < n-1$ ввиду ассоциативности имеем

$$\begin{aligned}(x_1 \dots x_k)(x_{k+1} \dots x_n) &= (x_1 \dots x_k)\big((x_{k+1} \dots x_{n-1})x_n\big) = \\ &= \Big((x_1 \dots x_k)(x_{k+1} \dots x_{n-1})\Big)x_n = \\ &= \bigg(\dots \bigg(\Big(\dots (x_1x_2) \dots x_k\Big)x_{k+1}\bigg) \dots x_{n-1}\bigg)x_n,\end{aligned}$$

т.е. снова получили левонормированную расстановку скобок. К такому же виду приводится правая часть равенства (IV.1). □

IV.12. Определение. Пусть $\circ$ и $*$ — две бинарные операции, заданные на множестве X. Операция $*$ называется *дистрибутивной* относительно операции $\circ$, если

$$\forall x, y, z \in X \quad (x \circ y) * z = (x * z) \circ (y * z), \quad z * (x \circ y) = (z * x) \circ (z * y).$$

IV.13. В случае некоммутативной операции нужно различать дистрибутивность справа (первое из приведённых соотношений) и слева (второе соотношение); если же операция коммутативна, то понятия дистрибутивности слева и справа совпадают.

IV.14. Примеры.

1. Умножение вещественных чисел дистрибутивно относительно сложения:

$$(a+b)c = ac + bc, \quad c(a+b) = ca + cb.$$

Здесь опущены некоторые скобки, что возможно благодаря соглашению о приоритете операций: сначала выполняются все умножения, а затем сложения.

2. Логические операции конъюнкция и дизъюнкция взаимно дистрибутивны:
$$A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C),$$
$$A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C).$$
3. Операции объединения и пересечения множеств взаимно дистрибутивны:
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C),$$
$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C).$$
4. Сложение вещественных чисел не является дистрибутивным относительно умножения: $(a \cdot b) + c \neq (a + c) \cdot (b + c)$.

Б. Нейтральные и обратимые элементы.

IV.15. Определение. Пусть $*$ — бинарная операция на множестве X. Элемент $e \in X$ называется *нейтральным* относительно операции $*$, если
$$\forall x \in X \quad e * x = x * e = x.$$
В случае некоммутативной операции e различают левый нейтральный элемент, определяемый соотношением $e * x = x$, и правый нейтральный элемент, определяемый соотношением $x * e = x$ для каждого $x \in X$.

IV.16. Примеры.
1. Нуль 0 является нейтральным элементом относительно сложения чисел, а единица 1 — нейтральным элементом относительно умножения чисел.
2. Ложное высказывание 0 является нейтральным элементом относительно дизъюнкции, а истинное высказывание 1 — нейтральным элементом относительно конъюнкции:
$$A \vee 0 = A, \quad A \wedge 1 = A.$$
3. Пустое множество является нейтральным элементом относительно операции объединения множеств, а универсальное множество — нейтральным элементом относительно операции пересечения:
$$A \cup \varnothing = A, \quad A \cap U = A.$$

IV.17. Предложение. *Если нейтральный элемент относительно бинарной операции $*$ существует, то он единствен, и, следовательно, все левые и правые нейтральные элементы относительно $*$ с ним совпадают.*

Доказательство. Пусть e и e' — нейтральные элементы относительно операции $*$. Тогда $e' = e' * e = e$, т.е. $e' = e$. □

IV.18. Определение. Пусть $*$ — бинарная операция на множестве X, обладающая нейтральным элементом e. Элемент x' называется *обратным* к элементу $x \in X$ относительно операции $*$, если $x' * x = e = x * x'$. В этом случае элемент x называется *обратимым*, а элементы x и x' — *взаимно обратными.*

IV.19. Очевидно, нейтральный элемент e является обратным самому себе.

IV.20. Примеры.
1. Относительно сложения вещественных чисел обратным к числу $x \in \mathbb{R}$ является противоположное число $-x$.
2. Относительно умножения вещественных чисел обратным к ненулевому числу $x \in \mathbb{R}$ является число $1/x$; число нуль не имеет обратного относительно умножения.

IV.21. Предложение. *Пусть $*$ — ассоциативная операция, x и y — обратимые элементы, x' и y' — их обратные.*

1. *Элемент x' единствен.*
2. *Элемент $x * y$ обратим; его обратным является $y' * x'$.*
3. *Элемент, обратный к x', также обратим, причём $(x')' = x$.*

Доказательство. 1. Пусть u и v — два элемента, обратных к x относительно операции $*$, т.е.

$$x * u = e = u * x, \quad x * v = e = v * x.$$

В силу ассоциативности операции $*$ имеем

$$u = u * e = u * (x * v) = (u * x) * v = e * v = v.$$

2. В силу ассоциативности $*$ имеем

$$(x * y) * (y' * x') = \big((x * y) * y'\big) * x' = \big(x * (y * y')\big) * x' = (x * e) * x' = x * x' = e.$$

Аналогично доказывается, что $(y' * x') * (x * y) = e$.

3. Пользуясь предыдущим фактом, получаем

$$(x')' * x' = (x * x')' = e' = e. \qquad \square$$

IV.22. Пусть $*$ — бинарная операция на множестве X. Подмножество $Y \subset X$ называется *замкнутым* относительно операции $*$, если

$$\forall x, y \in Y \quad x * y \in Y.$$

IV.23. Примеры.

1. Подмножество в $\mathbb{Z}$, состоящее из всех чётных чисел, замкнуто относительно операций сложения и умножения целых чисел.
2. Подмножество в $\mathbb{Z}$, состоящее из всех нечётных чисел, замкнуто относительно операции умножения целых чисел, но не замкнуто относительно операции сложения.

2. Простейшие алгебраические системы

IV.24. *Алгебраическая система* — это множество, на котором задан набор операций и отношений. В этом разделе мы рассмотрим некоторые простейшие и в то же время важные алгебраические системы: группы, ко́льца и поля́.

A. Группы.

IV.25. Определение. Непустое множество G с заданной на нём бинарной операцией $*\colon G \times G \to G$ называется *группой*, если выполняются следующие свойства:

Г1: операция $*$ ассоциативна, т.е.

$$\forall a, b, c \in G \quad a * (b * c) = (a * b) * c;$$

Г2: операция $*$ обладает *нейтральным элементом*:

$$\exists e \in G \; \forall a \in G \quad a * e = e * a = a;$$

Г3: каждый элемент множества G обратим:

$$\forall a \in G \; \exists b \in G \quad a * b = b * a = e;$$

обратный элемент для a обозначается a^{-1}.

Требования **Г1**–**Г3** называются *аксиомами группы*.

IV.26. Группа называется *абелевой*[1], если операция в ней коммутативна. В этом случае операцию обычно называют *сложением* и обозначают символом $+$; такая форма записи групповой операции называется *аддитивной*. Нейтральный элемент абелевой группы называют *нулём* и обозначают 0, а элемент, обратный к $a \in G$, называют *противоположным* и обозначают $-a$. Символом $a - b$ обозначают такой элемент $c \in G$, что $a = b + c$; при заданных $a, b \in G$ этот элемент единствен (докажите!).

IV.27. Для неабелевых групп используется альтернативный способ записи, мультипликативный; в этом случае операцию называют *умножением* и обозначают символом $\cdot$, который в ряде случаев вовсе опускают, записывая ab вместо $a \cdot b$ или $a * b$. Элемент, обратный к a, обозначают a^{-1}.

IV.28. Примеры.

1. Множество, состоящее из единственного элемента e и снабжённое операцией $*$, действующей по правилу $e*e = e$, является группой; элемент e является нейтральным.
2. Множество $\{1, -1\}$ является абелевой группой относительно операции умножения.
3. Множество вещественных чисел с операцией сложения $(\mathbb{R}, +)$ является абелевой группой и называется *аддитивной группой* $\mathbb{R}$. Нейтральным элементом является нуль, а обратным к числу $a \in \mathbb{R}$ — противоположное число $-a \in \mathbb{R}$. Аналогично определяются аддитивные группы $\mathbb{Q}$, $\mathbb{C}$.
4. Множество ненулевых вещественных чисел с операцией умножения $(\mathbb{R} \setminus \{0\}, \cdot)$ является абелевой группой и называется *мультипликативной группой* $\mathbb{R}^*$. Нейтральным элементом является единица, а обратным к числу $a \in \mathbb{R}^*$ — число $a^{-1} \in \mathbb{R}^*$. Аналогично определяются мультипликативные группы чисел $\mathbb{Q}^*$, $\mathbb{C}^*$.
5. Множество *всех* рациональных (вещественных, комплексных) чисел не образует группу относительно умножения, поскольку элемент 0 не имеет обратного относительно этой операции.
6. Множество целых чисел с операцией сложения $(\mathbb{Z}, +)$ — абелева группа. Нейтральным элементом является нуль, а обратным (противоположным, см. п. IV.26) к числу $a \in \mathbb{Z}$ — число $-a \in \mathbb{Z}$. Обратите внимание, что множество целых чисел не является группой относительно операции умножения несмотря на ассоциативность этой операции и наличие нейтрального элемента (единицы), поскольку обратимыми элементами относительно операции умножения являются лишь ± 1.

Читателю рекомендуется привести аналогичные примеры групп и не-групп самостоятельно.

IV.29. Определение. Непустое подмножество $H \subset G$ группы G называется *подгруппой* в G, если оно замкнуто относительно умножения в группе G и взятия обратного элемента, т.е. если выполняются следующие условия:

(i) $ab \in H$ для всех $a, b \in H$,

(ii) $a^{-1} \in H$ для любого $a \in H$.

Поскольку $aa^{-1} = e$, отсюда следует, что $e \in H$. Очевидно, $\{e\}$ и G являются подгруппами группы G; эти подгруппы называют *тривиальными*.

IV.30. Примеры.

[1] В честь норвежского математика Н. Абеля (N. Abel, 1802–1829).

1. Группа $(\mathbb{R}_+, \cdot)$ положительных вещественных чисел с операцией умножения является подгруппой мультипликативной группы $\mathbb{R}^*$.
2. Множество $2\mathbb{Z}$ чётных целых чисел образует подгруппу группы $\mathbb{Z}$ из примера IV.28.6. Множество нечётных целых чисел подгруппой не является (почему?).
3. В аддитивной группе $\mathbb{R}$ имеется цепочка подгрупп $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$.
4. В мультипликативной группе $\mathbb{R}^*$ имеется цепочка подгрупп $\{\pm 1\} \subset \mathbb{Q}^* \subset \mathbb{R}^*$.

Б. Кольца.

IV.31. Определение. *Кольцом* называется непустое множество R с двумя заданными на нём бинарными операциями, называемыми *сложением* и *умножением* (знаки операций $+$ и $\cdot$ соответственно; $\cdot$ обычно опускается), которые обладают следующими свойствами:

K1: относительно сложения $+$ множество R образует абелеву группу, называемую *аддитивной группой кольца*; нейтральный элемент этой группы называется *нулём* и обозначается 0;

K2: умножение $\cdot$ дистрибутивно относительно сложения: для любых $a, b, c \in R$ имеют место соотношения

$$(a+b)\cdot c = a\cdot c + b\cdot c, \quad c\cdot(a+b) = c\cdot a + c\cdot b.$$

Требования **K1** и **K2** называются *аксиомами кольца*. Если, кроме того, операция умножения коммутативна (ассоциативна), то кольцо называется *коммутативным* (*ассоциативным*); отметим, что некоммутативные и неассоциативные кольца встречаются весьма часто. Если операция умножения обладает нейтральным элементом, то он называется *единицей* кольца, а само кольцо — *кольцом с единицей*.

IV.32. Возможна ситуация, при которой произведение ненулевых элементов a, b кольца́ равно нулю: $ab = 0$; такие ко́льца называют *кольцами с делителями нуля. Кольцо без делителей нуля* — это кольцо, в котором из $ab = 0$ вытекает $(a = 0) \vee (b = 0)$. В кольце R без делителей нуля имеет место закон сокращения: если $ac = bc$ (или $ca = cb$) и $c \neq 0$, то $a = b$.

IV.33. Определение. Непустое подмножество $S \subset R$ кольца R называется *подкольцом* в R, если оно замкнуто относительно операций сложения, взятия противоположного элемента и умножения:

(1) $x + y \in S$ для любых $x, y \in S$;
(2) $-x \in S$ для любого $x \in S$;
(3) $x \cdot y \in S$ для любых $x, y \in S$.

В этом случае кольцо R называют *расширением* кольца S.

IV.34. Примеры.

1. Множество $\mathbb{Z}$ целых чисел с операциями сложения и умножения является коммутативным ассоциативным кольцом с единицей $(\mathbb{Z}, +, \cdot)$.
2. Множество $2\mathbb{Z}$ чётных целых чисел с операциями сложения и умножения является коммутативным ассоциативным кольцом без единицы $(2\mathbb{Z}, +, \cdot)$.
3. Цепочка подгрупп аддитивной группы $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$ (см. пример IV.30.3) является одновременно и цепочкой подколец.

4. Множество всех числовых функций $f\colon \mathbb{R} \to \mathbb{R}$ является коммутативным ассоциативным кольцом с единицей относительно обычных операций сложения и умножения функций. Это кольцо обладает делителями нуля; например, произведение функций
$$f(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geqslant 0, \end{cases} \qquad g(x) = \begin{cases} 1, & x < 0, \\ 0, & x \geqslant 0 \end{cases}$$
есть функция, тождественно равная нулю.
5. Множество многочленов от переменной x с вещественными коэффициентами является коммутативным и ассоциативным кольцом с единицей относительно операций сложения и умножения многочленов; это кольцо обозначается $\mathbb{R}[x]$. Это кольцо не имеет делителей нуля, т.е. произведение ненулевых многочленов является ненулевым многочленом (см. п. 2.56, с. 35).
6. На множестве $\mathbb{R} \times \mathbb{R}$ введём операции
$$(x_1,\ y_1) + (x_2,\ y_2) \stackrel{\text{def}}{=\!=} (x_1 + x_2,\ y_1 + y_2),$$
$$(x_1,\ y_1) \cdot (x_2,\ y_2) \stackrel{\text{def}}{=\!=} (x_1 x_2,\ y_1 y_2).$$
Легко проверить, что относительно этих операций множество $\mathbb{R}\times\mathbb{R}$ является коммутативным ассоциативным кольцом с единицей $(1, 1)$, обладающее делителями нуля:
$$(x, 0) \cdot (0, y) = (0, 0).$$

B. Поля.

IV.35. Определение. *Полем* называется коммутативное ассоциативное кольцо с единицей, в котором любой ненулевой элемент обратим.

IV.36. Примеры.
1. Множества рациональных чисел $\mathbb{Q}$ и вещественных чисел $\mathbb{R}$ являются примерами полей, хорошо знакомых читателю из школьного курса. Эти поля называются *числовыми полями.*
2. Кольцо $\mathbb{Z}$ целых чисел полем не является, поскольку в нём обратимы только элементы ± 1.
3. Менее тривиальным примером числового поля является множество $\mathbb{Q}[\sqrt{2}] \stackrel{\text{def}}{=\!=} \bigl\{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\bigr\}$ (проверьте самостоятельно).
4. Множество рациональных функций, т.е. функций вида $f(x) = P(x)/Q(x)$, где $P(x)$ и $Q(x)$ — многочлены от переменной x, также является полем относительно обычных операций сложения и умножения функций.

IV.37. В любом поле
$$ab = 0 \quad \Rightarrow \quad (a = 0) \vee (b = 0).$$
Иными словами, в любом поле отсутствуют делители нуля. Действительно, если $a \neq 0$, то, умножив обе части равенства $ab = 0$ на a^{-1}, получим $b = 0$.

IV.38. Определение. Подмножество $\mathbb{L} \subset \mathbb{K}$ поля $\mathbb{K}$ называется *подполем* в $\mathbb{K}$, если оно замкнуто относительно операций сложения, вычитания (т.е. операции, обратной к сложению), умножения и деления (т.е. операции, обратной к умножению); очевидно, $\mathbb{L}$ при этом само является полем. Поле $\mathbb{K}$ при этом называется *расширением* поля $\mathbb{L}$. Например, поле вещественных чисел $\mathbb{R}$ является расширением поля рациональных чисел $\mathbb{Q}$.

3. Кольца вычетов и поля вычетов

А. Бинарная операция, согласованная с отношением эквивалентности. Пусть на множестве X заданы одновременно бинарная операция $*$ и отношение эквивалентности $\sim$. Напомним, что отношение эквивалентности определяет разбиение множества X на непересекающиеся классы эквивалентных элементов, а множество полученных классов называется фактор-множеством множества X по отношению $\sim$ и обозначается $X/{\sim}$ (см. теорему II.36, с. 308).

IV.39. Бинарная операция $*$ и отношение $\sim$ на множестве X называются *согласованными*, если

$$\forall x, x', y, y' \in X \quad (x \sim x') \wedge (y \sim y') \;\Rightarrow\; x * y \sim x' * y'.$$

В этом случае на фактор-множестве $X/{\sim}$ можно определить бинарную операцию, которую будем обозначать тем же символом $*$, по правилу

$$[x] * [y] = [x * y].$$

Для того чтобы найти результат применения операции $*$ к классам $[x]$ и $[y]$, нужно в каждом из них взять по произвольному представителю, например, $x \in [x]$ и $y \in [y]$, найти $x * y$ и объявить результатом класс $[x * y]$, содержащий результат вычисления. Тот факт, что полученный класс не зависит от выбора представителей, обеспечивается согласованностью операции $*$ и отношения эквивалентности $\sim$.

IV.40. Все свойства операции $*$ на множестве X, имеющие характер тождеств, например, коммутативность, ассоциативность, наличие нейтрального элемента и обратного элемента, наследуются операцией на фактор-множестве. Например, если операция на X — сложение, обладающее нейтральным элементом — нулём 0, то класс эквивалентности $[0]$ является нейтральным элементом (нулём) в $X/{\sim}$; если $(-x)$ — элемент, противоположный элементу $x \in X$, то класс $[-x]$ является элементом фактор-множества $X/{\sim}$, противоположным элементу $[x]$.

Б. Определение колец вычетов.

IV.41. Два целых числа x и y называются *сравнимыми по модулю* m, обозначение $x \equiv y \pmod m$, если разность $x - y$ делится на m. Очевидно, $x \equiv y \pmod m$ тогда и только тогда, когда числа x и y при делении на m дают одинаковый остаток.

IV.42. Докажем, что отношение $\equiv \pmod m$ сравнимости по модулю является отношением эквивалентности.

Рефлексивность: $x \equiv x \pmod m$, поскольку $x - x = 0$ делится на m.

Симметричность: $x \equiv y \pmod m$ означает, что $x - y$ делится на m, а потому и $y - x = -(x - y)$ делится на m, т.е. $y \equiv x \pmod m$.

Транзитивность: соотношения $x \equiv y \pmod m$ и $y \equiv z \pmod m$ означают, что числа $x - y$ и $y - z$ делятся на m; но тогда и число $x - z = (x - y) + (y - z)$ также делится на m, т.е. $x \equiv z \pmod m$.

IV.43. Так как при делении целых чисел на m возможные остатки суть 0, 1, $\ldots$, $m - 1$ (всего имеется m различных остатков), существует ровно m классов эквивалентности по отношению $\equiv \pmod m$; они называются *классами вычетов* по модулю m. Соответствующее фактор-множество состоит из m элементов (классов вычетов) и обозначается $\mathbb{Z}_m$:

$$[0]_m = \{x \in \mathbb{Z} \mid x \equiv 0 \pmod m\} = \{0 + mk,\; k \in \mathbb{Z}\},$$

$$[1]_m = \{x \in \mathbb{Z} \mid x \equiv 1 \pmod m\} = \{1 + mk,\ k \in \mathbb{Z}\}, \quad \dots,$$
$$[m-1]_m = \{x \in \mathbb{Z} \mid x \equiv m-1 \pmod m\} = \{(m-1) + mk,\ k \in \mathbb{Z}\}.$$

(Если модуль сравнения m зафиксирован и в процессе рассуждений не меняется, то для краткости будем писать $[x]$ вместо $[x]_m$.)

IV.44. Отметим, что каждый класс эквивалентности может быть обозначен многими разными способами. Например, классы эквивалентности по модулю 2 суть

$$[0]_2 = [2]_2 = [4]_2 = [-2]_2 = [-4]_2 = \dots,$$
$$[1]_2 = [3]_2 = [5]_2 = [-1]_2 = [-3]_2 = \dots,$$

а по модулю 3 —

$$[0]_3 = [3]_3 = [6]_3 = [-3]_3 = [-6]_3 = \dots,$$
$$[1]_3 = [4]_3 = [7]_3 = [-2]_3 = [-5]_3 = \dots,$$
$$[2]_3 = [5]_3 = [8]_3 = [-1]_3 = [-4]_3 = \dots.$$

IV.45. Любая совокупность чисел, взятых по одному из каждого класса вычетов, называется *полной системой вычетов* по модулю m. Например, полными системами вычетов по модулю m являются

$$0,\ 1,\ \dots,\ m-1 \qquad \text{и} \qquad 1,\ 2,\ \dots,\ m.$$

В случае нечётного модуля, $m = 2k+1$, одной из возможных полных систем вычетов является

$$-k,\ -k+1,\ \dots,\ -2,\ -1,\ 0,\ 1,\ 2,\ \dots,\ k-1,\ k.$$

IV.46. Проверим, что введённое отношение эквивалентности согласовано с операциями сложения и умножения на $\mathbb{Z}$. Пусть

$$a \equiv a' \pmod m \quad \Leftrightarrow \quad a - a' = m\alpha \quad \Leftrightarrow \quad a' = a + m\alpha, \quad \alpha \in \mathbb{Z},$$
$$b \equiv b' \pmod m \quad \Leftrightarrow \quad b - b' = m\beta \quad \Leftrightarrow \quad b' = b + m\beta, \quad \beta \in \mathbb{Z}.$$

Складывая последние равенства, получим

$$a' + b' = (a+b) + m(\alpha + \beta) \quad \Leftrightarrow \quad a + b \equiv a' + b' \pmod m.$$

Для умножения проверка аналогична:

$$a'b' = \big(a + m\alpha\big)\big(b + m\beta\big) = ab + m\underbrace{\big(a\beta + b\alpha + m\alpha\beta\big)}_{\in\mathbb{Z}} \equiv ab \pmod m.$$

IV.47. Таким образом, на множестве классов эквивалентности $\mathbb{Z}_m$ можно определить операции сложения и умножения элементов:

$$[a] + [b] \stackrel{\text{def}}{=\!=} [a+b], \quad [a]\cdot[b] \stackrel{\text{def}}{=\!=} [a \cdot b], \tag{IV.2}$$

после чего оно становится кольцом, называемым *кольцом вычетов* по модулю m.

IV.48. ***Кольцо*** $\mathbb{Z}_2$**.** Кольцо вычетов по модулю 2 состоит из двух элементов — $[0]$ и $[1]$, первый из которых представляет собой класс всех чисел, делящихся на 2 без остатка, т.е. из чётных чисел, а второй класс $[1]$ состоит из нечётных чисел. Операции сложения и умножения соответствуют естественным представлениям о сумме и произведении чётных и нечётных чисел, например, сумма двух нечётных чисел есть чётное число ($[1]+[1]=[0]$) и т. д. Таблицы сложения и умножения в кольце $\mathbb{Z}_2$ имеют следующий вид:

$$\begin{array}{|c||cc|}\hline + & [0] & [1] \\ \hline [0] & [0] & [1] \\ [1] & [1] & [0] \\ \hline \end{array}\,, \qquad \begin{array}{|c||cc|}\hline \cdot & [0] & [1] \\ \hline [0] & [0] & [0] \\ [1] & [0] & [1] \\ \hline \end{array}\,. \tag{IV.3}$$

Таблица умножения показывает, что ненулевой элемент $[1]$ кольца $\mathbb{Z}_2$ обратим, $[1]^{-1}=[1]$, так что кольцо $\mathbb{Z}_2$ является полем. Это простейший пример конечного поля.

IV.49. *Кольцо* $\mathbb{Z}_3$*.* Кольцо вычетов по модулю 3 имеет следующие таблицы сложения и умножения:

$$\begin{array}{|c||ccc|}\hline + & [0] & [1] & [2] \\ \hline [0] & [0] & [1] & [2] \\ [1] & [1] & [2] & [0] \\ [2] & [2] & [0] & [1] \\ \hline \end{array}\,, \qquad \begin{array}{|c||ccc|}\hline \cdot & [0] & [1] & [2] \\ \hline [0] & [0] & [0] & [0] \\ [1] & [0] & [1] & [2] \\ [2] & [0] & [2] & [1] \\ \hline \end{array}\,. \tag{IV.4}$$

Здесь также ненулевые элементы обратимы:

$$[1]^{-1}=[1], \quad [2]^{-1}=[2],$$

поэтому $\mathbb{Z}_3$ — поле.

IV.50. *Кольцо* $\mathbb{Z}_4$*.* В кольце вычетов по модулю 4 таблицы сложения и умножения следующие:

$$\begin{array}{|c||cccc|}\hline + & [0] & [1] & [2] & [3] \\ \hline [0] & [0] & [1] & [2] & [3] \\ [1] & [1] & [2] & [3] & [0] \\ [2] & [2] & [3] & [0] & [1] \\ [3] & [3] & [0] & [1] & [2] \\ \hline \end{array}\,, \qquad \begin{array}{|c||cccc|}\hline \cdot & [0] & [1] & [2] & [3] \\ \hline [0] & [0] & [0] & [0] & [0] \\ [1] & [0] & [1] & [2] & [3] \\ [2] & [0] & [2] & [0] & [2] \\ [3] & [0] & [3] & [2] & [1] \\ \hline \end{array}\,. \tag{IV.5}$$

В кольце $\mathbb{Z}_4$ имеются делители нуля: $[2]\cdot[2]=[0]$. Это кольцо полем не является.

B. Поля вычетов.

IV.51. Предложение. *Кольцо вычетов $\mathbb{Z}_m$ является полем тогда и только тогда, когда m — простое число.*

Доказательство. Предположим, что число m — составное, т.е. $m=kl$, где $1<k,l<m$. Тогда оба класса $[k]_m$ и $[l]_m$ ненулевые, но

$$[k]_m\cdot[l]_m=[kl]_m=[m]_m=[0]_m,$$

т.е. в кольце $\mathbb{Z}_m$ имеются делители нуля, и потому оно не является полем.

Теперь предположим, что число m простое. Пусть $[a]_m\neq[0]_m$, т.е. число a не делится на m. Попытаемся найти обратный элемент для $[a]_m$ подбором, умножая $[a]_m$ по очереди на все элементы кольца. Получим элементы

$$[0]_m,\ [a]_m,\ [2a]_m,\ \ldots,\ [(m-1)a]_m. \tag{IV.6}$$

Докажем, что все эти элементы различны. Действительно, если $[ka]_m=[la]_m$, где $0\leqslant k<l\leqslant m-1$, то $[(l-k)a]_m=0$, т.е. число $(l-k)a$ делится на m,

что невозможно, поскольку ни $l-k$, ни a не делятся на m. Таким образом, в последовательности элементов (IV.6) встречаются все элементы кольца $\mathbb{Z}_m$, в том числе $[1]_m$, что означает обратимость элемента $[a]_m$. $\square$

Г. Характеристика поля.

IV.52. В полях вычетов мы встречаемся с новым явлением, не имевшим места в числовых полях $\mathbb{Q}$ и $\mathbb{R}$: в поле $\mathbb{Z}_p$ (p — простое число) имеет место равенство

$$\underbrace{1+1+\dots+1}_{p \text{ слагаемых}} = 0.$$

IV.53. Определение. Пусть $\mathbb{K}$ — произвольное поле. Наименьшее натуральное число p, для которого в поле $\mathbb{K}$ выполняется равенство

$$\underbrace{1+1+\dots+1}_{p \text{ слагаемых}} = 0, \tag{IV.7}$$

называется *характеристикой* этого поля и обозначается $\operatorname{char}\mathbb{K}$; если такого p не существует, то по определению полагаем $\operatorname{char}\mathbb{K}=0$.

IV.54. Таким образом, поля вычетов $\mathbb{Z}_p$ (p — простое число) имеют характеристику p, а для числовых полей $\operatorname{char}\mathbb{Q}=\operatorname{char}\mathbb{R}=0$.

IV.55. Если $\operatorname{char}\mathbb{K}=p$, то для любого $a\in\mathbb{K}$ имеем

$$\underbrace{a+a+\dots+a}_{p \text{ слагаемых}} = a\cdot(\underbrace{1+1+\dots+1}_{p \text{ слагаемых}}) = 0.$$

IV.56. В произвольном поле $\mathbb{K}$ определим умножение на натуральное число k как умножение на элемент $\underbrace{1+1+\dots+1}_{k \text{ слагаемых}}$. Тогда деление на k можно понимать как деление на этот элемент. В случае, когда k делится на характеристику поля, указанный элемент равен нулю и деление невозможно. Например, в поле $\mathbb{Z}_2$ невозможно деление на $2 \stackrel{\text{def}}{=\!=} [1]_2+[1]_2=[0]_2$.

IV.57. Предложение. *Если характеристика поля не равна нулю, то она является простым числом.*

Доказательство. Пусть $\operatorname{char}\mathbb{K}=p=kl$, где $1<k,l<n$. Тогда

$$\underbrace{1+1+\dots+1}_{p \text{ слагаемых}} = (\underbrace{1+1+\dots+1}_{k \text{ слагаемых}})\cdot(\underbrace{1+1+\dots+1}_{l \text{ слагаемых}}) = 0,$$

так что

$$\text{либо}\quad \underbrace{1+1+\dots+1}_{k \text{ слагаемых}} = 0,\qquad \text{либо}\quad \underbrace{1+1+\dots+1}_{l \text{ слагаемых}} = 0,$$

что противоречит определению характеристики. $\square$

4. Гомоморфизмы и изоморфизмы

IV.58. Пусть на множествах X и X' заданы бинарные операции $\circ$ и $\bullet$ соответственно. Будем говорить, что отображение $f : X \to X'$ *сохраняет операцию*, если

$$\forall x,y\in X \quad f(x\circ y)=f(x)\bullet f(y).$$

IV.59. Определение. Рассмотрим две алгебраические системы с одинаковым количеством операций: множество X с операциями $\circ, *, \ldots$ и множество Y с операциями $\bullet, \star, \ldots$. Отображение $f: X \to Y$ называется *гомоморфизмом* этих алгебраических систем, если оно сохраняет все операции, т.е. для любых $x_1, x_2 \in X$ имеем

$$f(x_1 \circ x_2) = f(x_1) \bullet f(x_2), \quad f(x_1 * x_2) = f(x_1) \star f(x_2), \quad \ldots.$$

IV.60. Определение. Взаимно однозначный гомоморфизм называется *изоморфизмом*. Алгебраические системы, между которыми существует изоморфизм, называют *изоморфными*.

IV.61. Отношение изоморфности алгебраических систем обладает всеми свойствами отношения эквивалентности: оно рефлексивно, симметрично и транзитивно (см. замечание III.3, с. 316).

IV.62. Примеры.

1. Рассмотрим отображение групп $(\mathbb{Z}, +)$ (см. пример IV.28.6) и $(\{\pm 1\}, \cdot)$ (см. пример IV.28.2), заданное правилом
$$f(m) = \begin{cases} 1, & \text{если } m \text{ чётно}, \\ -1, & \text{если } m \text{ нечётно}. \end{cases}$$
Поскольку сумма двух целых чисел одинаковой чётности чётна, а разной чётности — нечётна; поэтому имеем соотношение
$$\forall m, n \in \mathbb{Z} \quad f(m+n) = f(m) \cdot f(n),$$
означающее, что $f\colon \mathbb{Z} \to \{\pm 1\}$ является гомоморфизмом групп.
2. Группа $\{\pm 1\}$ (см. пример IV.28.2) изоморфна аддитивной группе кольца вычетов $\mathbb{Z}_2$ (см. пример IV.48); изоморфизм $f\colon \{\pm 1\} \to \mathbb{Z}_2$ задаётся правилом
$$f(1) = [0]_2, \quad f(-1) = [1]_2.$$
3. Отображение f аддитивной группы $\mathbb{R}$ (см. пример IV.28.3) на мультипликативную группу $\mathbb{R}_+$ положительных вещественных чисел (см. пример IV.30.1), заданное формулой $f\colon x \mapsto a^x$, где $a > 0$, является изоморфизмом:
$$\forall x_1, x_2 \in \mathbb{R} \quad f(x_1 + x_2) = a^{x_1 + x_2} = a^{x_1} \cdot a^{x_2} = f(x_1) \cdot f(x_2).$$
4. Отображение $f\colon \mathbb{Z} \to 2\mathbb{Z}$ аддитивной группы целых чисел (см. пример IV.28.6) в аддитивную группу чётных чисел (см. пример IV.30.2), заданное правилом $f(m) = 2m$ для всех $m \in \mathbb{Z}$, является гомоморфизмом групп:
$$\forall m, n \in \mathbb{Z} \quad f(m+n) = 2(m+n) = 2m + 2n = f(m) + f(n).$$
5. То же самое отображение $f\colon \mathbb{Z} \to 2\mathbb{Z}$, $f(m) = 2m$, не является гомоморфизмом *колец* $(\mathbb{Z}, +, \cdot)$ и $(2\mathbb{Z}, +, \cdot)$ (см. примеры IV.34.1 и IV.34.2), поскольку
$$f(mn) = 2(mn) \neq (2m)(2n) = f(m)f(n).$$
6. Определим отображение $f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ кольца $\mathbb{R} \times \mathbb{R}$ из примера IV.34.6 на кольцо $\mathbb{R}$ формулой $f\big((x, y)\big) = x$. Это отображение является гомоморфизмом колец:
$$f\big((x_1, y_1) + (x_2, y_2)\big) = f\big((x_1 + x_2, y_1 + y_2)\big) = \\ = x_1 + x_2 = f\big((x_1, y_1)\big) + f\big((x_2, y_2)\big),$$

$$f\big((x_1, y_1) \cdot (x_2, y_2)\big) = f\big((x_1x_2, y_1y_2)\big) = \\ = x_1x_2 = f\big((x_1, y_1)\big) \cdot f\big((x_2, y_2)\big).$$

7. Формулы (IV.2) показывают, что отображение
$$f\colon \mathbb{Z} \to \mathbb{Z}_m, \quad a \mapsto [a]_m$$
является гомоморфизмом колец $\mathbb{Z}$ и $\mathbb{Z}_m$.

IV.63. В алгебре мы интересуемся только теми свойствами алгебраических систем, которые выражаются в терминах операций, заданных на этих системах. Две изоморфные алгебраические системы являются «одинаково устроенными» в том смысле, что любое утверждение, сформулированное только в терминах операций, справедливо в одной из этих систем тогда и только тогда, когда оно справедливо в другой. Поэтому безразлично, какую из изоморфных друг другу алгебраических систем изучать: все они являются различными реализациями (моделями) одного и того же абстрактного объекта. Однако различные модели одной и той же алгебраической системы могут обладать специфическими свойствами, облегчающими их исследование; например, при изучении модели геометрического происхождения можно использовать геометрические методы и т. п.

Методические рекомендации

В основе успешного изучения математических дисциплин лежит *понимание*. Невозможно выучить математику «наизусть»: знание и умение решать задачи придёт только после понимания. Поэтому начинать нужно с работы над определениями и терминами; каждое новое понятие должно быть *усвоено*, и только в этом случае возможно его продуктивное использование в дальнейшем.

Математический текст нельзя читать как художественную литературу, требуется *конспектирование*. Доказательство теоремы можно понять, только повторив его самостоятельно, записав собственноручно.

Изучение математики невозможно без решения задач. Однако недостаточно просто получить верный ответ; более важным является *овладение эффективными методами*. Неоценимую помощь в этом могут оказать пособия [5, 6, 18].

Нужно иметь две отдельных тетради для конспектирования теоретического материала и для решения задач.

Метод грубой силы (или лобовой атаки) — это прямой подход к решению задачи, обычно основанный непосредственно на формулировке задачи и определениях используемых в этой задаче концепций. Примером может служить вычисление степени числа при помощи умножения этого числа на себя нужное количество раз. Практически все задачи, встречающиеся в курсе аналитической геометрии, могут быть решены этим методом, однако в большинстве случаев такое решение не будет ни самым быстрым, ни самым красивым; более того, как показывает практика, такое решение, как ни странно, часто оказывается ошибочным. За те столетия, которые прошли с момента возникновения метода координат, математики разработали множество быстрых и эффективных алгоритмов, некоторые из которых будут изучаться в настоящем курсе. Овладение этими методами является одной из основных целей курса, поэтому незнание или неумение пользоваться ими свидетельствует о неудовлетворительном результате обучения. Поэтому хочется предостеречь студентов, приступающих к изучению математики в университете: не нужно изобретать своих методов решения задач, за вас уже всё изобрели Ньютон, Гаусс и Коши, вам нужно только выучить всё это и тем самым заработать себе право на своё собственное слово в науке.

Литература

1. *Акопян А. В., Заславский А. А.* Геометрические свойства кривых второго порядка. — М.: Изд-во МЦНМО, 2007. — 136 с.
2. *Александров П. С.* Курс аналитической геометрии и линейной алгебры. — М.: Наука, 1979. — 512 с.
3. *Александров П. С.* Лекции по аналитической геометрии, пополненные необходимыми сведениями из алгебры. — М.: Наука, 1968. — 912 с.
4. *Беклемишев Д. В.* Курс аналитической геометрии и линейной алгебры. — М.: Физматлит, 2005. — 304 с.
5. *Беклемишева Л. А., Петрович А. Ю., Чубаров И. А.* Сборник задач по аналитической геометрии и линейной алгебре. — М.: Физматлит, 2004. — 496 с.
6. *Бортаковский А. С., Пантелеев А. В.* Аналитическая геометрия в примерах и задачах. — М.: Высшая школа, 2005. — 496 с.
7. *Винберг Э. Б.* Курс алгебры. — М.: Факториал Пресс, 2001. — 544 с.
8. *Ефимов Н. В.* Краткий курс аналитической геометрии. — М.: Физматлит, 2005. — 240 с.
9. *Ильин В. А., Ким Г. Д.* Линейная алгебра и аналитическая геометрия. — М.: Изд-во МГУ, 1998. — 320 с.
10. *Ильин В. А., Позняк Э. Г.* Аналитическая геометрия. — М.: Наука. Физматлит, 1999. — 224 с.
11. *Ильин В. А., Позняк Э. Г.* Линейная алгебра. — М.: Наука. Физматлит, 1999. — 296 с.
12. *Кадомцев С. Б.* Аналитическая геометрия и линейная алгебра. — М.: Физматлит, 2003. — 160 с.
13. *Ким Г. Д., Крицков Л. В.* Алгебра и аналитическая геометрия. Теоремы и задачи. Т. I. — М.: Планета знаний, 2007. — 469 с.
14. *Клетеник Д. В.* Сборник задач по аналитической геометрии. — М.: Наука, 1980. — 240 с.
15. *Кострикин А. И.* Введение в алгебру. Ч. 1. Основы алгебры. — М.: Физматлит, 1994. — 320 с.
16. *Моденов П. С.* Аналитическая геометрия. — М.: Изд-во МГУ, 1967. — 698 с.

17. *Моденов П. С., Пархоменко А. С.* Сборник задач по аналитической геометрии. — М.: Наука, 1976. — 384 с.
18. *Овчинников А. В.* Алгебра и геометрия в вопросах и задачах. Кн.1: Основы алгебры и аналитической геометрии. — М.: ЛЕНАНД, 2016. — 288 с.
19. *Постников М. М.* Лекции по геометрии. Семестр I. Аналитическая геометрия. — М.: Наука, 1979. — 336 с.
20. *Розенфельд Б. А.* Аполлоний Пергский. — М.: Изд-во МЦНМО, 2004. — 176 с.
21. *Цубербиллер О. Н.* Задачи и упражнения по аналитической геометрии. — М.: Лань, 2003. — 336 с.

Именной указатель

Предметный указатель

Учебное издание

Овчинников Алексей Витальевич

**Алгебра и геометрия для студентов-физиков.
Лекционный курс. Семестр 1**

Подписано в печать 1.06.2016 г.
Формат 60x90/16. Объем 22,5 п.л. Тираж 300 экз.
Заказ №

Оригинал-макет подготовлен с использованием
издательской системы LaTeX 2_{ε}.

Физический факультет МГУ им. М. В. Ломоносова
119991 Москва, ГСП-1, Ленинские горы, д. 1, стр. 2

Отпечатано в ППП «Типография «Наука»
121099, Москва, Шубинский пер., 6

ISBN 978-5-8279-0138-9

9 785827 901389 >