А. В. Овчинников

АЛГЕБРА И ГЕОМЕТРИЯ
Лекционный курс

Семестр 1

МОСКВА

2015

Оглавление

ГЛАВА 1

Элементарная аналитическая геометрия

1. Аксиоматика геометрии

А. Система аксиом Евклида—Гильберта. Традиционно считается, что родоначальниками геометрии как науки являются древние греки, перенявшие у египтян ремесло землемерия и измерения объёмов тел и превратившие его в строгую научную дисциплину. От набора рецептов античные геометры перешли к установлению общих закономерностей, составили первые систематические и доказательные труды по геометрии, центральное место среди которых занимают «Начала» Евклида, составленные около 300 г. до н.э. Более двух тысячелетий этот труд считался образцовым изложением в духе аксиоматического метода.

Строгий логический анализ основ геометрии, предпринятый Д. Гильбертом в конце XIX в., показал недостаточность аксиоматики, содержащейся в «Началах», и Гильберт предложил свою систему аксиом. В этой системе основными понятиями являются точка, прямая и плоскость. Основные отношения между понятиями:

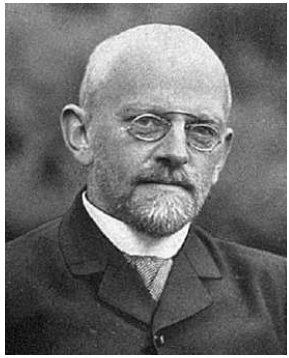
(1) инцидентность («точка лежит на прямой», «прямая проходит через точку» и т. п.); свойства отношения инцидентности описываются восемью аксиомами, четыре из которых планиметрические и четыре — стереометрические;
(2) порядок (понятие «лежать между») — 4 аксиомы;
(3) конгруэнтность (движение, равенство)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

Система аксиом Гильберта обладает рядом недостатков, важнейшими среди которых являются следующие:

(1) рассуждения, основанные на гильбертовой аксиоматике, чрезвычайно сложны и, как правило, требуют хорошего пространственного воображения;
(2) гильбертова система аксиом трудно обобщается на многомерный случай: при попытке обобщения происходит добавление новых исходных понятий и аксиом;
(3) эта система нигде в математике не используется, кроме элементарной геометрии, и потому практическая ценность её, особенно в современной математике, крайне низка.

Существуют и другие системы аксиом, в основе которых могут лежать как другие отношения между основными понятиями (например,

Евклид
(III в. до н.э.)



Д. Гильберт
(1862–1943)

Г. Вейль
(1885–1955)

в системе аксиом Шура вместо конгруэнтности используется движение, а в системе Кагана — расстояние), так и другие основные понятия. Важнейшей среди них является *система аксиом Вейля*, в которой основными понятиями являются точка и вектор, а прямые и плоскости определяются как множества точек. Различные части системы аксиом Вейля используются практически во всех разделах математики, и эта универсальность делает аксиоматику Вейля исключительно важной. Отметим, что все системы аксиом логически эквивалентны, выводятся одна из другой, так что аксиомы одной системы являются теоремами другой.

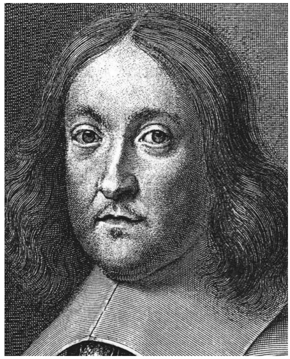
В. Что такое аналитическая геометрия. Аналитическая геометрия — это раздел геометрии, в котором геометрические фигуры и их свойства исследуются средствами алгебры, что становится возможным благодаря так называемому методу координат. При этом точки описываются наборами чисел, более сложные геометрические фигуры — уравнениями и неравенствами, так что геометрическая задача превращается в алгебраическую.

Идея координат и уравнения линии появилась ещё в древнегреческой математике; например, Архимед и Аполлоний 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 древнегреческой алгебры и слабого интереса к линиям, отличных от прямой и окружности, метод координат развития не получил.

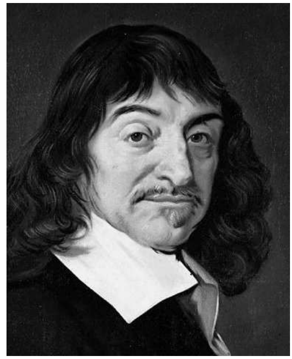
Первые шаги в развитии метода координат связаны с именами Виета, Ферма, Декарта (конец эпохи Возрождения, XVI–XVII вв.).

Мощь нового метода быстро сделала его весьма популярным, и исследования в области геометрии и анализа в Новое время (XVII–XVIII вв.) уже немыслимы без его применения. Своему последующему триумфу метод координат обязан таким гигантам, как Ньютон, Лейбниц, Эйлер.

Ф. Виет
(1540–1603)



П. Ферма
(1601–1665)



Р. Декарт
(1596–1650)

В XVIII столетии аналитическая геометрия, используя мощные методы математического анализа, завоевала новые вершины, и дала начало новому разделу математики — дифференциальной геометрии.

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.

2. Некоторые понятия из школьного курса геометрии

Предполагается, что из курса геометрии средней школы читателю известны основные геометрические понятия и факты. Напомним некоторые из них.

Точки обозначаем прописными латинскими буквами (A, B, ...), прямые — строчными латинскими буквами (a, b, ...), плоскости — строчными греческими буквами (α, β, ...). Прямая, проходящая через точки A и B, обозначается также (AB); плоскость, проходящая через точки A, B и C, обозначается (ABC).

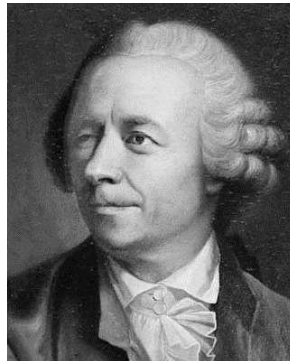
Термин «параллельность» будем использовать в расширенном смысле: две прямые a и b называются параллельными (обозначение $a \parallel b$), если они лежат в одной плоскости и не имеют общих точек либо если они совпадают. Аналогично, две плоскости α и β называются параллельными (обозначение $\alpha \parallel \beta$), если они не имеют общих точек либо совпадают.

Плоскость α и прямая a называются параллельными, если они либо не имеют общих точек, либо прямая лежит в плоскости. Будем считать по определению, что если $\alpha \parallel a$, то $a \parallel \alpha$ и наоборот.

Отрезок $[AB]$ — это множество, состоящее из точек A и B и всех точек прямой (AB), лежащих между A и B. Каждому отрезку $[AB]$ ставится в соответствие неотрицательное число, называемое длиной этого отрезка

И. Ньютон
(1643–1727)

Г. Лейбниц
(1646–1716)



Л. Эйлер
(1707–1783)

(или расстоянием между точками A и B) и обозначаемое $|AB|$. Длина отрезка обладает следующими свойствами:

(1) $|AB| = |BA|$ для любых точек A и B;
(2) $|AB| \geqslant 0$ для любых точек A и B, причём $|AB| = 0$ тогда и только тогда, когда $A = B$;
(3) $|AB| \leqslant |AC| + |CB|$ для любых точек A, B и C, причём $|AB| = |AC| + |CB|$ тогда и только тогда, когда $C \in [AB]$.

Любая точка A, лежащая на прямой a, разбивает эту прямую на два луча a_1 и a_2 с началом в точке A; эти лучи называются дополнительными друг к другу. Точка A принадлежит обоим лучам. Две точки $B \neq A$ и $C \neq A$ принадлежат одному лучу тогда и только тогда, когда $A \notin [BC]$, и принадлежат дополнительным лучам только тогда, когда $A \in [BC]$. Луч с началом в точке A, на котором лежит точка $B \neq A$, обозначается $[AB)$.

Два луча, лежащие на одной прямой, называются одинаково направленными, если их пересечение является лучом, и противоположно направленными в противном случае. Так, лучи $[CA)$ и $[AB)$ на рис. 1.1 одинаково направлены, а лучи $[CB)$ и $[AC)$ — противоположно направлены.

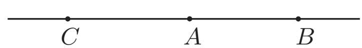


Рис. 1.1.

Каждая прямая a, лежащая в плоскости α, разбивает эту плоскость на две полуплоскости α_1 и α_2; будем говорить, что эти полуплоскости определяются прямой a. Сама прямая a принадлежит обеим полуплоскостям. Две точки A и B лежат в одной полуплоскости, определяемой прямой a, тогда и только тогда, когда $[AB] \cap a = \varnothing$ (см. рис. 1.2).

Два луча $[AB)$ и $[CD)$, лежащие на параллельных прямых, лежат также в одной плоскости. Лучи $[AB)$ и $[CD)$ называются *одинаково направленными*, или *сонаправленными* (обозначение $[AB) \uparrow\uparrow [CD)$), если

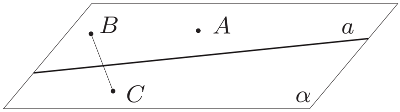


Рис. 1.2.

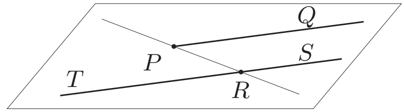


Рис. 1.3.

они лежат в одной полуплоскости, определяемой прямой (AC), и *противоположно направленными* (обозначение $[AB) \uparrow\downarrow [CD)$), если они лежат в разных полуплоскостях. Так, на рис. 1.3 $[PQ) \uparrow\uparrow [RS)$, $[PQ) \uparrow\downarrow [RT)$.

Если на прямой a выбраны три точки A, B, C (A лежит между B и C), то лучи $[AB)$ и $[AC)$ противоположно направлены; любой другой луч, лежащий на этой прямой, сонаправлен либо с $[AB)$, либо с $[AC)$. Таким образом, *на прямой a можно определить два взаимно противоположных направления* (ориентации), каждое из которых представляет собой множество всех лучей, сонаправленных либо с $[AB)$, либо с $[AC)$.

Если на прямой l выбрано одно из двух возможных направлений, то эта прямая называется *ориентированной прямой*, или *осью*, и обозначается $\vec{l}$.

На чертежах положительным направлением считается направление слева направо или снизу вверх. Разумеется, никакого точного математического смысла такой термин не имеет, а используется лишь для наглядности.

3. Направленные отрезки

Направленным отрезком $\overrightarrow{AB}$ называется упорядоченная пара точек A и B, первая A из которых называется началом, а вторая B — концом направленного отрезка $\overrightarrow{AB}$. 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 $\mathbf{0}_A$.

Направленный отрезок $\overrightarrow{AB}$ называют также *связанным вектором*, а его начало A — точкой приложения.

Говорят, что направленный отрезок $\overrightarrow{AB}$ параллелен прямой a (плоскости α), если $(AB) \parallel a$ (соответственно, $(AB) \parallel \alpha$) либо $\overrightarrow{AB} = \mathbf{0}_A$.

Направленные отрезки $\overrightarrow{A_1B_1}, \ldots, \overrightarrow{A_kB_k}$ называются *коллинеарными* (*компланарными*), если все они параллельны одной и той же прямой (соответственно, плоскости).

Длиной $|\overrightarrow{AB}|$ направленного отрезка $\overrightarrow{AB}$ называется длина отрезка $[AB]$; часто для краткости вместо $|\overrightarrow{AB}|$ пишут $|AB|$. Очевидно, длина нулевого направленного отрезка равна нулю.

Осью координат называется ось, на которой зафиксированы некоторая точка O, называемая

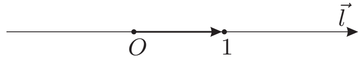


Рис. 1.4.

началом координат, и масштабный направленный отрезок, длина которого полагается равной единице; часто этот отрезок называют *ортом* рассматриваемой оси (см. рис. 1.4).

Ненулевые направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{CD}$ называются *одинаково направленными*, или *сонаправленными* (обозначение $\overrightarrow{AB} \uparrow\uparrow \overrightarrow{CD}$), если сонаправлены лучи $[AB)$ и $[CD)$, и *противоположно направленными* (обозначение $\overrightarrow{AB} \uparrow\downarrow \overrightarrow{CD}$), если лучи $[AB)$ и $[CD)$ имеют противоположное направление.

Направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{CD}$ называются *равными*, если середины отрезков $[AD]$ и $[BC]$ совпадают; обозначение $\overrightarrow{AB} = \overrightarrow{CD}$. На рис. 1.5 изображены все возможные случаи взаимного расположения равных направленных отрезков.

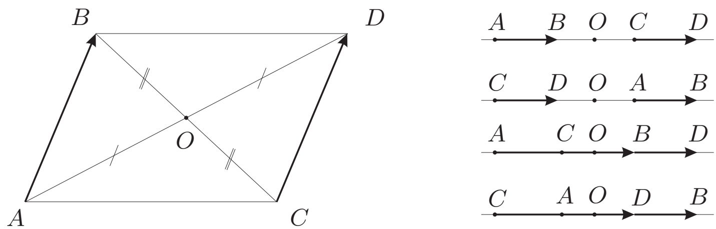


Рис. 1.5. Равные направленные отрезки

Из свойств параллелограмма следует, что ненулевые направленные отрезки, не лежащие на одной прямой, равны тогда и только тогда, когда четырёхугольник $ABDC$ — параллелограмм.

Отметим, что понятие равенства направленных отрезков отличается от понятия равенства чисел: два числа называются равными, если они совпадают; два направленных отрезка могут быть равными, но при этом не совпадать, будучи отложенными от разных точек. Более того, понятие равенства направленных отрезков может быть введено иначе, чем это было сделано выше, и в результате получится объект с совершенно иными свойствами (см. п. В лекции III, с. 39).

1.1. Теорема. *Направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{CD}$ равны тогда и только тогда, когда они имеют одинаковые длины и одинаковые направления:*

$$\overrightarrow{AB} = \overrightarrow{CD} \iff |\overrightarrow{AB}| = |\overrightarrow{CD}| \quad \text{и} \quad \overrightarrow{AB} \uparrow\uparrow \overrightarrow{CD}.$$

Кроме того, в этом случае $\overrightarrow{BC} = \overrightarrow{AD}$.

1.2. Теорема (об откладывании направленного отрезка)**.** *Для любого направленного отрезка $\overrightarrow{AB}$ и любой точки C существует единственная точка D такая, что $\overrightarrow{AB} = \overrightarrow{CD}$. Иными словами, направленный отрезок можно перенести в любую точку (или отложить от любой точки).*

Эти теоремы мы доказывать не будем, поскольку техника «синтетических» доказательств (т.е. доказательств в духе школьного курса геометрии, основанных на гильбертовой системе аксиом), лежит за рамками наших интересов.

4. Ориентированные углы

Пусть A, B, O — три точки в плоскости α, не лежащие на одной прямой. Прямая (OA) разбивает плоскость α на две полуплоскости; обозначим через α_1 ту из них, которая содержит точку B. Аналогично, прямая (OB) разбивает плоскость α на две полуплоскости, и мы обозначим через α_2 ту из них, которая содержит точку A. Пересечение полуплоскостей α_1 и α_2 называется *выпуклым углом* между лучами $[OA)$ и $[OB)$ и обозначается $\angle AOB$. Итак, $\angle AOB = \alpha_1 \cap \alpha_2$. Точка O называется вершиной угла $\angle AOB$, лучи $[OA)$ и $[OB)$ — его сторонами. Каждому углу ставится в соответствие его угловая мера (величина), измеряемая в градусах или радианах; в первом случае величина угла принимает значения в промежутке от 0° до 180°, во втором — от 0 до π.

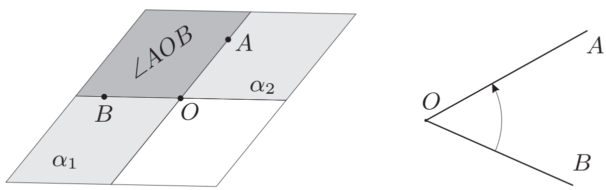


Рис. 1.6.

Углом $\angle AOB$ называют также совокупность двух лучей $[OA)$ и $[OB)$, имеющих общее начало.

Ориентированным углом $\angle AOB$ называется *упорядоченная* совокупность двух лучей $[OA)$ и $[OB)$, имеющих общее начало. В обозначении ориентированного угла $\angle AOB$ сначала указывается первая сторона угла.

Каждый угол может быть ориентирован одним из двух возможных способов. Откладывая до лекции III (см. с. 51) детальное изложение вопроса об ориентации, отметим, что выбрав на плоскости ориентированный угол (ненулевой и не развернутый), мы можем задать ориентацию плоскости (подобно тому как выбор одного из двух возможных направлений на прямой задаёт на этой прямой ориентацию) «по часовой стрелке» или «против часовой стрелки».

Будем считать ориентированный угол *положительным*, если кратчайший поворот от его первой стороны в направлении второй стороны производится *против часовой стрелки*; в противном случае ориентированный угол считается отрицательным. Как и в случае направления на прямой, эти термины не имеют точного математического смысла и используются лишь для наглядности.

Две пересекающиеся прямые образуют четыре попарно равных угла; величина меньшего из них называется величиной угла между прямыми.

Таким образом, угол между прямыми принимает значения от 0° до 90° (или от 0 до $\pi/2$).

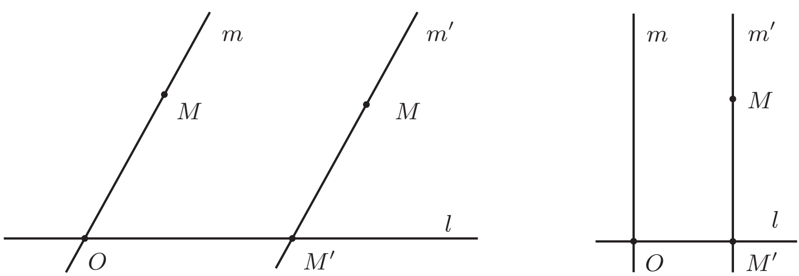
Очевидным образом определяются понятия угла и ориентированного угла между двумя направленными отрезками, между направленным отрезком и осью, между двумя осями.

5. Проектирование

В элементарной геометрии важную роль играют разнообразные операции проектирования. Будем рассматривать следующие виды проекций.

A. Проекция точки M на прямую l параллельно прямой m. Пусть на плоскости заданы две прямые l и m, пересекающиеся в точке O. *Проекцией точки M на прямую l параллельно прямой m* называется точка M' пересечения прямой l и прямой m', проходящей[1] через M параллельно m (см. рис. 1.7(a)). Будем обозначать проекцию M' точки M на прямую l параллельно прямой m следующим образом:

$$M' = \mathrm{Pr}_l^{\parallel m} M.$$



(a) (b)

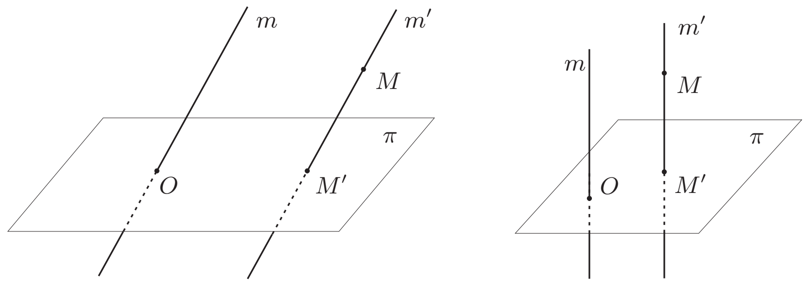
Рис. 1.7. (a) Проекция точки M на прямую l параллельно прямой m; (b) ортогональная проекция точки M на прямую l

Если прямые l и m перпендикулярны, то проекция называется *ортогональной* (см. рис. 1.7(b)) и обозначается $\mathrm{Pr}_l^{\perp}$ или Pr_l.

B. Проекция точки M на плоскость π параллельно прямой m. Пусть в пространстве заданы плоскость π и прямая m, пересекающиеся в точке O. *Проекцией точки M на плоскость π параллельно прямой m* называется точка M' пересечения плоскости π и прямой m', проходящей через M параллельно m (см. рис. 1.8(a)). Будем обозначать проекцию M' точки M на плоскость π параллельно прямой m следующим образом:

$$M' = \mathrm{Pr}_\pi^{\parallel m} M.$$

[1] Если точка M лежит на прямой m, то $m' = m$, и проекцией точки M является точка O.



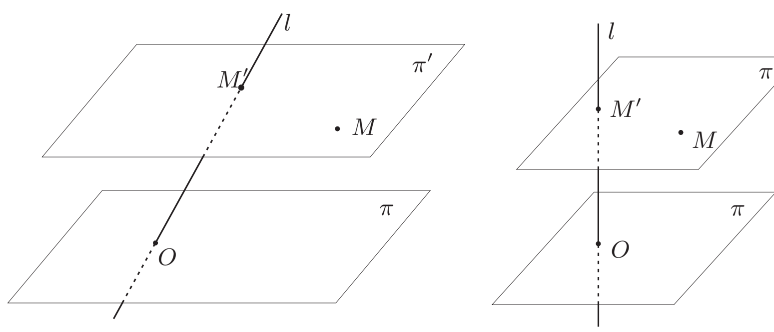
(a) (b)

Рис. 1.8. (a) Проекция точки M на плоскость π параллельно прямой m; (b) ортогональная проекция точки M на плоскость π

Если плоскость π и прямая m перпендикулярны, то проекция называется *ортогональной* (см. рис. 1.8(b)) и обозначается $\mathrm{Pr}_{\pi}^{\perp}$ или Pr_{π}.

C. Проекция точки M на прямую l параллельно плоскости π. Пусть в пространстве заданы прямая l и плоскость π, пересекающиеся в точке O. *Проекцией точки M на прямую l параллельно плоскости π* называется точка M' пересечения прямой l и плоскости π', проходящей через M параллельно π (см. рис. 1.9(a)). Будем обозначать проекцию M' точки M на прямую l параллельно плоскости π следующим образом:

$$M' = \mathrm{Pr}_{l}^{\parallel \pi} M.$$



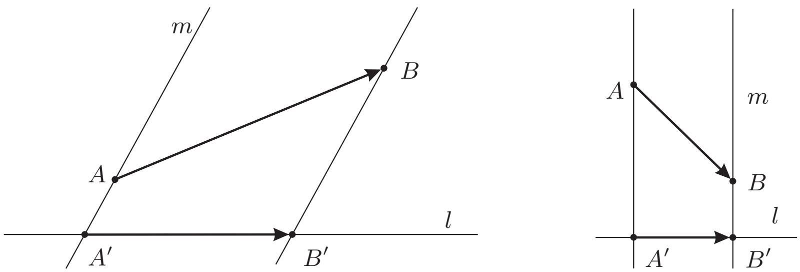
(a) (b)

Рис. 1.9. (a) Проекция точки M на прямую l параллельно плоскости π; (b) ортогональная проекция точки M на прямую l

Если прямая l и плоскость π перпендикулярны, то проекция называется *ортогональной* (см. рис. 1.9(b)) и обозначается $\mathrm{Pr}_{l}^{\perp}$ или Pr_{l}.

D. Проекция направленного отрезка $\overrightarrow{AB}$ на прямую l и на ось координат $\vec{l}$ параллельно прямой m (планиметрическая ситуация). *Проекцией направленного отрезка $\overrightarrow{AB}$ на прямую l параллельно прямой m* называется *направленный отрезок* $\overrightarrow{A'B'}$, где A' и B' — проекции точек A и B на прямую l параллельно прямой m (см. рис. 1.10). Будем обозначать проекцию $\overrightarrow{A'B'}$ направленного отрезка $\overrightarrow{AB}$ на прямую l параллельно прямой m следующим образом:

$$\overrightarrow{A'B'} = \mathrm{Pr}_l^{\parallel m}\,\overrightarrow{AB}.$$



(a) (b)

Рис. 1.10. (a) Проекция направленного отрезка $\overrightarrow{AB}$ на прямую l параллельно прямой m; (b) ортогональная проекция направленного отрезка $\overrightarrow{AB}$ на прямую l

Проекцией направленного отрезка $\overrightarrow{AB}$ на ось координат $\vec{l}$ параллельно прямой m называется *число*

$$\mathrm{pr}_{\vec{l}}^{\parallel m}\,\overrightarrow{AB} \equiv \overline{AB} \stackrel{\text{def}}{=\!=} \begin{cases} |\overrightarrow{A'B'}|, & \text{если } \overrightarrow{A'B'} \uparrow\uparrow \vec{l}, \\ -|\overrightarrow{A'B'}|, & \text{если } \overrightarrow{A'B'} \uparrow\downarrow \vec{l}, \end{cases}$$

где $\overrightarrow{A'B'} = \mathrm{Pr}_l^{\parallel m}\,\overrightarrow{AB}$ — проекция направленного отрезка $\overrightarrow{AB}$ на прямую l параллельно прямой m.

Так, на рис. 1.11 проекция вектора $\overrightarrow{AB}$ на ось $\vec{l}$ равна 2, а проекция вектора $\overrightarrow{CD}$ на ось $\vec{l}$ равна -1.

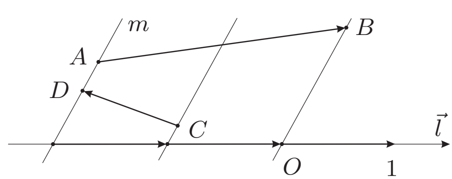


Рис. 1.11.

Если прямые l и m перпендикулярны, то обе эти проекции называются ортогональными и обозначаются $\mathrm{Pr}_l^{\perp}$ и $\mathrm{pr}_{\vec{l}}^{\perp}$ (значок $\perp$ также может быть опущен).

Легко видеть, что ортогональная проекция направленного отрезка $\overrightarrow{AB}$ на ось координат $\vec{l}$ равна

$$\mathrm{pr}_{\vec{l}}\overrightarrow{AB} = \left|\overrightarrow{AB}\right| \cdot \cos\varphi,$$

где φ — величина угла между осью $\vec{l}$ и направленным отрезком $\overrightarrow{AB}$ (см. рис. 1.12).

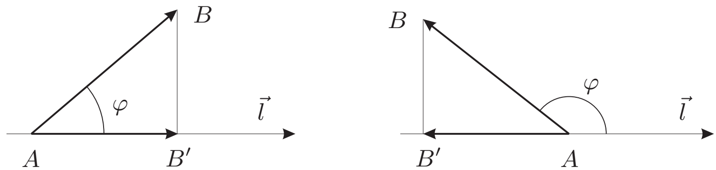


Рис. 1.12. Ортогональная проекция направленного отрезка на ось координат

Обратите внимание на различие понятий проекции направленного отрезка на прямую и на ось координат: первое из указанных понятий представляет собой направленный отрезок, второе — число.

E. Проекция направленного отрезка $\overrightarrow{AB}$ на прямую l и на ось координат $\vec{l}$ параллельно плоскости π (стереометрическая ситуация). *Проекцией направленного отрезка $\overrightarrow{AB}$ на прямую l параллельно плоскости π* называется *направленный отрезок $\overrightarrow{A'B'}$*, где A' и B' — проекции точек A и B на прямую l параллельно плоскости π (см. рис. 1.13). Будем обозначать проекцию $\overrightarrow{A'B'}$ направленного отрезка $\overrightarrow{AB}$ на прямую l параллельно плоскости π следующим образом:

$$\overrightarrow{A'B'} = \mathrm{Pr}_l^{\parallel\pi}\, \overrightarrow{AB}.$$

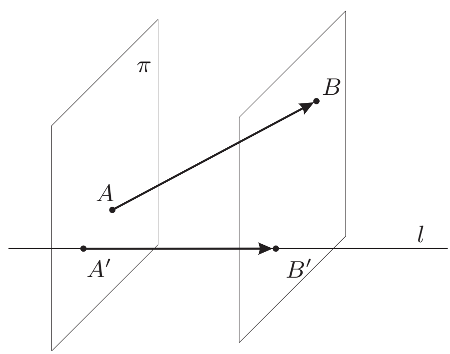


Рис. 1.13. Проекция направленного отрезка на прямую l параллельно плоскости π

Проекцией направленного отрезка $\overrightarrow{AB}$ на ось координат $\vec{l}$ параллельно плоскости π называется *число*

$$\mathrm{pr}_{\vec{l}}^{\parallel\pi} \equiv \overline{AB} \stackrel{\text{def}}{=\!=} \begin{cases} \ \ |\overrightarrow{AB}|, & \text{если } \overrightarrow{A'B'} \uparrow\uparrow \vec{l}, \\ -|\overrightarrow{AB}|, & \text{если } \overrightarrow{A'B'} \uparrow\downarrow \vec{l}, \end{cases}$$

где $\overrightarrow{A'B'} = \mathrm{Pr}_l^{\parallel\pi}$ — проекция направленного отрезка $\overrightarrow{AB}$ на прямую l параллельно плоскости π.

Если прямая l и плоскость π перпендикулярны, то оба указанных вида проекций называются ортогональными и обозначаются $\mathrm{Pr}_l^{\perp}$ (или Pr_l) и $\mathrm{pr}_{\vec{l}}^{\perp}$ (или $\mathrm{pr}_{\vec{l}}$).

F. Проекция направленного отрезка $\overrightarrow{AB}$ на плоскость π параллельно прямой m. Проекцией направленного отрезка $\overrightarrow{AB}$ на плоскость π параллельно прямой m называется направленный отрезок $\overrightarrow{A'B'}$, где A' и B' — проекции точек A и B на плоскость π параллельно прямой m (см. рис. 1.14). Обозначение:

$$\overrightarrow{A'B'} = \mathrm{Pr}_{\pi}^{\parallel m}\, \overrightarrow{AB}.$$

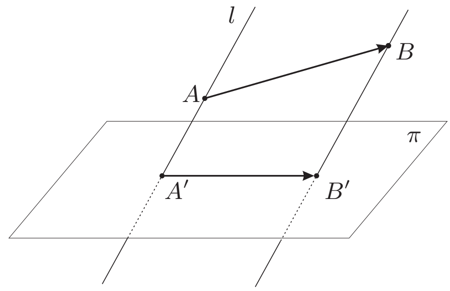


Рис. 1.14. Проекция направленного отрезка на плоскость π параллельно прямой m

G. Простейшие свойства проекций.

1. Проекция направленного отрезка $\overrightarrow{AB}$ равна нулю (т.е. является нулевым направленным отрезком) тогда и только тогда, когда он параллелен той прямой или плоскости, вдоль которой производится проектирование (см. рис. 1.15).
2. Проекции направленного отрезка $\overrightarrow{AB}$ на две параллельные прямые (плоскости) равны между собой (см. рис. 1.16).
3. Проекции равных направленных отрезков на одну и ту же прямую (плоскость, ось координат) равны.

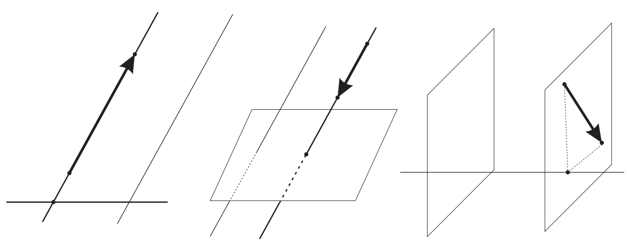


Рис. 1.15.

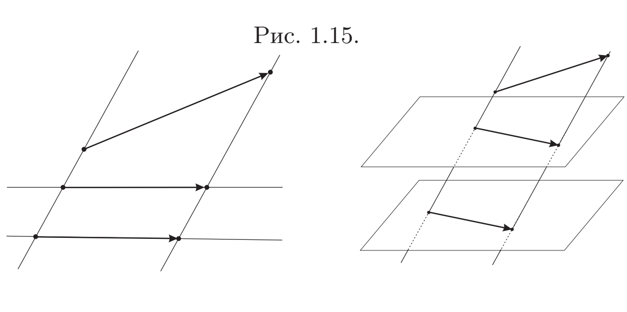


Рис. 1.16.

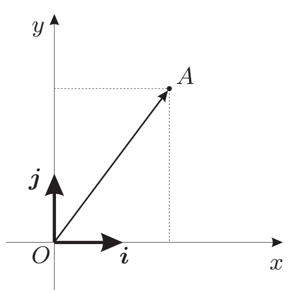
6. Системы координат

Системой координат называется совокупность геометрических фигур на плоскости (в пространстве), позволяющих однозначно описать положение каждой точки плоскости (пространства) с помощью набора из нескольких чисел. Система координат позволяет описывать геометрические объекты алгебраическими средствами.

Рассмотрим наиболее употребительные системы координат на плоскости и в пространстве.

А. Декартова прямоугольная система координат. *Декартова прямоугольная система координат* Oxy на плоскости ($Oxyz$ в пространстве) состоит из двух (соответственно, трёх) взаимно перпендикулярных осей координат с общим началом O и ортами $\boldsymbol{i}$, $\boldsymbol{j}$ (и $\boldsymbol{k}$ для пространства) одинаковой длины. Оси координат обозначаются Ox, Oy и Oz и называются *осью абсцисс*, *осью ординат* и *осью аппликат* соответственно. В пространственном случае плоскости Oxy, Oyz и Oxz называются координатными плоскостями.

Для каждой точки A пространства рассмотрим направленный отрезок $\overrightarrow{OA}$, называемый *радиус-вектором* точки A. Ортогональные проекции x, y и z радиус-вектора $\overrightarrow{OA}$ на оси координат Ox, Oy и Oz соответственно (или, равносильно, проекции на оси координат параллельно координатным плоскостям) называются *координатами точки* A (и также *координатами направленного отрезка* $\overrightarrow{OA}$); x — *абсциссой*, y — *ординатой*, z — *аппликатой*. Запись $A(x, y, z)$ означает, что точка A имеет координаты x, y и z; для направленного отрезка используется запись $\overrightarrow{OA} = (x, y, z)$. Планиметрические определения формулируются аналогично.



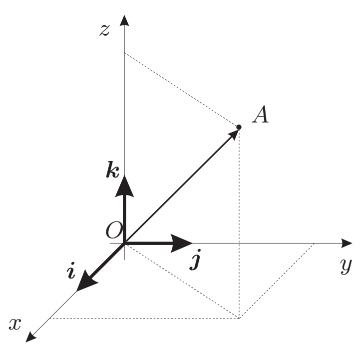


Рис. 1.17. Декартова прямоугольная система координат на плоскости и в пространстве

Используются также другие обозначения: оси координат обозначаются Ox_1, Ox_2, Ox_3, орты — $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ координаты точки — x_1, x_2, x_3.

B. 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):

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

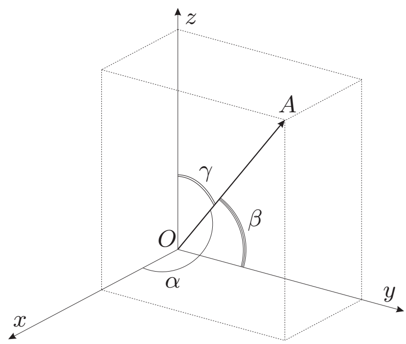


Рис. 1.18.

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

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

отсюда получаем

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

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

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1.$$

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 $\overrightarrow{BC}$ (такой направленный отрезок $\overrightarrow{OA}$ существует в силу теоремы 1.2).

C. Декартова косоугольная система координат. *Декартова косоугольная система координат* на плоскости (в пространстве) состоит из двух (соответственно, трёх) осей координат с общим началом O и ортами $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ (и $\boldsymbol{e}_3$ для пространства). Углы между осями координат не обязательно должны быть прямыми, а орты могут иметь различные длины. Косоугольные системы координат удобны во многих геометрических задачах, в частности, в задачах геометрической кристаллографии.

Координатами точки A (и направленного отрезка $\overrightarrow{OA}$) называются проекции направленного отрезка $\overrightarrow{OA}$ на оси координат параллельно координатным плоскостям (и аналогично в планиметрическим случае).

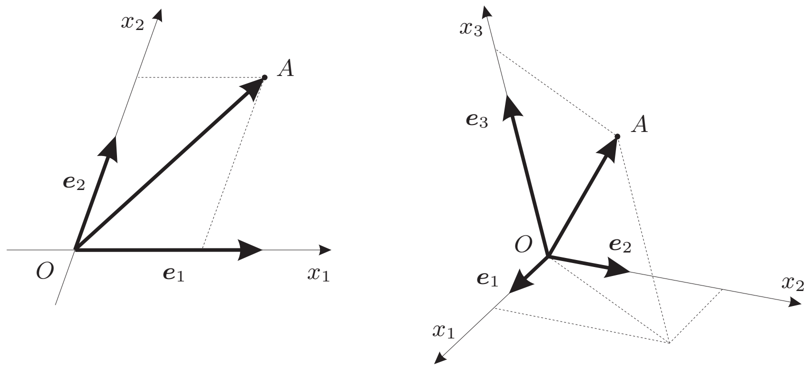


Рис. 1.19. Декартова косоугольная система координат на плоскости и в пространстве

D. Координаты направленного отрезка. 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.

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

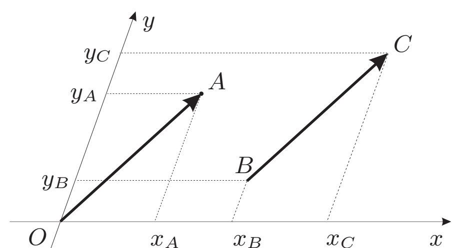


Рис. 1.20. Координаты направленного отрезка

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

так что

$$\overrightarrow{BC} = \big(x_C - x_B,\ y_C - y_B,\ z_C - z_B\big).$$

E. Полярная система координат на плоскости. Помимо декартовых координат, при решении различных задач на плоскости широко используются так называемые *полярные координаты*, определяемые следующим образом.

На плоскости фиксируется луч $[OP)$, начало которого называется *полюсом* (а сам луч — *полярной осью*). Положение любой точки A плоскости описывается упорядоченной парой чисел (r, φ), где r — расстояние от точки A до полюса, а φ — ориентированный угол между полярной осью и направленным отрезком $\overrightarrow{OA}$ (угол отсчитывается против часовой стрелки). Очевидно,

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

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

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

Значение угла φ для самой точки O не определено: она вполне задаётся значением $r = 0$.

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, то легко получить формулы, выражающие декартовы координаты через полярные:

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

Формулы обратного перехода имеют вид

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

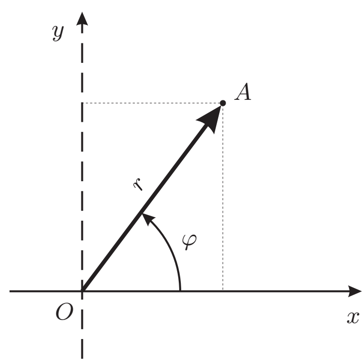
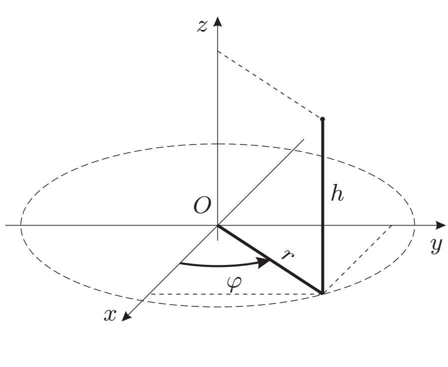


Рис. 1.21. Полярные координаты на плоскости

Во многих задачах удобно разрешить полярному радиусу 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 $(-r, \varphi + \pi)$ (см. с. 28).

F. Цилиндрическая система координат в пространстве. 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 проекции точки A на координатную плоскость Oxy, а h — проекция радиус-вектора $\overrightarrow{OA}$ на ось Oz.



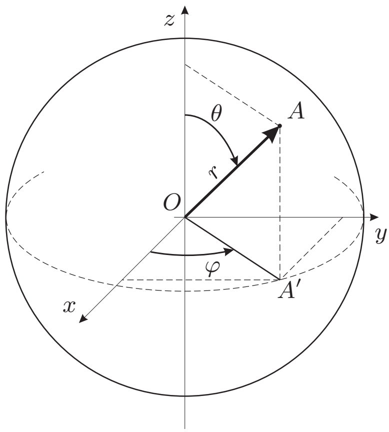


Рис. 1.22. Сферические координаты

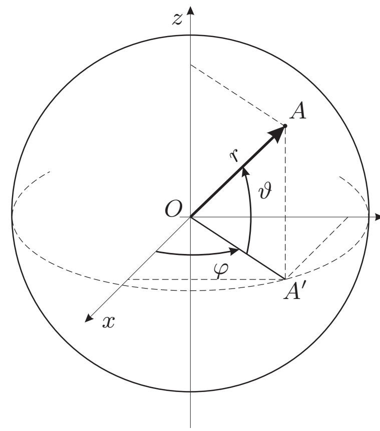


Рис. 1.23. Географические координаты

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

$$\begin{cases} x = r\cos\varphi, \\ y = r\sin\varphi, \\ z = h, \end{cases} \qquad \begin{cases} r = \sqrt{x^2+y^2}, \\ \cos\varphi = \dfrac{x}{\sqrt{x^2+y^2}}, \\ \sin\varphi = \dfrac{y}{\sqrt{x^2+y^2}}, \\ h = z. \end{cases}$$

G. Сферическая система координат в пространстве. Сферические координаты точки A удобно определить, 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. 1.22). Диапазоны изменения значений сферических координат:

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

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

$$x = r\cos\varphi\sin\theta, \quad y = r\sin\varphi\sin\theta, \quad z = r\cos\theta.$$

Часто используется другой вариант сферических координат, в котором вместо зенитного угла θ используется угол ϑ, отсчитываемый не от оси Oz, а от проекции $\overrightarrow{OA'}$ радиус-вектора $\overrightarrow{OA}$ на плоскость Oxy и называемый (географической) *широтой*. В этом случае азимутальный угол 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, незначительно отличаются от соответствующих формул для сферических координат:

$$x = r\cos\phi\cos\vartheta, \quad y = r\sin\phi\cos\vartheta, \quad z = r\sin\vartheta.$$

7. Простейшие задачи аналитической геометрии

A. Вычисление расстояния между двумя точками. Пусть на плоскости задана прямоугольная декартова система координат.

Легко видеть, что расстояние между двумя точками $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$ на плоскости может быть вычислено по формуле

$$|M_1M_2| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

являющейся простым следствием теоремы Пифагора (см. рис. 1.24).

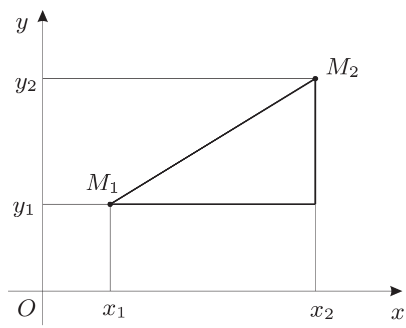


Рис. 1.24.

Совершенно аналогично можно получить формулу для пространственного случая: расстояние между точками $M_1(x_1, y_1, z_1)$ и $M_2(x_2, y_2, z_2)$ вычисляется по формуле

$$|M_1M_2| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

B. 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.

Спроектировав точки A, B и M на любую из координатных осей, например, на ось Ox, и обозначив проекции через A_1, B_1 и M_1 соответственно, получим, что точка M_1 делит отрезок A_1B_1, в том же отношении $\alpha : \beta$, считая от точки A_1. Решая уравнение

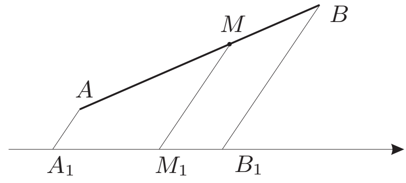


Рис. 1.25. Деление отрезка в заданном отношении

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

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

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

C. 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, полученным выше, находим координаты центра масс:

$$x_B = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}, \quad y_B = \frac{m_1 y_1 + m_2 y_2}{m_1 + m_2}, \quad z_B = \frac{m_1 z_1 + m_2 z_2}{m_1 + m_2}.$$

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. Ещё раз применяя формулы деления отрезка, получаем для координат точки C:

$$x_C = \frac{(m_1 + m_2)x_B + m_3 x_3}{(m_1 + m_2) + m_3} = \frac{m_1 x_1 + m_2 x_2 + m_3 x_3}{m_1 + m_2 + m_3}$$

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 можно доказать следующие формулы для координат центра масс:

$$x = \frac{m_1 x_1 + \cdots + m_k x_k}{m_1 + \cdots + m_k}, \; y = \frac{m_1 y_1 + \cdots + m_k y_k}{m_1 + \cdots + m_k}, \; z = \frac{m_1 z_1 + \cdots + m_k z_k}{m_1 + \cdots + m_k}.$$

8. Уравнения линий и поверхностей

A. Общие рассуждения. 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.

Уравнение линии на плоскости — это уравнение вида

$$F(x, y) = 0,$$

каждое решение (x, y) которого представляет собой прямоугольные декартовы координаты некоторой точки линии, причём для каждой точки линии найдется некоторое решение данного уравнения.

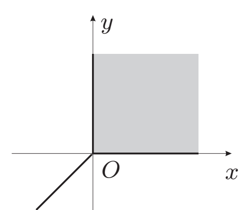


Рис. 1.26.

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. 1.26): $x - |x| - y + |y| = 0$.

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

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

Если точка $A(x, y)$ удовлетворяет уравнению $F(x, y) = 0$, то точка $A'(-x, y)$, симметричная точке A относительно оси Oy, также удовлетворяет этому уравнению. Таким образом, линия, определяемая уравнением $F(x, y) = 0$, симметрична относительно оси Oy. Аналогично, если

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

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

Конечно, вместо прямоугольных декартовых координат можно использовать любые другие.

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

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

При дополнительных условиях на функцию $G(x, y, z)$ получаются поверхности некоторых специальных типов.

1.3. Пример. Рассмотрим уравнение поверхности, не содержащее переменной z;

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

Пусть (x_0, y_0) — решение этого уравнения. Очевидно, все точки (x_0, y_0, z) пространства с различными значениями z имеют одну и ту же ортогональную проекцию на плоскость Oxy, а именно точку (x_0, y_0), так что все они лежат на одной прямой, параллельной оси Oz. Таким образом, рассматриваемое уравнение определяет *цилиндрическую поверхность* с образующей, параллельной оси Oz; это же уравнение является одновременно уравнением направляющей — линии в плоскости Oxy.

1.4. Пример. Поверхность называется *конической* с центром в точке C, если вместе с любой своей точкой M_0 она целиком содержит прямую (CM_0). Докажем, что уравнение $G(x, y, z) = 0$, в котором G является однородной функцией степени p, т.е. удовлетворяет условию

$$G(tx, ty, tz) = t^p G(x, y, z) \quad \text{для всех } t \in \mathbb{R},$$

задаёт коническую поверхность с центром в начале координат. Действительно, если точка $M_0(x_0, y_0, z_0)$ лежит на поверхности, т.е. $G(x_0, y_0, z_0) = 0$, то в силу однородности функции G точка (tx_0, ty_0, tz_0) также лежит на этой поверхности при любом значении $t \in \mathbb{R}$. Но множество точек $\{(tx_0, ty_0, tz_0) \mid t \in \mathbb{R}\}$ является прямой, проходящую через начало координат (при $t = 0$) и точку M_0 (при $t = 1$).

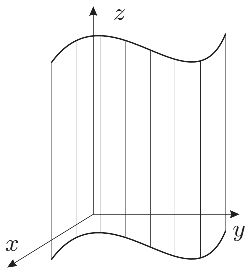


Рис. 1.27.
Цилиндрическая поверхность

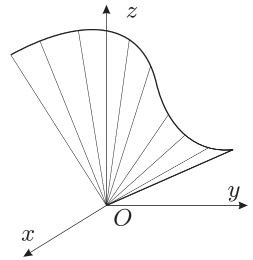


Рис. 1.28.
Коническая поверхность

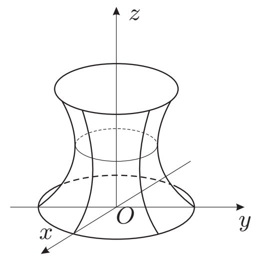


Рис. 1.29.
Поверхность вращения

1.5. Пример. Уравнение вида $F\left(\sqrt{x^2+y^2}, z\right) = 0$ определяет *поверхность вращения*,, сечение которой плоскостью Oxz — линия $F(|x|, z) = 0$ в этой плоскости — симметрично относительно оси Oz. Каждая из двух симметричных частей этой линии, называемая *главным меридианом* поверхности, при вращении вокруг оси Oz образует рассматриваемую поверхность.

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).

Вместо уравнений можно рассматривать неравенства; это позволяет описывать области на плоскости и в пространстве.

В дальнейшем, если не оговорено иное, под координатами точки будем понимать её декартовы прямоугольные координаты.

B. Уравнения прямых на плоскости. Как известно из школьного курса, прямая на плоскости задаётся линейным уравнением

$$Ax + By = C.$$

Используются также другие виды уравнения прямой (систематическое изучение прямых на плоскости будет проведено в лекции **??**).

1. Уравнение с угловым коэффициентом имеет вид

$$y = kx + b,$$

где k — угловой коэффициент прямой, геометрический смысл которого — тангенс угла между осью Ox и прямой (см. рис. 1.30):

$$k = \operatorname{tg}\alpha.$$

Такое уравнение описывает функциональную зависимость переменной y от аргумента x и широко используется в анализе. Для целей геометрии такое уравнение не слишком удобно: во-первых, в нём нарушено равноправие переменных x и y; во-вторых, геометрический смысл коэффициента k довольно сложен; в-третьих, таким уравнением невозможно задать прямую, параллельную оси ординат.

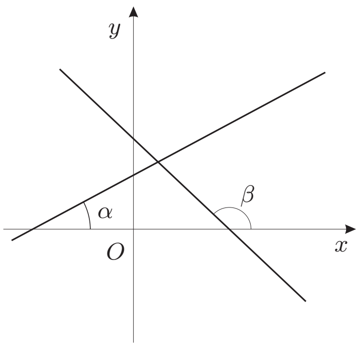


Рис. 1.30. Прямые, заданные уравнением с угловым коэффициентом

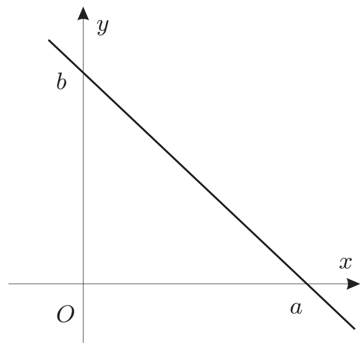


Рис. 1.31. Прямые, заданные уравнением в отрезках

2. Уравнение прямой «в отрезках» имеет вид

$$\frac{x}{a} + \frac{y}{b} = 1.$$

Геометрический смысл чисел a и b очевиден: это соответственно абсцисса точки пересечения прямой с осью Ox и ордината точки пересечения прямой с осью Oy (см. рис. 1.31).

C. Окружность. Окружность радиуса r с центром в начале координат имеет уравнение

$$x^2 + y^2 = r^2,$$

а окружность радиуса R с центром в точке $C(a, b)$ — уравнение

$$(x - a)^2 + (y - b)^2 = R^2$$

(см. рис. 1.32). Каждое из этих уравнений выражает тот факт, что произвольная точка (x, y), лежащая на окружности, удалена от центра окружности на одно и то же расстояние.

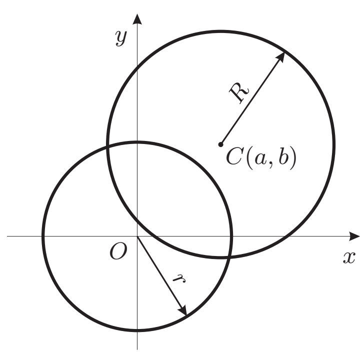


Рис. 1.32. Окружности

D. Парабола и гипербола. В школьном курсе парабола и гипербола задавались соответственно уравнениями

$$y = ax^2, \quad y = \frac{a}{x}$$

(опять функциональные зависимости y от x!) и при различных значениях a выглядят, как изображено на рис. 1.33. В лекции **??** эти линии будут охарактеризованы их геометрическими свойствами.

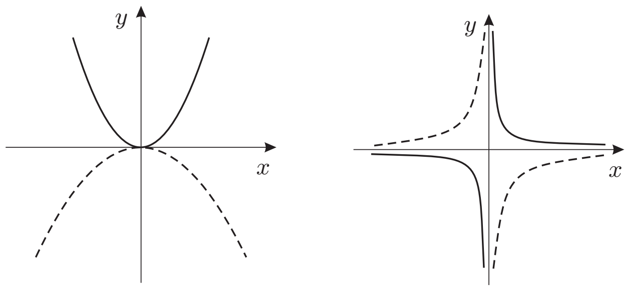
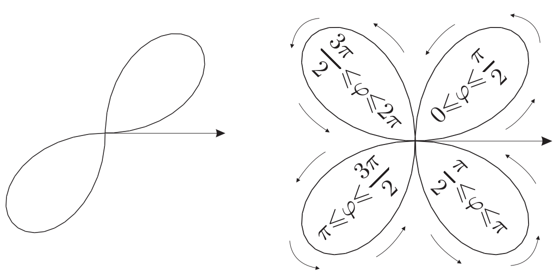


Рис. 1.33. Парабола и гипербола

Е. Примеры линий, заданных уравнениями в полярных координатах. Кривые, задаваемые в полярных координатах уравнениями вида $r = \cos k\varphi$ и $r = \sin k\varphi$, называются *розами.*

Рассмотрим уравнение $r = \sin 2\varphi$, определяющее некоторую линию в полярных координатах. Если на полярный наложено условие $r \geqslant 0$, то из неравенства $\sin 2\varphi \geqslant 0$ получаем следующую область изменения полярного угла φ: $\varphi \in \left[0; \dfrac{\pi}{2}\right] \cup \left[\pi; \dfrac{3\pi}{2}\right]$. Соответствующая линия изображена на рис. 1.34 слева; она называется *двухлепестковой розой.* На рис. 1.34 справа изображена *четырёхлепестковая роза*, описываемая тем же самым уравнением, но уже без ограничения $r \geqslant 0$ (см. п. 6.Е, с. 21).



(a) (b)

Рис. 1.34. Розы: (a) двухлепестковая; (b) четырёхлепестковая

Линия, заданная уравнением

$$(x^2 + y^2)^2 = a^2(x^2 - y^2),$$

называется *лемнискатой Бернулли.* Переходя к полярным координатам, преобразуем уравнение к виду

$$r = \sqrt{a^2 \cos 2\varphi},$$

после чего её нетрудно нарисовать (см. рис. 1.35).

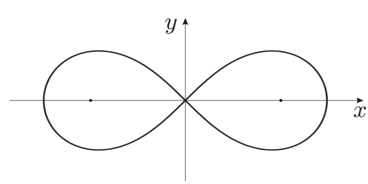


Рис. 1.35. Лемниската Бернулли

F. Уравнения некоторых поверхностей. Из школьного курса читателю известны уравнения плоскости

$$Ax + By + Cz = D$$

и сферы радиуса r

$$x^2 + y^2 + z^2 = r^2, \quad (x-a)^2 + (y-b)^2 + (z-c)^2 = r^2$$

(первое из этих уравнений описывает сферу с центром в начале координат, второе — с центром в точке (a, b, c)).

Уравнение $x^2 + y^2 = r^2$, описывающее окружность на плоскости, является также уравнением круговой цилиндрической поверхности с образующими, параллельными оси Oz (см. пример 1.3).

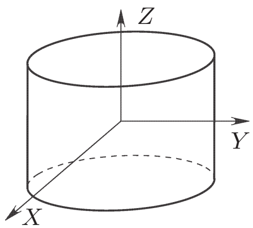


Рис. 1.36. Цилиндр вращения

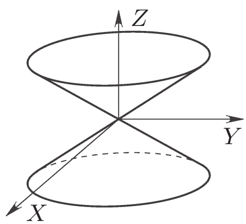


Рис. 1.37. Конус вращения

Уравнение $x^2+y^2 = z^2$ описывает круговую коническую поверхность с вершиной в начала координат (см. пример 1.4). И цилиндр, и конус являются поверхностями вращения (вокруг оси Oz; см. пример 1.5).)

9. Параметрические уравнения

A. Параметрические уравнения линий. 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.

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.

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

$$x = R\cos t, \quad y = R\sin t,$$
$$0 \leqslant t < 2\pi \ (\operatorname{mod} 2\pi).$$

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).

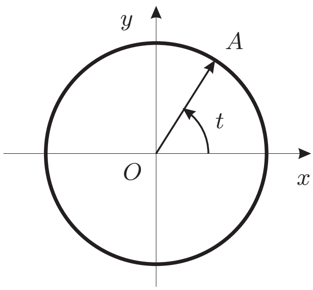


Рис. 1.38. Параметризация окружности

В качестве примера параметрического задания пространственной линии рассмотрим *винтовую линию*, т.е. траекторию точки, совершающей два одновременных движения: равномерное вращение с угловой скоростью ω в плоскости Oxy по окружности радиуса R и равномерное поступательное движение вдоль оси Oz со скоростью c (см. рис. 1.39). Запись закона движения и даёт параметрические уравнения винтовой линии:

$$x = R\cos\omega t,$$
$$y = R\sin\omega t,$$
$$z = ct.$$

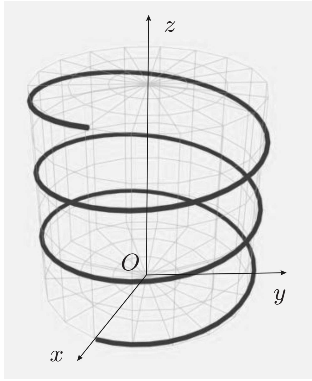


Рис. 1.39. Винтовая линия

В. Параметрическое задание поверхностей. Поверхности также можно задавать параметрическими уравнениями вида

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

где параметры u, v изменяются независимо в некоторой области D вспомогательной координатной плоскости Ouv.

Например, сферу радиуса R с центром в начале координат, заданную уравнением

$$x^2 + y^2 + z^2 = R^2,$$

можно представить параметрически следующим образом:

$$x = R\cos u \sin v,$$
$$y = R\sin u \sin v,$$
$$z = R\cos v,$$
$$0 \leqslant u < 2\pi, \quad 0 \leqslant v \leqslant \pi.$$

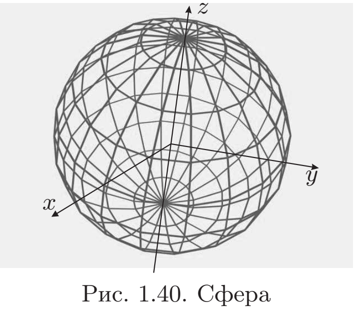


Рис. 1.40. Сфера

Линии, на которых значение u постоянно, представляют собой меридианы сферы, линии, на которых постоянно значение v, — параллели.

10. Пересечения и проекции

Линии в пространстве можно задавать как пересечение двух поверхностей:

$$\begin{cases} F(x,y,z) = 0, \\ G(x,y,z) = 0. \end{cases}$$

Проекцией[1] точки $M(x,y,z)$ на плоскость Oxy является точка $N(x,y)$. Таким образом, операция проектирования на координатную плоскость сводится к игнорированию одной из координат.

Если линия задана как пересечение двух поверхностей $F(x,y,z) = 0$ и $G(x,y,z) = 0$, то уравнение её проекции на плоскость Oxy получается исключением z из этих уравнений.

1.6. Пример. Найдём уравнение проекции на плоскость Oxy линии пересечения сферы и плоскости:

$$x^2 + y^2 + z^2 = 1, \quad x + y + z = 1.$$

Исключая из данных уравнений переменную z, получаем уравнение

$$x^2 + y^2 + (1 - (x+y))^2 = 1 \quad \Leftrightarrow \quad x^2 + xy + y^2 - x - y = 0.$$

1.7. Пример. *Кривая Вивиани*[2] представляет собой пересечение цилиндра радиуса R и сферы радиуса $2R$, центр которой лежит на поверхности цилиндра.

Получим параметрические уравнения кривой Вивиани. Уравнение сферы имеет вид

$$x^2 + y^2 + z^2 = 4R^2,$$

а цилиндра —

$$(x-R)^2 + y^2 = R^2 \iff x^2 + y^2 = 2Rx.$$

Из этих уравнений можно исключить y:

$$z^2 = 4R^2 - 2Rx.$$

Полученное уравнение описывает проекцию кривой Вивиани на плоскость Oxz; ясно, что эта проекция представляет собой параболу.

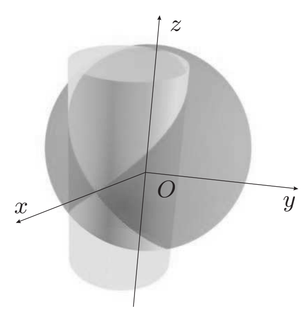


Рис. 1.41. Кривая Вивиани

Кроме того, очевидно, что проекцией кривой Вивиани на плоскость Oxy является окружность $x^2 + y^2 = 2Rx$ — направляющая цилиндра. Параметризуем эту окружность, введя в плоскости

[1] Речь идёт об ортогональной проекции в случае прямоугольной декартовой системы координат и о проекции параллельно оси Oz в случае косоугольной системы.

[2] В. Вивиани (1622–1703) — итальянский физик и математик, ученик Галилея и Торричелли.

Oxy полярные координаты (полярный угол мы обозначаем здесь через t):

$$x = r\cos t, \quad y = r\sin t.$$

Имеем

$$x^2 + y^2 = 2Rx \iff r^2 = 2Rr\cos t \iff r = 2R\cos t.$$

Теперь можно записать выражения для x и y через параметр t:

$$x = r\cos t = 2R\cos^2 t = R(1+\cos 2t), \quad y = r\sin t = 2R\sin t\cos t = R\sin 2t.$$

Полная окружность $x^2 + y^2 = 2Rx$ получается, если $0 \leqslant t \leqslant \pi$. Теперь можно найти выражение для z:

$$z^2 = 4R^2 - 2Rx = 4R^2\sin^2 t \iff z = \pm 2R\sin t.$$

Знак $\pm$ описывает тот факт, что каждой точке направляющей цилиндра соответствует две точки на кривой Вивиани: одна на верхней, а другая на нижней полусфере. Этот знак можно опустить, если разрешить параметру t изменяться в промежутке $0 \leqslant t < 2\pi$; в этом случае окружность — направляющая цилиндра — обходится дважды, причём первому обороту соответствуют точки кривой Вивиани, лежащие на верхней полусфере, а второму — на нижней.

Итак, окончательный результат:

$$\begin{cases} x = R(1+\cos 2t), \\ y = R\sin 2t, \\ z = 2R\sin t, \end{cases} \qquad 0 \leqslant t < 2\pi.$$

Проекция кривой Вивиани на плоскость Oyz имеет параметрические уравнения

$$\begin{cases} y = R\sin 2t, \\ z = 2R\sin t, \end{cases} \qquad 0 \leqslant t < 2\pi.$$

Линия, описываемая этими параметрическими уравнениями, является траекторией точки, участвующей в двух гармонических колебаниях во взаимно перпендикулярных направлениях: вдоль оси Oz с частотой (угловой скоростью) 1 и вдоль оси Oy с частотой, вдвое большей. Эта линия — одна из так называемых *фигур Лиссажу*[1] — изображена на рис. 1.42. Фигуры Лиссажу можно наблюдать на экране осциллографа при изучении колебаний.

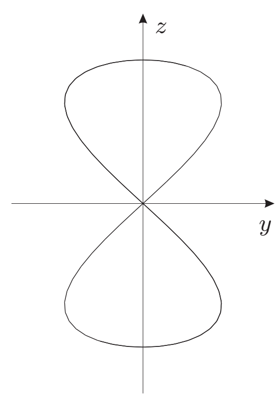


Рис. 1.42. Фигура Лиссажу

11. Упражнения

1.1. Запишите уравнение прямой $Ax + By = C$, где $C > 0$, в полярных координатах.

[1] Ж. А. Лиссажу (1822–1880) — французский математик.

1.1. Используя формулы $x = r\cos\varphi$, $y = r\sin\varphi$, получим
$$Ar\cos\varphi + Br\sin\varphi = C$$
откуда
$$r\sqrt{A^2+B^2}\left(\frac{A}{\sqrt{A^2+B^2}}\cos\varphi + \frac{B}{\sqrt{A^2+B^2}}\sin\varphi\right) = C.$$
Введём угол θ, определяемый соотношениями
$$\cos\theta = \frac{A}{\sqrt{A^2+B^2}}, \quad \sin\theta = \frac{B}{\sqrt{A^2+B^2}}$$
(такой угол определяется однозначно с точностью до $2\pi n$). Таким образом, получаем
$$r\sqrt{A^2+B^2}\,(\cos\theta\cos\varphi + \cos\theta\sin\varphi) = C \iff r = \frac{C}{\sqrt{A^2+B^2}\cos(\varphi-\theta)}.$$
Диапазон изменения угла φ должен быть таким, чтобы
$$\cos(\varphi-\theta) > 0;$$
отсюда находим
$$-\frac{\pi}{2} < \varphi - \theta < \frac{\pi}{2} \iff \theta - \frac{\pi}{2} < \varphi < \theta + \frac{\pi}{2}.$$

1.2. Изобразите линию, имеющую в полярных координатах уравнение $r = k\varphi$, $k > 0$. (Эта линия называется спиралью Архимеда.)

1.2. Данную линию можно рассматривать как траекторию точки, равномерно движущейся по лучу, который совершает равномерное вращение вокруг своего начала. На рис. **??** (см. с. **??**) изображен фрагмент линии $r = \varphi$, отвечающий изменению φ от 0 до 4π.

1.3. Получите параметрические уравнения окружности радиуса R с центром на оси Ox, проходящей через начало координат.

1.3. Поскольку центр C окружности имеет координаты $C(R,0)$, её уравнение в декартовых координатах имеет вид
$$(x-R)^2 + y^2 = R^2 \iff x^2 + y^2 = 2Rx.$$
Вводя полярные координаты r, φ, связанные с декартовыми формулами $x = r\cos\varphi$, $y = r\sin\varphi$, найдём
$$(r\cos\varphi)^2 + (r\sin\varphi)^2 = 2Rr\cos\varphi \iff r = 2R\cos\varphi.$$
Поскольку $r \geqslant 0$, то полярный угол φ должен быть изменяться в интервале, где $\cos\varphi \geqslant 0$, т.е. $\varphi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

1.4. Циклоидой называется траектория точки обода катящегося по прямой колеса. Получите параметрические уравнения циклоиды, взяв в качестве параметра угол θ поворота колеса.

1.4. Обозначим радиус колеса через R. Найдём координаты точки P обода колеса, которая в исходном положении колеса находилась в начале координат. При повороте колеса на угол θ без проскальзывания длина дуги ATP окружности, изображающей колесо, равна расстоянию OA, на которое перемещается центр C колеса вдоль оси Ox: $|OA| = R\theta$. Из чертежа ясно, что $|CQ| = -R\sin\theta$, $|PQ| = -R\cos\theta$, так что интересующие нас координаты точки P обода будут равны

$$x = R(\theta - \sin\theta), \quad y = R(1 - \cos\theta).$$

Эти уравнения и представляют собой параметрические уравнения циклоиды; первая арка циклоиды соответствует изменению параметра θ от 0 до 2π.

1.5. Точки P и Q делят отрезок $[AB]$ в отношении $\alpha : \beta : \gamma$. Найдите, в каком отношении делит отрезок $[AB]$ середина отрезка $[PQ]$.

1.6. Изобразите в полярной системе координат линии, заданные уравнениями $r = \cos 3\varphi$, $r = \sin 3\varphi$, в предположении, что (a) на полярный радиус наложено ограничение $r \geqslant 0$; (b) в отсутствие такого ограничения.

1.7. Изобразите в полярной системе координат линию, заданную уравнением $r = \cos 4\varphi$ в предположении, что (a) на полярный радиус наложено ограничение $r \geqslant 0$; (b) в отсутствие такого ограничения.

1.8. В полярной системе координат даны две точки $A_1(r_1, \varphi_1)$ и $A_2(r_2, \varphi_2)$. Найдите расстояние между ними.

1.9. Составьте уравнение окружности в полярных координатах, если её центр находится в точке $C(r_0, \varphi_0)$, а радиус равен R.

1.10. Составьте уравнение линии (в прямоугольных декартовых координатах), точки которой обладают следующим свойством: произведение расстояний от каждой точки линии до двух данных точек $F_1(-a; 0)$ и $F_2(a; 0)$ есть постоянная величина a^2. Эта линия называется лемнискатой Бернулли. Запишите уравнение этой линии в полярных координатах. Постройте линию, пользуясь её полярным уравнением.

1.11. Астроидой называется линия, получаемая следующим образом. Прямоугольник, две стороны которого лежат на осях координат, изменяется так, что его диагональ сохраняет постоянную величину a. Астроида — это траектория основания перпендикуляра, опущенного из вершины прямоугольника, противоположной началу координат, на его диагональ. Составьте уравнение астроиды в декартовой системе координат. Используя аналогию полученного уравнения с уравнением окружности, составьте параметрические уравнения астроиды.

ГЛАВА 2

Векторы и операции над ними

1. Отношения эквивалентности

Пусть X — некоторое множество. Рассмотрим его декартов квадрат $X^2 \equiv X \times X$, т.е. множество, элементами которого являются всевозможные упорядоченные пары элементов из X:

$$X^2 \equiv X \times X \stackrel{\text{def}}{=\!=} \{(x, y) \mid x \in X,\ y \in X\}.$$

Например, $\mathbb{R}^2$ — множество всех упорядоченных пар вещественных чисел, которое можно рассматривать как координатную плоскость Oxy.

2.1. Определение. *Отношением* R на множестве X называется любое подмножество R в X^2; если элемент (x, y) из X^2 принадлежит R, то говорят, что *элементы* x *и* y *множества* X *находятся в отношении* R. Для обозначения последнего факта чаще всего используется запись вида $x * y$, где $*$ — какой-либо специальный символ, используемый в качестве знака рассматриваемого отношения.

Приведём примеры отношений:

(1) отношение равенства $x = y$ на множестве вещественных чисел;
(2) отношение $x < y$ на множестве вещественных чисел;
(3) отношение $x \leqslant y$ на множестве вещественных чисел;
(4) отношение $x \,\vdots\, y$ на множестве целых чисел, выражаемое словами «x без остатка делится на y»;
(5) отношение $a \parallel b$ параллельности прямых;[1]
(6) отношение равенства $\overrightarrow{AB} = \overrightarrow{CD}$ на множестве направленных отрезков;
(7) отношение сонаправленности лучей $[AB) \uparrow\uparrow [CD)$;
(8) отношение подобия треугольников: $\triangle ABC \sim \triangle PQR$;
(9) отношение на множестве людей, задаваемое следующим образом: $A \bowtie B$ тогда и только тогда, когда A и B имеют хотя бы одного общего родителя (т.е. являются братьями или сёстрами или братом и сестрой).

2.2. Определение. Отношение $*$ на множестве X называется

(i) рефлексивным, если для любого $x \in X$ имеем $x * x$;

[1] Напомним, что в предыдущей лекции мы договорились (см. с. 8) считать совпадающие прямые (плоскости) параллельными.

(ii) симметричным, если для любых $x, y \in X$ из $x * y$ следует $y * x$;
(iii) транзитивным, если для любых $x, y, z \in X$ из $x * y$ и $y * z$ следует $x * z$.

Приведённые выше отношения обладают следующими свойствами:

	рефлекс.	симметр.	транзитивн.
$x = y$	+	+	+
$x < y$	−	−	+
$x \leqslant y$	+	−	+
$x \,\vdots\, y$	+	−	+
$a \parallel b$	+	+	+
$\overrightarrow{AB} = \overrightarrow{CD}$	+	+	+
$[AB) \uparrow\uparrow [CD)$	+	+	+
$\triangle ABC \sim \triangle PQR$	+	+	+
$A \bowtie B$	+	+	−

2.3. Определение. Отношение, обладающее свойствами рефлексивности, симметричности и транзитивности, называется *отношением эквивалентности.*

Среди приведённых примеров таковыми являются отношения равенства чисел и направленных отрезков, сонаправленности лучей, параллельности, подобия. При этом эквивалентность (т.е. равенство) чисел означает их совпадение; для других отношений эквивалентности ситуация несколько сложнее. Так, два равных направленных отрезка могут и не совпадать (если они отложены от разных точек), но у них есть нечто общее: их длины равны и направления совпадают. Аналогично, две параллельных прямых могут не совпадать, но имеют «общее направление», что отчётливо видно на чертежах. Подобные треугольники имеют разные размеры, но их «внешний вид» одинаков, и это то, что делает их эквивалентными. Таким образом, отношение эквивалентности позволяет выделить общие черты у различных объектов (не обязательно математических).

Для обозначения отношений эквивалентности используют обычно знаки $=$, $\equiv$, $\approx$, $\sim$, $\simeq$, $\cong$ и т. п.

2.4. Определение. Пусть на множестве X задано отношение эквивалентности $\sim$. Рассмотрим произвольный элемент $x \in X$ и множество всех элементов, эквивалентных ему:

$$[x] \overset{\text{def}}{=\!=} \{y \in X \mid y \sim x\}.$$

Это множество называется *классом эквивалентности элемента x.*

Например, если на множестве всех треугольников введено отношение подобия и $\triangle ABC$ — равносторонний, то классом эквивалентности $[\triangle ABC]$ является множество всех равносторонних треугольников.

Докажем, что для двух различных $x_1, x_2 \in X$ классы $[x_1]$ и $[x_2]$ либо не пересекаются, либо совпадают. Предположим, что $[x_1] \cap [x_2] \neq \varnothing$, т.е. $\exists y \in [x_1] \cap [x_2]$. Тогда $x_1 \sim y$ и $x_2 \sim y$, откуда $y \sim x_2$ (симметричность) и $x_1 \sim x_2$ (транзитивность), т.е. $x_1 \in [x_2]$, а также $x_2 \sim x_1$ (симметричность), т.е. $x_2 \in [x_1]$, так что $[x_1] = [x_2]$. Итак, если два класса эквивалентности имеют хотя бы один общий элемент, то они полностью совпадают. Кроме того, очевидно, каждый элемент $x \in X$ принадлежит задаваемому им классу, $x \in [x]$ в силу рефлексивности. Таким образом, мы доказали следующую теорему.

2.5. Теорема. *Любое отношение эквивалентности $\sim$ порождает разбиение множества A на непересекающиеся классы эквивалентности своих элементов*:

$$X = \bigcup_{a \in A} [a].$$

Верно и обратное утверждение (более того, оно очевидно).

2.6. Теорема. *Рассмотрим произвольное множество X, и представим его в виде объединения непересекающихся подмножеств*:

$$X = \bigcup_{\alpha \in A} X_\alpha, \quad X_\alpha \cap X_\beta = \varnothing \text{ при } \alpha \neq \beta.$$

Определим на множестве X отношение $\approx$ следующим образом:

$$x \approx y \iff x \text{ и } y \text{ лежат в одном } X_\alpha,$$

Отношение $\approx$ является отношением эквивалентности.

2.7. Определение. Рассмотрим отношение эквивалентности $\sim$ на множестве X. Множество

$$X/\!\sim \stackrel{\text{def}}{=\!=} \{[a] \mid a \in A\},$$

элементами которого являются классы эквивалентности $[x]$ относительно $\sim$, называется *фактор-множеством множества X по отношению $\sim$*.

Очевидно, что каждый класс эквивалентности $[x]$ определяется любым своим элементом, и для того чтобы составить представление об устройстве класса, достаточно взять из него любой элемент, называемый представителем этого класса.

2.8. Пример. Рассмотрим ещё один важный пример. На множестве $\mathbb{N}^2$ упорядоченных пар натуральных чисел введём отношение $\doteq$ следующим образом:

$$(a, b) \doteq (c, d) \iff ad = bc.$$

Проверим, что это отношение является отношением эквивалентности.

1. Рефлексивность: $(a, b) \doteq (a, b)$, поскольку $ab = ba$.

2. Симметричность: если $(a, b) \doteq (c, d)$, т.е. $ad = bc$, то $(c, d) = (a, b)$, поскольку $cb = da$.

3. 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

$$adf = bcf, \quad bcf = bde \quad \Rightarrow \quad adf = bde \quad \Rightarrow$$
$$\Rightarrow \quad af = be \iff (a, b) \doteq (e, f).$$

Записывая упорядоченную пару натуральных чисел (a, b) в виде $\dfrac{a}{b}$, а закон эквивалентности — в виде

$$\frac{a}{b} \doteq \frac{c}{d} \iff ad = bc,$$

мы немедленно узнаём в этих парах рациональные дроби. Отношение эквивалентности пар (дробей) позволяет сокращать дроби, классом эквивалентности каждой пары является множество всех равных между собой дробей, изображающих одно и то же рациональное число, 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, 6/9 и т. д.

2. Понятие вектора

А. Определение вектора. Рассмотрим множество направленных отрезков (на плоскости или в пространстве). Мы уже установили выше, что отношение равенства направленных отрезков, введённое в лекции 1, является отношением эквивалентности. Класс эквивалентности направленного отрезка $[\overrightarrow{AB}]$ относительно этого отношения эквивалентности, т.е. множество всех направленных отрезков, равных $\overrightarrow{AB}$, называется *вектором* (или *свободным вектором*), порождённым направленным отрезком $\overrightarrow{AB}$.

Таким образом, вектор — это не один направленный отрезок, а бесконечное множество (класс эквивалентности) направленных отрезков, равных между собой. Как уже отмечалось, для того чтобы составить представление об устройстве класса эквивалентности, достаточно рассмотреть представитель этого класса, т.е. любой из направленных отрезков, входящий в соответствующий класс. По этой причине векторы часто изображаются стрелками, не отличимыми по виду от направленных отрезков. Грубо говоря, разница между векторами и направленными отрезками состоит в том, что у направленного вектора имеется точка приложения, а у вектора её нет, так что вектор можно трактовать как стрелку, которая может свободно перемещаться по плоскости (пространству), не меняя своего направления и длины.

Векторы обычно обозначаются строчными полужирными буквами: $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ и т. д.

Отложить вектор $\boldsymbol{a}$ от точки A означает построить направленный отрезок $\overrightarrow{AB}$ с началом в точке A, входящий в класс эквивалентности, описываемый вектором $\boldsymbol{a}$, или, что то же, такой, что $[\overrightarrow{AB}] = \boldsymbol{a}$.

Во многих случаях термины «вектор» и «направленный отрезок, полученный откладыванием вектора от точки», считают синонимами и пишут, например, $\boldsymbol{a} = \overrightarrow{AB}$. Такие вольности речи вполне допустимы, если не приводят к недоразумениям.

Как указывалось ранее, равные направленные векторы, отложенные от различных точек, не совпадают. Для векторов же ситуация точно такая, как для чисел: равные векторы — это просто один и тот же вектор (класс эквивалентности, состоящий из одних и тех же направленных отрезков).

Нулевой вектор — это класс эквивалентности нулевого направленного отрезка; обозначение нулевого вектора $\mathbf{0}$.

Множество всех векторов на прямой будем обозначать V_1, на плоскости — V_2, в пространстве будем обозначать V_3.

Длина вектора $\boldsymbol{a}$ — это длина любого направленного отрезка, являющегося его представителем; обозначение $|\boldsymbol{a}|$.[1] Аналогичным образом, мы можем говорить о сонаправленных и противоположно направленных векторах, о коллинеарных и компланарных векторах и т. п.

Чтобы определить проекцию вектора $\boldsymbol{a}$ на прямую l или на плоскость π, рассмотрим произвольный представитель этого вектора $\overrightarrow{AB}$ (т.е. направленный отрезок $\overrightarrow{AB}$, полученный откладыванием данного вектора $\boldsymbol{a}$ от произвольной точки A), спроектируем его на прямую l или на плоскость π и возьмём класс эквивалентности получившегося направленного отрезка-проекции:

$$\text{если } \boldsymbol{a} = \left[\overrightarrow{AB}\right], \quad \text{то } \operatorname{Pr}\boldsymbol{a} = \left[\operatorname{Pr}\overrightarrow{AB}\right],$$

где Pr означает любой вид проекции на прямую или на плоскость. Проекция вектора $\boldsymbol{a}$ на ось $\vec{l}$ определяется как проекция на эту ось любого представителя этого вектора:

$$\text{если } \boldsymbol{a} = \left[\overrightarrow{AB}\right], \quad \text{то } \operatorname{pr}_{\vec{l}}\boldsymbol{a} = \operatorname{pr}_{\vec{l}}\overrightarrow{AB}.$$

B. О других разновидностях векторов. Понятие векторной величины возникло в связи с рассмотрением в физике таких понятий, как скорость, ускорение, перемещение и т. п. Понятие вектора (свободного вектора) в геометрии несколько отличается от физических векторных величин.

Наше определение вектора базировалось на определении равенства направленных отрезков, сформулированном в лекции 1 (см. с. 11). Если равенство направленных отрезков определить как-либо иначе, то и понятие вектора (как класса эквивалентности) получится другим.

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. Связанный (или приложенный) вектор — это в точности то же самое, что направленный отрезок.

---

[1] Иногда длину вектора $\boldsymbol{a}$ обозначают $\|\boldsymbol{a}\|$, чтобы избежать путаницы с модулем числа и определителем матрицы.

Именно таким вектором является скорость материальной точки в механике: вектор скорости «прикреплён» к частице, и его начальная точка перемещается вместе с частицей.

Теперь будем считать два направленных отрезка равными, если они лежат на одной прямой, имеют одинаковое направление и равные длины. Класс эквивалентности направленных отрезков относительно такого отношения эквивалентности приводит нас 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).

3. Линейные операции над векторами

A. Определение и свойства линейных операций.

2.9. Определение. *Суммой* двух векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ называется вектор, обозначаемый $\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b}$, определяемый одним из следующих двух способов (см. рис. 2.1).

1. Правило треугольника. Отложим вектор $\boldsymbol{a}$ от точки A, получим направленный отрезок $\overrightarrow{AB}$; отложим вектор $\boldsymbol{b}$ от точки B, получим направленный отрезок $\overrightarrow{BC}$. Класс эквивалентности $\boldsymbol{c} = [\overrightarrow{AC}]$ направленного отрезка $\overrightarrow{AC}$ называется суммой векторов $\boldsymbol{a}$ и $\boldsymbol{b}$.
2. Правило параллелограмма. Отложим векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ от точки O, получим направленные отрезки $\overrightarrow{OA}$ и $\overrightarrow{OB}$. Проведём через точку A прямую, параллельную (OB), а через точку B — прямую, параллельную (OA). Пусть C — точка пересечения построенных прямых. Класс эквивалентности $\boldsymbol{c} = [\overrightarrow{OC}]$ направленного отрезка $\overrightarrow{OC}$ называется суммой векторов $\boldsymbol{a}$ и $\boldsymbol{b}$.

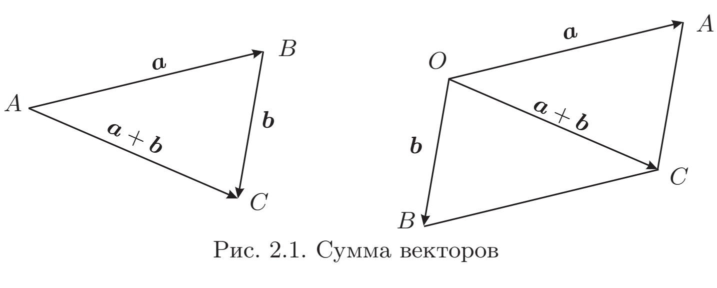


Рис. 2.1. Сумма векторов

2.10. Определение. *Произведением вектора* $\boldsymbol{a}$ *на вещественное число* α называется вектор $\alpha\boldsymbol{a}$, определяемый следующим образом: длина вектора $\alpha\boldsymbol{a}$ равна

$$\big\|\alpha\boldsymbol{a}\big\| = |\alpha| \cdot \|\boldsymbol{a}\|,$$

а его направление совпадает с направлением вектора $\boldsymbol{a}$, если $\alpha > 0$, и противоположно направлению вектора $\boldsymbol{a}$, если $\alpha < 0$. Если же $\alpha = 0$, то $0 \cdot \boldsymbol{a} = \boldsymbol{0}$.

Операции сложения векторов и умножения вектора на число, называются *линейными операциями.*

Приведённые определения нуждаются в проверке корректности.

2.11. Предложение. *Определения суммы векторов и произведения вектора на число корректны, т.е. результат их выполнения не зависит от выбора представителей векторов.*

Доказательство. Докажем корректность определения суммы векторов $\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{c}$ по правилу треугольника (корректность определения произведения вектора на число докажите самостоятельно). Требуется доказать, что если в качестве исходной взять другую точку A_1, то после проведения всех необходимых построений получится представитель $\overrightarrow{A_1C_1}$ того же вектора $\boldsymbol{c}$. Рассмотрим случай, когда $A_1 \notin (AB)$ (см. рис. 2.2(a)); второй случай рассмотрите самостоятельно, используя рис. 2.2(b).

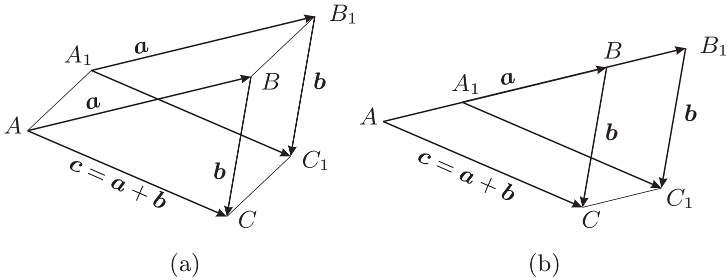


Рис. 2.2. Корректность определения суммы векторов

Поскольку направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{A_1B_1}$ являются представителями одного и того же вектора $\boldsymbol{a}$, они равны, так что по теореме 1.1 (см. с. 11) получаем $\overrightarrow{AA_1} = \overrightarrow{BB_1}$. Аналогично из соотношения $\overrightarrow{BC} = \overrightarrow{B_1C_1}$ находим $\overrightarrow{BB_1} = \overrightarrow{CC_1}$. Таким образом, $\overrightarrow{AA_1} = \overrightarrow{CC_1}$, откуда вытекает, что $\overrightarrow{AC} = \overrightarrow{A_1C_1}$, т.е. направленные отрезки $\overrightarrow{AC}$ и $\overrightarrow{A_1C_1}$ лежат в одном классе эквивалентности. □

2.12. Теорема. *Линейные операции над векторами обладают следующими свойствами*:

B1: *коммутативность сложения*: *для любых векторов* $\boldsymbol{a}$ *и* $\boldsymbol{b}$

$$\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{b}+\boldsymbol{a};$$

B2: *ассоциативность сложения*: *для любых векторов* $\boldsymbol{a}$, $\boldsymbol{b}$ *и* $\boldsymbol{c}$

$$(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}=\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c});$$

B3: *свойство нулевого вектора*: *для любого вектора* $\boldsymbol{a}$

$$\boldsymbol{a}+\boldsymbol{0}=\boldsymbol{a};$$

B4: *существование противоположного вектора*: *для любого вектора* $\boldsymbol{a}$ *существует такой вектор* $\boldsymbol{a}'$, *что*

$$\boldsymbol{a}+\boldsymbol{a}'=\boldsymbol{0};$$

B5: *свойство единицы*: *для любого вектора* $\boldsymbol{a}$
$$1 \cdot \boldsymbol{a} = \boldsymbol{a};$$

B6: *ассоциативность умножения на число*: *для любого вектора* $\boldsymbol{a}$ *и любых чисел* α *и* β
$$(\alpha\beta)\,\boldsymbol{a} = \alpha\,(\beta\boldsymbol{a})\,;$$

B7: *дистрибутивность относительно сложения векторов*: *для любых векторов* $\boldsymbol{a}$ *и* $\boldsymbol{b}$ *и любого числа* α
$$\alpha\,(\boldsymbol{a} + \boldsymbol{b}) = \alpha\boldsymbol{a} + \alpha\boldsymbol{b};$$

B8: *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. Сложим векторы по правилу параллелограмма, отложив их от произвольной точки O, построив направленные отрезки $\overrightarrow{OA}$ и $\overrightarrow{OB}$, параллелограмм со сторонами $\overrightarrow{OA}$ и $\overrightarrow{OB}$ и проведя диагональ параллелограмма $\overrightarrow{OC}$. Так как этот параллелограмм и его диагональ $\overrightarrow{OC}$ не зависят от того, в каком порядке даны векторы $\boldsymbol{a}$ и $\boldsymbol{b}$, то сложение векторов коммутативно.

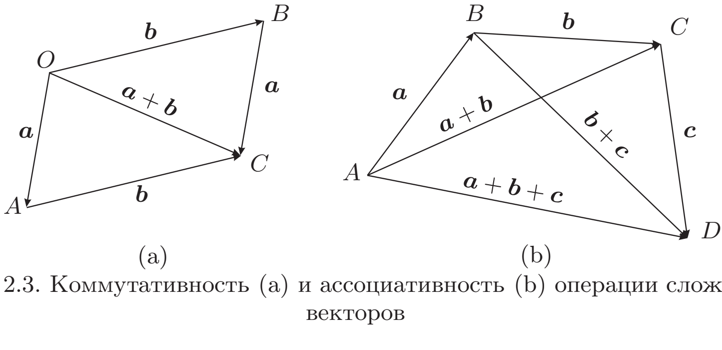


Рис. 2.3. Коммутативность (a) и ассоциативность (b) операции сложения векторов

Докажем ассоциативность сложения. Отложим данные векторы $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ следующим образом:
$$\boldsymbol{a} = \overrightarrow{AB}, \quad \boldsymbol{b} = \overrightarrow{BC}, \quad \boldsymbol{c} = \overrightarrow{CD}.$$
Тогда
$$\boldsymbol{b} + \boldsymbol{c} = \overrightarrow{BC} + \overrightarrow{CD} = \overrightarrow{BD}, \quad \boldsymbol{a} + \big(\boldsymbol{b} + \boldsymbol{c}\big) = \overrightarrow{AB} + \overrightarrow{BD} = \overrightarrow{AD},$$
$$\boldsymbol{a} + \boldsymbol{b} = \overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}, \quad \big(\boldsymbol{a} + \boldsymbol{b}\big) + \boldsymbol{c} = \overrightarrow{AC} + \overrightarrow{CD} = \overrightarrow{AD},$$
что и требовалось.

Доказательство коммутативности и ассоциативности сложения проиллюстрировано на рис. 2.3. □

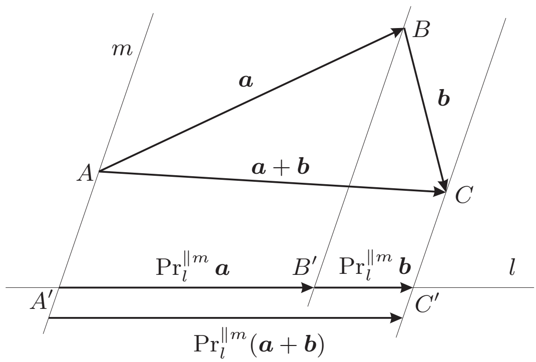


Рис. 2.4.

В. Связь проекций и линейных операций.

2.13. Теорема (связь проекций и линейных операций)**.** 1. *Проекция на прямую или на плоскость суммы векторов равна сумме проекций слагаемых*:

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

2. *Проекция на прямую или на плоскость произведения вектора* $\boldsymbol{a}$ *на число равна произведению на это число проекции вектора* $\boldsymbol{a}$:

$$\mathrm{Pr}(\alpha\boldsymbol{a}) = \alpha\,\mathrm{Pr}\,\boldsymbol{a}.$$

3. *Проекция на ось суммы векторов равна сумме проекций слагаемых*:

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

4. *Проекция на ось произведения вектора* $\boldsymbol{a}$ *на число равна произведению на это число проекции вектора* $\boldsymbol{a}$:

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

Доказательство. Отложив вектор $\boldsymbol{a}$ от точки A, получим направленный отрезок $\overrightarrow{AB}$; затем, отложив вектор $\boldsymbol{b}$ от точки B, получим направленный отрезок $\overrightarrow{BC}$; направленный отрезок

$$\overrightarrow{AC} = \overrightarrow{AB} + \overrightarrow{BC}$$

является представителем вектора $\boldsymbol{a}+\boldsymbol{b}$. Для проекций этих направленных отрезков имеем

$$\mathrm{Pr}\,\overrightarrow{AB} = \overrightarrow{A'B'}, \quad \mathrm{Pr}\,\overrightarrow{BC} = \overrightarrow{B'C'},$$

$$\mathrm{Pr}\,\overrightarrow{AC} = \overrightarrow{A'C'} = \overrightarrow{A'B'} + \overrightarrow{B'C'} = \mathrm{Pr}\,\overrightarrow{AB} + \mathrm{Pr}\,\overrightarrow{BC}$$

независимо от вида проекции. Доказательство проиллюстрировано на рис. 2.4–2.6. □

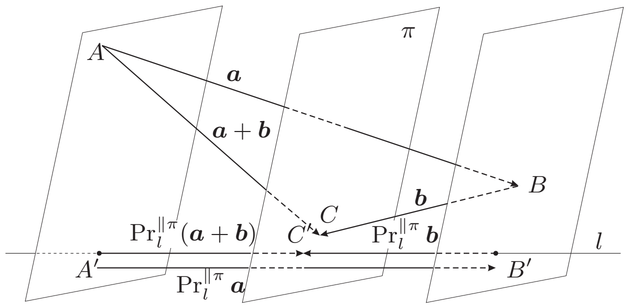


Рис. 2.5.

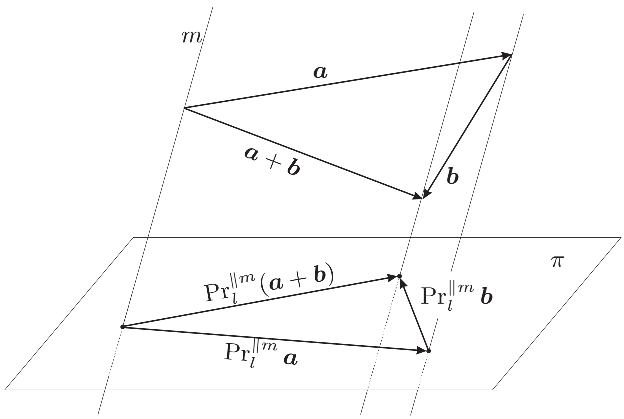


Рис. 2.6.

C. Коллинеарные и компланарные векторы. Два вектора называются *коллинеарными*, если коллинеарны направленные отрезки, являющиеся их представителями. Если коллинеарные векторы отложить от общего начала, получим направленные отрезки, лежащие на одной прямой.

Три вектора называются *компланарными*, если компланарны направленные отрезки, являющиеся их представителями. Если компланарные векторы отложить от общего начала, получим направленные отрезки, лежащие в одной плоскости. (Отметим, что два любые вектора компланарны.)

2.14. Теорема. 1. *Для того, чтобы два вектора $\boldsymbol{a}, \boldsymbol{b}$ были коллинеарны, необходимо и достаточно, чтобы существовали такие числа α, β, не равные одновременно нулю, что*

$$\alpha\boldsymbol{a} + \beta\boldsymbol{b} = \boldsymbol{0}.$$

2. *Для того, чтобы три вектора $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ были компланарны, необходимо и достаточно, чтобы существовали такие числа α, β, γ не равные одновременно нулю, что*

$$\alpha\boldsymbol{a} + \beta\boldsymbol{b} + \gamma\boldsymbol{c} = \boldsymbol{0}.$$

Доказательство. Докажем второе утверждение теоремы.

1. Пусть векторы $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ компланарны. Отложив их от одной точки, видим, что один из них можно выразить через два остальных, например,

$$\boldsymbol{a} = x\boldsymbol{b} + y\boldsymbol{c}.$$

Положив

$$\alpha = 1, \quad \beta = -x, \quad \gamma = -y,$$

получим требуемое равенство.

2. Пусть в равенстве

$$\alpha\boldsymbol{a} + \beta\boldsymbol{b} + \gamma\boldsymbol{c} = \boldsymbol{0}$$

один из коэффициентов отличен от нуля, например, $\alpha \neq 0$. Тогда можно записать

$$\boldsymbol{a} = -\frac{\beta}{\alpha}\boldsymbol{b} - \frac{\gamma}{\alpha}\boldsymbol{c},$$

и векторы оказываются компланарными. □

Обратная теорема формулируется следующим образом.

2.15. Теорема. 1. *Для того, чтобы два вектора $\boldsymbol{a}, \boldsymbol{b}$ были неколлинеарны, необходимо и достаточно, чтобы равенство*

$$\alpha\boldsymbol{a} + \beta\boldsymbol{b} = \boldsymbol{0}$$

было возможно лишь при $\alpha = \beta = 0$.

2. *Для того, чтобы три вектора $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ были некомпланарны, необходимо и достаточно, чтобы равенство*

$$\alpha\boldsymbol{a} + \beta\boldsymbol{b} + \gamma\boldsymbol{c} = \boldsymbol{0}$$

было возможно лишь при $\alpha = \beta = \gamma = 0$.

4. Базис и координаты

Базисом на плоскости будем называть произвольный упорядоченный набор двух ненулевых неколлинеарных векторов $\boldsymbol{a}_1, \boldsymbol{a}_2$.

Если на плоскости задан какой-либо базис $\boldsymbol{a}_1, \boldsymbol{a}_2$, то любой вектор на этой плоскости можно представить в виде комбинации

$$\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2.$$

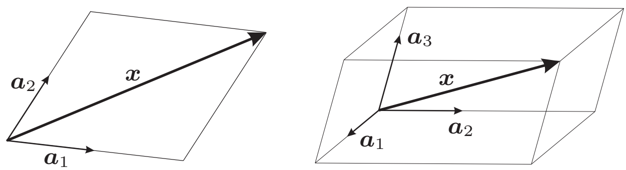
Это соотношение называется *разложением вектора $\boldsymbol{x}$ по базису $\boldsymbol{a}_1, \boldsymbol{a}_2$*, а числа x_1, x_2 — *координатами вектора $\boldsymbol{x}$ в базисе $\boldsymbol{a}_1, \boldsymbol{a}_2$*. Легко видеть, что это определение полностью согласуется с определением координат направленного отрезка относительно косоугольной декартовой системы координат, сформулированным в лекции 1. Таким образом, координаты вектора суть не что иное, как проекции на координатные оси любого направленного отрезка, являющегося представителем данного вектора.

Базисом в пространстве будем называть произвольный упорядоченный набор трёх ненулевых некомпланарных векторов $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$.

Любой вектор пространства можно представить в виде комбинации векторов базиса

$$\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + x_3\boldsymbol{a}_3;$$

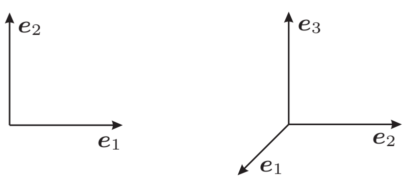
это соотношение называется *разложением вектора* $\boldsymbol{x}$ *по базису* $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$, а числа x_1, x_2, x_3 — *координатами вектора* $\boldsymbol{x}$ *в базисе* $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$.



(a) (b)

Рис. 2.7. Разложение вектора по базису (a) на плоскости; (b) в пространстве

В аналитической геометрии используются преимущественно ортонормированные базисы, т.е. базисы, состоящие из единичных попарно ортогональных (перпендикулярных) векторов.



U V

Рис. 2.8. Ортонормированный базис (a) на плоскости; (b) в пространстве

2.16. Теорема. *Разложение вектора по базису единственно, т.е. набор координат векторов в данном базисе определен однозначно.*

Доказательство. Предположим, что вектор $\boldsymbol{x}$ имеет в базисе $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ два различных набора координат:

$$\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + x_3\boldsymbol{a}_3 = y_1\boldsymbol{a}_1 + y_2\boldsymbol{a}_2 + y_3\boldsymbol{a}_3.$$

Вычитая второе разложение из первого, получим

$$\mathbf{0} = (x_1 - y_1)\boldsymbol{a}_1 + (x_2 - y_2)\boldsymbol{a}_1 + (x_3 - y_3)\boldsymbol{a}_3.$$

Так как векторы базиса некомпланарны, то это равенство возможно лишь при нулевых коэффициентах (см. теорему 2.15), т.е.

$$x_1 = y_1, \quad x_2 = y_2, \quad x_3 = y_3.$$

Теорема доказана. □

Таким образом, зафиксировав на плоскости (в пространстве) некоторый базис и раскладывая любые векторы по этому базису, мы можем превращать задачи о векторах в задачи о столбцах чисел.

Используемые обозначения. Мы будем пользоваться следующими обозначениями.

Базис на плоскости обозначается $\boldsymbol{i}, \boldsymbol{j}$ или $\boldsymbol{e}_1, \boldsymbol{e}_2$, а в пространстве — $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ или $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ (в большинстве задач этот базис считается ортонормированным). Системы координат обозначаются следующим образом:

$$O\boldsymbol{i}\boldsymbol{j} \iff Oxy \iff O\boldsymbol{e}_1\boldsymbol{e}_2 \iff Ox^1x^2,$$
$$O\boldsymbol{i}\boldsymbol{j}\boldsymbol{k} \iff Oxyz \iff O\boldsymbol{e}_1\boldsymbol{e}_2\boldsymbol{e}_3 \iff Ox^1x^2x^3.$$

Разложение вектора $\boldsymbol{a}$ по базису $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ записываем в виде

$$\boldsymbol{a} = a_x\boldsymbol{i} + a_y\boldsymbol{j} + a_z\boldsymbol{k} = a^1\boldsymbol{e}_1 + a^2\boldsymbol{e}_2 + a^3\boldsymbol{e}^3.$$

Используется также запись

$$\boldsymbol{a} = (a_x, a_y, a_z) = (a^1, a^2, a^3),$$

в которой вектор и набор его координат отождествляются. Подчеркнём, что это отождествление зависит от выбора базиса.

Для обозначения координат точки используем запись $A(x, y, z)$ или $A(\boldsymbol{r})$, где $\boldsymbol{r}$ обозначает радиус-вектор точки A, т.е. направленный отрезок $\overrightarrow{OA}$, где O — начало координат. Обратите внимание, что радиус-вектор точки является *связанным* (отложенным) вектором, т.е. направленным отрезком, но можно рассмотреть его класс эквивалентности и получить из него свободный вектор, хотя в большинстве задач это бессмысленно.[1]

5. Скалярное произведение векторов

А. Определение и свойства скалярного произведения. Напомним, что *ортогональная проекция вектора* $\boldsymbol{a}$ *на координатную ось* $\vec{l}$ — это число

$$\mathrm{pr}_{\vec{l}}\boldsymbol{a} = \begin{cases} \left|\mathrm{Pr}_l\, \boldsymbol{a}\right|, & \text{если } \mathrm{Pr}_l\, \boldsymbol{a} \uparrow\uparrow \vec{l}, \\ -\left|\mathrm{Pr}_l\, \boldsymbol{a}\right|, & \text{если } \mathrm{Pr}_l\, \boldsymbol{a} \uparrow\downarrow \vec{l}. \end{cases}$$

Любой вектор $\boldsymbol{l}$, сонаправленный с осью $\vec{l}$, называется *направляющим вектором оси*; проекция на ось $\vec{l}$ обозначается, наряду с $\mathrm{pr}_{\vec{l}}$, символом $\mathrm{pr}_{\boldsymbol{l}}$.

Подчеркнём, что согласно принятым определениям проекция вектора на прямую является *вектором*, а проекция вектора на ось — *числом.*

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 $\boldsymbol{l}$).

[1] Примером осмысленной задачи, в которой радиус-вектор точки удобно рассматривать как свободный вектор, является задача о сдвиге начала координат (см. гл. **??**).

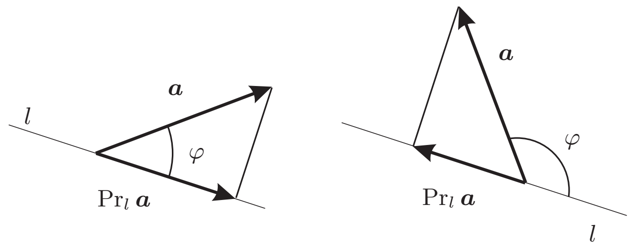


Рис. 2.9. Ортогональная проекция вектора на прямую

2.17. Определение. *Скалярным произведением двух векторов* $\boldsymbol{a}$ *и* $\boldsymbol{b}$ называется число

$$(\boldsymbol{a}, \boldsymbol{b}) = |\boldsymbol{a}| \cdot |\boldsymbol{b}| \cdot \cos\varphi = \mathrm{pr}_{\boldsymbol{b}}\, \boldsymbol{a} \cdot |\boldsymbol{b}|,$$

где φ — угол между векторами $\boldsymbol{a}$ и $\boldsymbol{b}$.

Два вектора называются *ортогональными*, если их скалярное произведение равно нулю. Очевидно, нулевой вектор ортогонален любому другому, а ненулевые векторы ортогональны тогда и только тогда, когда они перпендикулярны.

2.18. Теорема. *Скалярное произведение векторов обладает следующими свойствами*:

СП1: *линейность по каждому аргументу* (*сомножителю*):

$$(\alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2, \boldsymbol{b}) = \alpha_1(\boldsymbol{a}_1, \boldsymbol{b}) + \alpha_2(\boldsymbol{a}_2, \boldsymbol{b}),$$
$$(\boldsymbol{a}, \beta_1\boldsymbol{b}_1 + \beta_2\boldsymbol{b}_2) = \beta_1(\boldsymbol{a}, \boldsymbol{b}_1) + \beta_2(\boldsymbol{a}, \boldsymbol{b}_2),$$

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 числа*;

СП2: *симметричность* (*коммутативность*):

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

СП3: *положительная определённость*: *для любого вектора* $\boldsymbol{a}$

$$(\boldsymbol{a}, \boldsymbol{a}) \geqslant 0,$$

причём $(\boldsymbol{a}, \boldsymbol{a}) = 0$ *тогда и только тогда, когда* $\boldsymbol{a} = \boldsymbol{0}$.

Доказательство. Требуемые свойства очевидным образом вытекают из определения и свойств проекций (см. теорему 2.13). Например,

$$(\alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2, \boldsymbol{b}) = \mathrm{pr}_{\boldsymbol{b}}\Big(\alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2\Big) \cdot |\boldsymbol{b}| =$$
$$= \alpha_1\,\mathrm{pr}_{\boldsymbol{b}}\,\boldsymbol{a}_1 \cdot |\boldsymbol{b}| + \alpha_2\,\mathrm{pr}_{\boldsymbol{b}}\,\boldsymbol{a}_2 \cdot |\boldsymbol{b}| = \alpha_1(\boldsymbol{a}_1, \boldsymbol{b}) + \alpha_2(\boldsymbol{a}_2, \boldsymbol{b}),$$

что и требовалось. □

2.19. Теорема. 1. *Проекция вектора* $\boldsymbol{a}$ на прямую, *параллельную вектору* $\boldsymbol{b}$*, равна*

$$\operatorname{Pr}_{\boldsymbol{b}} \boldsymbol{a} = \frac{(\boldsymbol{a}, \boldsymbol{b})}{(\boldsymbol{b}, \boldsymbol{b})} \boldsymbol{b}. \tag{2.1}$$

2. *Проекция вектора* $\boldsymbol{a}$ на ось, *коллинеарную вектору* $\boldsymbol{b}$*, равна*

$$\operatorname{pr}_{\boldsymbol{b}} \boldsymbol{a} = \frac{(\boldsymbol{a}, \boldsymbol{b})}{|\boldsymbol{b}|}. \tag{2.2}$$

Доказательство. Обозначим $\operatorname{Pr}_{\boldsymbol{b}} \boldsymbol{a}$ через $\boldsymbol{c}$, а вектор $\boldsymbol{a} - \boldsymbol{c}$ — через $\boldsymbol{d}$; очевидно, $\boldsymbol{c} = \lambda \boldsymbol{b}$, где коэффициент λ подлежит определению, а вектор $\boldsymbol{d}$ ортогонален $\boldsymbol{b}$. Имеем:

$$0 = (\boldsymbol{d}, \boldsymbol{b}) = (\boldsymbol{a} - \boldsymbol{c}, \boldsymbol{b}) = (\boldsymbol{a}, \boldsymbol{b}) - \lambda(\boldsymbol{b}, \boldsymbol{b}),$$

откуда

$$\lambda = \frac{\boldsymbol{a}, \boldsymbol{b}}{(\boldsymbol{b}, \boldsymbol{b})} \quad \Rightarrow \quad \operatorname{Pr}_{\boldsymbol{b}} \boldsymbol{a} = \lambda \boldsymbol{b} = \frac{\boldsymbol{a}, \boldsymbol{b}}{(\boldsymbol{b}, \boldsymbol{b})} \boldsymbol{b}.$$

Формула (2.2) вытекает непосредственно из определения 2.17. □

В. Ортогональные базисы. Напомним, что *ортонормированным базисом* называется базис, векторы которого имеют единичную длину (нормированы) и попарно ортогональны. Если отказаться от требования нормированности, получим определение *ортогонального базиса.*

Ортогональные и ортонормированные базисы особенно удобны для практических вычислений, поскольку все формулы аналитической геометрии выглядят в них весьма просто.

2.20. Теорема. *Скалярное произведение векторов* $\boldsymbol{a}$ *и* $\boldsymbol{b}$ *выражается через их координаты* (a_x, a_y, a_z) *и* (b_x, b_y, b_z) *относительно произвольного ортонормированного базиса* $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ *формулой*

$$(\boldsymbol{a}, \boldsymbol{b}) = a_x b_x + a_y b_y + a_z b_z.$$

Если обозначить через A *и* B *соответственно столбцы* $(a_x, a_y, a_z)^T$ *и* $(b_x, b_y, b_z)^T$ *координат векторов* $\boldsymbol{a}$ *и* $\boldsymbol{b}$*, то*

$$(\boldsymbol{a}, \boldsymbol{b}) = A^T B.$$

Доказательство. Подставляя вместо $\boldsymbol{a}$ и $\boldsymbol{b}$ их разложения по ортонормированному базису $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ и пользуясь свойством линейности скалярного произведения, получаем

$$\begin{aligned}
(\boldsymbol{a}, \boldsymbol{b}) &= (a_x \boldsymbol{i} + a_y \boldsymbol{j} + a_z \boldsymbol{k},\ b_x \boldsymbol{i} + b_y \boldsymbol{j} + b_z \boldsymbol{k}) = \\
&= (a_x \boldsymbol{i},\ b_x \boldsymbol{i} + b_y \boldsymbol{j} + b_z \boldsymbol{k}) + \\
&\qquad + (a_y \boldsymbol{j},\ b_x \boldsymbol{i} + b_y \boldsymbol{j} + b_z \boldsymbol{k}) + \\
&\qquad + (a_x \boldsymbol{k},\ b_x \boldsymbol{i} + b_y \boldsymbol{j} + b_z \boldsymbol{k}) = \\
&= a_x b_x \underbrace{(\boldsymbol{i}, \boldsymbol{i})}_{=1} + a_x b_y \underbrace{(\boldsymbol{i}, \boldsymbol{j})}_{=0} + a_x b_z \underbrace{(\boldsymbol{i}, \boldsymbol{k})}_{=0} +
\end{aligned}$$

$$
\begin{aligned}
&+ a_y b_x \underbrace{(\boldsymbol{j}, \boldsymbol{i})}_{=0} + a_y b_y \underbrace{(\boldsymbol{j}, \boldsymbol{j})}_{=1} + a_y b_z \underbrace{(\boldsymbol{j}, \boldsymbol{k})}_{=0} + \\
&+ a_z b_x \underbrace{(\boldsymbol{k}, \boldsymbol{i})}_{=0} + a_z b_y \underbrace{(\boldsymbol{k}, \boldsymbol{j})}_{=0} + a_z b_z \underbrace{(\boldsymbol{k}, \boldsymbol{k})}_{=1} = \\
&= a_x b_x + a_y b_y + a_z b_z.
\end{aligned}
$$

Указанное вычисление можно записать короче, если воспользоваться следующей системой обозначений: векторы ортонормированного базиса обозначим $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$, координаты векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ — через (a_1, a_2, a_3) и (b_1, b_2, b_3) соответственно. Заметив, что

$$
(\boldsymbol{e}_j, \boldsymbol{e}_k) = \begin{cases} 1, & \text{если } j = k, \\ 0, & \text{если } j \neq k, \end{cases}
$$

можем записать, пользуясь свойством линейности скалярного умножения,

$$
\begin{aligned}
(\boldsymbol{a}, \boldsymbol{b}) &= \left(\sum_{j=1}^{3} a_j \boldsymbol{e}_j,\ \sum_{k=1}^{3} b_k \boldsymbol{e}_k \right) = \sum_{j=1}^{3} a_j \left(\boldsymbol{e}_j,\ \sum_{k=1}^{3} b_k \boldsymbol{e}_k \right) = \\
&= \sum_{j=1}^{3} a_j \sum_{k=1}^{3} b_k (\boldsymbol{e}_j, \boldsymbol{e}_k) = \sum_{j=1}^{3} \sum_{k=1}^{3} a_j b_k (\boldsymbol{e}_j, \boldsymbol{e}_k) = \sum_{j=1}^{3} a_j b_j,
\end{aligned}
$$

поскольку в сумме остаются только те слагаемые, в которых $j = k$. □

Скалярный квадрат вектора — это скалярное произведение вектора на себя:

$$
(\boldsymbol{a}, \boldsymbol{a}) = |\boldsymbol{a}|^2.
$$

Ясно, что длину вектора можно выразить через его скалярный квадрат:

$$
|\boldsymbol{a}| = \sqrt{(\boldsymbol{a}, \boldsymbol{a})} = \sqrt{a_x^2 + a_y^2 + a_z^2},
$$

а угол между векторами — найти по формуле

$$
\cos\varphi = \frac{(\boldsymbol{a}, \boldsymbol{b})}{|\boldsymbol{a}| \cdot |\boldsymbol{b}|} = \frac{a_x b_x + a_y b_y + a_z b_z}{\sqrt{a_x^2 + a_y^2 + a_z^2}\,\sqrt{b_x^2 + b_y^2 + b_z^2}}.
$$

2.21. Теорема. *Координаты (a_1, a_2, a_3) произвольного вектора $\boldsymbol{a}$ относительно ортогонального базиса $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ вычисляются по формулам*

$$
a_1 = \frac{(\boldsymbol{a}, \boldsymbol{e}_1)}{(\boldsymbol{e}_1, \boldsymbol{e}_1)}, \quad a_2 = \frac{(\boldsymbol{a}, \boldsymbol{e}_2)}{(\boldsymbol{e}_2, \boldsymbol{e}_2)}, \quad a_3 = \frac{(\boldsymbol{a}, \boldsymbol{e}_3)}{(\boldsymbol{e}_3, \boldsymbol{e}_3)}. \tag{2.3}
$$

В частности, координаты (a_x, a_y, a_z) вектора $\boldsymbol{a}$ относительно ортонормированного базиса $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ вычисляются по формулам

$$
a_x = (\boldsymbol{a}, \boldsymbol{i}), \quad a_y = (\boldsymbol{a}, \boldsymbol{j}), \quad a_z = (\boldsymbol{a}, \boldsymbol{k}) \tag{2.4}
$$

Формулы (2.3) и (2.4) называются формулами Гиббса.

Доказательство. Умножим скалярно обе части разложения
$$\boldsymbol{a} = a_1\boldsymbol{e}_1 + a_2\boldsymbol{e}_2 + a_3\boldsymbol{e}_3$$
на вектор $\boldsymbol{e}_1$ и воспользуемся свойствами линейности скалярного произведения и ортогональности базиса:
$$(\boldsymbol{a}, \boldsymbol{e}_1) = a_1(\boldsymbol{e}_1, \boldsymbol{e}_1) + a_2\underbrace{(\boldsymbol{e}_2, \boldsymbol{e}_1)}_{=0} + a_3\underbrace{(\boldsymbol{e}_3, \boldsymbol{e}_1)}_{=0},$$
откуда
$$a_1 = \frac{(\boldsymbol{a}, \boldsymbol{e}_1)}{(\boldsymbol{e}_1, \boldsymbol{e}_1)}.$$
Остальные формулы получаются аналогично. □

6. Векторное и смешанное произведения векторов

А. Левые и правые базисы. Как указывалось в лекции 1, выбор ориентации *на прямой* означает выбор одного из двух возможных направлений на ней. Если прямая изображена на чертеже горизонтальной линией, то положительным направлением на ней обычно считают направление «слева направо»; для вертикальных прямых положительным считается направление «снизу вверх». Никакого инвариантного (не зависящего от чертежа) математического смысла такие понятия на имеют: поворот рисунка на 180° превращает положительное направление в отрицательное.

Выбор ориентации на плоскости осуществляется указанием некоторого ориентированного угла (отличного от 0° и 180°), т.е. направления вращения «по часовой стрелке» или «против часовой стрелки»: ориентированный угол считается *положительным*, если кратчайший поворот от его первой стороны в направлении второй стороны производится *против часовой стрелки.*

2.22. Определение. Базис *на плоскости* $\big(\boldsymbol{i}, \boldsymbol{j}\big)$ будем называть *правым*, если кратчайший поворот, переводящий вектор $\boldsymbol{i}$ в вектор $\boldsymbol{j}$, осуществляется против часовой стрелки (см. рис. 2.10).

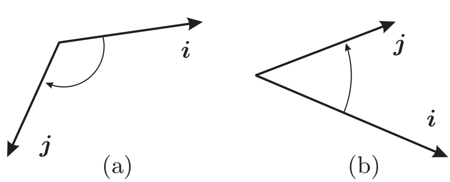


Рис. 2.10. Базис на плоскости: (a) левый, (b) правый

2.23. Определение. Базис *в пространстве* $\big(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\big)$ называется *правым*, если выполнено одно из следующих эквивалентных между собой условий (см. рис. 2.11):

(1) если смотреть из конца вектора $\boldsymbol{k}$, то кратчайший поворот, переводящий вектор $\boldsymbol{i}$ в вектор $\boldsymbol{j}$, осуществляется против часовой стрелки;

(2) векторы $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ удовлетворяют правилу буравчика: если вращать буравчик в направлении поворота, переводящего (кратчайшим образом) вектор $\boldsymbol{j}$ в вектор $\boldsymbol{j}$, то поступательное движение буравчика происходит в направлении вектора $\boldsymbol{k}$;

(3) векторы $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ удовлетворяют правилу правой руки: их расположение совпадает с естественным положением большого, указательного и среднего пальцев правой руки.

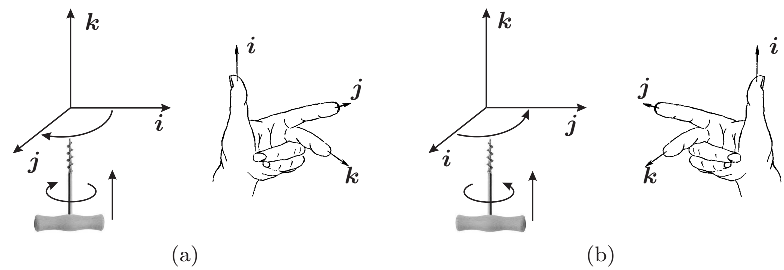


Рис. 2.11. Базис в пространстве: (a) левый, (b) правый

2.24. Замечание. 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).

2.25. Замечание. 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.

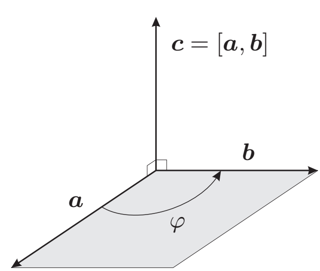
B. Векторное произведение векторов.

2.26. Определение. *Векторным произведением* 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 — угол между векторами;

(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.



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 $\boldsymbol{b}$:

$$S_{\boldsymbol{ab}} = \Big|[\boldsymbol{a}, \boldsymbol{b}]\Big|.$$

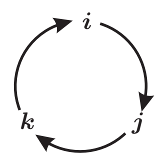
2.27. Предложение (критерий коллинеарности двух векторов)**.** *Векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ коллинеарны тогда и только тогда, когда $[\boldsymbol{a}, \boldsymbol{b}] = \boldsymbol{0}$.*

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. $[\boldsymbol{a}, \boldsymbol{b}] = \boldsymbol{0}$. □

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»).

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:

II множ. / I 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



(2.5)

Для векторов *левого* ортонормированного базиса все знаки $+$ и $-$ должны быть изменены на противоположные.

В дальнейшем мы будем пользоваться исключительно правыми базисами.

2.28. Предложение. *Операция векторного умножения антикоммутативна, т.е. для любых векторов $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{E}$ выполняется соотношение*

$$[\boldsymbol{a}, \boldsymbol{b}] = -[\boldsymbol{b}, \boldsymbol{a}].$$

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. □

С. Смешанное произведение векторов.

2.29. 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 $\boldsymbol{c}$:

$$(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \stackrel{\text{def}}{=\!=} \big(\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big).$$

2.30. Теорема (критерий компланарности трёх векторов)**.** *Векторы $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ компланарны тогда и только тогда, когда $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \mathbf{0}$.*

Доказательство. Необходимость. Пусть векторы $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ компланарны. Будем считать, что $\boldsymbol{a} \neq \mathbf{0}$ 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 $[\boldsymbol{b}, \boldsymbol{c}]$ перпендикулярен 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 $[\boldsymbol{b}, \boldsymbol{c}] = \mathbf{0}$, либо $\cos\varphi = 0$, где φ — угол между векторами $\boldsymbol{a}$ и $[\boldsymbol{b}, \boldsymbol{c}]$. Это означает, что либо $\boldsymbol{a} = \mathbf{0}$, либо $\boldsymbol{b}$ и $\boldsymbol{c}$ коллинеарны, либо $\boldsymbol{a}$ параллелен плоскости π, содержащей векторы $\boldsymbol{b}$ и $\boldsymbol{c}$. Во всех трёх случаях векторы $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ компланарны. □

2.31. Теорема. *Смешанное произведение некомпланарных векторов $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ равно по абсолютной величине объёму $V_{\boldsymbol{abc}}$ параллелепипеда, построенного на этих векторах, отложенных от одной точки, причём*

$$(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \begin{cases} V_{\boldsymbol{abc}}, & \textit{если } \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \textit{ — правая тройка,} \\ -V_{\boldsymbol{abc}}, & \textit{если } \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \textit{ — левая тройка.} \end{cases}$$

Доказательство. Отложив векторы $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ от одной точки, получим параллелепипед, объём которого можно найти по формуле

$$V_{\boldsymbol{abc}} = S_{\boldsymbol{bc}} \cdot h,$$

где $S_{\boldsymbol{bc}}$ — площадь основания, численно равная длине вектора $\Big|[\boldsymbol{b}, \boldsymbol{c}]\Big|$, а h — высота параллелепипеда, равная $|\boldsymbol{a}|\cos\alpha$, где α — угол между векторами $\boldsymbol{a}$ и $[\boldsymbol{b}, \boldsymbol{c}]$. Итак,

$$V_{\boldsymbol{abc}} = \Big|[\boldsymbol{b}, \boldsymbol{c}]\Big| \cdot |\boldsymbol{a}| \cdot \cos\alpha = \Big|\big(\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big)\Big| = \Big|(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})\Big|.$$

Знак смешанного произведения $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ определяется только знаком $\cos\alpha$, но $\cos\alpha > 0$ тогда и только тогда, когда векторы $\boldsymbol{c}$ 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. □

2.32. Предложение. *Для любых векторов $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$*

$$\Big(\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\Big) = \Big([\boldsymbol{a}, \boldsymbol{b}], \boldsymbol{c}\Big). \tag{2.6}$$

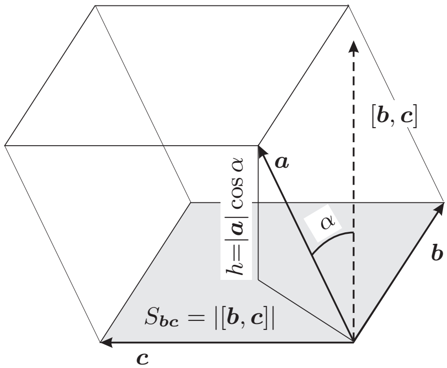


Рис. 2.12. Объём параллелепипеда

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

$$\Big([\boldsymbol{a}, \boldsymbol{b}], \boldsymbol{c}\Big) = \Big(\boldsymbol{c}, [\boldsymbol{a}, \boldsymbol{b}]\Big) = (\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{b}) = V_{\boldsymbol{cab}} = V_{\boldsymbol{abc}} = (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \Big(\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\Big),$$

что и требовалось доказать. □

2.33. Следствие. *Для любых векторов* $\boldsymbol{a}$, $\boldsymbol{b}$ *и* $\boldsymbol{c}$

$$(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = (\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{a}) = (\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{b}) = -(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{b}) = -(\boldsymbol{b}, \boldsymbol{a}, \boldsymbol{c}) = -(\boldsymbol{c}, \boldsymbol{b}, \boldsymbol{a}). \quad (2.7)$$

Доказательство. Например, имеем

$$(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \Big(\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\Big) = \Big(\boldsymbol{a}, -[\boldsymbol{c}, \boldsymbol{b}]\Big) = -(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{b}).$$

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 (см. замечание 2.25). □

2.34. 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 (2.7). □

2.35. 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 свойства **СП3** скалярного произведения (см. с. 48) следует $\boldsymbol{d} = \boldsymbol{0}$, что и требовалось. □

D. Вычисление векторного и смешанного произведений в ортонормированном базисе. Пусть векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ заданы координатами относительно *правого* ортонормированного базиса $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$:

$$\boldsymbol{a} = a_x\boldsymbol{i} + a_y\boldsymbol{j} + a_z\boldsymbol{k} = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix}, \quad \boldsymbol{b} = b_x\boldsymbol{i} + b_y\boldsymbol{j} + b_z\boldsymbol{k} = \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix}.$$

Пользуясь линейностью векторного произведения (предложение 2.35) и таблицей (2.5) векторного умножения для векторов ортонормированного базиса, для векторного произведения $[\boldsymbol{a}, \boldsymbol{b}]$ получаем:

$$[\boldsymbol{a}, \boldsymbol{b}] = \Big[a_x\boldsymbol{i} + a_y\boldsymbol{j} + a_z\boldsymbol{k},\ b_x\boldsymbol{i} + b_y\boldsymbol{j} + b_z\boldsymbol{k}\Big] =$$
$$= a_xb_x \underbrace{[\boldsymbol{i}, \boldsymbol{i}]}_{=\boldsymbol{0}} + a_xb_y \underbrace{[\boldsymbol{i}, \boldsymbol{j}]}_{=\boldsymbol{k}} + a_xb_z \underbrace{[\boldsymbol{i}, \boldsymbol{k}]}_{=-\boldsymbol{j}} +$$
$$+ a_yb_x \underbrace{[\boldsymbol{j}, \boldsymbol{i}]}_{=-\boldsymbol{k}} + a_yb_y \underbrace{[\boldsymbol{j}, \boldsymbol{j}]}_{=\boldsymbol{0}} + a_yb_z \underbrace{[\boldsymbol{j}, \boldsymbol{k}]}_{=\boldsymbol{i}} +$$
$$+ a_zb_x \underbrace{[\boldsymbol{k}, \boldsymbol{i}]}_{=\boldsymbol{j}} + a_zb_y \underbrace{[\boldsymbol{k}, \boldsymbol{j}]}_{=-\boldsymbol{i}} + a_zb_z \underbrace{[\boldsymbol{k}, \boldsymbol{k}]}_{=\boldsymbol{0}}$$
$$= \boldsymbol{i}\Big(a_yb_z - a_zb_y\Big) + \boldsymbol{j}\Big(a_zb_x - a_xb_z\Big) + \boldsymbol{k}\Big(a_xb_y - a_yb_z\Big).$$

Для удобства запоминания этой формулы введём следующее обозначение:

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} \overset{\text{def}}{=\!=} AD - BC;$$

указанное выражение и его запись в виде двухрядной таблицы из чисел называется определителем второго порядка. Определители используются в математике весьма широко; их изучению посвящена лекция 11.

Полученная формула для вычисления векторного произведения при помощи определителей записывается в виде

$$[\boldsymbol{a}, \boldsymbol{b}] = \boldsymbol{i} \begin{vmatrix} a_y & a_z \\ b_y & b_z \end{vmatrix} - \boldsymbol{j} \begin{vmatrix} a_x & a_z \\ b_z & b_z \end{vmatrix} + \boldsymbol{k} \begin{vmatrix} a_x & a_y \\ b_z & b_y \end{vmatrix}. \qquad (2.8)$$

Запись можно ещё сократить, введя понятие определителя третьего порядка:

$$\begin{vmatrix} A & B & C \\ D & E & F \\ G & H & I \end{vmatrix} \overset{\text{def}}{=\!=} A \begin{vmatrix} E & F \\ H & I \end{vmatrix} - B \begin{vmatrix} D & F \\ G & I \end{vmatrix} + C \begin{vmatrix} D & E \\ G & H \end{vmatrix}$$

(обратите внимание на знак «$-$» во втором слагаемом). Тогда формула для векторного произведения примет вид

$$[\boldsymbol{a}, \boldsymbol{b}] = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a_x & a_y & a_z \\ b_z & b_y & b_z \end{vmatrix}. \qquad (2.9)$$

В случае *левого* ортонормированного базиса все векторные произведения базисных векторов имеют противоположный знак, поэтому формула для векторного произведения имеет вид

$$[\boldsymbol{a}, \boldsymbol{b}] = - \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a_x & a_y & a_z \\ b_z & b_y & b_z \end{vmatrix}.$$

2.36. Теорема. *Смешанное произведение векторов $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ выражается через их координаты относительно ортонормированного базиса $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ формулой*

$$(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \pm \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix}, \qquad (2.10)$$

где знак «+» выбирается в случае правого базиса, а знак «$-$» — в случае левого.

Доказательство. Введём обозначение

$$\boldsymbol{d} = [\boldsymbol{b}, \boldsymbol{c}] = \pm \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix} = \begin{pmatrix} d_x \\ d_y \\ d_z \end{pmatrix}.$$

Имеем

$$(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = (\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]) = a_x d_x + a_y d_y + a_z d_z =$$

$$= \pm \left(a_1 \begin{vmatrix} b_y & b_z \\ c_y & c_z \end{vmatrix} - a_2 \begin{vmatrix} b_x & b_z \\ c_x & c_z \end{vmatrix} + a_3 \begin{vmatrix} b_y & b_z \\ c_y & c_z \end{vmatrix} \right) = \pm \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix},$$

что и требовалось доказать. □

Е. 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.*

2.37. Теорема. *Имеет место следующая формула*:

$$\big[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big] = \boldsymbol{b}(\boldsymbol{a}, \boldsymbol{c}) - \boldsymbol{c}(\boldsymbol{a}, \boldsymbol{b}). \tag{2.11}$$

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

$$\big[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]\big] = \alpha\big(\boldsymbol{b}(\boldsymbol{a}, \boldsymbol{c}) - \boldsymbol{c}(\boldsymbol{a}, \boldsymbol{b})\big).$$

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. □

2.38. Теорема. *Справедливо следующее тождество, называемое тождеством Якоби*:

$$[\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{c}]] + [\boldsymbol{b}, [\boldsymbol{c}, \boldsymbol{a}]] + [\boldsymbol{c}, [\boldsymbol{a}, \boldsymbol{b}]] = \boldsymbol{0}.$$

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. □

F. Аксиальные и полярные векторы. Векторные произведения базисных векторов $\boldsymbol{i}$, $\boldsymbol{j}$ и $\boldsymbol{k}$ различны для правого и левого базисов, например, $[\boldsymbol{i}, \boldsymbol{j}] = \boldsymbol{k}$ для правого базиса и $[\boldsymbol{i}, \boldsymbol{j}] = \boldsymbol{k}$ для левого. Поскольку левый базис получается зеркальным отражением из правого, мы заключаем, что векторное произведение меняет направление при зеркальном отражении. Векторы, обладающие таким свойством, называются *аксиальными* векторами, в отличие от обычных — *полярных* векторов.

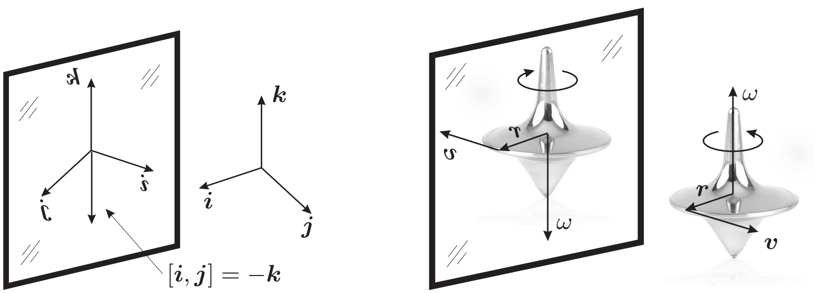


Рис. 2.13.

Простейшим примером аксиального вектора является вектор ω угловой скорости вращающегося тела, величина («длина») которого равна углу поворота тела вокруг центра вращения за единицу времени, а направление определяется правилом буравчика: если вектор угловой скорости ω отложен от центра вращения, то для наблюдателя, находящегося на конце вектора ω, выглядит происходящим против часовой стрелки. При переходе от правого базиса к левому вектор угловой скорости меняет направление: вектор угловой скорости зеркального изображения вращающегося волчка направлен в сторону, противоположную направлению вектора угловой скорости самого волчка. Линейная же скорость точки волчка, которая связана с угловой скоростью формулой $\boldsymbol{v} = [\omega, \boldsymbol{r}]$, где $\boldsymbol{r}$ — радиус-вектор точки относительно центра вращения, является обычным (полярным) вектором, т.е. отражается в зеркале обычным, «нормальным» образом.

7. Упражнения

2.1. Выразите длину x биссектрисы AD внутреннего угла A треугольника ABC через длины c и b сторон AB и AC и угол α между ними.

2.1. Полагая $\overrightarrow{AB} = \boldsymbol{c}$, $\overrightarrow{AC} = \boldsymbol{b}$, $\overrightarrow{AD} = \boldsymbol{x}$, имеем

$$\frac{BD}{DC} = \frac{c}{b} \iff b \cdot BD = c \cdot DC,$$

откуда

$$b\cdot\overrightarrow{BD}=c\cdot\overrightarrow{DC}\iff b(\boldsymbol{x}-\boldsymbol{c})=c(\boldsymbol{b}-\boldsymbol{x})\ \Rightarrow\ \boldsymbol{x}=\frac{1}{b+c}\big(b\boldsymbol{c}+c\boldsymbol{b}\big).$$

Поэтому

$$x=\sqrt{(\boldsymbol{x},\boldsymbol{x})}=\sqrt{\frac{b^2c^2+c^2b^2+2bc(\boldsymbol{b},\boldsymbol{c})}{(b+c)^2}}=\sqrt{\frac{2b^2c^2+2b^2c^2\cos\alpha}{(b+c)^2}}=\frac{2bc\cos\frac{\alpha}{2}}{b+c}.$$

2.2. Найдите длину d диагонали OD параллелепипеда, зная длины его рёбер $OA=a$, $OB=b$, $OC=c$ и углы $\angle BOC=\alpha$, $\angle COA=\beta$, $\angle AOB=\gamma$. Найдите также углы α', β' и γ' между диагональю OD и рёбрами OA, OB, OC.

2.2. Имеем $\boldsymbol{d}=\boldsymbol{a}+\boldsymbol{b}+\boldsymbol{c}$, откуда

$$d^2=\big(\boldsymbol{a}+\boldsymbol{b}+\boldsymbol{c},\ \boldsymbol{a}+\boldsymbol{b}+\boldsymbol{c}\big)=(\boldsymbol{a},\boldsymbol{a})+(\boldsymbol{b},\boldsymbol{b})+(\boldsymbol{c},\boldsymbol{c})+2(\boldsymbol{a},\boldsymbol{b})+2(\boldsymbol{b},\boldsymbol{c})+2(\boldsymbol{a},\boldsymbol{c})=$$
$$=a^2+b^2+c^2+2ab\cos\gamma+2bc\cos\alpha+2ac\cos\beta.$$

Далее,

$$\cos\alpha'=\frac{(\boldsymbol{a},\boldsymbol{d})}{ad}=\frac{(\boldsymbol{a},\ \boldsymbol{a}+\boldsymbol{b}+\boldsymbol{c})}{ad}=\frac{a^2+ab\cos\gamma+ac\cos\beta}{ad}$$

и аналогично для двух других углов.

2.3. Докажите тождество

$$(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})=\begin{vmatrix}(\boldsymbol{a},\boldsymbol{x}) & (\boldsymbol{a},\boldsymbol{y}) & (\boldsymbol{a},\boldsymbol{z})\\ (\boldsymbol{b},\boldsymbol{x}) & (\boldsymbol{b},\boldsymbol{y}) & (\boldsymbol{b},\boldsymbol{z})\\ (\boldsymbol{c},\boldsymbol{x}) & (\boldsymbol{c},\boldsymbol{y}) & (\boldsymbol{c},\boldsymbol{z})\end{vmatrix}.$$

2.3. Используя формулу для двойного векторного произведения, получаем

$$\Big[[\boldsymbol{a},\boldsymbol{b}],[\boldsymbol{p},\boldsymbol{c}]\Big]=\boldsymbol{p}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})-\boldsymbol{c}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{p})=\boldsymbol{b}(\boldsymbol{p},\boldsymbol{c},\boldsymbol{a})-\boldsymbol{a}(\boldsymbol{p},\boldsymbol{c},\boldsymbol{b}).$$

Полагая здесь $\boldsymbol{p}=[\boldsymbol{x},\boldsymbol{y}]$, находим

$$[\boldsymbol{x},\boldsymbol{y}](\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=\boldsymbol{c}\Big(\boldsymbol{a},\boldsymbol{b},[\boldsymbol{x},\boldsymbol{y}]\Big)+\boldsymbol{b}\Big([\boldsymbol{x},\boldsymbol{y}],\boldsymbol{c},\boldsymbol{a}\Big)-\boldsymbol{a}\Big([\boldsymbol{x},\boldsymbol{y}],\boldsymbol{c},\boldsymbol{b}\Big)=$$
$$=\boldsymbol{c}\Big([\boldsymbol{a},\boldsymbol{b}],[\boldsymbol{x},\boldsymbol{y}]\Big)+\boldsymbol{b}\Big([\boldsymbol{x},\boldsymbol{y}],[\boldsymbol{c},\boldsymbol{a}]\Big)-\boldsymbol{a}\Big([\boldsymbol{x},\boldsymbol{y}],[\boldsymbol{c},\boldsymbol{b}]\Big)=$$
$$=\boldsymbol{c}\Big((\boldsymbol{a},\boldsymbol{x})(\boldsymbol{b},\boldsymbol{y})-(\boldsymbol{a},\boldsymbol{y})(\boldsymbol{b},\boldsymbol{x})\Big)+\boldsymbol{b}\Big((\boldsymbol{x},\boldsymbol{c})(\boldsymbol{y},\boldsymbol{a})-(\boldsymbol{x},\boldsymbol{a})(\boldsymbol{y},\boldsymbol{c})\Big)-$$
$$-\boldsymbol{a}\Big((\boldsymbol{x},\boldsymbol{c})(\boldsymbol{y},\boldsymbol{b})-(\boldsymbol{x},\boldsymbol{b})(\boldsymbol{y},\boldsymbol{c})\Big).$$

Умножая обе части этого равенства скалярно на z, имеем

$$(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})=(\boldsymbol{c},\boldsymbol{z})\Big((\boldsymbol{a},\boldsymbol{x})(\boldsymbol{b},\boldsymbol{y})-(\boldsymbol{a},\boldsymbol{y})(\boldsymbol{b},\boldsymbol{x})\Big)+$$
$$+(\boldsymbol{b},\boldsymbol{z})\Big((\boldsymbol{x},\boldsymbol{c})(\boldsymbol{y},\boldsymbol{a})-(\boldsymbol{x},\boldsymbol{a})(\boldsymbol{y},\boldsymbol{c})\Big)-(\boldsymbol{a},\boldsymbol{z})\Big((\boldsymbol{x},\boldsymbol{c})(\boldsymbol{y},\boldsymbol{b})-(\boldsymbol{x},\boldsymbol{b})(\boldsymbol{y},\boldsymbol{c})\Big)=$$

$$
= \begin{vmatrix} (\boldsymbol{a},\boldsymbol{x}) & (\boldsymbol{a},\boldsymbol{y}) & (\boldsymbol{a},\boldsymbol{z}) \\ (\boldsymbol{b},\boldsymbol{x}) & (\boldsymbol{b},\boldsymbol{y}) & (\boldsymbol{b},\boldsymbol{z}) \\ (\boldsymbol{c},\boldsymbol{x}) & (\boldsymbol{c},\boldsymbol{y}) & (\boldsymbol{c},\boldsymbol{z}) \end{vmatrix}.
$$

Доказанное соотношение выражает формулу определителя произведения матриц порядка 3.

2.4. Найдите объём V параллелепипеда, зная длины его рёбер $OA = a$, $OB = b$, $OC = c$ и углы $\angle BOC = \alpha$, $\angle COA = \beta$, $\angle AOB = \gamma$.

2.4. Используя тождество, доказанное в задаче **??**, получаем

$$
V = \big|(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})\big| = \sqrt{(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})^2} = \sqrt{\begin{vmatrix} (\boldsymbol{a},\boldsymbol{a}) & (\boldsymbol{a},\boldsymbol{b}) & (\boldsymbol{a},\boldsymbol{c}) \\ (\boldsymbol{b},\boldsymbol{a}) & (\boldsymbol{b},\boldsymbol{b}) & (\boldsymbol{b},\boldsymbol{c}) \\ (\boldsymbol{c},\boldsymbol{a}) & (\boldsymbol{c},\boldsymbol{b}) & (\boldsymbol{c},\boldsymbol{c}) \end{vmatrix}} =
$$

$$
= \sqrt{\begin{vmatrix} a^2 & ab\cos\gamma & ac\cos\beta \\ ab\cos\gamma & b^2 & bc\cos\alpha \\ ac\cos\beta & bc\cos\alpha & c^2 \end{vmatrix}} = abc\sqrt{\begin{vmatrix} 1 & \cos\gamma & 1\cos\beta \\ \cos\gamma & 1 & \cos\alpha \\ \cos\beta & \cos\alpha & 1 \end{vmatrix}} =
$$

$$
= abc\sqrt{1 + 2\cos\alpha\cos\beta\cos\gamma - \cos^2\alpha - \cos^2\beta - \cos^2\gamma}.
$$

2.5. Найдите векторы $\boldsymbol{x}$ и $\boldsymbol{y}$, если известны их сумма $\boldsymbol{a} \neq \boldsymbol{0}$, скалярное произведение p и векторное произведение $\boldsymbol{b} \neq \boldsymbol{0}$, при условии, что векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ ортогональны. Докажите, что если векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ не ортогональны, то задача не имеет решений.

2.5. Если существуют такие векторы $\boldsymbol{x}$ и $\boldsymbol{y}$, что $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{a}$ и $[\boldsymbol{x},\boldsymbol{y}] = \boldsymbol{b}$, то

$$
(\boldsymbol{a},\boldsymbol{b}) = (\boldsymbol{x}+\boldsymbol{y}, [\boldsymbol{x},\boldsymbol{y}]) = (\boldsymbol{x},\boldsymbol{x},\boldsymbol{y}) + (\boldsymbol{y},\boldsymbol{x},\boldsymbol{y}) = 0,
$$

поэтому при $(\boldsymbol{a},\boldsymbol{b}) \neq 0$ задача неразрешима.

Выражая $\boldsymbol{y}$ из соотношения $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{a}$ и подставляя в уравнения $(\boldsymbol{x},\boldsymbol{y}) = p$ и $[\boldsymbol{x},\boldsymbol{y}] = \boldsymbol{b}$, получаем

$$
(\boldsymbol{x}, \boldsymbol{a} - \boldsymbol{x}) = p, \quad [\boldsymbol{x},\boldsymbol{a}] = \boldsymbol{b}.
$$

Из последнего уравнения следует, что $\boldsymbol{x} \perp \boldsymbol{b}$; поскольку $\boldsymbol{a} \perp \boldsymbol{b}$, то вектор $\boldsymbol{x}$ компланарен векторам $\boldsymbol{a}$ и $[\boldsymbol{a},\boldsymbol{b}]$:

$$
\boldsymbol{x} = \lambda\boldsymbol{a} + \mu[\boldsymbol{a},\boldsymbol{b}].
$$

Подставляя это разложение в уравнение $[\boldsymbol{x},\boldsymbol{a}] = \boldsymbol{b}$, получаем

$$
\big[\lambda\boldsymbol{a} + \mu[\boldsymbol{a},\boldsymbol{b}],\ \boldsymbol{a}\big] = \boldsymbol{b} \iff \mu\big[[\boldsymbol{a},\boldsymbol{b}]\boldsymbol{a}\big] = \boldsymbol{b} \iff \mu(\boldsymbol{a},\boldsymbol{a})\boldsymbol{b} = \boldsymbol{b},
$$

так что

$$
\mu = \frac{1}{a^2}, \quad \boldsymbol{x} = \lambda\boldsymbol{a} + \frac{[\boldsymbol{a},\boldsymbol{b}]}{a^2}.
$$

Подставляя это выражение для $\boldsymbol{x}$ в соотношение $(\boldsymbol{x}, \boldsymbol{a} - \boldsymbol{x}) = p$, находим

$$
\lambda a^2 - \lambda^2 a^2 - \frac{b^2}{a^2} = p,
$$

откуда
$$\lambda = \frac{a^2 \pm \sqrt{a^4 - 4(b^2 + pa^2)}}{2a^2}.$$
Если $a^4 < 4(b^2 + pa^2)$, то задана не имеет решений. Если $a^4 = 4(b^2 + pa^2)$, то $\lambda = 1/2$ и задача имеет единственное решение
$$\boldsymbol{x} = \frac{\boldsymbol{a}}{2} + \frac{[\boldsymbol{a}, \boldsymbol{b}]}{a^2}, \quad \boldsymbol{y} = \frac{\boldsymbol{a}}{2} - \frac{[\boldsymbol{a}, \boldsymbol{b}]}{a^2}.$$
Если $a^4 = 4(b^2 + pa^2)$, то задача имеет два решения:
$$\boldsymbol{x} = \frac{a^2 + \sqrt{a^4 - 4(b^2 + pa^2)}}{2a^2}\boldsymbol{a} + \frac{[\boldsymbol{a}, \boldsymbol{b}]}{a^2}, \ \boldsymbol{y} = \frac{a^2 - \sqrt{a^4 - 4(b^2 + pa^2)}}{2a^2}\boldsymbol{a} - \frac{[\boldsymbol{a}, \boldsymbol{b}]}{a^2}$$
или
$$\boldsymbol{x} = \frac{a^2 - \sqrt{a^4 - 4(b^2 + pa^2)}}{2a^2}\boldsymbol{a} + \frac{[\boldsymbol{a}, \boldsymbol{b}]}{a^2}, \ \boldsymbol{y} = \frac{a^2 + \sqrt{a^4 - 4(b^2 + pa^2)}}{2a^2}\boldsymbol{a} - \frac{[\boldsymbol{a}, \boldsymbol{b}]}{a^2}.$$

2.6. Пусть $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ — некомпланарные векторы. Найдите такой вектор $\boldsymbol{x}$, что $(\boldsymbol{a}, \boldsymbol{x}) = \alpha$, $(\boldsymbol{b}, \boldsymbol{x}) = \beta$, $(\boldsymbol{c}, \boldsymbol{x}) = \gamma$.

2.7. Даны три компланарных вектора $\boldsymbol{x}$, $\boldsymbol{a}$, $\boldsymbol{b}$, причём векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ неколлинеарны. Выразите коэффициенты разложения вектора $\boldsymbol{x}$ по векторам $\boldsymbol{a}$ и $\boldsymbol{b}$.

2.8. Даны три вектора $\overrightarrow{OA} = \boldsymbol{a}$, $\overrightarrow{OB} = \boldsymbol{b}$, $\overrightarrow{OC} = \boldsymbol{c}$, причём векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ неколлинеарны. Пусть H — ортогональная проекция точки C на плоскость OAB. Найдите вектор $\overrightarrow{CH}$.

ГЛАВА 3

Комплексные числа. Многочлены

1. Числовые поля

Из школьного курса алгебры хорошо известна ситуация, когда квадратное уравнение $ax^2 + bx + c = 0$ не имеет корней: это случается, когда дискриминант уравнения $D = b^2 - 4ac$ отрицателен. Из этой ситуации возможен формальный выход: достаточно ввести в рассмотрение «числа» вида $a + b\sqrt{-1}$, как такие квадратные уравнения становятся разрешимыми; более того, для «корней», полученных таким образом, имеют место все свойства, присущие вещественным корням, например, теорема Виета. Разумеется, вопрос о существовании подобного рода «чисел» при этом остаётся открытым.

Числовое поле $\mathbb{K}$ — это множество чисел, замкнутое[1] относительно четырёх арифметических операций: сложения, вычитания, умножения, деления на ненулевое число, обладающих привычными свойствами:

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

Примерами числовых полей могут служить множества рациональных чисел $\mathbb{Q}$ и множество вещественных чисел $\mathbb{R}$. Множества натуральных чисел $\mathbb{N}$ и целых чисел $\mathbb{Z}$ не являются числовыми полями, поскольку в них не выполняется операция деления (а в $\mathbb{N}$ ещё и вычитание).

Числовое поле $\mathbb{K}'$ называется *расширением* числового поля $\mathbb{K}$, если $\mathbb{K} \subset \mathbb{K}'$; при этом поле $\mathbb{K}$ называется *подполем* поля $\mathbb{K}'$. Например, $\mathbb{Q} \subset \mathbb{R}$,

---

[1]Это означает, что результатом операции над произвольными элементами множества $\mathbb{K}$ также является элемент множества $\mathbb{K}$.

т.е. поле $\mathbb{Q}$ является подполем поля $\mathbb{R}$, а поле $\mathbb{R}$ является расширением поля $\mathbb{Q}$.

Пусть $\mathbb{K}$ — числовое поле. *Многочленом* степени n (от одной переменной x) над полем $\mathbb{K}$ называется функция вида

$$P(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \dots + a_{n-2}x^2 + a_{n-1}x + a_n,$$

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

Степень многочлена $\deg P(x)$ — это наибольшая из степеней входящих в его состав одночленов. Степень нулевого многочлена (т.е. многочлена, все коэффициенты которого нулевые) считается равной $-\infty$. В дальнейшем при необходимости будем указывать степень многочлена при помощи нижнего индекса: запись $P_n(x)$ обозначает многочлен степени n.

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

Отметим, что один и тот же многочлен можно рассматривать как элемент разных множеств $\mathbb{K}[x]$; 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 $\pm\sqrt{2}$.

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

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. Фактически 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.

2. Определение поля $\mathbb{C}$

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

3.1. 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 $K = \mathbb{C}$.

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 подмножество

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

Сначала докажем, что

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

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

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:

$$(a+ib)+(c+id)=(a+c)+i(b+d), \tag{3.1}$$

$$(a+ib)\cdot(c+id)=(ac-bd)+i(ad+bc); \tag{3.2}$$

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

Будем искать элемент, *противоположный* к $a+ib$, в виде $x+iy$; для этого требуется решить уравнение $(a+ib)+(x+iy)=0$:

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

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

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

Будем искать элемент, *обратный* к $a+ib\neq 0$, в виде $x+iy$; для этого нужно решить уравнение $(a+ib)\cdot(x+iy)=1$:

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

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

$$(a+ib)^{-1}=\frac{a-ib}{a^2+b^2}, \tag{3.3}$$

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!). На практике деление комплексных чисел осуществляют, домножая числитель $a+ib$ и знаменатель $c+id$ на число $c-id$; в результате в знаменателе образуется вещественной 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 $c+id$ (см. ниже).

Поскольку 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, $K = \mathbb{C}$.

Предыдущее исследование подсказывает, как *доказать существование* поля комплексных чисел. Обозначим через $\mathbb{C}$ множество $\mathbb{R}\times\mathbb{R}$ всех упорядоченных пар вещественных чисел и введём на нём операции сложения и умножения формулами, подсказанными соотношениями (3.1) и (3.2):

$$(a,b)+(c,d) = (a+c,\ b+d), \tag{3.4}$$

$$(a,b)\cdot(c,d) = (ac-bd,\ ad+bc). \tag{3.5}$$

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

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

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 в $\mathbb{C}$. Формула (3.3) подсказывает, как должен выглядеть элемент, обратный к $(a,b)\neq(0;0)$:

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

непосредственная проверка (умножение (a,b) на $(a,b)^{-1}$) доказывает справедливость этой формулы. Таким образом, $\mathbb{C}$ является *полем.*

Далее,

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

т.е. операции над парами 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* $\mathbb{C}$.

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

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

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

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

Комплексное число $a - ib$ называется *сопряжённым* к числу $z = a + ib$ и обозначается $\bar{z}$; очевидно,

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

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

Непосредственным вычислением проверяется следующее утверждение.

3.2. 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*).

Итак, мы доказали существование поля $\mathbb{C}$, построив его модель, состоящую из упорядоченных пар вещественных чисел, арифметические операции над которыми определяются формулами (3.4) и (3.5). В дальнейшем будут построены и другие модели поля $\mathbb{C}$.

B. Тригонометрическая форма записи. Формула Муавра. Комплексное число $z = x + iy$ можно изобразить точкой (x, y) координатной плоскости либо радиус-вектором этой точки (т.е. вектором с началом в $(0, 0)$ и концом в (x, y)). При сложении чисел соответствующие векторы складываются (по правилу параллелограмма); разность чисел $z_2 - z_1$ изображается вектором с началом в точке z_1 и концом в точке z_2.

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.

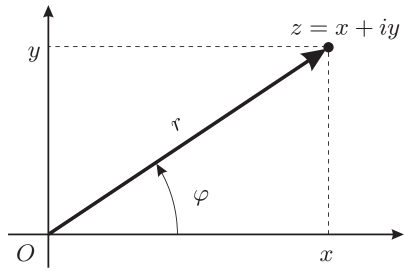


Рис. 3.1.

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,

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

Аргумент числа 0 не определён.

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*:

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

Перемножим два комплексных числа, записанных в тригонометрической форме:

$$\begin{aligned} r_1(\cos\varphi_1 + i\sin\varphi_1) \cdot r_2(\cos\varphi_2 + i\sin\varphi_2) &= \\ = r_1 r_2\Big(\cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2 + i\cos\varphi_1\sin\varphi_2 + i\sin\varphi_1\cos\varphi_2\Big) &= \\ = r_1 r\Big(\cos(\varphi_1 + \varphi_2) + i\sin(\varphi_1 + \varphi_2)\Big). &\quad (3.6) \end{aligned}$$

Таким образом,

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

Отсюда очевидным образом получается так называемая *формула Муавра*:

$$\Big(\cos\varphi + i\sin\varphi\Big)^n = \cos n\varphi + i\sin n\varphi, \tag{3.7}$$

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

C. Формула Эйлера. Показательная форма записи. Рассмотрим функцию $f(\varphi) = \cos\varphi + i\sin\varphi$. Из (3.6) 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.* Она является *определением* функции $e^{i\varphi}$ и потому не доказывается, хотя математический анализ доставляет множество доводов в пользу целесообразности такого определения.

Нетрудно доказать, что

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

Действительно, если $z_1 = x-1+iy_1$ и $z_2 = x_2+iy_2$, то

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

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

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

в частности, она периодична с периодом $2i\pi$:

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

D. Возведение комплексного числа в степень. Тригонометрическая и показательная формы записи удобны при возведении комплексных чисел в степень:

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

Эта формула была ранее доказана при $n\in\mathbb{N}$, но легко убедиться, что она справедлива и при $n\in\mathbb{Z}$. Действительно, поскольку

$$\frac{1}{\cos\varphi + i\sin\varphi} = \frac{\cos\varphi - i\sin\varphi}{(\cos\varphi + i\sin\varphi)(\cos\varphi - i\sin\varphi)} = \cos\varphi - i\sin\varphi,$$

получаем

$$\Big[r(\cos\varphi + i\sin\varphi)\Big]^{-n} = \left(\frac{1}{r(\cos\varphi + i\sin\varphi)}\right)^n = r^{-n}(\cos\varphi - i\sin\varphi)^n =$$

$$= r^{-n}(\cos n\varphi - \sin n\varphi) = r^{-n}\big(\cos(-n\varphi) + \sin(-n\varphi)\big).$$

Те же выкладки в показательной форме намного короче:

$$\frac{1}{e^{i\varphi}} = \frac{e^{-i\varphi}}{e^{i\varphi}\cdot e^{-i\varphi}} = e^{-i\varphi}, \quad \left(re^{i\varphi}\right)^{-n} = r^{-n}\left(e^{-i\varphi}\right)^n = r^{-n}e^{-in\varphi}.$$

3. Извлечение корней из комплексных чисел

А. Формула извлечения корня. Число w называется *корнем n-й степени из числа z*, если $w^n = z$:

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

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. Имеем:

$$\left(Re^{i\Phi}\right)^n = re^{i\varphi} \quad \Longleftrightarrow \quad R^n e^{in\Phi} = re^{i\varphi} \quad \Longleftrightarrow$$

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

Таким образом, получается не одно, а несколько значений корня, однако различными будут только те, которые отвечают значениям $k = 0, 1, \ldots, n-1$. Геометрически эти корни изображаются вершинами правильного n-угольника, вписанного в окружность радиуса $r^{1/n}$.

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

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

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

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

Таким образом, мы доказали следующую теорему.

3.3. Теорема. *Существует ровно n различных значений корня n-й степени из 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 числа k принимают значения $0, 1, \ldots, n-1$.

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.

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

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

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

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

где берётся арифметическое значение корня. Учитывая первое уравнение 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 из положительного вещественного числа. Найденные числа x и y удовлетворяют первому уравнению системы $x^2 - y^2 = a$. Проверим выполнение второго уравнения $2xy = b$:

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

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

$$\sqrt{a+ib} = \pm \left(\sqrt{\frac{\sqrt{a^2+b^2}+a}{2}} + i \operatorname{sign} b \sqrt{\frac{\sqrt{a^2+b^2}-a}{2}} \right). \tag{3.8}$$

[1] Эта функция определяется следующим образом:

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

4. Элементарные функции комплексного аргумента

А. Тригонометрические функции. 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:

$$\cos z \stackrel{\text{def}}{=\!=} \frac{e^{iz} + e^{-iz}}{2}, \quad \sin z \stackrel{\text{def}}{=\!=} \frac{e^{iz} - e^{-iz}}{2i}. \tag{3.9}$$

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.

В. Гиперболические функции. Гиперболические функции определяются равенствами, похожими на (3.9):

$$\operatorname{ch} z \stackrel{\text{def}}{=\!=} \frac{e^{z} + e^{-z}}{2}, \quad \operatorname{sh} z \stackrel{\text{def}}{=\!=} \frac{e^{z} - e^{-z}}{2}. \tag{3.10}$$

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

Сравнивая формулы (3.9) и (3.10), находим связь между гиперболическими и тригонометрическими функциями:

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

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

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

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

$$\operatorname{sh} x + \operatorname{sh} y = \frac{1}{i}(\sin ix + \sin iy) =$$

$$
\begin{aligned}
&= \frac{1}{i}\cdot 2\sin\frac{i(x+y)}{2}\cos\frac{i(x-y)}{2} = \\
&= \frac{2}{i}\cdot i\,\mathrm{sh}\,\frac{x+y}{2}\,\mathrm{ch}\,\frac{x-y}{2} = 2\,\mathrm{sh}\,\frac{x+y}{2}\,\mathrm{ch}\,\frac{x-y}{2}.
\end{aligned}
$$

$$
\cos(x+iy) = \cos x\cos iy - \sin x\sin iy = \cos x\,\mathrm{ch}\,y - i\sin x\,\mathrm{sh}\,y,
$$

$$
\sin(x+iy) = \sin x\cos iy + \cos x\sin iy = \sin x\,\mathrm{ch}\,y + i\cos x\,\mathrm{sh}\,y.
$$

C. Логарифм комплексного числа. Представим комплексное число z в показательной форме:

$$
z = r\,e^{i\varphi} = e^{\ln r}e^{i\varphi} = e^{\ln r + i\varphi}.
$$

Эта запись показывает, что логарифмом комплексного числа z естественно считать комплексное число $\ln r + i\varphi$. (Это в некоторой степени объясняет «логарифмическое поведение» аргумента: $\arg(z_1z_2) = \arg z_1 + \arg z_2$, точнее, $\mathrm{Arg}(z_1z_2) = \mathrm{Arg}\,z_1 + \mathrm{Arg}\,z_2$.)

Логарифмическая функция, введённая таким образом, определена для всех комплексных чисел $z \neq 0$. Однако в силу многозначности аргумента логарифм также оказывается «многозначной функцией» (этот оксюморон часто встречается в теории функции комплексного переменного). Итак, принимаем следующее определение:

$$
\mathrm{Ln}\,z \overset{\mathrm{def}}{=\!=} \ln|z| + \mathrm{Arg}\,z.
$$

Можно было бы установить однозначность, выбирая в приведённом определении лишь главное значение аргумента, но это приводит к ряду неудобств. Например, свойство логарифма

$$
\mathrm{Ln}(z_1z_2) = \mathrm{Ln}\,z_1 + \mathrm{Ln}\,z_2,
$$

так же, как и соответствующее свойство аргумента,

$$
\mathrm{Arg}(z_1z_2) = \mathrm{Arg}\,z_1 + \mathrm{Arg}\,z_2,
$$

имеет место лишь с учётом многозначности.

Например,

$$
\begin{aligned}
&\mathrm{Ln}\,1 = \ln|1| + i\,\mathrm{Arg}\,1 = 2\pi ik, && k\in\mathbb{Z},\\
&\mathrm{Ln}\,i = \ln|i| + i\,\mathrm{Arg}\,i = i\left(\frac{\pi}{2} + 2\pi k\right), && k\in\mathbb{Z},
\end{aligned}
$$

D. Степенная функция. Определим степенную функцию z^α комплексного аргумента z с произвольным комплексным показателем α следующим образом:

$$
z^\alpha = \exp(\alpha\,\mathrm{Ln}\,z).
$$

В силу многозначности логарифма степенная функция также оказывается многозначной.

Вычислим, например, значение i^i:

$$
i^i = \exp(i\cdot\mathrm{Ln}\,i) = \exp\left(-\frac{\pi}{2} - 2\pi k\right), \quad k\in\mathbb{Z}.
$$

Получили довольно неожиданный результат: «очень мнимое» выражение i^i принимает бесконечный набор вещественных значений!

5. Многочлены

А. Кольцо многочленов. Множество многочленов $\mathbb{K}[x]$ замкнуто относительно операций сложения, вычитания и умножения; это означает, что результатом выполнения указанных операций над многочленами является также многочлен. При этом для степени суммы и произведения многочленов $P(x)$ и $Q(x)$ имеют место соотношения

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

Будем говорить, что многочлен $P(x)$ *делится* на многочлен $D(x)$, если существует такой многочлен $Q(x)$, что $P(x) = D(x)Q(x)$; при этом $Q(x)$ называется *частным* от деления $P(x)$ на $D(x)$:

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

Ясно, что

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

Деление многочленов осуществляется при помощи алгоритма «деления уголком».

Деление одного многочлена на другой в обычном смысле слова в кольце $\mathbb{K}[x]$, как правило, невозможно. Однако возможно деление с остатком, весьма похожее на деление с остатком в множестве целых чисел.

3.4. Теорема (теорема о делении с остатком)**.** *Пусть $P(x), D(x) \in \mathbb{K}[x]$, причём $D(x) \neq 0$. 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

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

где $a_0 \neq 0$, $b_0 \neq 0$. 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 $\deg\big(P(x) - D(x)Q(x)\big) < \deg D(x)$. Он и будет неполным частным от деления $P(x)$ на $D(x)$, а многочлен $R(x) = P(x) - D(x)Q(x)g$ будет остатком.

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

$$P(x) = D(x)Q_1(x) + R_1(x) = D(x)Q_2(x) + R_2(x),$$

где

$$Q_1(x) \neq Q_2(x), \quad \deg R_1(x) < \deg D(x), \quad \deg R_2(x) < \deg D(x),$$

и, следовательно,

$$\deg\big(R_1(x) - R_2(x)\big) < \deg D(x).$$

Тогда

$$R_1(x) - R_2(x) = \big(Q_2(x) - Q_1(x)\big)D(x),$$

откуда

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

Полученное противоречие показывает, что $Q_1(x) = Q_2(x)$ и, стало быть, $R_1(x) = R_2(x)$. □

Таким образом, множество многочленов не является полем, поскольку в нём не всегда выполнима операция деления. Множество, замкнутое относительно операций сложения, вычитания и умножения, но незамкнутое относительно операции деления (точнее, в котором такая операция не может быть корректно определена), называется *кольцом.* Итак, *множество* $\mathbb{K}[x]$ *многочленов является кольцом.* Ещё одним хорошо знакомым читателю примером кольца является множество целых чисел. В дальнейшем нам встретятся другие примеры колец.

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

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

Поэтому остаток R при делении многочлена $P(x)$ на линейный двучлен $x - c$ равен значению многочлена $P(x)$ в точке c:

$$R = P(c).$$

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

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

Достаточность. Пусть $P(c) = 0$. Тогда, положив $x = c$ в равенстве $P(x) = (x-c)Q(x)+R$, получим $R = P(c) = 0$, так что $P(x) = (x-c)Q(x)$. □

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

$$a_0x^n + a_1x^{n-1} + a_2x^{n-2} \cdots + a_{n-1}x + a_n =$$
$$(x - c)\Big(b_0x^{n-1} + b_1x^{n-2} + b_2x^{n-3} \cdots + b_{n-2}x + b_{n-1}\Big) + R =$$
$$= b_0x^n+(b_1-cb_0)x^{n-1}+(b_2-cb_1)x^{n-2}+\cdots+(b_{n-1}+cb_{n-2})x+(R-cb_{n-1}).$$

Приравнивая коэффициенты при одинаковых степенях x, получаем

$$\begin{aligned} b_0 &= a_0, \\ b_1 &= cb_0 + a_1, \\ b_2 &= cb_1 + a_2, \\ &\ldots\ldots\ldots\ldots\ldots \\ b_{n-1} &= cb_{n-2} + a_{n-1}, \\ r &= cb_{n-1} + a_n. \end{aligned} \tag{3.11}$$

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

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

Элементы нижней строки этой таблицы вычисляются последовательно по формулам (3.11): $b_0 = a_0$, а каждый последующий элемент равен сумме предыдущего элемента, умноженного на c, и элемента, находящегося над ним.

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

6. Корни многочлена

A. Количество корней многочлена.

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

Доказательство. Применим индукцию по степени многочлена. 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. $\square$

Теорема 3.6 утверждает, что число корней многочлена степени $\leqslant n$ не превосходит n. Для любого $n \in \mathbb{N}$ можно указать многочлен степени n, имеющий ровно n корней. Например, многочлен $(x-1)(x-2)\dots(x-n) \in \mathbb{R}[x]$ имеет n корней $1, 2, \dots, n$. В то же время существуют многочлены, число корней которых меньше их степени; например, многочлен второй степени $x^2 + 1 \in \mathbb{R}[x]$ вообще не имеет корней (обратите внимание, что многочлен $x^2 + \underline{1 \in \mathbb{C}[x]}$ имеет корни $\pm i$). Поэтому естествен интерес к вопросам о существовании корней многочлена и о точном их числе; этот вопрос анализируется ниже.

B. Теорема единственности.

3.7. Теорема (теорема единственности)**.** *Существует не более одного многочлена степени не выше n, принимающего в данных (различных) точках $c_1, \dots, c_{n+1}$ заданные значения. Иными словами, многочлен степени не выше n однозначно определяется своими значениями в $n+1$ точках.*

Доказательство. Предположим, что существуют два многочлена $P(x)$ и $Q(x)$ степени не выше n, принимающие равные значения в точках $c_1, \dots, c_{n+1}$. Рассмотрим многочлен $S(x) = P(x) - Q(x)$. Его степень также не выше n, причём $S(c_j) = 0$, $j = 1, \dots, n+1$, т.е. $c_1, \dots, c_{n+1}$ — корни многочлена $h(x)$. По теореме 3.6 получаем $S(x) = 0$, т.е. $P(x) = Q(x)$. $\square$

3.8. Теорема (теорема о равенстве многочленов)**.** *Два многочлена $P(x) = a_0x^n + a_1x^{n-1} + \dots + a_n$ и $Q(x) = b_0x^n + b_1x^{n-1} + \dots + b_n$ над числовым полем $\mathbb{K}$ тождественно равны тогда и только тогда, когда они имеют равные коэффициенты при одинаковых степенях переменной:*

$$a_0 = b_0, \quad a_1 = b_1, \quad \dots a_n = b_n.$$

В частности, тождественно равными могут быть лишь многочлены равных степеней.

Доказательство. Очевидно, если коэффициенты многочленов $P(x)$ и $Q(x)$ попарно равны, то $P(x) \equiv Q(x)$. Обратно, пусть многочлены тождественно равны, т.е. для любого $c_0 \in \mathbb{K}$ имеем $P(c) = Q(c)$. Введём обозначение $n = \max\{\deg f, \deg g\}$. Так как числовое поле $\mathbb{K}$ бесконечно, в нём найдутся $n+1$ различных элементов $c_1, \ldots, c_{n+1}$. Согласно предположению, многочлены $P(x)$ и $Q(x)$ принимают одинаковые значения в каждой из точек c_j, $j = 1, \ldots, n+1$, так что по теореме 3.7 получаем $P(x) = Q(x)$. $\square$

C. Кратность корня. 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

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

3.9. Определение. *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,

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

Если $k = 1$, корень c называется *простым*, если $k > 1$ — *k-кратным*.

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 $P(x)$.

3.10. Теорема. *Если число c является корнем кратности p многочлена $P(x)$, то оно является корнем кратности $p - 1$ его производной $P'(x)$.*

Доказательство. Согласно условию имеем

$$P(x) = (x - c)^p Q(x), \quad Q(c) \neq 0.$$

Продифференцируем это равенство:

$$P'(x) = p(x-c)^{p-1}Q(x) + (x-c)^p Q'(x) = (x-c)^{p-1}\Big[pQ(x) + (x-c)Q'(x)\Big].$$

Таким образом, кратность корня c многочлена $P'(x)$ не ниже $p - 1$. Проверим, что она не может быть равна p, т.е. что значение выражения в квадратных скобках при $x = c$ отлично от нуля:

$$\Big[pQ(x) + (x - c)Q'(x)\Big]\Big|_{x=c} = pQ(c) \neq 0,$$

что и требовалось. $\square$

Следующее утверждение уточняет теорему 3.6.

3.11. Теорема. *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.*

Доказательство. Пусть c_1 — корень многочлена $P(x)$; тогда
$$P(x) = (x - c_1)P_1(x), \quad P_1(x) \in \mathbb{K}[x].$$
Пусть c_2 — корень многочлена P_1; тогда
$$P_1(x) = (x - c_2)P_2(x), \quad P_2(x) \in \mathbb{K}[x],$$
и, следовательно,
$$P(x) = (x - c_1)(x - c_2)P_2(x).$$
Продолжая подобным образом, в конце концов представим многочлен в виде
$$P(x) = (x - c_1)(x - c_2)\dots(x - c_m)Q(x),$$
где многочлен $Q(x) \in \mathbb{K}[x]$ не имеет корней в числовом поле $\mathbb{K}$. Числа $c_1, \dots, c_m$ — это все корни многочлена $P(x)$. В самом деле, для любого $c \in \mathbb{K}$ имеем
$$P(c) = (c - c_1)(c - c_2)\dots(c - c_m)Q(c).$$
Поскольку $Q(c) \neq 0$, то $P(c) = 0$ тогда и только тогда, когда $c = c_j$ при некотором j. Таким образом, число корней многочлена $P(x)$ не превосходит m. Если среди корней c_j, $j = 1, 2 \dots, m$, имеются кратные, то, очевидно, каждый корень подсчитывался столько раз, какова его кратность. □

D. Основная теорема алгебры.

3.12. 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 опускаем, поскольку его нельзя провести чисто алгебраическими средствами: требуется применение методов математического анализа. Наиболее короткое и изящное доказательство будет приведено в курсе теории функций комплексной переменной.

Легко доказать, что *каждый многочлен степени n в поле $\mathbb{C}$ имеет ровно n корней, если каждый корень считать столько раз, какова его кратность*. Действительно, рассмотрим многочлен[2] $P_n(z)$. Согласно основной теореме алгебры он имеет корень $z = c_1$ и может быть представлен в виде
$$P_n(z) = (z - c_1)P_{n-1}(z).$$
Многочлен $P_{n-1}(z)$ также имеет корень $z = c_2$, так что
$$P_n(z) = (z - c_1)(z - c_2)P_{n-2}(z).$$
Продолжая процедуру, получаем, что многочлен $P_n(z)$ допускает разложение вида
$$A_n(z) = a(z - c_1)(z - c_2)\cdots(z - c_n),$$
причём среди корней $c_1, \dots, c_n$ могут быть и совпадающие.

---

[1] Традиционно переменная, принимающая комплексные значения, обозначается буквой z.

[2] Здесь нижние индексы одновременно различают многочлены между собой и указывают их степени.

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

3.13. Теорема. *Пусть $P(z)$ — многочлен с вещественными коэффициентами. Тогда для любого $z \in \mathbb{C}$ имеем*

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

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

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

Поэтому

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

что и требовалось доказать. □

3.14. Теорема. *Если $P(z)$ — многочлен с вещественными коэффициентами, число c является его корнем, то сопряжённое число $\bar{z}$ также является корнем многочлена $P(z)$.*

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

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:

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

так что он не имеет вещественных корней.

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 кольца $\mathbb{K}[x]$, называется *приводимым*; в противном случае, т.е. когда такого разложения не существует, многочлен называется *неприводимым*. Неприводимые многочлены в кольце $\mathbb{K}[x]$ являются аналогами простых чисел в кольце $\mathbb{Z}$ целых чисел.

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

$$P(x) = a(x - c_1)^{\alpha_1} \cdots (x - c_s)^{\alpha_s} (x + p_1 x + q_1)^{\beta_1} \cdots (x + p_r x + q_r)^{\beta_r}.$$

Таким образом, в кольце $\mathbb{R}[x]$ многочленов с вещественыыми коэффициентами неприводимыми являются лишь многочлены первой степени и квадратные трёхчлены с отрицательным дискриминантом.

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

7. Упражнения

ГЛАВА 4

Матрицы и операции над ними

1. Матрицы

А. Основные определения. *Матрицей* размера $m \times n$ (или $(m \times n)$-матрицей) над множеством X называется 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:

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

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

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}.$$

Используются также следующие сокращённые обозначения, смысл которых вполне очевиден:

$$A = (a_j^i)_n^m, \quad B = (b_{ij})_{mn}.$$

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

В дальнейшем полезным окажется также следующее обозначение элементов матрицы: если $A = (a_j^i)_n^m$ или $B = (b_{ij})_{mn}$, то $[A]_j^i = a_j^i$ и

$[B]_{ij} = b_{ij}$; квадратные скобки здесь обозначают «операцию извлечения элемента из матрицы».

Множество всех матриц размера $m \times n$, элементы которых принадлежат множеству X, обозначается $X^{m\times n}$. Мы будем работать в основном с матрицами, элементами которых являются элементы некоторого поля $\mathbb{K}$ (как правило $\mathbb{K} = \mathbb{Q}$, $\mathbb{R}$ или $\mathbb{C}$); множество таким матриц обозначаются $\mathbb{K}^{m\times n}$.

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

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

и называется *арифметическим пространством строк.*

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

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

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

Во многих случаях удобно считать, что структурными элементами матрицы $A \in X^{m\times n}$ являются не отдельные её элементы, а целые строки или столбцы. В этом случае мы заключаем матрицу в квадратные скобки и пользуемся нумерацией по системе «верхний-нижний индекс»; столбцы матрицы A при этом будем обозначать A_j, $j = 1, \ldots, n$, а строки — A^i, $i = 1, \ldots, n$:

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

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

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

где

$$\begin{aligned} A^1 &= \begin{pmatrix} a_1^1 & a_2^1 & \cdots & a_n^1 \end{pmatrix}, \\ A^2 &= \begin{pmatrix} a_1^2 & a_2^2 & \cdots & a_n^2 \end{pmatrix}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ A^m &= \begin{pmatrix} a_1^m & a_2^m & \cdots & a_n^m \end{pmatrix}. \end{aligned}$$

Более общим является понятие *блочной матрицы*, т.е. матрицы, элементы которой в свою очередь также 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

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

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

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 строк равно количеству столбцов,

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

Порядком квадратной матрицы называют количество её строк или столбцов. *Главная диагональ* квадратной матрицы состоит из элементов $a_1^1, a_2^2, \ldots, a_n^n$, а *побочная диагональ* — из элементов a_n^1, a_{n-1}^2, $\ldots$, a_2^{n-1}, a_1^n. Сумма элементов главной диагонали называется *следом* 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,

$$\begin{pmatrix} a_1^1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & a_2^2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & a_3^3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{n-1}^{n-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & a_n^n \end{pmatrix} = \operatorname{diag}\left(a_1^1, a_2^2, a_3^3, \ldots, a_{n-1}^{n-1}, a_n^n\right);$$

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

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

В. Линейные операции над матрицами. В дальнейшем рассматриваем матрицы, элементами которых являются элементы произвольного числового поля $\mathbb{K}$.

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

$$a^i_j = b^i_j \quad \forall i = 1, \ldots, m, \ \forall j = 1, \ldots, n.$$

Сумма матриц $A = (a^i_j)^m_n$ *и* $B = (b^i_j)^m_n$ одинакового размера $m \times n$ и *произведение матрицы* $A = (a^i_j)^m_n$ *на число* α определяются следующим образом:

$$[A + B]^i_j = [A]^i_j + [B]^i_j, \quad [\alpha A]^i_j = \alpha [A]^i_j \quad \forall i = 1, \ldots, m, \ \forall j = 1, \ldots, n.$$

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 тех же местах.

4.1. Теорема. *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:*

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

(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 теоремы очевидным образом вытекает из определения линейных операций над матрицами.

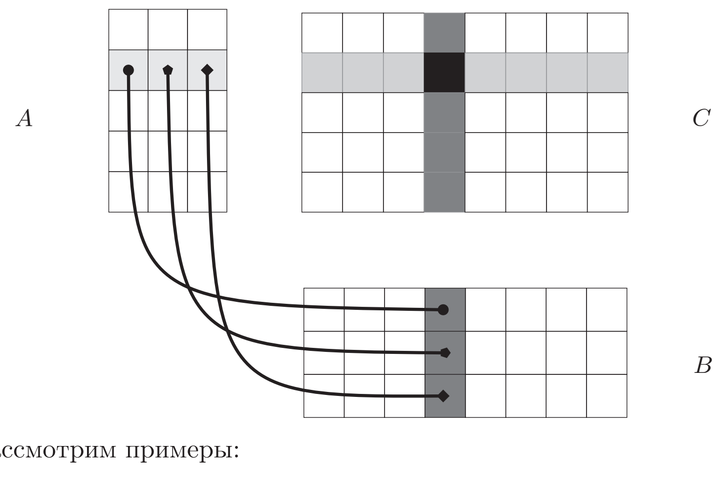
Если имеется несколько матриц $A, B, C, \ldots \in \mathbb{K}^{m\times n}$ и такое же количество числовых коэффициентов $\alpha, \beta, \gamma, \ldots \in \mathbb{K}$, то выражение $\alpha A+\beta B+\gamma C+\ldots$ называется линейной комбинацией матриц $A, B, C, \ldots$ с коэффициентами $\alpha, \beta,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

2. Произведение матриц

А. Определение и примеры. *Произведением* матриц $A \in \mathbb{K}^{m\times s}$ и $B \in \mathbb{K}^{s\times n}$ называется матрица $AB \in \mathbb{K}^{m\times n}$, элементы которой вычисляются по формуле

$$[AB]^k_j = \sum_{l=1}^{s} a^k_l\, b^l_j.$$

Таким образом, элемент $[AB]^k_j$, расположенный в k-й строке и j-м столбце матрицы AB, равен сумме попарных произведений элементов k-й строки матрицы A и элементов j-го столбца матрицы B. Ясно, что перемножить матрицы можно лишь в том случае, когда *количество столбцов первой из перемножаемых матриц равно количеству строк второй.* Следующая схема иллюстрирует правило умножения матриц:



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.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \left(\begin{array}{c|c} 1\cdot 1 + 2\cdot 3 + 3\cdot 5 & 1\cdot 2 + 2\cdot 4 + 3\cdot 6 \\ 4\cdot 1 + 5\cdot 3 + 6\cdot 5 & 4\cdot 2 + 5\cdot 4 + 6\cdot 6 \end{array} \right) = \begin{pmatrix} 22 & 28 \\ 49 & 64 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \left(\begin{array}{c|c|c} 1\cdot 1 + 2\cdot 4 & 1\cdot 2 + 2\cdot 5 & 1\cdot 3 + 2\cdot 6 \\ 3\cdot 1 + 4\cdot 4 & 3\cdot 2 + 4\cdot 5 & 3\cdot 3 + 4\cdot 6 \\ 5\cdot 1 + 6\cdot 4 & 5\cdot 2 + 6\cdot 5 & 5\cdot 3 + 6\cdot 6 \end{array} \right) = \begin{pmatrix} 9 & 12 & 15 \\ 19 & 26 & 33 \\ 29 & 40 & 51 \end{pmatrix}.$$

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

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

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

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

Такие матрицы называются *коммутирующими.*

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

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

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

$$\delta^i_j = \begin{cases} 1, & \text{если } i = j, \\ 0, & \text{если } i \neq j; \end{cases} \tag{4.1}$$

δ^i_j называется *символом Кронекера.*

4.2. Теорема. *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.*

Доказательство. Докажем ассоциативность. Пусть

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

Проверим совпадение 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 $(AB)C$ равны.

Проверим совпадение соответствующих элементов матриц:

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

Остальные свойства доказываются аналогично. □

4.3. Теорема. *Множество* $\mathbb{K}^{n\times n}$ *квадратных матриц с элементами из поля* $\mathbb{K}$ *является ассоциативным некоммутативным кольцом с единицей, в котором имеются делители нуля.*

Для доказательства достаточно проверить, что выполнены все требования определения кольца; проделайте это самостоятельно.

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

$$A^0 = \mathbb{1}, \quad A^d = \underbrace{A \cdots A}_{d \text{ раз}}.$$

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

$$f(A) = a_0 A^d + a_1 A^{d-1} + \cdots + a_{d-1} A + a_d \mathbb{1}.$$

B. Структура произведения матриц. 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

$$A = \Big[A_1, \ A_2, \ \ldots, \ A_p\Big];$$

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:

$$C = \begin{Vmatrix} C_1 & C_2 & \ldots & C_n \end{Vmatrix}$$

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 следующую теорему (утверждения (3) и (4) теоремы, касающиеся строк матрицы $C = AB$, докажите самостоятельно).

4.4. Теорема. *Пусть* $A = (a_l^j)_p^m$, $B = (b_k^l)_n^p$.

(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:*

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

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

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

C. Обратная матрица.

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

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

Матрица, для которой существует обратная, называется *обратимой*.

Не все матрицы обратимы; очевидно, для нулевой матрицы обратная не существует. Вопрос о существовании обратной матрицы будет решён позже (см. лекцию 11, раздел 5, с. 142).

4.6. Теорема. *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 4.5.

(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

$$(A^{-1})^{-1} \cdot A^{-1} = (AA^{-1})^{-1} = \mathbb{1}^{-1} = \mathbb{1},$$

что и требовалось. □

Элементарными методами (методом грубой силы) можно доказать следующую теорему.

4.7. Теорема. *Матрица*

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

обратима тогда и только тогда, когда $ad - bc \neq 0$, и обратная к ней вычисляется по формуле

$$A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{4.2}$$

Если $ab - bc = 0$, то матрица A обратной не имеет.

Доказательство. Будем искать обратную матрицу в виде

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

Перемножая эту матрицу с исходной, находим

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

Приравняем полученную матрицу единичной матрице:

$$\begin{pmatrix} ax+cy & bx+dy \\ au+cv & bu+dv \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

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

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

Если же $ad - bc = 0$, то система будет разрешима лишь когда $b = d = 0$, т.е. если исходная матрица имеет вид

$$\begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix}.$$

Но в таком случае

$$\begin{pmatrix} x & y \\ u & v \end{pmatrix} \begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix} = \begin{pmatrix} * & 0 \\ * & 0 \end{pmatrix},$$

т.е. единичная матрица в произведении получиться не может.

Решая вторую систему, получим

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

Таким образом, обратная матрица равна

$$\begin{pmatrix} x & y \\ u & v \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

что и требовалось доказать. □

D. Транспонирование. Матрицей, *транспонированной* по отношению к матрице $A = (a_j^k)_n^m \in \mathbb{K}^{m \times n}$, называется матрица $A^T \in \mathbb{K}^{n \times m}$, элементы которой суть

$$[A^T]_k^j = [A]_j^k.$$

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

Например,

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

4.8. Теорема. *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:

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

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

$$(A^{-1})^T A^T = (AA^{-1})^T = \mathbb{1}^T = \mathbb{1},$$

что и требовалось. □

Использование операции транспонирования позволяет экономить место при записи столбцов, например, вместо

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

можно написать $(1, 2, 3, 4)^T$.

E. Симметричные и кососимметричные матрицы. Матрица A называется *симметричной*, если $A = A^T$, и *кососимметричной* (антисимметричной), если $A = -A^T$. 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 $A\mathbb{K}^{n\times n}$.

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

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

4.9. Теорема. *Любую квадратную матрицу A можно единственным образом представить 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:

$$A_1^T = \left(\frac{1}{2}(A + A^T)\right)^T = \frac{1}{2}\big(A^T + (A^T)^T\big) = \frac{1}{2}\big(A^T + A\big) = A_1,$$
$$A_2^T = \left(\frac{1}{2}(A - A^T)\right)^T = \frac{1}{2}\big(A^T - (A^T)^T\big) = \frac{1}{2}\big(A^T - A\big) = -A_2;$$

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 $A = A_1 + A_2 = B_1 + B_2$, где A_1, B_1 симметричны, а A_2, B_2 кососимметричны. Тогда

$$\underbrace{A_1 - B_1}_{\text{симм.}} = \underbrace{B_2 - A_2}_{\text{кососимм.}}.$$

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 $A_1 - B_1 = B_2 - A_2 = O$, т.е. $A_1 = B_1$ и $A_2 = B_2$. $\square$

3. Упражнения

ГЛАВА 5

Арифметическое пространство $\mathbb{K}^n$, I Системы линейных уравнений

Арифметическое n-мерное пространство $\mathbb{K}^n$ является самой простой моделью векторного пространства, на примере которой могут быть проиллюстрированы все основные понятия и факты линейной алгебры. В лекциях 5 и 7 излагается теория пространства $\mathbb{K}^n$, с которой тесно связана теория определителей, изложенная в лекции 11.

1. Пространство $\mathbb{K}^n$ и его подпространства

Напомним (см. лекцию 4), что арифметическим пространством $\mathbb{K}^n$ называется множество всех вектор-столбцов высоты n (т.е. матриц, состоящих из единственного n-элементного столбца):

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

На множестве $\mathbb{K}^n$ определены линейные операции: сложение столбцов и умножение столбцов на числа из поля $\mathbb{K}$, обладающие свойствами, перечисленными в теореме 4.1 лекции 4 (см. с. 86). Отметим, что операция умножения столбца на столбец не определена, поскольку не выполняется требование согласования размеров (см. определение произведения матриц в п. 2 лекции 4, с. 87).

Вся теория, излагаемая ниже, может быть дословно перенесена на вектор-строки. Для краткости в дальнейшем будем употреблять термин «вектор», подразумевая вектор-столбец; все формулы записываются именно для вектор-столбцов.

А. Линейные комбинации. Линейная зависимость.

5.1. Определение. *Семейством* (*упорядоченным набором*) 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 множества X, который называется при этом k-*м членом семейства*, а всё семейство обозначается при этом $(x_1, x_2, \ldots, x_n)$.

Если $\{j_1, j_2, \ldots, j_s\} \subset \{1, 2, \ldots, n\}$, то отображение

$$f\colon \{j_1, j_2, \ldots, j_s\} \to X,$$

определённое тем же правилом, что f, но лишь на элементах множества $\{j_1, j_2, \ldots, j_s\}$, определяет подсемейство $(x_{j_1}, x_{j_2}, \ldots, x_{j_s})$ семейства $(x_1, x_2, \ldots, x_n)$.

5.2. Определение. Пусть дано семейство векторов $X_1, X_2, \ldots, X_k \in \mathbb{K}^n$ и семейство чисел $\alpha^1, \alpha^2, \ldots, \alpha^k \in \mathbb{K}$. *Линейной комбинацией* векторов $X_1, X_2, \ldots, X_k$ с коэффициентами $\alpha^1, \alpha^2, \ldots, \alpha^k$ называется выражение вида

$$\alpha^1 X_1 + \alpha^2 X_2 + \cdots + \alpha^k X_k \equiv \sum_{s=1}^{k} \alpha^s X_s. \tag{5.1}$$

Для компактной записи линейной комбинации введём следующие матрицы:

$$X = \bigl[X_1, X_2, \ldots, X_k\bigr], \quad \mathfrak{A} = \bigl(\alpha^1, \alpha^2, \ldots, \alpha^k\bigr)^T.$$

Тогда, очевидно, линейная комбинация (5.1) может быть записана как произведение матриц $\boldsymbol{X}$ и $\mathfrak{A}$:

$$\alpha^1 \boldsymbol{x}_1 + \alpha^2 \boldsymbol{x}_2 + \cdots + \alpha^k \boldsymbol{x}_k = \boldsymbol{X}\mathfrak{A}. \tag{5.2}$$

Линейная комбинация называется *тривиальной*, если все её коэффициенты равны нулю:

$$\alpha^1 = \alpha^2 = \cdots = \alpha^k = 0.$$

Ясно, что тривиальная линейная комбинация произвольных векторов равна нулевому вектору.

5.3. Определение. *Линейной оболочкой* векторов $X_1, X_2, \ldots, X_k$ называется множество всех линейных комбинаций этих векторов:

$$L(X_1, X_2, \ldots, X_k) \stackrel{\text{def}}{=\!=} \Bigl\{\alpha^1 X_1 + \alpha^2 X_2 + \cdots + \alpha^k X_k \Bigm| \alpha^1, \alpha^2, \ldots, \alpha^k \in \mathbb{K}\Bigr\}.$$

5.4. Определение. Семейство векторов $X_1, \ldots, X_k$ называется *линейно зависимым*, если существует нетривиальная линейная комбинация этих векторов, равная нулевому вектору.

Семейство векторов $X_1, \ldots, X_k$ называется *линейно независимым*, если обращение линейной комбинации этих векторов в нулевой вектор возможно лишь в том случае, если эта линейная комбинация тривиальна.

5.5. Замечание. 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».

Линейно зависимые и линейно независимые векторы обладают следующими свойствами.

5.6. Теорема.

1. *Любое семейство с повторениями*[1] *линейно зависимо.*
2. *Если в семействе векторов $X_1, \ldots, X_k$ имеется нулевой вектор O, то это семейство линейно зависимо.*
3. *Семейство векторов $X_1, \ldots, X_k$ 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 $X_1, \ldots, X_k, X_{k+1}, \ldots, X_r$ имеется линейно зависимое подсемейство $X_1, \ldots, X_k$, то и всё семейство также линейно зависимо.*
5. *Любая часть линейно независимого семейства векторов является линейно независимым подсемейством.*
6. *Если семейство векторов $X_1, \ldots, X_k$ линейно независимо, а семейство векторов $X_1, \ldots, X_k, X$ линейно зависимо, то X является линейной комбинацией векторов $X_1, \ldots, X_k$.*
7. *Если семейство векторов $X_1, \ldots, X_k$ линейно независимо, а вектор X нельзя через них выразить, то семейство векторов $X_1, \ldots, X_k, X$ линейно независимо.*

Доказательство. 1. Пусть, например, в семействе векторов $X_1, \ldots, X_k$ первый и второй векторы совпадают, $X_1 = X_2$. Тогда нетривиальная линейная комбинация

$$1 \cdot X_1 + (-1) \cdot X_2 + 0 \cdot X_3 + \cdots + 0 \cdot X_k$$

равна нулевому вектору.

2. Пусть в семействе векторов $X_1, \ldots, X_k$ один вектор нулевой, например, $X_k = O$. Нетривиальная линейная комбинация

$$0 \cdot X_1 + 0 \cdot X_2 + \cdots + 1 \cdot X_k$$

равна, очевидно, нулевому вектору.

3. Пусть семейство векторов $X_1, \ldots, X_k$ линейно зависимо; тогда существует их нетривиальная линейная комбинация, равная нулевому вектору:

$$\alpha^1 X_1 + \alpha^2 X_2 + \cdots + \alpha^k X_k = O.$$

Среди коэффициентов этой линейной комбинации имеются ненулевые; для определённости будем считать, что $\alpha^k \neq 0$; тогда

$$X_k = -\frac{\alpha^1}{\alpha^k} X_1 - \frac{\alpha^2}{\alpha^k} X_2 - \cdots - \frac{\alpha^{k-1}}{\alpha^k} X_{k-1},$$

что и требовалось.

[1] Это означает, что среди членов семейства имеются одинаковые.

Обратно, если

$$X_k = \beta^1 X_1 + \beta^2 X_2 + \dots + \beta^{k-1} X_{k-1},$$

то

$$(-1)X_k + \beta^1 X_1 + \beta^2 X_2 + \dots + \beta^{k-1} X_{k-1} = \mathbf{0},$$

это нетривиальная линейная комбинация, равная нулевому вектору, т.е. семейство $X_1, \dots, X_k$ линейно зависимо.

4. Если подсемейство $X_1, \dots, X_k$ линейно зависимо, то существует линейная комбинация

$$\alpha^1 X_1 + \dots + \alpha^k X_k = O,$$

в которой имеется хотя бы один ненулевой коэффициент. Если теперь к этой линейной комбинации добавить тривиальную линейную комбинацию векторов $X_{k+1}, \dots, X_r$, то получится нетривиальная линейная комбинация

$$\overbrace{\underbrace{\alpha^1 X_1 + \alpha^2 X_2 + \dots + \alpha^k X_k}_{\text{нетрив. л.к.}} + \underbrace{0 \cdot X_{k+1} + \dots 0 \cdot X_r}_{\text{трив. л.к.}}}^{\text{нетрив. л.к.}} = O,$$

что и требовалось.

Утверждение 5 непосредственно следует из предыдущего (рассуждение от противного).

6. Поскольку семейство векторов $X_1, \dots, X_k, X$ линейно зависимо, выполняется соотношение

$$\alpha^1 X_1 + \dots + \alpha^k X_k + \alpha X = O. \tag{5.3}$$

Докажем, что $\alpha \neq 0$. Действительно, в противном случае получили бы

$$\alpha^1 X_1 + \dots + \alpha^k X_k = O$$

и, следовательно, $\alpha^1 = \dots = \alpha^k = 0$ в силу линейной независимости семейства векторов $X_1, \dots, X_k$. Значит, равенство (5.3) выполняется лишь при нулевых значениях всех коэффициентов, что противоречит линейной зависимости семейства векторов $X_1, \dots, X_k, X$. Итак, в (5.3) имеем $\alpha \neq 0$, откуда

$$X = -\frac{\alpha^1}{\alpha} X_1 - \dots - \frac{\alpha^k}{\alpha} X_k.$$

Утверждение 7 непосредственно следует из предыдущего (рассуждение от противного). □

B. Подпространства в $\mathbb{K}^n$. Введём теперь фундаментальное понятие подпространства.

5.7. Определение. Множество $\mathcal{P}$ векторов пространства $\mathbb{K}^n$ называется *подпространством*, если для любых векторов $X, Y \in \mathcal{P}$ их произвольная линейная комбинация $\alpha X + \beta Y$ также принадлежит этому множеству:

$$\forall X, Y \in \mathcal{P}\ \forall \alpha, \beta \in \mathbb{K}^n \quad \alpha X + \beta Y \in \mathcal{P}.$$

Проверьте самостоятельно, что множество векторов с *нулевой* первой компонентой является подпространством, а множество векторов с *положительной* первой компонентой — не является.

Очевидно, само пространство $\mathbb{K}^n$ и множество $\{O\}$, состоящее из одного нулевого вектора, являются подпространствами, которые называются *тривиальными*.

Важный пример подпространства пространства $\mathbb{K}^n$ представляет линейная оболочка векторов.

5.8. Теорема. *Линейная оболочка $L(X_1, X_2, \dots, X_k)$ произвольных векторов $X_1, X_2, \dots, X_k$ пространства $\mathbb{K}^n$ является подпространством в $\mathbb{K}^n$.*

Доказательство. Тот факт, что векторы X и Y лежат в указанной линейной оболочке, означает, что

$$X = \sum_{i=1}^{k} \alpha^i X_i, \quad Y = \sum_{i=1}^{k} \beta^i X_i,$$

где α_i, β_i — некоторые числа. Рассмотрим произвольную линейную комбинацию векторов X и Y и убедимся, что она также лежит в линейной оболочке $L(X_1, X_2, \dots, X_k)$:

$$\alpha X + \beta Y = \alpha \left(\sum_{i=1}^{k} \alpha^i X_i \right) + \beta Y = \left(\sum_{i=1}^{k} \beta^i X_i \right) = \sum_{i=1}^{k} (\alpha\alpha^i + \beta\beta^i) X_i,$$

что и требовалось. □

C. Базис подпространства.

5.9. Определение. Пусть $\mathcal{P}$ — ненулевое подпространство в $\mathbb{K}^n$. Линейно независимое семейство векторов $X_1, \dots, X_r$ называется *базисом* подпространства $\mathcal{P}$, если любой вектор $X \in \mathcal{P}$ может быть представлен в виде линейной комбинации векторов $X_1, \dots, X_r$:

$$\forall X \in \mathcal{P} \ \exists \alpha^1, \dots \alpha^r : \quad X = \alpha^1 X_i + \dots + \alpha^r X_r.$$

Числа $\alpha^1, \dots \alpha^r$ называются координатами вектора X относительно базиса $X_1, \dots, X_r$.

5.10. Замечание. Очевидно, если векторы $X_1, \dots, X_r$ образуют базис в подпространстве $\mathcal{P}$, то линейная оболочка этих векторов совпадает с $\mathcal{P}$:

$$\mathcal{P} = L(X_1, \dots, X_r)$$

(если это не ясно, прочитайте ещё раз определения подпространства, линейной оболочки и базиса).

5.11. Теорема. *Разложение вектора по базису единственно, т.е. набор координат векторов в данном базисе определен однозначно.*

Доказательство. Предположим, что в базисе $X_1, X_2, \ldots, X_r$ вектор X имеет два различных набора координат:

$$X = \alpha^1 X_1 + \alpha^2 X_2 + \cdots + \alpha^r X_r = \beta^1 X_1 + \beta^2 X_2 + \cdots + \beta^r \boldsymbol{a}_r.$$

Вычитая второе разложение из первого, получим

$$O = (\alpha^1 - \beta^1) X_1 + (\alpha^2 - \beta^2) X_1 + \cdots + (\alpha^r - \beta^r) X_r.$$

Так как векторы базиса линейно независимы, то это равенство возможно лишь при нулевых коэффициентах, т.е.

$$\alpha^1 = \beta^1, \quad \alpha^2 = \beta^2, \quad \alpha^r = \beta^r.$$

Теорема доказана. □

Возникают следующие естественные вопросы:

(1) любое ли подпространство в $\mathbb{K}^n$ обладает базисом?
(2) если да, то единствен ли базис в данном подпространстве?
(3) если базис не единствен, то одинаково ли количество векторов, составляющих различные базисы подпространства?

Для того чтобы ответить на эти вопросы, нам сначала придётся разработать теорию систем линейных уравнений, чем мы и займёмся в следующем разделе.

2. Системы линейных уравнений

А. Основные определения. Рассмотрим систему, состоящую из m линейных уравнений с n неизвестными:

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

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

$$A = \begin{pmatrix} a_1^1 & a_2^1 & \ldots & a_n^1 \\ a_1^2 & a_2^2 & \ldots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^m & a_2^m & \ldots & a_n^m \end{pmatrix} = \Big[A_1, A_2, \ldots, A_n\Big], \quad B = \begin{pmatrix} b^1 \\ b^2 \\ \vdots \\ b^m \end{pmatrix},$$

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

тогда систему можно записать в матричном виде

$$AX = B,$$

а также в виде
$$A_1x^1 + A_2x^2 + \cdots + A_nx^n = B.$$

Набор чисел $x^1, x^2, \ldots, x^n$ (столбец $X = (x^1, x^2, \ldots, x^n)^T$) 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.

B. Однородные системы. Система линейных уравнений называется *однородной*, если правые части всех уравнений равны нулю:
$$\begin{cases} a_1^1x^1 + a_2^1x^2 + \cdots + a_n^1x^n = 0, \\ a_1^2x^1 + a_2^2x^2 + \cdots + a_n^1x^n = 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_1^mx^1 + a_2^mx^2 + \cdots + a_n^mx^n = 0; \end{cases}$$
в матричной записи
$$AX = O \quad \text{или} \quad A_1x^1 + A_2x^2 + \cdots + A_nx^n = O.$$

Однородная система всегда совместна: числа $x^1 = x^2 = \cdots = x^n = 0$ (или, что то же, нулевой столбец) образуют её решение, называемое *тривиальным*. Однако однородная система может иметь и нетривиальные (т.е. ненулевые) решения.

5.12. Теорема. *Однородная система имеет нетривиальное решение тогда и только тогда, когда столбцы её основной матрицы линейно зависимы.*

Доказательство. 1. Пусть однородная система имеет нетривиальное решение $x^1, \ldots, x^n$; это означает, что
$$A_1x^1 + A_2x^2 + \cdots + A_nx^n = O,$$
где хотя бы один из коэффициентов x^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.
$$A_1x^1 + A_2x^2 + \cdots + A_nx^n = O,$$
где хотя бы один из коэффициентов x^k ненулевой. Этот набор коэффициентов и образует нетривиальное решение однородной системы. □

5.13. Теорема. *Если X_1, X_2 — два решения однородной системы $AX = 0$, то любая их линейная комбинация также является решением этой системы. Иными словами, множество всех решений однородной системы образует подпространство в пространстве столбцов $\mathbb{K}^n$.*

Доказательство. Столбцы X_1, X_2 являются решениями системы, т.е.

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

Пусть c^1, c^2 — произвольные числа. Рассмотрим линейную комбинацию столбцов X_1, X_2 с коэффициентами c^1, c^2: $X = c^1X_1 + c^2X_2$. Имеем:

$$AX = A(c^1X_1 + c^2X_2) = c^1 \underbrace{AX_1}_{=O} + c^2 \underbrace{AX_2}_{=O} = O,$$

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

5.14. Замечание. По существу, имеется всего два способа задания подпространств в $\mathbb{K}^n$: как множества решений однородной системы (теорема 5.13) и как линейной оболочки заданного набора векторов (теорема 5.8).

5.15. Определение. *Фундаментальное семейство решений* (ФСР) однородной системы — это такое семейство (т.е. упорядоченный набор, см. определение 5.1, с. 97) линейно независимых решений $X_1, X_2, \ldots, X_s$, что любое решение данной системы можно представить в виде их линейной комбинации:

$$X = c^1X_1 + c^2X_2 + \cdots + c^sX_s,$$

где $c^1, c^2, \ldots, c^s$ — произвольные числа. Приведённое выражение, позволяющее с помощью надлежащего выбора чисел $c^1, c^2, \ldots, c^s$ получить любое решение однородной системы, называется *общим решением* данной системы.

Согласно определению 5.9 (см. с. 101) фундаментальное семейство решений однородной системы линейных уравнений является базисом в подпространстве пространства $\mathbb{K}^n$, состоящем из всех решений этой системы.

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

$$\Phi = \Big[X_1,\ X_2,\ \ldots,\ X_s\Big].$$

Общее решение системы выражается через фундаментальную матрицу по формуле

$$X = \Phi \begin{pmatrix} c^1 \\ c^2 \\ \vdots \\ c^s \end{pmatrix}.$$

C. Неоднородные системы. Система $AX = B$ называется *неоднородной*, если $B \neq O$. Часто неоднородную систему $AX = B$ рассматривают вместе с однородной системой $AX = O$.

5.16. Теорема. *Если X_1, X_2 — два решения неоднородной системы $AX = B$, то их разность $X_1 - X_2$ является решением 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. □

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*.

D. Системы упрощённого вида. Неизвестная x^k называется *базисной*, если она входит только в одно уравнение системы.

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] Таким образом, полагаем, что в системе упрощённого вида количество базисных неизвестных равно количеству уравнений в системе.

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

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):

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

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

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

[1]Например, в системе

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

базисными неизвестными являются x^1, x^2 (они входят только в первое уравнение) и x^3 (она входит только во второе), однако в этом случае лишь одну из неизвестных x^1 или x^2 будем относить к числу базисных.

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

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

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

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

Столбцы, фигурирующие в последнем выражении, образуют фундаментальное семейство решений данной однородной системы, которая называется *нормальным ФСР (НФСР), соответствующим базисным неизвестным* $x^1, \ldots, x^r$ *и свободным неизвестным* $x^{r+1}, \ldots, x^n$: каждый столбец, входящий в НФСР, получается, если значение одной из свободных неизвестных взять равным 1, а остальные положить равными нулю (соответствующие элементы столбцов выделены). Нормальная фундаментальная матрица рассматриваемой системы составлена из столбцов

НФСР:

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

выделенные элементы этой матрицы образуют единичную матрицу порядка n. Общее решение системы выражается через нормальную фундаментальную матрицу формулой

$$X = \Phi \begin{pmatrix} c^1 \\ \vdots \\ c^{n-r} \end{pmatrix}.$$

Ситуация в случае неоднородной системы аналогична: переписав упрощённую неоднородную систему

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

в виде

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

обнаруживаем, что её общее решение можно записать следующим образом:

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

(выделены свободные неизвестные). Фигурирующее здесь частное решение неоднородной системы $(b^1, \dots, b^r, 0, \dots, 0)^T$ называется *базисным решением, отвечающим базисным неизвестным* $x^1, \dots, x^r$ *и свободным неизвестным* $x^{r+1}, \dots, x^n$.

Пример 5.1. Рассмотрим однородную систему упрощённого вида

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

Базисными неизвестными являются x^1, x^2, x^4 (они выделены), свободными — x^3, x^5. Запишем систему в виде

$$\begin{cases} x^1 = -3x^3 - x^5, \\ x^2 = -4x^3 - 2x^5, \\ x^4 = -x^5. \end{cases}$$

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 данной однородной системы уравнений состоит из столбцов

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

Здесь выделены ячейки таблицы, в которых находятся значения свободных неизвестных; матрица, составленная из выделенных элементов, является единичной матрицей. Нормальная фундаментальная матрица данной системы

$$\Phi = \begin{pmatrix} -3 & -1 \\ -4 & -2 \\ \boxed{1} & \boxed{0} \\ 0 & -1 \\ \boxed{0} & \boxed{1} \end{pmatrix}.$$

Обратите внимание, что выделенные элементы нормальной фундаментальной матрицы образуют единичную матрицу порядка 2. Общее решение системы представляется через нормальную фундаментальную матрицу по формуле

$$X = \begin{pmatrix} -3 & -1 \\ -4 & -2 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c^1 \\ c^2 \end{pmatrix}.$$

Пример 5.2. Теперь рассмотрим неоднородную систему упрощенного вида:

$$\begin{cases} \boxed{x^1} & & + & 3x^3 & & + & x^5 & = & 1, \\ & \boxed{x^2} & + & 4x^3 & & + & 2x^5 & = & 2, \\ & & & & \boxed{x^4} & + & 2x^5 & = & 3; \end{cases}$$

базисные неизвестные выделены. Переписав систему в виде

$$\begin{cases} x^1 = 1 - 3x^3 - x^5, \\ x^2 = 2 - 4x^3 - 2x^5, \\ x^4 = 3 - x^5, \end{cases}$$

получаем её общее решение:

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

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

3. Алгоритм Гаусса—Жордана

А. Элементарные преобразования. Итак, если система линейных уравнений 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.

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.*

Иногда к элементарным преобразованиям причисляют

(4) удаление из системы уравнения вида

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

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—Жордана.

B. Матрицы упрощённого вида. Говорят, что матрица имеет *упрощённый вид*, если она является расширенной матрицей СЛУ 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:

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

ведущие столбцы и ведущие элементы выделены.

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) все остальные столбцы являются линейными комбинациями предшествующих базисных столбцов.

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

Шаг № k.

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

(1) Среди строк матрицы с номерами $k, \ldots, m$ выбираем 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) Переставляем ведущую строку на k-е место.
(4) Вычитаем 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.

Пример 5.3. Приведём к упрощённому виду матрицу

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

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:

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

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). Выполняем элементарное преобразование: к четвёртой строке добавляем ведущую (первую), умноженную на (-3):

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

Первый столбец матрицы теперь выглядит как первый столбец 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:

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

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). Выполняем следующие элементарные преобразования:

(1) к первой строке прибавляем ведущую (вторую) строку, умноженную на $(-1/2)$;
(2) к третьей строке прибавляем ведущую (вторую) строку, умноженную на 2;
(3) к четвёртой строке прибавляем ведущую (вторую) строку, умноженную на $1/2$.

Результат выполнения:

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

Второй столбец матрицы теперь выглядит как второй столбец 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:

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

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

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

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

Результат выполнения:

$$\begin{pmatrix} 1 & 0 & 2 & \frac{1}{2} & 2 & 0 \\ 0 & 1 & -3 & -1 & -4 & 1 \\ 0 & 0 & 0 & \boxed{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \Longrightarrow \begin{pmatrix} 1 & 0 & 2 & 0 & 1 & -1 \\ 0 & 1 & -3 & 0 & -2 & 3 \\ 0 & 0 & 0 & \boxed{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Четвёртый столбец матрицы теперь выглядит как третий столбец 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:

$$\begin{pmatrix} \boxed{1} & 0 & 2 & 0 & 1 & -1 \\ 0 & \boxed{1} & -3 & 0 & -2 & 3 \\ 0 & 0 & 0 & \boxed{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

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:

$$A_3 = 2A_1 - 3A_2, \quad A_5 = A_1 - 2A_2 + 2A_4, \quad A_6 = -A_1 + 3A_2 + 2A_4.$$

Коэффициенты этих линейных комбинаций суть не что иное, как элементы соответствующих столбцов преобразованной матрицы.

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

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

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

$$\begin{pmatrix} -2 \\ 1 \\ -1 \\ 0 \end{pmatrix} = -\begin{pmatrix} 0 \\ 2 \\ 0 \\ 3 \end{pmatrix} + 3\begin{pmatrix} -2 \\ 1 \\ -1 \\ 1 \end{pmatrix} + 2\begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix}.$$

Разумеется, этот факт не случаен; он будет объяснён в п. В (см. теорему 7.2(2), с. 147).

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.

5.17. Теорема (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. □

D. «Правило прямоугольников». В рассмотренном выше примере вычисления с дробями доставляли определённые неудобства, которых можно избежать, если применить следующую модификацию алгоритма Гаусса[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:

(1) первую строку матрицы умножим на ведущий элемент c;

(2) 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 (разделив её на r), получим

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

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

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

(1) третью строку матрицы умножим на ведущий элемент c;

(2) вторую (ведущую) строку умножим на z;

в результате получаем матрицу

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

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.

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 двух других элементов матрицы, r и a, расположенных в двух других противоположных углах того же прямоугольника. Для остальных элементов ситуация аналогична:

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

$$x \to cx - za: \begin{pmatrix} p & q & r & s \\ a & b & \boldsymbol{c} & d \\ x & y & z & u \end{pmatrix}, \qquad u \to cu - zd: \begin{pmatrix} p & q & r & s \\ a & b & \boldsymbol{c} & d \\ x & y & z & u \end{pmatrix}.$$

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

(1) в матрице выбирается ведущий элемент a_t^s;
(2) ведущая строка (строка A^s) переписывается без изменений;
(3) все элементы ведущего столбца (столбца A_t), кроме ведущего элемента, заменяются нулями;
(4) каждый элемент a_j^i матрицы, не принадлежащий ведущей строке или ведущему столбцу, заменяется на число $a_j^i a_t^s - a_t^i a_j^s$.

Схематическая иллюстрация:

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

Пример 5.4. Приведём к упрощённому виду матрицу из примера 5.3,

$$\begin{pmatrix} 0 & -2 & 6 & 2 & 8 & -2 \\ 2 & 1 & 1 & 0 & 0 & 1 \\ 0 & -1 & 3 & 1 & 4 & -1 \\ 3 & 1 & 3 & 0 & 1 & 0 \end{pmatrix},$$

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:

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

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:

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

Шаг 3. В качестве ведущей строки выберем третью; после преобразований получаем

$$\begin{pmatrix} 2 & 0 & 4 & 1 & 4 & 0 \\ 0 & 1 & -3 & -1 & -4 & 1 \\ 0 & 0 & 0 & \boxed{-1} & -2 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 2 & 0 & 4 & 1 & 4 & 0 \\ 0 & 1 & -3 & -1 & -4 & 1 \\ 0 & 0 & 0 & \boxed{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow$$

$$\Rightarrow \begin{pmatrix} 2 & 0 & 4 & 0 & 2 & -2 \\ 0 & 1 & -3 & 0 & -2 & 3 \\ 0 & 0 & 0 & \boxed{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} \boxed{1} & 0 & 2 & 0 & 1 & -1 \\ 0 & \boxed{1} & -3 & 0 & -2 & 3 \\ 0 & 0 & 0 & \boxed{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

в последней матрице выделены базисные столбцы.

Е. Алгоритм Гаусса. 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]:

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

Матрица ступенчатого вида имеет следующую структуру:

(1) все нулевые строки располагаются ниже всех ненулевых;
(2) *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. 111) 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.

4. Упражнения

[1] В англоязычной литературе он называется «row echelon form».

ГЛАВА 6

Определители

ОПРЕДЕЛИТЕЛЬ, или детерминант, — одно из важнейших понятий алгебры, имеющее разнообразные применения во всех разделах математики. Впервые идеи, близкие к понятию определителя, возникли в конце XVII в. в работах Г. Ф. Лейбница, связанных с исследованием систем линейных уравнений, однако само понятие определителя, его свойства и методы вычисления появились лишь в середине XVII в. в работах швейцарского математика Г. Крамера. Сам термин «определитель» («детерминант») введён К. Ф. Гауссом. Методы Крамера сразу же получили широкое распространение и быстро нашли многочисленные приложения как в самой математике, так и в родственных науках — астрономии и механике.

1. Мотивировка определения. Определитель второго порядка

А. Система двух уравнений с двумя неизвестными. Формулы Крамера. Рассмотрим систему двух линейных уравнений с двумя неизвестными:

$$\begin{cases} ax + by = p, \\ cx + dy = q, \end{cases} \tag{6.1}$$

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, найдем

$$+\begin{cases} ax + by = p \\ cx + dy = q \end{cases} \left| \begin{array}{l} \cdot\, d \\ \cdot\, (-b) \end{array} \right. \quad \Longrightarrow \quad (ad - bc)x = pd - qb.$$

Аналогично, умножая первое уравнение на $-c$, второе на a и складывая полученные уравнения, найдем

$$+\begin{cases} ax + by = p \\ cx + dy = q \end{cases} \left| \begin{array}{l} \cdot\, (-c) \\ \cdot\, a \end{array} \right. \quad \Longrightarrow \quad (ad - bc)y = qa - pc.$$

Если $ad - bc \neq 0$, то система имеет единственное решение

$$x = \frac{pd - qb}{ad - bc}, \quad y = \frac{qa - pc}{ad - bc}.$$

Если $ad - bc = 0$, то возможны следующие ситуации:

(i) если хотя бы одно из чисел $pd - qb$ и $qa - pc$ отлично от нуля, то система решений не имеет;

(ii) оба числа $pd-qb$ и $qa-pc$ равны нулю; тогда уравнения системы пропорциональны, и система может иметь бесконечно много решений (каждое из которых получается, если одну из неизвестных взять произвольно, а вторую вычислить с помощью любого из уравнений системы), а может и не иметь решений.[1]

B. Определитель второго порядка. Запишем коэффициенты системы в виде (2×2)-матрицы

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

которая называется основной матрицей системы. Поставим в соответствие этой матрице число $ad - bc$; оно называется определителем (детерминантом) матрицы A и обозначается

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \det A = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Этот определитель называют определителем второго порядка (по количеству строк и столбцов матрицы A).

Таким образом, определитель второго порядка можно рассматривать как функцию

$$\det : \mathbb{K}^{2\times 2} \to \mathbb{K},$$

областью определения которой является множество всех (2×2)-матриц, либо как функцию

$$\det : \mathbb{K}^2 \times \mathbb{K}^2 \to \mathbb{K},$$

двух аргументов, каковыми являются двухэлементные столбцы.

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 матрица A_x (соответственно, A_y) получается из матрицы A заменой первого (соответственно, второго) столбца на столбец, состоящий из свободных членов уравнений. Полученные формулы называются *формулами Крамера.*

---

[1]Например, система

$$\begin{cases} 0x + 0y = 1, \\ 0x + 0y = 1; \end{cases}$$

здесь все три определителя равны нулю, но решений, очевидно, нет.

6.1. Теорема. *Определитель второго порядка* $\det A$ *обладает следующими свойствами*:

(1) *линейность по первому столбцу*:

$$\begin{vmatrix} a_1+a_2 & b \\ c_1+c_2 & d \end{vmatrix} = \begin{vmatrix} a_1 & b \\ c_1 & d \end{vmatrix} + \begin{vmatrix} a_2 & b \\ c_2 & d \end{vmatrix}; \qquad \begin{vmatrix} \alpha a & b \\ \alpha c & d \end{vmatrix} = \alpha \begin{vmatrix} a & b \\ c & d \end{vmatrix};$$

кроме того, имеет место аналогичное свойство линейности по второму столбцу;

(2) *кососимметричность*: *при перестановке столбцов определитель меняет знак,*

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = -\begin{vmatrix} b & a \\ d & c \end{vmatrix};$$

(3) *нормировка*: *определитель единичной матрицы равен* 1,

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

Доказательство производится непосредственной проверкой, например,

$$\begin{vmatrix} a_1+a_2 & b \\ c_1+c_2 & d \end{vmatrix} = (a_1+a_2)d - (c_1+c_2)b = \\ = (a_1 d - c_1 b) + (a_2 d - c_2 b) = \begin{vmatrix} a_1 & b \\ c_1 & d \end{vmatrix} + \begin{vmatrix} a_2 & b \\ c_2 & d \end{vmatrix}.$$

Из этих основных свойств определителя можно вывести ряд новых свойств, полезных при вычислениях.

1. Непосредственно проверяется что, *определитель с двумя одинаковыми столбцами равен нулю*:

$$\begin{vmatrix} a & a \\ c & c \end{vmatrix} = ac - ac = 0.$$

Однако этот факт может быть получен как следствие кососимметричности (см. ниже доказательство предложения 6.3.2).

Легко проверить, что *определитель с пропорциональными столбцами равен нулю*: для этого достаточно, пользуясь свойством линейности, вынести за знак определителя коэффициент пропорциональности, после чего получится определитель с одинаковыми столбцами.

2. *Определитель второго порядка не изменится, если к любому из его столбцов прибавить другой столбец, умноженный на произвольное число.*

Действительно,

$$\begin{vmatrix} a+\alpha b & b \\ c+\alpha d & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \alpha \underbrace{\begin{vmatrix} b & b \\ d & d \end{vmatrix}}_{=0} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

3. *Определитель не изменяется при транспонировании матрицы*:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix}.$$

Это означает, что строки и столбцы определителя равноправны: любое утверждение, справедливое для столбцов, будет справедливым и для строк.

Рассмотрим два несложных примера:

$$\begin{vmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{vmatrix} = \cos^2\alpha + \sin^2\alpha = 1.$$

Теперь вычислим определитель

$$\begin{vmatrix} 12345 & 12347 \\ 24691 & 24695 \end{vmatrix}.$$

Вычтем из второй строки удвоенную первую строку:

$$\begin{vmatrix} 12345 & 12347 \\ 24691 - 2\cdot 12345 & 24695 - 2\cdot 12347 \end{vmatrix} =$$

$$= \begin{vmatrix} 12345 & 12347 \\ 1 & 1 \end{vmatrix} = 12345\cdot 1 - 12347\cdot 1 = -2.$$

2. Определитель 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. Рассмотрим квадратную $(n\times n)$-матрицу $A \in \mathbb{K}^{n\times n}$ с элементами из некоторого поля $\mathbb{K}$. 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

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

6.2. Определение. Определитель, или детерминант, матрицы A — это числовая функция столбцов этой матрицы

$$\det : \underbrace{\mathbb{K}^n \times \dots \times \mathbb{K}^n}_{n} \to \mathbb{K},$$

обозначаемая

$$\det A = \bigl|A\bigr| = \det\bigl[A_1, A_2, \dots, A_n\bigr] = \Bigl|A_1, A_2, \dots, A_n\Bigr|$$

и обладающая следующими свойствами (скопированными со свойств определителя второго порядка):

(1) *полилинейность*, т.е. линейность по каждому аргументу; например, свойство линейности по первому аргументу заключается в выполнении соотношения

$$\Bigl|\alpha' A_1' + \alpha'' A_1'',\ A_2,\ \dots,\ A_n\Bigr| = \alpha'\Bigl|A_1',\ A_2,\ \dots,\ A_n\Bigr| + \alpha''\Bigl|A_1'',\ A_2,\ \dots,\ A_n\Bigr|;$$

(2) *кососимметричность*: при перестановке любых двух соседних столбцов определитель меняет знак на противоположный,

$$\Bigl|A_1',\ \dots,\ A_k,\ A_{k+1},\ \dots,\ A_n\Bigr| = -\Bigl|A_1',\ \dots,\ A_{k+1},\ A_k,\ \dots,\ A_n\Bigr|;$$

(3) *нормировка*: определитель единичной матрицы $\mathbb{1} \in \mathbb{K}^{n\times n}$ равен единице,

$$\det \mathbb{1} = 1.$$

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 определяющих свойств, которыми должен обладать определитель, если он существует.

6.3. Предложение.

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:

$$= -\left|A_1,\ \ldots,\ A_k,\ \underbrace{A_{j+1},\ A_{j+2},\ \ldots,\ A_{k-1}}_{m+1\ \text{столбцов}},\ A_j,\ \ldots,\ A_n\right|,$$

что и требовалось.

2. Допустим, что у определителя имеются одинаковые столбцы. Если их переставить местами, то, с одной стороны, знак определителя изменится на противоположный в силу свойства кососимметричности, а с другой стороны — определитель не изменится, поскольку переставленные столбцы одинаковы. Таким образом, если обозначить значение рассматриваемого определителя через A, то $A = -A$, откуда $A = 0$.

Отметим, что равенство $A = 0$ вытекает из равенства $A = -A$ только в случае, если характеристика поля $\mathbb{K}$, которому принадлежат элементы определителя, не равна 2. В случае же $\operatorname{char}\mathbb{K} = 2$ равенство $A = -A$ эквивалентно $2A = 0$ и выполняется при любом A. Свойство кососимметричности определяется в этом случае как равенство нулю определителя с одинаковыми столбцами.

3. Проверим, что определитель не изменится, если к его первому столбцу прибавить линейную комбинацию остальных столбцов:

$$\left|A_1 + \alpha^2 A_2 + \cdots + \alpha^n A_n,\ A_2,\ \ldots,\ A_n\right| =$$
$$= \left|A_1,\ A_2,\ \ldots,\ A_n\right| + \alpha^2 \underbrace{\left|A_2,\ A_2,\ \ldots,\ A_n\right|}_{=0} + \cdots + \alpha^n \underbrace{\left|A_n,\ A_2,\ \ldots,\ A_n\right|}_{=0} =$$
$$= \left|A_1,\ A_2,\ \ldots,\ A_n\right|,$$

что и требовалось. □

В. Формулы Крамера. Мы также можем получить формулы для решения системы n линейных уравнений с n неизвестными (формулы Крамера), ещё не зная способов вычисления определителя!

Рассмотрим систему n линейных уравнений с n неизвестными:

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

Введя в рассмотрение основную и расширенную матрицы этой системы

$$A = \begin{pmatrix} a_1^1 & a_2^1 & \dots & a_n^1 \\ a_1^2 & a_2^2 & \dots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^n & a_2^n & \dots & a_n^n \end{pmatrix}, \quad \tilde{A} = \Big[A,\ B\Big] = \begin{pmatrix} a_1^1 & a_2^1 & \dots & a_n^1 & b^1 \\ a_1^2 & a_2^2 & \dots & a_n^2 & b^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1^n & a_2^n & \dots & a_n^n & b^n \end{pmatrix},$$

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

$$x^1A_1 + x^2A_2 + \dots + x^nA_n = B \tag{6.3}$$

или

$$AX = B$$

(см. лекцию 5, п. 2, с. 102)

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 (6.3) имеем

$$\Delta_1 = \Big|B,\ A_2,\ \dots,\ A_n\Big| =$$

$$= \Big|x^1A_1 + x^2A_2 + \dots + x^nA_n,\ A_2,\ \dots,\ A_n\Big| =$$

$$= x^1 \cdot \underbrace{\begin{vmatrix} A_1, & A_2, & \ldots, & A_n \end{vmatrix}}_{=\Delta} + + x^2 \cdot \underbrace{\begin{vmatrix} A_2, & A_2, & \ldots, & A_n \end{vmatrix}}_{=0} + \cdots + x^n \cdot \underbrace{\begin{vmatrix} A_n, & A_2, & \ldots, & A_n \end{vmatrix}}_{=0}$$

откуда, при условии $\Delta \neq 0$, получаем

$$x^1 = \frac{\Delta_1}{\Delta}.$$

Аналогично получаются формулы для остальных неизвестных:

$$x^k = \frac{\Delta_k}{\Delta}, \quad k = 1, 2, \ldots, n,$$

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.

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.

C. Перестановки. Понятие группы. Пусть N — n-элементное множество; будем считать, что $N = \{1, 2, \ldots, n\}$. *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.

Если переставить местами столбцы этой таблицы, сохраняя все соответствия $i \mapsto \sigma(i)$, то получим другую запись той же перестановки, например

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 2 & 1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 4 & 2 \\ 2 & 4 & 3 & 1 \end{bmatrix}.$$

Для простоты принято записывать перестановки так, чтобы числа в верхнем ряду таблицы находились в «естественном порядке», т.е. $1, 2, \ldots, n$.

Перестановка ε, заданная формулой $\varepsilon(i) = i$ для любого $i \in N$, т.е.

$$\varepsilon = \begin{bmatrix} 1 & 2 & \ldots & n \\ 1 & 2 & \ldots & n \end{bmatrix},$$

называется *тождественной*.

6.4. Предложение. *Количество всех перестановок n-элементного множества N равно*

$$n! = 1 \cdot 2 \cdots (n-1) \cdot n \tag{6.4}$$

(это число называется факториалом натурального числа n).

Доказательство. Подсчитаем количество различных вариантов заполнить вторую строку таблицы значений перестановки. В качестве $\sigma(1)$ может быть выбрано любое из чисел $1, 2, \ldots, n$ (всего n вариантов); в качестве $\sigma(2)$ — любое из оставшихся (всего $n-1$ вариантов) и т. д. Перемножая числа возможных вариантов, приходим к (6.4) $\square$

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:

$$\varepsilon = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix},$$
$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}.$$

Ещё раз подчеркнём что каждый элемент множества S_n является отображением $N \to N$. Введём на S_n бинарную операцию композиции отображений: если $\sigma : N \to N$ и $\tau : N \to N$ — 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

$$\sigma = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}, \quad \tau = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix},$$

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 (см. теорему II.22, с. 330), 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,

$$
\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{bmatrix}, \quad \sigma^{-1} = \begin{bmatrix} 3 & 1 & 4 & 2 \\ 1 & 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{bmatrix}.
$$

Мы подошли к чрезвычайно важному понятию, которое встречается во всех разделах математики.

6.5. Определение. Рассмотрим множество G с заданной на нём бинарной операцией, т.е. отображением $*\colon G \times G \to G$, которое каждой упорядоченной паре (a, b) элементов множества G ставит в соответствие элемент $c \in G$, называемый композицией элементов a и b и обозначаемый $c = a * b$. Множество G называется *группой*, если операция $*$ обладает следующими свойствами:

Г1: операция $*$ ассоциативна, т.е.

$$\forall a, b, c \in G: \; a * (b * c) = (a * b) * c;$$

Г1: существует элемент $e \in G$, обладающий свойством

$$\forall a \in G: \; a * e = e * a = a;$$

этот элемент называется нейтральным элементом группы;

Г1: каждый элемент множества G обратим относительно операции $*$, т.е.

$$\forall a \in G \; \exists b \in G: \; a * b = b * a = e;$$

элемент b называется *обратным* к элементу a и обозначается a^{-1}.

Требования **Г1–Г3** называются аксиомами группы.

6.6. Замечание. В различных конкретных ситуациях операция композиции в группе может называться сложением, умножением или как-то иначе и обозначаться знаками $+$, $\cdot$, $\circ$, $\times$ и т. д. Нейтральный элемент e может называться нулём (особенно если групповая операция называется сложением), единицей или как-то иначе. Обратный элемент также может называться по-другому; так, если групповая операция называется сложением, то обратный элемент принято называть противоположным и обозначать $(-a)$.

6.7. Пример. Очевидными, хорошо знакомыми читателю примерами групп являются:

(a) множество целых чисел с операцией сложения: $(\mathbb{Z}, +)$; нейтральным элементом является число 0, «обратным» для $x \in \mathbb{Z}$ является $-x$ («обратный» элемент принято называть в данной ситуации противоположным);

(b) множество ненулевых рациональных чисел с операцией умножения: $(\mathbb{Q} \setminus \{0\}, \cdot)$; нейтральным элементом является число 1, обратным для $x \in \mathbb{Q} \setminus \{0\}$ является $1/x$; аналогичный пример — $(\mathbb{R}, +)$;

(c) множество геометрических векторов с операцией сложения: $(V, +)$; нейтральным элементом является нулевой вектор $\mathbf{0}$, «обратным» для $\boldsymbol{a} \in V$ является противоположный вектор $-\boldsymbol{a}$;

(d) множество $(m \times n)$-матриц с операцией сложения: $(\mathbb{K}^{m \times n}, +)$.

Отметим, что множество $\mathbb{Q}$ *всех* рациональных чисел не образует группу относительно умножения, поскольку элемент 0 не имеет обратного относительно этой операции. Читателю рекомендуется привести аналогичные примеры групп и не-групп самостоятельно.

Таким образом, множество S_n всех перестановок n-множества, снабжённое операцией композиции (умножения) перестановок, является группой, которая называется *симметрической группой.*

Введём важное понятие знака перестановки.

6.8. Определение. Пусть i, j — два различных элемента множества $N = \{1, 2, \ldots, n\}$, $i \neq j$. Будем говорить, что пары $(i, \sigma(i))$ и $(j, \sigma(j))$ *образуют инверсию* в перестановке $\sigma \in S_n$, если разности $i-j$ и $\sigma(i)-\sigma(j)$ имеют разные знаки.

Ясно, что если перестановка задана как двухрядная таблица, верхний ряд которой записан в естественном 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:

$$(3, 2),\ (3, 1),\ (5, 2),\ (5, 1),\ (5, 4),\ (2, 1).$$

6.9. 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 $-$.

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:

$$\operatorname{sgn}(\sigma)=\prod_{\substack{\{i,j\}\subset N\\ i\neq j}}\operatorname{sign}\left(\frac{i-j}{\sigma(i)-\sigma(j)}\right),$$

где символ sign означает знак вещественного числа:

$$\operatorname{sign}x=\begin{cases}-1, & \text{если } x<0,\\ \ \ 0, & \text{если } x=0,\\ \ \ 1, & \text{если } x>0.\end{cases} \tag{6.5}$$

Очевидно, что для любых $x,y\in\mathbb{R}$ имеем

$$\operatorname{sign}(xy)=\operatorname{sign}x\cdot\operatorname{sign}y.$$

6.10. Теорема. *Знак произведения двух перестановок равен произведению знаков этих перестановок:*

$$\operatorname{sgn}(\sigma\tau)=\operatorname{sgn}\sigma\cdot\operatorname{sgn}\tau. \tag{6.6}$$

Доказательство. Перестановку σ можно представить в виде

$$\sigma=\begin{bmatrix}1 & 2 & \dots & n\\ \sigma(1) & \sigma(2) & \dots & \sigma(n)\end{bmatrix}=\begin{bmatrix}\tau(1) & \tau(2) & \dots & \tau(n)\\ \sigma(\tau(1)) & \sigma(\tau(2)) & \dots & \sigma(\tau(n))\end{bmatrix},$$

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

$$\begin{aligned}\operatorname{sgn}\sigma\cdot\operatorname{sgn}\tau&=\prod_{\substack{\{i,j\}\subset N\\ i\neq j}}\operatorname{sign}\frac{\tau(i)-\tau(j)}{\sigma(\tau(i))-\sigma(\tau(j))}\cdot\prod_{\substack{\{i,j\}\subset N\\ i\neq j}}\operatorname{sign}\frac{i-j}{\tau(i)-\tau(j)}=\\ &=\prod_{\substack{\{i,j\}\subset N\\ i\neq j}}\operatorname{sign}\frac{\tau(i)-\tau(j)}{\sigma(\tau(i))-\sigma(\tau(j))}\cdot\operatorname{sign}\frac{i-j}{\tau(i)-\tau(j)}=\\ &=\prod_{\substack{\{i,j\}\subset N\\ i\neq j}}\operatorname{sign}\left(\frac{\tau(i)-\tau(j)}{\sigma(\tau(i))-\sigma(\tau(j))}\cdot\frac{i-j}{\tau(i)-\tau(j)}\right)=\end{aligned}$$

$$= \prod_{\substack{\{i,j\}\subset N\\ i\neq j}} \operatorname{sign}\left(\frac{i-j}{\sigma(\tau(i))-\sigma(\tau(j))}\right) = \operatorname{sgn}(\sigma\tau),$$

что и требовалось доказать. $\square$

6.11. Следствие. *Взаимно обратные перестановки имеют один и тот же знак.*

Доказательство. Поскольку $\sigma^{-1}\sigma = \varepsilon$ и $\operatorname{sgn}\varepsilon = 1$, имеем

$$1 = \operatorname{sgn}\varepsilon = \operatorname{sgn}(\sigma^{-1}\sigma) = \operatorname{sgn}\sigma^{-1}\cdot\operatorname{sgn}\sigma$$

так что $\operatorname{sgn}\sigma^{-1} = \operatorname{sgn}\sigma$. $\square$

D. Существование и единственность определителя. Формула полного развёртывания определителя. Рассмотрим столбцы единичной матрицы

$$\boldsymbol{e}_1 = \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \quad \boldsymbol{e}_2 = \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}, \quad \ldots, \quad \boldsymbol{e}_n = \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}.$$

Очевидно, любой n-элементный столбец можно представить в виде линейной комбинации этих столбцов, например,

$$X = \begin{pmatrix}x^1\\x^2\\\vdots\\x_n\end{pmatrix} = x^1\boldsymbol{e}_1 + x^2\boldsymbol{e}_2 + \cdots + x^n\boldsymbol{e}_n = \sum_{k=1}^{n} x^k\boldsymbol{e}_k. \qquad (6.7)$$

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.

Найдём сначала общее выражение полилинейной кососимметрической функции $F(A_1, A_2, \ldots, A_n)$ столбцов квадратной матрицы A через элементы этих столбцов.

Разложим каждый столбец матрицы A в линейную комбинацию вида (6.7) 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

$$F\Big(A_1,\ A_2,\ \ldots,\ A_n\Big) = c \cdot \sum_{\sigma \in S_n} \operatorname{sgn}\sigma \cdot a_1^{\sigma_1} \cdot a_2^{\sigma_2} \cdots a_n^{\sigma_n}, \tag{6.8}$$

где суммирование производится по всевозможным перестановкам $\sigma \in S_n$, т.е. сумма состоит из $n!$ слагаемых. Таким образом, если полилинейная кососимметрическая функция столбцов квадратной матрицы существует, то она *единственна* и задаётся формулой (6.8).

Для доказательства *существования* достаточно убедиться, что функция $F(A_1, A_2, \ldots, A_n)$, заданная выражением (6.8), является полилинейной и кососимметрической.

Проверим, например, линейность по первому аргументу:

$$\begin{aligned}
&F\Big(\alpha' A_1' + \alpha'' A_1'',\ A_2,\ \ldots,\ A_n\Big) = \\
&\qquad = c \cdot \sum_{\sigma \in S_n} \operatorname{sgn}\sigma \cdot \Big(\alpha' a_1'^{\sigma_1} + \alpha'' a_1''^{\sigma_1}\Big) \cdot a_2^{\sigma_2} \cdots a_n^{\sigma_n} = \\
&= c\alpha' \sum_{\sigma \in S_n} \operatorname{sgn}\sigma \cdot a_1'^{\sigma_1} \cdot a_2^{\sigma_2} \cdots a_n^{\sigma_n} + c\alpha'' \sum_{\sigma \in S_n} \operatorname{sgn}\sigma \cdot a_1''^{\sigma_1} \cdot a_2^{\sigma_2} \cdots a_n^{\sigma_n} = \\
&\qquad\qquad = \alpha' \cdot F\Big(A_1',\ A_2,\ \ldots,\ A_n\Big) + \alpha'' \cdot F\Big(A_1'',\ A_2,\ \ldots,\ A_n\Big).
\end{aligned}$$

Для любого другого аргумента проверка совершенно аналогична.

Чтобы проверить кососимметричность, переставим в матрице A два соседних столбца A_k и A_{k+1} и запишем $F(A_1, \ldots, A_{k+1}, A_k, \ldots, A_n)$:

$$\begin{aligned}
&F\Big(A_1,\ \ldots,\ A_{k-1},\ A_{k+1},\ A_k,\ A_{k+2},\ \ldots,\ A_n\Big) = \\
&\qquad = c \operatorname{sgn}(\sigma_1, \ldots, \sigma_k, \sigma_{k+1}, \ldots, \sigma_n) \cdot a_1^{\sigma_1} \cdots a_{k+1}^{\sigma_k} \cdot a_k^{\sigma_{k+1}} \cdots a_n^{\sigma_n}.
\end{aligned} \tag{6.9}$$

[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 $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$.

Рассмотрим какое-либо слагаемое из этой суммы:

$$\operatorname{sgn}(\sigma_1, \ldots, \sigma_k, \sigma_{k+1}, \ldots, \sigma_n) \cdot a_1^{\sigma_1} \cdots \underline{a_{k+1}^{\sigma_k}} \cdot \underline{a_k^{\sigma_{k+1}}} \cdots a_n^{\sigma_n}.$$

Переставив местами подчёркнутые сомножители, получим

$$\operatorname{sgn}(\sigma_1, \ldots, \sigma_k, \sigma_{k+1}, \ldots, \sigma_n) \cdot a_1^{\sigma_1} \cdots a_k^{\sigma_{k+1}} \cdot a_{k+1}^{\sigma_k} \cdots a_n^{\sigma_n};$$

это произведение входит в сумму (6.8) с противоположным знаком, поскольку

$$\operatorname{sgn}(\sigma_1, \ldots, \sigma_{k+1}, \sigma_k, \ldots, \sigma_n) = -\operatorname{sgn}(\sigma_1, \ldots, \sigma_k, \sigma_{k+1}, \ldots, \sigma_n).$$

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 (6.9), входит в формулу (6.8) с противоположным знаком, что завершает доказательство кососимметричности.

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 (6.8), получаем для определителя следующую *формулу полного развёртывания*, выражающую $\det A$ через элементы матрицы A:

$$\det A = \begin{vmatrix} A_1, & A_2, & \ldots, & A_n \end{vmatrix} = \cdot \sum_{\sigma \in S_n} \operatorname{sgn} \sigma \cdot a_1^{\sigma_1} \cdot a_2^{\sigma_2} \cdots a_n^{\sigma_n}. \qquad (6.10)$$

(Условие нормировки вытекает из очевидного факта, что в случае единичной матрицы сумма в (6.10) 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 6.2 свойствами, существует, единственна и выражается через элементы матрицы A формулой (6.10). Кроме того, мы установили следующий факт.

6.12. Предложение. *Если $F(A)$ — какая-либо полилинейная кососимметрическая функция столбцов квадратной матрицы A, то*

$$F(A) = F(\mathbb{1}) \cdot \det A.$$

6.13. Замечание. Формула полного развёртывания определителя (6.10) для практических вычислений непригодна. Действительно, попробуем оценить время, необходимое для вычисления определителя с помощью этой формулы. Правая часть в (6.10) 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 более простые операции. Например, при $n = 100$, оценив значение $100!$ с помощью формулы Стирлинга (**??**) (см. лекцию III, с. 335), получим

$100!\cdot 99 > 10^{153}$. 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).

E. Инвариантность определителя относительно транспонирования. Формула полного развёртывания определителя позволяет доказать свойство инвариантности определителя относительно операции транспонирования матрицы.

6.14. Теорема.

$$\det A^T = \det A.$$

Доказательство. Обозначив элементы матрицы A^T через $\tilde{a}_j^k$, 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 $\tilde{a}_j^k = a_k^j$

$$= \sum_{\sigma\in S_n} \operatorname{sgn}\sigma \cdot a_{\sigma_1}^1\cdot a_{\sigma_2}^2\cdots a_{\sigma_n}^n =$$

переставим сомножители так, чтобы нижние индексы (номера столбцов) расположились в натуральном 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 6.11, с. 132) условие $\sigma\in S_n$ эквивалентно условию $\tau = \sigma^{-1}\in S_n$, получаем

$$= \sum_{\tau\in S_n} \operatorname{sgn}\tau \cdot a_1^{\tau_1}\cdot a_2^{\tau_2}\cdots a_n^{\tau_n} = \det A,$$

что и требовалось доказать. □

В частности, мы получаем, что *определитель представляет собой полилинейную кососимметрическую функцию строк матрицы.*

F. Определители специального вида.

6.15. Предложение. *Определитель треугольной матрицы равен произведению её диагональных элементов.*

Доказательство. Докажем утверждение для нижнетреугольной матрицы:

$$\begin{vmatrix} a_1^1 & 0 & 0 & \ldots & 0 \\ a_1^2 & a_2^2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1^n & a_2^n & a_3^n & \ldots & a_n^n \end{vmatrix} = a_1^1 \cdot a_2^2 \cdots a_n^n.$$

Произведение диагональных элементов матрицы $a_1^1 \cdot a_2^2 \cdots a_n^n$ входит в сумму (6.10) со знаком плюс, так как соответствует тривиальной перестановке. Все остальные слагаемый в этой сумме равны нулю. Действительно, для нижнетреугольной матрицы, у которой $a_k^j = 0$ при $j < k$, из $a_1^{\sigma_1} a_2^{\sigma_2} \ldots a_n^{\sigma_n} \neq 0$ следует

$$\sigma_1 \geqslant 1, \quad \sigma_2 \geqslant 2, \quad \ldots, \quad \sigma_n \geqslant n.$$

Поскольку $\sigma_1 + \sigma_2 + \cdots + \sigma_n = 1 + 2 + \cdots + n$, это возможно лишь при $\sigma_1 = 1$, $\sigma_2 = 2$, …, $\sigma_n = n$. □

6.16. Предложение (определитель с нулевым углом)**.** *Пусть квадратная матрица A имеет блочную структуру*

$$A = \begin{bmatrix} B & D \\ O & C \end{bmatrix},$$

где B и C — квадратные блоки. Тогда

$$\begin{vmatrix} B & D \\ O & C \end{vmatrix} = |B| \cdot |C|. \tag{6.11}$$

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 6.12 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 6.12,

$$G(C) = G(\mathbb{1}) \cdot |C|.$$

Здесь

$$G(\mathbb{1}) = \begin{vmatrix} \mathbb{1} & D \\ O & \mathbb{1} \end{vmatrix};$$

это определитель треугольной матрицы с единицами на диагонали, который, разумеется, равен единице. Итак,

$$\det A = F(B) = F(\mathbb{1}) \cdot |B| = G(C) \cdot |B| = G(\mathbb{1}) \cdot |C| \cdot |B| = |C| \cdot |B|,$$

что и требовалось доказать. □

3. Разложение определителя по столбцам и строкам

А. Алгебраические дополнения и миноры. 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

$$\det A = \Big|A_1,\ A_2,\ \ldots,\ A_n\Big| = \Big|\underbrace{a_1^1\boldsymbol{e}_1 + a_1^2\boldsymbol{e}_2 + \cdots + a_1^n\boldsymbol{e}_n}_{=A_1},\ A_2,\ \ldots,\ A_n\Big| =$$

$$= a_1^1\underline{\Big|\boldsymbol{e}_1,\ A_2,\ \ldots,\ A_n\Big|} + a_1^2\underline{\Big|\boldsymbol{e}_2,\ A_2,\ \ldots,\ A_n\Big|} + \cdots + a_1^n\underline{\Big|\boldsymbol{e}_n,\ A_n,\ \ldots,\ A_n\Big|}.$$

Подчеркнутые определители называются *алгебраическими дополнениями* (АД) элементов $a_1^1, a_1^2, \ldots, a_1^n$ первого столбца матрицы A; обозначим их $\mathcal{A}_1^1, \mathcal{A}_1^2, \ldots, \mathcal{A}_1^n$ соответственно. Очевидно, эти величины *не зависят от элементов первого столбца матрицы A.*

Полученная формула называется *разложением определителя по первому столбцу*:

$$\det A = \sum_{k=1}^{n} a_1^k \mathcal{A}_1^k.$$

Она может служить средством вычисления определителя, если известен способ нахождения алгебраических дополнений.

Разумеется, вместо первого столбца можно выбрать любой другой:

$$\det A = \Big|A_1,\ \ldots,\ A_j,\ \ldots,\ A_n\Big| =$$

$$= \Big|A_1,\ \ldots,\ \underbrace{a_j^1\boldsymbol{e}_1 + a_j^2\boldsymbol{e}_2 + \cdots + a_j^n\boldsymbol{e}_n}_{=A_j},\ \ldots,\ A_n\Big| =$$

$$= a_j^1 \cdot \underbrace{\left| A_1,\ \ldots,\ \boldsymbol{e}_1,\ \ldots,\ A_n \right|}_{=\mathcal{A}_j^1} + a_j^2 \cdot \underbrace{\left| A_1,\ \ldots,\ \boldsymbol{e}_2,\ \ldots,\ A_n \right|}_{=\mathcal{A}_j^2} + \cdots +$$

$$+ \ a_j^n \cdot \underbrace{\left| A_1,\ \ldots,\ \boldsymbol{e}_n,\ \ldots,\ A_n \right|}_{=\mathcal{A}_j^n} = \sum_{k=1}^{n} a_j^k \mathcal{A}_j^k.$$

Ясно, что алгебраические дополнения элементов j-го столбца не зависят от элементов этого столбца.

Для алгебраического дополнения $\mathcal{A}_1^1$ элемента a_1^1 по формуле (6.11) 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.

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 $(-1)^{k-1}$:

$$\mathcal{A}_j^k = \underbrace{(-1)^{j-1} \cdot (-1)^{k-1}}_{=(-1)^{j+k}} \cdot \begin{vmatrix} 1 & a_1^k & \dots & a_{j-1}^k & a_{j+1}^k & \dots & a_n^k \\ 0 & a_1^1 & \dots & a_{j-1}^1 & a_{j+1}^1 & \dots & a_n^1 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_1^{k-1} & \dots & a_{j-1}^{k-1} & a_{j+1}^{k-1} & \dots & a_n^{k-1} \\ 0 & a_1^{k+1} & \dots & a_{j-1}^{k+1} & a_{j+1}^{k+1} & \dots & a_n^{k+1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_1^n & \dots & a_{j-1}^n & a_{j+1}^n & \dots & a_n^n \end{vmatrix}.$$

Как было показано выше, из полученного определителя порядка n можно удалить первую строку и первый столбец, в результате чего образуется определитель порядка $n-1$, который, очевидно, получается из исходного определителя $\det A$ вычёркиванием k-й строки и j-го столбца. Этот определитель называется *минором* элемента a_j^k и обозначается M_j^k (линии изображают вычеркнутые строку и столбец):

$$M_j^k = \left|\begin{array}{ccc|ccc} a_1^1 & \dots & a_{j-1}^1 & a_{j+1}^1 & \dots & a_n^1 \\ \vdots & & \vdots & \vdots & & \vdots \\ a_1^{k-1} & \dots & a_{j-1}^{k-1} & a_{j+1}^{k-1} & \dots & a_n^{k-1} \\ \hline a_1^{k+1} & \dots & a_{j-1}^{k+1} & a_{j+1}^{k+1} & \dots & a_n^{k+1} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_1^n & \dots & a_{j-1}^n & a_{j+1}^n & \dots & a_n^n \end{array}\right|.$$

Таким образом, мы нашли выражение алгебраического дополнения $\mathcal{A}_j^k$ элемента a_j^k через соответствующий минор M_j^k:

$$\boxed{\mathcal{A}_j^k = (-1)^{j+k} M_j^k.}$$

Выпишем ещё раз формулу разложения определителя по j-му столбцу:

$$|A| = \sum_{k=1} a_j^k \mathcal{A}_j^k = \sum_{k=1} (-1)^{k+j} a_j^k M_j^k. \tag{6.12}$$

Аналогично получается формула разложения определителя по строке:

$$\det A = \sum_{j=1}^{n} a_j^k \mathcal{A}_j^k = \sum_{j=1}^{n} (-1)^{k+j} a_j^k M_j^k. \tag{6.13}$$

Итак, *определитель равен сумме попарных произведений элементов произвольно выбранного столбца (или строки) на их алгебраические дополнения.*

Каждая из формул (6.12), (6.13) приводят вычисление определителя n-го порядка к вычислению n определителей, порядок каждого из которых на единицу меньше.

B. Фальшивое разложение определителя. Рассмотрим формулу (6.12) разложения определителя матрицы A по j-му столбцу. Если заменить в ней элементы a_j^k $(k = 1, 2, \ldots, n)$ j-го столбца матрицы A на произвольные числа $b^1, \ldots, b^n$, то, очевидно, получится разложение по j-му столбцу определителя новой матрицы A', полученной из матрицы A заменой j-го столбца $A_j = (a_j^1, a_j^2, \ldots, a_j^n)^T$ на столбец $B = (b^1, b^2, \ldots, b^n)$. 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.

6.17. Теорема (фальшивое разложение определителя)**.** *Сумма попарных произведений элементов произвольно выбранного столбца на алгебраические дополнения элементов другого столбца равна нулю*:

$$\sum_{k=1} a_j^k \mathcal{A}_p^k = 0, \quad \textit{если } j \neq p.$$

Аналогичное утверждение имеет место и для строк:

$$\sum_{j=1} a_j^k \mathcal{A}_j^q = 0, \quad \textit{если } k \neq q.$$

Объединяя эти формулы с (6.12) и (6.13), получаем:

$$\sum_{k=1} a_j^k \mathcal{A}_p^k = |A| \cdot \delta_{jp} = \begin{cases} |A|, & \text{если } j = p, \\ 0, & \text{если } j \neq p, \end{cases} \tag{6.14}$$

$$\sum_{j=1} a_j^k \mathcal{A}_j^q = |A| \cdot \delta_{kq} = \begin{cases} |A|, & \text{если } k = q, \\ 0, & \text{если } k \neq q. \end{cases} \tag{6.15}$$

4. Важные теоремы об определителях

A. Критерий равенства определителя нулю.

6.18. Теорема. *Определитель матрицы A равен нулю тогда и только тогда, когда её столбцы (строки) линейно зависимы.*

Доказательство. Необходимость. Пусть $\det A = 0$. Рассмотрим однородную систему линейных уравнений

$$AX = O \iff x^1 A_1 + x^2 A_2 + \dots + x^n A_n = O.$$

Формулы Крамера к этой системе неприменимы, так что либо она несовместна, либо решение не единственно. Поскольку, очевидно, однородная система совместна всегда (имеет тривиальное решение $X = O$), реализуется вторая возможность, т.е. существуют такие числа $(x^1, x^2, \dots, x^n)$, среди которых имеется хотя бы одно ненулевое, что $x^1 A_1 + x^2 A_2 + \dots + x^n A_n = O$, что и означает линейную зависимость столбцов матрицы A.

Достаточность. Пусть столбцы матрицы линейно зависимы. По теореме 5.6(2) (см. лекцию 5, с. 99) один из них равен линейной комбинации остальных. Вычитая из этого столбца указанную линейную комбинацию, получим определитель с нулевым столбцом, который, очевидно, равен нулю. □

B. Определитель произведения матриц.

6.19. Теорема. *Если A и B — квадратные матрицы одного порядка, то*

$$|AB| = |A| \cdot |B|.$$

Доказательство. Пусть $C = AB$; согласно теореме 4.4, 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 матрицы $|C|$ получаем

$$|C| = \Big|C_1, \ \dots, \ C_n\Big| = \left|\sum_{l_1=1}^{n} A_{l_1} \; b_1^{l_1}, \ \cdots \sum_{l_n=1}^{n} A_{l_n} \; b_n^{l_n}\right| =$$

$$= \sum_{l_1=1}^{n} \cdots \sum_{l_1=1}^{n} b_1^{l_1} \cdots b_n^{l_n} \cdot \Big|A_{l_1}, \ \dots, \ A_{l_n}\Big|.$$

В этой сумме, состоящей из n^n слагаемых, обращаются в нуль все слагаемые, в состав которых в качестве множителя входят определители

$\Big|A_{l_1},\ \ldots,\ A_{l_n}\Big|$ 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

$$|C| = \sum_{\sigma\in S_n} b_1^{\sigma_1}\cdots b_n^{\sigma_n}\cdot\Big|A_{\sigma_1},\ \ldots,\ A_{\sigma_n}\Big|.$$

Поскольку

$$\Big|A_{\sigma_1},\ \ldots,\ A_{\sigma_n}\Big| = \operatorname{sgn}\sigma\cdot\Big|A_1,\ \ldots,\ A_n\Big|,$$

получаем далее

$$|C| = \underbrace{\Big|A_{\sigma_1},\ \ldots,\ A_{\sigma_n}\Big|}_{=\|A|}\sum_{\sigma\in S_n}\operatorname{sgn}\sigma b_1^{\sigma_1}\cdots b_n^{\sigma_n} = |A|\cdot|B|,$$

что и требовалось доказать. □

5. Обратная матрица

6.20. Определение. Квадратная матрица называется невырожденной, или неособой, если её определитель отличен от нуля.

6.21. Теорема. *Квадратная матрица обратима тогда и только тогда, когда она невырождена.*

Доказательство. Необходимость. Пусть матрица A обратима и A^{-1} — её обратная. По теореме 6.19 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 матрицей для матрицы A; иногда используется обозначение $\operatorname{adj} A$ (adjoint).

Нетрудно найти произведение матриц A и $A^\vee$: для любого элемента этого произведения имеем

$$\big[A\cdot A^\vee\big]_j^k = \sum_{p=1}^n [A]_p^k\,[A^\vee]_j^p = \sum_{p=1}^n a_p^k\,\mathcal{A}_p^j = \det A\cdot\delta_{kj}$$

(мы воспользовались формулой (6.15)). Таким образом,

$$A\cdot A^\vee = |A|\cdot\mathbb{1},$$

так что при $|A| \neq 0$ имеем

$$A^{-1} = \frac{1}{|A|} A^{\vee}, \tag{6.16}$$

т.е. мы не только доказали существование обратной матрицы для A, но и получили формулу для её нахождения.

Отметим, что в случае, когда $|A| = 0$, произведение матрицы A на её присоединённую равно нулевой матрице. □

Формула (6.16) 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 $n-1$.

ГЛАВА 7

Арифметическое пространство $\mathbb{K}^n$, II
Базис и размерность. Ранг матрицы

1. Дальнейшие приложения алгоритма Гаусса—Жордана

A. Элементарные преобразования и умножение матриц. Элементарные преобразования строк матрицы можно рассматривать со следующей точки зрения.

7.1. Теорема. *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. Поскольку при умножении матриц каждый столбец матрицы-произведения равен произведению матрицы-первого сомножителя на соответствующий столбец матрицы-второго сомножителя (см. теорему 4.4(2) лекции 4, с. 90), достаточно проверить утверждение теоремы для матрицы A, состоящей из одного столбца. Если R — перестановка строк с номерами s и t, имеем

$$R(\mathbb{1}) = R\left(\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}\right) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix},$$

$$R(\mathbb{1})\cdot A=\begin{pmatrix}1&\cdots&0&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&0&\cdots&1&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&1&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&0&\cdots&0&\cdots&1\end{pmatrix}\cdot\begin{pmatrix}a^1\\ \vdots\\ a^s\\ \vdots\\ a^t\\ \vdots\\ a^n\end{pmatrix}=\begin{pmatrix}a^1\\ \vdots\\ a^t\\ \vdots\\ a^s\\ \vdots\\ a^n\end{pmatrix}=R(A).$$

Обратимость матрицы $R(\mathbb{1})$ элементарного преобразования очевидна: обратная матрица, осуществляющая обратное элементарное преобразование, совпадает с исходной,

$$\begin{pmatrix}1&\cdots&0&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&0&\cdots&1&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&1&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&0&\cdots&0&\cdots&1\end{pmatrix}^{-1}=\begin{pmatrix}1&\cdots&0&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&0&\cdots&1&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&1&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&0&\cdots&0&\cdots&1\end{pmatrix}.$$

Если R — элементарное преобразование, состоящее в умножении строки с номером s на число $\lambda\neq 0$, то

$$R(\mathbb{1})=R\left(\begin{pmatrix}1&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots\\ 0&\cdots&1&\cdots&0\\ \vdots&&\vdots&&\vdots\\ 0&\cdots&0&\cdots&1\end{pmatrix}\right)=\begin{pmatrix}1&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots\\ 0&\cdots&\lambda&\cdots&0\\ \vdots&&\vdots&&\vdots\\ 0&\cdots&0&\cdots&1\end{pmatrix},$$

$$R(\mathbb{1})\cdot A=\begin{pmatrix}1&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots\\ 0&\cdots&\lambda&\cdots&0\\ \vdots&&\vdots&&\vdots\\ 0&\cdots&0&\cdots&1\end{pmatrix}\cdot\begin{pmatrix}a^1\\ \vdots\\ a^s\\ \vdots\\ a^n\end{pmatrix}=\begin{pmatrix}a^1\\ \vdots\\ \lambda a^s\\ \vdots\\ a^n\end{pmatrix}=R(A).$$

Обратимость матрицы $R(\mathbb{1})$ элементарного преобразования здесь также очевидна:

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & \lambda & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & \frac{1}{\lambda} & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

Наконец, для преобразования, состоящего в прибавлении строки с номером s, умноженной на число λ, к строке с номером t, имеем

$$R(\mathbb{1}) = R\left(\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}\right) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \lambda & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix},$$

$$R(\mathbb{1}) \cdot A = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \lambda & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \cdot \begin{pmatrix} a^1 \\ \vdots \\ a^s \\ \vdots \\ a^t \\ \vdots \\ a^n \end{pmatrix} = \begin{pmatrix} a^1 \\ \vdots \\ a^s \\ \vdots \\ a^t + \lambda a^s \\ \vdots \\ a^n \end{pmatrix} = R(A).$$

Обратное преобразование состоит в вычитании умноженной на λ строки с номером s из строки с номером t; этому преобразованию соответствует матрица

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & -\lambda & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix},$$

которая и является обратной для матрицы элементарного преобразования в данном случае.

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

$$(R_2 \circ R_1)(A) = R_2\bigl(R_1(A)\bigr) = R_2(\mathbb{1}) \cdot R_1(A) = R_2(\mathbb{1}) \cdot \bigl(R_1(\mathbb{1}) \cdot A\bigr) = \\ = \bigl(R_2(\mathbb{1}) \cdot R_1(\mathbb{1})\bigr) \cdot A = R_2\bigl(R_1(\mathbb{1})\bigr) \cdot A = \bigl(R_2 \circ R_1\bigr)(\mathbb{1}) \cdot A,$$

что и требовалось.

Итак, каждое преобразование матрицы с помощью элементарных преобразований строк эквивалентно умножению этой матрицы слева на некоторую обратимую квадратную матрицу. □

B. Сохранение линейной зависимости и независимости при элементарных преобразованиях.

7.2. 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*

$$\alpha^1 A_1 + \alpha^2 A_2 + \dots + \alpha^k A_n = O,$$

то соответствующие столбцы матрицы A' связаны такой же линейной зависимостью:

$$\alpha^1 A'_1 + \alpha^2 A'_2 + \dots + \alpha^k A'_n = O.$$

Доказательство. Обозначив через C матрицу выполненных элементарных преобразований, можем записать $A'_i = CA_i$.

(1) Пусть столбцы $A_1, \dots, A_n$ матрицы A линейно независимы. Предположим, что столбцы преобразованной матрицы A', напротив, линейно зависимы, т.е. существуют такие коэффициенты $\alpha^1, \dots, \alpha^n$, среди которых имеется хотя бы один ненулевой, что

$$\alpha^1 A'_1 + \alpha^2 A'_2 + \dots + \alpha^k A'_n = O.$$

Заменяя A'_i на CA_i и вынося за скобки матрицу C, получаем

$$\alpha^1 CA_1 + \alpha^2 CA_2 + \dots + \alpha^k CA_n = O \quad \Longleftrightarrow \quad C\Bigl(\alpha^1 A_1 + \alpha^2 A_2 + \dots + \alpha^k A_n\Bigr) = O.$$

Умножая обе части равенства слева на C^{-1}, находим

$$\alpha^1 A_1 + \alpha^2 A_2 + \dots + \alpha^k A_n = O,$$

что означает линейную зависимость столбцов матрицы A, противоречие.

(2) Умножая слева на C обе части равенства

$$\alpha^1 A_1 + \alpha^2 A_2 + \dots + \alpha^k A_n = O$$

и замечая, что $CA_i = A'_i$, получаем

$$\alpha^1 A'_1 + \alpha^2 A'_2 + \dots + \alpha^k A'_n = O,$$

что и требовалось. □

C. Вычисление обратной матрицы. При помощи алгоритма Гаусса—Жордана можно обращать матрицы. Действительно, если при помощи элементарных преобразований строк привести данную квадратную матрицу A к единичной, то, в силу теоремы 7.1 матрицей преобразования будет A^{-1}. На практике удобно поступить следующим образом: составить блочную матрицу $\left[A \mid \mathbb{1}\right]$ и элементарными преобразованиями строк привести её левый блок к единичной матрице; тогда в правом блоке образуется матрица A^{-1}:

$$\left[A \mid \mathbb{1}\,\right] \xrightarrow[\text{строк}]{\text{эл. преобр.}} \left[\mathbb{1} \mid A^{-1}\,\right].$$

Пример 7.1. Найдите матрицу, обратную к матрице

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

Решение:

$$\left(\begin{array}{ccc|ccc} \boxed{1} & 2 & 3 & 1 & 0 & 0 \\ 2 & 3 & 4 & 0 & 1 & 0 \\ 3 & 4 & 6 & 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & -2 & -3 & -3 & 0 & 1 \end{array}\right) \to$$

$$\to \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & \boxed{1} & 2 & 2 & -1 & 0 \\ 0 & -2 & -3 & -3 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & -1 & -3 & 2 & 0 \\ 0 & 1 & 2 & 2 & -1 & 0 \\ 0 & 0 & \boxed{1} & 1 & -2 & 1 \end{array}\right) \to$$

$$\to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -2 & 0 & 1 \\ 0 & 1 & 0 & 0 & 3 & -2 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right).$$

Обратная матрица равна

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

Аналогичную методику можно использовать для вычисления матрицы $A^{-1}B$:

$$\left[\,A \mid B\,\right] \xrightarrow[\text{строк}]{\text{эл. преобр.}} \left[\,\mathbb{1} \mid A^{-1}B\,\right].$$

Если требуется найти матрицу AB^{-1}, можно подобным образом использовать элементарные преобразования столбцов, однако на практике поступают следующим образом:

$$\left[\, B^T \mid A^T \,\right] \xrightarrow[\text{строк}]{\text{эл. преобр.}} \left[\, \mathbb{1} \mid (B^T)^{-1}A^T \,\right] \xrightarrow[\text{рование}]{\text{транспони-}} \Big((B^T)^{-1}A^T\Big)^T = AB^{-1}.$$

D. Составление однородной системы по заданной ФСР. Пусть имеется набор линейно независимых столбцов $X_1, \dots, X_p \in \mathbb{K}^n$. Составим систему линейных однородных уравнений, для которой данный набор столбцов образует фундаментальное семейство решений.

Пусть X — произвольное решение искомой системы; оно может быть разложено в линейную комбинацию столбцов $X_1, \dots, X_p$:

$$X = \alpha^1 X_1 + \dots + \alpha^p X_p.$$

Будем рассматривать это соотношение как систему неоднородных уравнений относительно неизвестных $\alpha^1, \dots, \alpha^p$. Если привести эту систему к упрощённому виду, получим

$$\left(\begin{array}{cccc|c} x_1^1 & x_2^1 & \cdots & x_p^1 & x^1 \\ x_1^2 & x_2^2 & \cdots & x_p^2 & x^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_1^k & x_2^k & \cdots & x_p^p & x^k \\ x_1^{k+1} & x_2^{k+1} & \cdots & x_p^{p+1} & x^{p+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_1^n & x_2^n & \cdots & x_p^n & x^n \end{array}\right) \Rightarrow \left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & * \\ 0 & 1 & \cdots & 0 & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & * \\ 0 & 0 & \cdots & 0 & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & * \end{array}\right).$$

Поскольку последний столбец матрицы представляет собой линейную комбинацию предыдущих столбцов, выражения, расположенные в правом нижнем углу последней матрицы, должны быть равны нулю; это и есть искомая система. Очевидно, количество уравнений в этой системе равно $n - p$.

Пример 7.2. Найдите 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.

Решение. Составим матрицу, столбцами которой являются данные векторы X_1 и X_2, а также вектор $(x^1, x^2, x^3, x^4)^T$ и приведём её к упрощённому виду:

$$\left(\begin{array}{cc|c} 1 & 5 & x^1 \\ 2 & 6 & x^2 \\ 3 & 7 & x^3 \\ 4 & 8 & x^4 \end{array}\right) \Rightarrow \left(\begin{array}{cc|c} 1 & 5 & x^1 \\ 0 & -4 & -6x^1 + 5x^2 \\ 0 & -8 & -7x^1 + 5x^3 \\ 0 & -12 & -8x^1 + 5x^4 \end{array}\right) \Rightarrow$$

$$\Rightarrow \left(\begin{array}{cc|c} -4 & 0 & 26x^1 - 25y \\ 0 & -4 & 5x^2 - 6x^1 \\ 0 & 0 & -20x^1 + 40x^2 - 20x^3 \\ 0 & 0 & -40x^1 + 60x^2 - 20x^4 \end{array}\right).$$

Для того, чтобы столбцы матрицы были линейно зависимы, требуется обращение в нуль выделенных элементов. Разделив эти выражения на (-20) и приравняв их к нулю, получаем искомую систему уравнений:

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

Замечание. Найденная система имеет упрощённый вид; базисными неизвестными являются x^3 и x^4:

$$\begin{cases} x^3 = -\ x^1 + 2x^2, \\ x^4 = -2x^1 + 3x^2 \end{cases}$$

Нормальное фундаментальное семейство решений, отвечающая этим базисным неизвестным, есть

$$Y_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ -2 \end{pmatrix}, \quad Y_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix}.$$

Нетрудно получить разложения векторов $X_1 = (1, 2, 3, 4)^T$ и $X_2(5, 6, 7, 8)^T$ (которые также образуют ФСР, но не являющуюся *нормальной*) в линейные комбинации векторов НФСР X_1 и X_2:

$$\begin{array}{c} \begin{array}{cccc} Y_1 & Y_2 & X_1 & X_2 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 5 \\ 0 & 1 & 2 & 6 \\ -1 & 2 & 3 & 7 \\ -2 & 3 & 4 & 8 \end{pmatrix} \end{array} \Rightarrow \begin{pmatrix} 1 & 0 & 1 & 5 \\ 0 & 1 & 2 & 6 \\ 0 & 2 & 4 & 12 \\ 0 & 3 & 6 & 18 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 1 & 5 \\ 0 & 1 & 2 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

т.е. $X_1 = Y_1 + 2Y_2$, $X_2 = 5Y_1 + 6Y_2$.

2. Базис и размерность

А. Дальнейшие свойства линейно зависимых и линейно независимых векторов. Свойства линейно зависимых и линейно независимых векторов, перечисленные в теореме 5.6 (см. лекцию 5, с. 99) будучи довольно очевидными, тривиально доказываются. Следующее же свойство отнюдь не тривиально.

Напомним, что *семейством* элементов множества мы называем упорядоченный набор элементов (см. определение 5.1, с. 97).

7.3. Теорема. *Если векторы $Y_1, \dots, Y_s$ принадлежат линейной оболочке векторов $X_1, \dots, X_r$ и $s > r$, то семейство векторов $Y_1, \dots, Y_s$ линейно зависимо.*

Доказательство. Каждый из векторов $Y_1, \dots, Y_s$ можно представить в виде линейной комбинации векторов $X_1, \dots, X_r$:

$$\begin{gathered} Y_1 = a_1^1 X_1 + a_1^2 X_2 + \dots + a_1^r X_r, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ Y_s = a_s^1 X_1 + a_s^2 X_2 + \dots + a_s^r X_r, \end{gathered}$$

где a_j^i — некоторые числа. Введя в рассмотрение матрицы

$$\underset{1\times s}{\boldsymbol{Y}} = \big[Y_1, Y_2, \dots, Y_s\big], \quad \underset{1\times r}{\boldsymbol{X}} = \big[X_1, X_2, \dots, X_r\big], \qquad \underset{r\times s}{A} = \begin{pmatrix} a_1^1 & a_2^1 & \dots & a_s^1 \\ a_1^2 & a_2^2 & \dots & a_s^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^r & a_2^r & \dots & a_s^r \end{pmatrix},$$

мы можем записать эти равенства в виде одного матричного соотношения

$$\boldsymbol{Y} = \boldsymbol{X} A. \tag{7.1}$$

Рассмотрим систему r однородных линейных уравнений с s неизвестными

$$\begin{cases} a_1^1 z^1 + a_2^1 z^2 + \dots + a_s^1 z^s = 0, \\ a_1^2 z^1 + a_2^2 z^2 + \dots + a_s^2 z^s = 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_1^r z^1 + a_2^r z^2 + \dots + a_s^r z^s = 0 \end{cases} \iff \quad AZ = O, \quad \text{где } Z = \begin{pmatrix} z^1 \\ z^2 \\ \vdots \\ z^s \end{pmatrix}.$$

Поскольку по условию $s > r$, т.е. количество неизвестных превышает количество уравнений, эта однородная система имеет нетривиальное решение (см. теорему 5.17 в лекции 5, с. 115). Обозначим это решение $Z_* = (z_*^1, z_*^2, \dots, z_*^s)^T$ и составим линейную комбинацию векторов $Y_1, \dots, Y_s$ с коэффициентами $z_*^1, z_*^2, \dots, z_*^s$. Эта линейная комбинация нетривиальна и в матричной записи имеет вид

$$z_*^1 Y_1 + z_*^2 Y_2 + \dots + z_*^s Y_s \equiv \boldsymbol{Y} Z_*.$$

В силу (7.1) имеем

$$\boldsymbol{Y} Z_* = (\boldsymbol{X} A) Z_* = \boldsymbol{X}(A Z_*) = \boldsymbol{X} O = \mathbf{0},$$

что и доказывает линейную зависимость векторов $Y_1, \dots, Y_s$. $\square$

7.4. Следствие. *Любая линейно независимая система векторов в пространстве $\mathbb{K}^n$ состоит не более чем из n векторов.*

Доказательство. Пространство $\mathbb{K}^n$ является линейной оболочкой n векторов

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

(эти векторы образуют так называемый *стандартный* базис пространства $\mathbb{K}^n$). Остаётся применить теорему 7.3. $\square$

В. Существование базиса. Размерность.

7.5. Теорема.

1. *Каждое ненулевое подпространство $\mathcal{P}$ в $\mathbb{K}^n$ обладает конечным базисом.*
2. *Все базисы подпространства $\mathcal{P}$ состоят из одного и того же числа векторов*; *это число называется размерностью подпространства $\mathcal{P}$ и обозначается* $\dim \mathcal{P}$.

Доказательство. По условию, $\mathcal{P}$ содержит хотя бы один ненулевой вектор X_1. Предположим, что в $\mathcal{P}$ найден линейно независимый набор векторов $X_1, \ldots, X_k$. Если их линейная оболочка $L(X_1, \ldots, X_k)$ не совпадает с подпространством $\mathcal{P}$, то выберем в $\mathcal{P}$ вектор $X_{k+1} \notin L(X_1, \ldots, X_k)$. Вектор X_{k+1} не является линейной комбинацией векторов $X_1, \ldots, X_k$, поэтому, в силу утверждения 7 теоремы 5.6 система $X_1, \ldots, X_k, X_{k+1}$ линейно независима. Процесс расширения линейно независимой системы не может продолжаться неограниченно, поскольку в силу следствия 7.4 любая линейно независимая система в $\mathbb{K}^n$ состоит не более чем из n векторов. Таким образом, при некотором натуральном $r \leqslant n$ линейно независимая система $X_1, \ldots, X_k, \ldots, X_r$ станет максимальной, т.е. при любом векторе $X \neq O$ система $X_1, \ldots, X_k, \ldots, X_r, X$ будет линейно зависимой. Согласно утверждению 6 теоремы 5.6 получим включение $X \in L(X_1, \ldots, X_r)$, так что $\mathcal{P} = L(X_1, \ldots, X_r)$, а векторы $X_1, \ldots, X_r$ образуют базис в $\mathcal{P}$.

Теперь предположим, что $Y_1, \ldots, Y_s$ — ещё один базис в $\mathcal{P}$. По теореме 7.3 имеем неравенство $s \leqslant r$. Меняя местами системы векторов $X_1, \ldots, X_r$ и $Y_1, \ldots, Y_s$, аналогично получаем неравенство $r \leqslant s$. Таким образом, $s = r$, и теорема доказана. $\square$

Итак, каждому подпространству $\mathcal{P}$ в $\mathbb{K}^n$ ставится в соответствие натуральное число $r \leqslant n$, называемое размерностью этого подпространства, $r = \dim \mathcal{P}$. В частности, $\dim \mathbb{K}^n = n$. По определению полагается, что размерность нулевого подпространства $\mathcal{P} = \{O\}$ равна нулю: $\dim\{O\} = 0$.

7.6. Теорема.

1. *Пусть векторы $Y_1, \ldots, Y_p$ принадлежат линейной оболочке векторов $X_1, \ldots, X_q$. Тогда*

$$L(Y_1, \ldots, Y_p) \subset L(X_1, \ldots, X_q).$$

2. *Пусть $\mathcal{P}$ и $\mathcal{Q}$ — два подпространства в $\mathbb{K}^n$, причём $\mathcal{Q} \subseteq \mathcal{P}$. Тогда* $\dim \mathcal{Q} \leqslant \dim \mathcal{P}$.
3. *Пусть $\mathcal{P}$ и $\mathcal{Q}$ — два подпространства в $\mathbb{K}^n$, причём $\mathcal{Q} \subseteq \mathcal{P}$ и* $\dim \mathcal{P} = \dim \mathcal{Q}$. *Тогда $\mathcal{P} = \mathcal{Q}$.*

Доказательство. 1. Достаточно показать, что любой вектор X, лежащий в $L(Y_1, \ldots, Y_p)$, лежит также и в $L(X_1, \ldots, X_q)$. Итак, пусть $X \in L(Y_1, \ldots, Y_p)$, т.е.

$$X = \beta^1 Y_1 + \cdots + \beta^p Y_p.$$

Поскольку $Y_1, \ldots, Y_p \in L(X_1, \ldots, X_q)$, имеем

$$Y_1 = \alpha_1^1 X_1 + \cdots + \alpha_1^q X_q, \ldots Y_p = \alpha_p^1 X_1 + \cdots + \alpha_p^q X_q.$$

Имеем

$$\begin{aligned} X &= \beta^1 Y_1 + \cdots + \beta^p Y_p = \\ &= \beta^1 \left(\alpha_1^1 X_1 + \cdots + \alpha_1^q X_q\right) + \cdots + \beta^p \left(\alpha_p^1 X_1 + \cdots + \alpha_p^q X_q\right) = \\ &= (\beta^1 \alpha_1^1 + \ldots \beta^p \alpha_p^1) X_1 + \cdots + (\beta^1 \alpha_1^q + \ldots \beta^p \alpha_p^q) X_q, \end{aligned}$$

а это означает, что $X \in L(X_1, \ldots, X_q)$, т.е. $L(Y_1, \ldots, Y_p) \subset L(X_1, \ldots, X_q)$.

2. Пусть $X_1, \ldots, X_r$ — базис в $\mathcal{P}$, т.е. $\dim \mathcal{P} = r$, а $Y_1, \ldots, Y_s$ — базис в $\mathcal{Q}$, т.е. $\dim \mathcal{Q} = r$. По теореме 7.3 имеем $s \leqslant r$, что и требовалось.

3. Пусть $Y_1, \ldots, Y_s$ — базис в $\mathcal{Q}$, т.е. $\dim \mathcal{Q} = s$. Предположим, что $\mathcal{Q} \subsetneqq \mathcal{P}$, т.е. существует такой вектор $Z \in \mathcal{P}$, что $Z \notin \mathcal{Q}$. Поскольку Z не может быть представлен в виде линейной комбинации векторов $Y_1, \ldots, Y_s$, делаем вывод, что семейство $Z, Y_1, \ldots, Y_s$ линейно независимо (см. утверждение 7 теоремы 5.6 из лекции 5, с. 99). Таким образом, $\mathcal{P}$ содержит не менее $s + 1$ линейно независимых векторов, т.е. $\dim \mathcal{P} \geqslant s+1$, противоречие с условием $\dim \mathcal{P} = \dim \mathcal{Q}$. □

C. Способы задания подпространств в $\mathbb{K}^n$. Как уже указывалось, имеется по существу два способа задать подпространство в $\mathbb{K}^n$:

(i) либо как линейную оболочку некоторого семейства столбцов (теорема 5.8),

(ii) либо как множество решений некоторой однородной системы линейных уравнений (теорема 5.13).

Чтобы перейти от задания подпространства способом (i) к заданию способом (ii), нужно составить однородную систему с заданным множеством решений; это делается с помощью алгоритма, описанного в разделе D (см. с. 149).

Переход от задания подпространства способом (ii) к заданию способом (i) — это не что иное, как решение заданной системы: её ФСР является базисом в рассматриваемом подпространстве.

3. Ранг матрицы

А. Пространства строк и столбцов матрицы. Рассмотрим матрицу

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

и введём обычные обозначения для её столбцов и строк:

$$A_1 = \begin{pmatrix} a_1^1 \\ a_1^2 \\ \vdots \\ a_1^m \end{pmatrix}, \ A_2 = \begin{pmatrix} a_2^1 \\ a_2^2 \\ \vdots \\ a_2^m \end{pmatrix}, \ \dots, \ A_n = \begin{pmatrix} a_n^1 \\ a_n^2 \\ \vdots \\ a_n^m \end{pmatrix}, \quad \begin{aligned} A^1 &= \left(a_1^1,\ a_2^1,\ \dots,\ a_n^1\right), \\ A^2 &= \left(a_1^2,\ a_2^2,\ \dots,\ a_n^2\right), \\ A^m &= \left(a_1^m,\ a_2^m,\ \dots,\ a_n^m\right). \end{aligned}$$

Рассмотрим линейные оболочки столбцов и строк матрицы A:

$$L(A_1, A_2, \dots, A_n), \quad L(A^1, A^2, \dots, A^m).$$

7.7. 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 лекции 5 (см. теорему 7.2, с. 147). 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 7.6 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 $L(A^1, A^2, \dots, A^m) \subset L(A'^1, A'^2, \dots, A'^m)$, откуда и следует требуемое равенство линейных оболочек. □

Замечание. 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.

Замечание. 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. **??**). Теперь становится понятным, что базисные строки и столбцы матрицы образуют соответственно базисы в линейных оболочках строк и столбцов, чем и объясняется их название.

B. Понятие ранга матрицы.

7.8. 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* $\operatorname{rk} A$.

Доказательство. Приведя матрицу A к упрощённому виду при помощи алгоритма Гаусса—Жордана:

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

видим, что количество базисных строк (равное максимальному количеству линейно независимых строк матрицы, т.е. размерности линейной оболочки строк) совпадает с количеством базисных столбцов, что и требовалось. □

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 (см. пример **??** в лекции 7, с. **??**). 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.

C. Совместность неоднородных систем. Теорема Кронекера—Капелли.

7.9. Теорема (теорема Кронекера—Капелли). *Неоднородная система $AX = B$ совместна тогда и только тогда, когда ранги её основной и расширенной матриц совпадают:*

$$\operatorname{rk} A = \operatorname{rk} \bigl[A \mid B\bigr].$$

Доказательство. Поскольку линейная оболочка столбцов основной матрицы системы содержится в линейной оболочке столбцов расширенной матрицы, $L(A_1, \ldots, A_n) \subset L(A_1, \ldots, A_n, B)$, имеем неравенство $\operatorname{rk} A \leqslant \operatorname{rk}[A \mid B]$, справедливое для любой системы (совместной или несовместной).

Записав систему $AX = B$ в виде

$$x^1 A_1 + x^2 A_2 + \cdots + x^n A_n = B,$$

видим, что совместность рассматриваемой системы эквивалентна тому факту, что столбец B принадлежит линейной оболочке столбцов матрицы A, т.е. $L(A_1, \ldots, A_n, B) \subset L(A_1, \ldots, A_n)$ (см. теорему 7.6) и потому $\operatorname{rk}[A \mid B] \leqslant \operatorname{rk} A$. Таким образом, для совместных систем и только для них $\operatorname{rk}[A \mid B] = \operatorname{rk} A$. □

При решении неоднородных систем методом Гаусса—Жордана использовать теорему Кронекера—Капелли нет необходимости, поскольку из упрощённого вида матрицы системы совместность или несовместность системы очевидна. У несовместных систем несовпадение рангов основной и расширенной матриц проявляется в появлении строк вида $(0, 0, \ldots, 0, 1)$, отвечающих несовместным уравнениям $0 \cdot x^1 + \cdots + 0 \cdot x^n = 1$.

D. Ранг произведения матриц.

7.10. Теорема. *Если существует произведение матриц A и B, то*

$$\operatorname{rk}(AB) \leqslant \operatorname{rk} A, \quad \operatorname{rk}(AB) \leqslant \operatorname{rk} B, \tag{7.2}$$

т.е. ранг произведения матриц не превышает ранга каждого из сомножителей.

Доказательство. Докажем первое неравенство. Для произвольного вектора $Y \in \operatorname{im}(AB)$ найдётся такой вектор X, что $Y = A(BX)$, т.е. $Y \in \operatorname{im} A$, так что $\operatorname{im}(AB) \subseteq \operatorname{im} A$ и поэтому $\dim \operatorname{im}(AB) \leqslant\sim \operatorname{im} A$.

Можно рассуждать иначе: 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 4.4 лекции 4, с. 90), так что

$$L(C_1, C_2, \ldots, C_s) \subseteq L(A_1, A_2, \ldots, A_n)$$

(см. утверждение 1 теоремы 7.6, с. 152). Поэтому

$$\dim L(C_1, C_2, \ldots, C_s) \leqslant \dim L(A_1, A_2, \ldots, A_n),$$

что и требовалось доказать. □

4. Теорема о базисном миноре

Выделим в прямоугольной матрице $A \in \mathbb{K}^{m\times n}$ несколько строк с номерами $i_1 < i_2 < \dots < i_k$ (всего k строк) и такое же количество столбцов с номерами $j_1 < j_2 < \dots < j_k$ (ясно, что $k \leqslant \min\{m, n\}$) и образуем из элементов на их пересечении квадратную матрицу порядка k:

$$\begin{pmatrix} a_1^1 & \dots & a_{j_1}^1 & \dots & a_{j_2}^1 & \dots & a_{j_k}^1 & \dots & a_n^1 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_1^{i_1} & \dots & a_{j_1}^{i_1} & \dots & a_{j_2}^{i_1} & \dots & a_{j_k}^{i_1} & \dots & a_n^{i_1} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_1^{i_2} & \dots & a_{j_1}^{i_2} & \dots & a_{j_2}^{i_2} & \dots & a_{j_k}^{i_2} & \dots & a_n^{i_2} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_1^{i_k} & \dots & a_{j_1}^{i_k} & \dots & a_{j_2}^{i_k} & \dots & a_{j_k}^{i_k} & \dots & a_n^{i_k} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_1^{m} & \dots & a_{j_1}^{m} & \dots & a_{j_2}^{m} & \dots & a_{j_k}^{m} & \dots & a_n^{m} \end{pmatrix} \Rightarrow \begin{pmatrix} a_{j_1}^{i_1} & \dots & a_{j_2}^{i_1} & \dots & a_{j_k}^{i_1} \\ \vdots & & \vdots & & \vdots \\ a_{j_1}^{i_2} & \dots & a_{j_2}^{i_2} & \dots & a_{j_k}^{i_2} \\ \vdots & & \vdots & & \vdots \\ a_{j_1}^{i_k} & \dots & a_{j_2}^{i_k} & \dots & a_{j_k}^{i_k} \end{pmatrix};$$

будем называть её подматрицей (порядка k) матрицы A, соответствующей выбранным строкам и столбцам (а сами выбранные строки и столбцы — образующими рассматриваемую подматрицу), и обозначать

$$A\begin{pmatrix} i_1\, i_2 \dots i_k \\ j_1\, j_2 \dots j_k \end{pmatrix} = \begin{pmatrix} a_{j_1}^{i_1} & \dots & a_{j_2}^{i_1} & \dots & a_{j_k}^{i_1} \\ \vdots & & \vdots & & \vdots \\ a_{j_1}^{i_2} & \dots & a_{j_2}^{i_2} & \dots & a_{j_k}^{i_2} \\ \vdots & & \vdots & & \vdots \\ a_{j_1}^{i_k} & \dots & a_{j_2}^{i_k} & \dots & a_{j_k}^{i_k} \end{pmatrix}.$$

Определитель этой подматрицы называется минором (порядка k) матрицы A, соответствующим выбранным строкам и столбцам (а сами выбранные строки и столбцы — образующими рассматриваемый минор).

7.11. Определение. Минор порядка r матрицы A называется базисным, если он отличен от нуля, а любой минор большего порядка равен нулю.

Разумеется, у матрицы может быть не один базисный минор. Например,

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

миноры

$$\begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} = -2, \quad \begin{vmatrix} 1 & 4 \\ 2 & 6 \end{vmatrix} = -2, \quad \begin{vmatrix} 3 & 4 \\ 4 & 6 \end{vmatrix} = 2$$

являются базисными.

Будем говорить, что подматрица $\hat{\hat{A}}$ матрицы A *окаймляет* подматрицу $\hat{A}$, если $\hat{A}$ получается из $\hat{\hat{A}}$ вычёркиванием одной крайней строки (первой или последней) и одного крайнего столбца.

7.12. Теорема (теорема о базисном миноре)**.**

1. *Столбцы (строки), образующие базисный минор, линейно независимы.*
2. *Любой столбец (строка) является линейной комбинацией столбцов (строк), образующих базисный минор.*

Таким образом, столбцы (строки), образующие базисный минор, являются базисными столбцами (строками), а порядок базисного минора матрицы равен её рангу.

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 6.18 лекции 11, с. 141).

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:

$$\begin{vmatrix} a_1^1 & \cdots & a_r^1 & a_k^1 \\ \vdots & & \vdots & \vdots \\ a_1^r & \cdots & a_r^r & a_k^r \\ a_1^s & \cdots & a_r^s & a_k^s \end{vmatrix}.$$

Этот минор равен нулю, так как при $s \leqslant r$ он содержит две одинаковые строки, а при $s > r$ является минором бóльшего порядка, нежели базисный минор. Разложив указанный минор по элементам последней строки,

получим
$$a_1^s M_1 + \dots + a_r^s M_r + a_k^s M = 0,$$
где $M_1, \dots, M_r$ — алгебраические дополнения первых r элементов последней строки, не зависящие от номера s этой строки, а M — базисный минор. Отсюда получаем
$$a_k^s = -\frac{M_1}{M} a_1^s - \dots - \frac{M_r}{M} a_r^s,$$
так что
$$A_k = -\frac{M_1}{M} A_1 - \dots - \frac{M_r}{M} A_r,$$
что и требовалось доказать. □

5. Линейные отображения, осуществляемые матрицами

Рассмотрим матрицу $A \in \mathbb{K}^{m\times n}$. Каждому столбцу $X \in \mathbb{K}^n$ мы можем поставить в соответствие столбец $Y \in \mathbb{K}^m$ по формуле $Y = AX$, причём
$$A(\alpha^1 X_1 + \alpha^2 X_2) = \alpha^1 AX_1 + \alpha^2 AX_2.$$
Таким образом, матрица A порождает линейное отображение
$$\boldsymbol{A}\colon\ \mathbb{K}^n \to \mathbb{K}^m, \quad X \mapsto AX. \tag{7.3}$$
В дальнейшем мы будем отождествлять матрицу A и порождаемое ею отображение $\boldsymbol{A}$.

7.13. Определение. *Ядром* матрицы A называется множество всех векторов $X \in \mathbb{K}^n$, которые при отображении (7.4) обращаются в нулевой вектор:
$$\ker A \overset{\text{def}}{=\!=} \Bigl\{X \in \mathbb{K}^n \Bigm| AX = O\Bigr\}. \tag{7.4}$$
Иными словами, $\ker A$ — это множество всех решений однородной системы уравнений $AX = O$ и в силу теоремы 5.13 (см. лекцию 5, с. 103) является подпространством в $\mathbb{K}^n$.

7.14. Определение. *Образом* матрицы A называется множество всех векторов $Y \in \mathbb{K}^m$, которые могут быть получены при отображении (7.4) из некоторого вектора $X \in \mathbb{K}^n$:
$$\operatorname{im} A \overset{\text{def}}{=\!=} \Bigl\{Y \in \mathbb{K}^m \Bigm| \exists X \in \mathbb{K}^n : Y = AX\Bigr\}. \tag{7.5}$$
Иными словами, $\operatorname{im} A$ — это множество значений отображения (7.3).

Согласно утверждению (1) теоремы 4.4 (см. лекцию 4, с. 90) столбец $Y = AX$ представляет собой линейную комбинацию столбцов матрицы A с коэффициентами, равными элементам столбца X. Следовательно, образ матрицы A является не чем иным, как линейной оболочкой её столбцов:
$$\operatorname{im} A = L\bigl(A_1, A_2, \dots, A_n\bigr),$$
и потому подпространством в $\mathbb{K}^m$. Таким образом, согласно определению ранга матрицы,
$$\dim \operatorname{im} A = \operatorname{rk} A.$$

Нахождение базиса в $\operatorname{im} A$ производится приведением матрицы A к упрощённому виду с помощью метода Гаусса—Жордана.

Анализ решения однородной системы $AX = O$, полученного методом Гаусса—Жордана, приводит к следующему результату.

7.15. Теорема. *Пусть $A \in \mathbb{K}^{m\times n}$. Тогда*

$$\dim\ker A + \dim\operatorname{im} A = n.$$

Доказательство. Размерность пространства решений системы $AX = O$, т.е. размерность ядра матрицы A, равно количеству векторов, составляющих ФСР этой системы, т.е. количеству произвольных постоянных в выражении общего решения, которое, в свою очередь, есть разность между количеством всех неизвестных и количеством базисных неизвестных. Поскольку количество базисных неизвестных равно количеству базисных столбцов матрицы A, т.е. её рангу, получаем требуемое утверждение. $\square$

6. Упражнения

ГЛАВА 8

Векторное пространство

Линейные операции над геометрическими векторами (см. гл. III) и над матрицами (см. гл. 4) обладают одинаковыми алгебраическими свойствами, установленными в теоремах 2.12 и 4.1. Это означает, что множество векторов и множество матриц, снабжённые линейными операциями, являются частными случаями некоторого более общего математического объекта — векторного пространства. Понятие векторного пространства образуется в результате абстрагирования от конкретной природы объектов, над которыми возможно выполнять подобные операции. В определении векторного пространства вещественные числа заменяются на произвольное числовое поле, а простейшие свойства операций «сложения векторов» и «умножения вектора на число» постулируются в качестве аксиом без указания конкретного вида этих операций.

Аксиомы векторного пространства составляют первую часть векторно-точечной аксиоматики Вейля, на основе которой в дальнейшем будет систематически изложена геометрия аффинных и евклидовых пространств.

1. Определение векторного пространства

8.1. Определение. *Векторное* (*линейное*) *пространство* над числовым полем $\mathbb{K}$ — это упорядоченная четвёрка $(\mathcal{V}, \mathbb{K}, +, \cdot)$, состоящая из следующих элементов:

(a) множества $\mathcal{V}$, элементы которого называются *векторами* и обозначаются полужирными строчными латинскими буквами $\boldsymbol{a}, \boldsymbol{b}, \ldots, \boldsymbol{x}, \boldsymbol{y}, \ldots$,

(b) поля $\mathbb{K}$, элементы которого называются *скалярами* и обозначаются греческими буквами $\alpha, \beta, \gamma, \ldots$ (мы будем заниматься только случаем числовых полей $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$,),

(c) бинарной операции на $\mathcal{V}$

$$+\colon\ \mathcal{V} \times \mathcal{V} \to \mathcal{V}, \quad (\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} + \boldsymbol{y},$$

называемой *сложением векторов*,

(d) бинарной операции

$$\cdot\colon\ \mathbb{K} \times \mathcal{V} \to \mathcal{V}, \quad (\alpha, \boldsymbol{x}) \mapsto \alpha \cdot \boldsymbol{x}$$

называемой *умножением вектора на скаляр.*

Для элементов четвёрки $(\mathcal{V}, \mathbb{K}, +, \cdot)$ должен быть выполнен следующий комплекс условий — аксиом векторного пространства:

V1. коммутативность сложения:

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}:\ \boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x};$$

V2. ассоциативность сложения:

$$\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{V}:\ (\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c} = \boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c});$$

V3. существование нулевого вектора:

$$\exists \boldsymbol{0} \in \mathcal{V}\ \forall x \in \mathcal{V}:\ \boldsymbol{x} + \boldsymbol{0} = \boldsymbol{x};$$

V4. существование противоположного вектора:

$$\forall \boldsymbol{x} \in \mathcal{V}\ \exists -\boldsymbol{x} \in \mathcal{V}:\ -\boldsymbol{x} + \boldsymbol{x} = \boldsymbol{0};$$

V5. свойство единицы поля:

$$\forall \boldsymbol{x} \in \mathcal{V}:\ 1 \cdot \boldsymbol{x} = \boldsymbol{x};$$

V6. ассоциативность умножения на число[1]:

$$\forall \boldsymbol{x} \in \mathcal{V},\ \forall \alpha, \beta \in \mathbb{K}:\ \bigl(\alpha \cdot \beta\bigr) \cdot \boldsymbol{x} = \alpha \cdot \bigl(\beta \cdot \boldsymbol{x}\bigr);$$

V7. дистрибутивность относительно сложения векторов:

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V},\ \forall \alpha \in \mathbb{K}:\ \alpha\bigl(\boldsymbol{x} + \boldsymbol{y}\bigr) = \alpha\boldsymbol{x} + \alpha\boldsymbol{y};$$

V8. дистрибутивность относительно сложения чисел[2]:

$$\forall \boldsymbol{x} \in \mathcal{V},\ \forall \alpha, \beta \in \mathbb{K}:\ \bigl(\alpha + \beta\bigr)\boldsymbol{x} = \alpha\boldsymbol{a} + \beta\boldsymbol{a}.$$

В большинстве случаев не вызывает недоразумений обозначение векторного пространства $(\mathcal{V}, \mathbb{K}, +, \cdot)$ одним лишь символом $\mathcal{V}$; в тех случаях, когда требуется подчеркнуть тот факт, что $\mathcal{V}$ является векторным пространством над полем $\mathbb{K}$, будем писать $\mathcal{V}(\mathbb{K})$.

8.2. Замечание. Очевидно, аксиомы **V1–V4** означают, что пара $(\mathcal{V}, +)$ (т.е. множество $\mathcal{V}$ с рассматриваемой на нём операцией сложения) является группой, в которой групповая операция коммутативна. Такие группы называются *коммутативными* или *абелевыми*.

8.3. Замечание. Обозначать нулевой элемент и операцию сложения в поле $\mathbb{K}$ и в множестве векторов $\mathcal{V}$ одинаковыми символами не вполне последовательно, но удобно и в большинстве случаев не приводит к недоразумениям.

8.4. Замечание. Многие общие понятия и теоремы многомерной алгебры и геометрии можно проиллюстрировать схематическими чертежами. Однако нужно помнить, что окружающее нас физическое пространство трёхмерно, и чертежи, изображающие двумерные или трёхмерные объекты, не всегда могут адекватно передавать информацию о многомерных объектах. Кроме того, физическое пространство является *евклидовым*,

---

[1] Обратите внимание, что $\cdot$ имеет разный смысл в разных частях этой формулы: в скобках в левой части $\cdot$ обозначает умножение элементов поля $\mathbb{K}$, а во всех остальных случаях — умножение вектора на скаляр.

[2] Здесь также знак $+$ в левой части равенства обозначает сложение скаляров — элементов поля $\mathbb{K}$, а в правой части — сложение векторов — элементов множества $\mathcal{V}$.

т.е. в нём определены не только сложение векторов и умножение на скаляры, но также длины векторов, углы между ними, площади и объёмы некоторых фигур и т. п. Наши чертежи несут принудительную информацию об этих «метрических» свойствах и мы машинально её воспринимаем, хотя в аксиомах векторного пространства эти свойства отсутствуют. До введения в векторном пространстве дополнительной структуры скалярного произведения, позволяющей осуществлять измерения длин и углов (см. лекцию **??**), нельзя представлять себе, что один вектор длиннее другого или что два вектора образуют прямой угол.

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 **V1**–**V8** векторного пространства составляют *первую группу аксиом Вейля.*

Из аксиом векторного пространства выводятся простейшие следствия. Обратите внимание, что при их доказательстве используется не конкретная специфика линейных операций, а лишь их свойства, зафиксированные в аксиомах.

8.5. 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 любого вектора* $\boldsymbol{x} \in \mathcal{V}$ *имеем* $0 \cdot \boldsymbol{x} = \mathbf{0}$.
4. *Для любого вектора* $\boldsymbol{x} \in \mathcal{V}$ *его противоположный* $-\boldsymbol{x}$ *выражается формулой* $-\boldsymbol{x} = (-1) \cdot \boldsymbol{x}$.
5. *Для любого вектора* $\boldsymbol{x} \in \mathcal{V}$ *и любого скаляра* $\alpha \in \mathbb{K}$ *имеют место соотношения* $-(\alpha \cdot \boldsymbol{x}) = (-\alpha) \cdot \boldsymbol{x} = \alpha \cdot (-\boldsymbol{x})$.
6. *Для любого скаляра* $\alpha \in \mathbb{K}$ *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 $\boldsymbol{x}$, т.е.

$$\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x} = \mathbf{0}, \quad \boldsymbol{x} + \boldsymbol{y}' = \boldsymbol{y}' + \boldsymbol{x} = \mathbf{0}.$$

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. Их цепочки равенств

$$\boldsymbol{x} + 0 \cdot \boldsymbol{x} \overset{\mathbf{V5}}{=\!=} 1 \cdot \boldsymbol{x} + 0 \cdot \boldsymbol{x} \overset{\mathbf{V8}}{=\!=} (1 + 0) \cdot \boldsymbol{x} = 1 \cdot \boldsymbol{x} \overset{\mathbf{V5}}{=\!=} \boldsymbol{x}$$

следует, что $0 \cdot \boldsymbol{x} = \mathbf{0}$.

4. Аналогично предыдущему, имеем

$$(-1) \cdot \boldsymbol{x} + \boldsymbol{x} \overset{\mathbf{V5}}{=\!=} (-1) \cdot \boldsymbol{x} + 1 \cdot \boldsymbol{x} \overset{\mathbf{V8}}{=\!=} (-1 + 1) \cdot \boldsymbol{x} = 0 \cdot \boldsymbol{x} \overset{\text{п. 3}}{=\!=} \mathbf{0},$$

т.е. $(-1) \cdot \boldsymbol{x}$ — вектор, противоположный $\boldsymbol{x}$.

5. Имеем

$$-(\alpha \boldsymbol{x}) \overset{\text{п. 4}}{=\!=} (-1)\cdot(\alpha\boldsymbol{x}) \overset{\mathbf{V6}}{=\!=} (-1\cdot\alpha)\cdot\boldsymbol{x} = (-\alpha)\cdot\boldsymbol{x},$$

$$\begin{aligned}-(\alpha \boldsymbol{x}) &\overset{\text{п. 4}}{=\!=} (-1)\cdot(\alpha\boldsymbol{x}) \overset{\mathbf{V6}}{=\!=} (-1\cdot\alpha)\cdot\boldsymbol{x} = \\ &= \big(\alpha\cdot(-1)\big)\cdot\boldsymbol{x} \overset{\mathbf{V6}}{=\!=} \alpha\cdot\big((-1)\cdot\boldsymbol{x}\big) \overset{\text{п. 4}}{=\!=} \alpha\cdot(-\boldsymbol{x}).\end{aligned}$$

6. Имеем

$$\alpha\cdot\boldsymbol{x} + \alpha\cdot\mathbf{0} \overset{\mathbf{V7}}{=\!=} \alpha\cdot(\boldsymbol{x}+\mathbf{0}) \overset{\mathbf{V3}}{=\!=} \alpha\cdot\boldsymbol{x},$$

т.е. $\alpha\cdot\mathbf{0} = \mathbf{0}$. □

2. Базис и размерность

А. Линейная комбинация, линейная оболочка, линейная зависимость и независимость. Понятия линейной комбинации, линейной оболочки, линейно зависимого и линейно независимого семейства векторов произвольного векторного пространства вводятся совершенно так же, как и для пространства $\mathbb{K}^n$. Повторим соответствующие определения и факты.

8.6. Определение. Пусть дано семейство (т.е. упорядоченный набор, см. определение 5.1, с. 97) векторов $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ векторного пространства $\mathcal{V}$ и семейство чисел $\alpha^1, \alpha^2, \ldots, \alpha^k \in \mathbb{K}$. *Линейной комбинацией* векторов $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ с коэффициентами $\alpha^1, \alpha^2, \ldots, \alpha^k$ называется выражение вида

$$\alpha^1\boldsymbol{x}_1 + \alpha^2\boldsymbol{x}_2 + \cdots + \alpha^k\boldsymbol{x}_k \equiv \sum_{s=1}^{k}\alpha^s\boldsymbol{x}_s. \tag{8.1}$$

Линейная комбинация называется *тривиальной*, если все её коэффициенты равны нулю:

$$\alpha^1 = \alpha^2 = \cdots = \alpha^k = 0.$$

Ясно, что тривиальная линейная комбинация произвольных векторов равна нулевому вектору.

8.7. Определение. *Линейной оболочкой* семейства векторов $X_1, X_2, \ldots, X_k$ называется множество всех линейных комбинаций этих векторов:

$$L(X_1, X_2, \ldots, X_k) \overset{\text{def}}{=\!=} \Big\{\alpha^1X_1 + \alpha^2X_2 + \cdots + \alpha^kX_k \;\Big|\; \alpha^1, \alpha^2, \ldots, \alpha^k \in \mathbb{K}\Big\}.$$

8.8. Определение. Семейство векторов $X_1, \ldots, X_k$ называется *линейно зависимым*, если существует нетривиальная линейная комбинация этих векторов, равная нулевому вектору.

Семейство векторов $X_1, \ldots, X_k$ называется *линейно независимым*, если обращение линейной комбинации этих векторов в нулевой вектор возможно лишь в том случае, если эта линейная комбинация тривиальна.

Используя терминологию наглядной геометрии, два линейно зависимых вектора часто называют *коллинеарными*, а три — *компланарными*.

8.9. Теорема (свойства линейно зависимых и линейно независимых векторов)**.**

1. *Любое семейство с повторениями линейно зависимо.*
2. *Если в семействе векторов* $\boldsymbol{x}_1, \ldots,_k$ *имеется нулевой вектор* $\mathbf{0}$*, то это семейство линейно зависимо.*
3. *Семейство векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ *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* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_r$ *имеется линейно зависимое подсемейство* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$*, то и всё семейство также линейно зависимо.*
5. *Любая часть линейно независимого семейства векторов является линейно независимым подсемейством.*
6. *Если семейство векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ *линейно независимо, а семейство векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{x}$ *линейно зависимо, то вектор* $\boldsymbol{x}$ *является линейной комбинацией векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$*.*
7. *Если семейство векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ *линейно независимо, а вектор* $\boldsymbol{x}$ *нельзя через них выразить, то семейство векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{x}$ *линейно независимо.*

Доказательство этой теоремы дословно повторяет доказательство теоремы 5.6 из лекции 5 (см. с. 99).

8.10. Теорема (см. теорему 7.3, с. 151)**.** *Если векторы* $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s$ *принадлежат линейной оболочке векторов* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ *и* $s > r$*, то семейство векторов* $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s$ *линейно зависимо.*

B. Базис.

8.11. Определение. Линейно независимое семейство векторов $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ векторного пространства $\mathcal{V}(\mathbb{K})$ называется *базисом*, если любой вектор $\boldsymbol{x} \in \mathcal{V}$ может быть представлен в виде линейной комбинации

$$\boldsymbol{x} = x^1\boldsymbol{e}_1 + \cdots + x^n\boldsymbol{e}_n, \quad x_1, \ldots, x_n \in \mathbb{K}, \tag{8.2}$$

которая называется *разложением вектора* $\boldsymbol{x}$ *по базису* $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. Коэффициенты $x^1, \ldots, x^n$ указанного разложения называются координатами вектора $\boldsymbol{x}$ в базисе $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.

8.12. Определение. Векторное пространство $\mathcal{V}$ называется *конечномерным*, если в нём существует базис, состоящий из конечного числа векторов.

8.13. Замечание. Можно рассматривать бесконечные семейства векторов и бесконечные линейные комбинации, однако в этом случае необходимо позаботиться о том, чтобы линейная комбинация, состоящая из бесконечного числа слагаемых, имела смысл. Сделать это можно лишь используя средства анализа. Раздел математики, в котором исследуются подобные вопросы, называется функциональным анализом. В этом семестре нас будут интересовать только конечномерные векторные пространства.

8.14. Замечание. Обратите внимание, что базис — это именно *семейство* (т.е. *упорядоченный набор*), а не множество векторов. Однако свойства полноты и линейной независимости сохранятся, если векторы семейства произвольным образом перенумеровать (переставить). Перенумеровав векторы базиса, мы снова получим базис, но другой.

Пусть $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ — некоторый базис в конечно мерном векторном пространстве $\mathcal{V}(\mathbb{K})$. Условимся векторы базиса записывать в виде *матрицы-строки* $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$, а координаты $x^1, \ldots, x^n \in \mathbb{K}$ вектора $\boldsymbol{x}$ относительно базиса — в виде столбца $X = \big(x^1, \ldots, x^n\big)^T$. Тогда разложение (8.2) запишется в виде матричного равенства

$$\boldsymbol{x} = \boldsymbol{E}X. \tag{8.3}$$

8.15. Теорема (см. теорему 5.11)**.** *Разложение вектора по базису единственно.*

C. Правило суммирования Эйнштейна. Векторы базиса нумеруются нижними индексами, а координаты вектора верхними именно для того, чтобы для записи сумм однотипных слагаемых вида (8.2) можно было использовать компактные матричные обозначения типа (8.3). Легко видеть, что выбранный способ нумерации полностью согласуется с правилом нумерации элементов матриц: верхний индекс нумерует строку, а нижний — столбец.

Кроме того, в алгебре используется и другой способ записи сумм вида (8.2), 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 считается, что имеет место *сумма* подобных однотипных выражений, в которой повторяющийся индекс принимает некоторый (заранее оговорённый) диапазон значений, хотя знак суммирования $\sum$ отсутствует.

Например, выражение $a_k b^k$ означает сумму $a_1 b^1 + a_2 b^2 + \cdots + a_n b^n$ (из контекста должно быть ясно или в явном виде оговорено, 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:

$$(a_k + b_k)c^k = (a_1 + b_1)c^1 + (a_2 + b_2)c^2 + \cdots + (a_n + b_n)c^n =$$

$$= a_1c^1 + a_2c^2 + \dots + a_nc^n + b_1c^1 + b_2c^2 + \dots + b_nc^n =$$
$$= a_kc^k + b_kc^k.$$

Читатель быстро привыкнет к этим обозначениям и оценит их компактность и удобство; сомнительные же и трудные случаи мы всегда будем особо оговаривать.

Напомним (см. формулу 4.1 в лекции 4, с. 88), что *символ Кронекера* — это обозначение для элементов единичной матрицы:

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

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 к замене индекса.*

D. Размерность.

8.16. Теорема. *Все базисы конечномерного векторного пространства $\mathcal{V}$ состоит из одинакового числа векторов. Это число называется размерностью пространства $\mathcal{V}$ и обозначается $\dim \mathcal{V}$; пространство $\mathcal{V}$ при этом называется n-мерным.*

Доказательство. Предположим, что в пространстве $\mathcal{V}$ имеется два базиса $\boldsymbol{E} = \big(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\big)$ и $\boldsymbol{F} = \big(\boldsymbol{f}_1, \dots, \boldsymbol{f}_m\big)$, причём $m > n$. Поскольку каждый из векторов $\boldsymbol{f}_1, \dots, \boldsymbol{f}_m$ лежит в линейной оболочке $L(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n) = \mathcal{V}$, согласно теореме 8.10 они линейно зависимы, противоречие. □

Наиболее распространённым методом нахождения размерности векторного пространства является подсчёт количества векторов в каком-либо базисе этого пространства или, что по существу то же самое, подсчёт количества независимых величин, через которые может быть однозначно выражен любой вектор рассматриваемого пространства: эти величины служат координатами вектора в некотором подходящем базисе. Позже будут разработаны и другие методы вычисления размерностей. Это очень важно, поскольку многие числовые инварианты в математике определяются именно как размерности (например, ранг матрицы); базисы же соответствующих пространств могут оказаться трудно вычислимыми или не имеющими особого смысла.

8.17. Замечание. Очевидно, что размерность n векторного пространства $\mathcal{V}$ — это натуральное число, обладающее следующим свойством: в $\mathcal{V}$ существуют линейно независимые семейства векторов, состоящие из n членов, но любое семейство, состоящее из $n + 1$ членов, является линейно зависимым.

E. Примеры.

8.18. Пример. Арифметическое пространство $\mathbb{K}^n$ является векторным пространством над полем $\mathbb{K}$; это пространство было детально изучено в предыдущих лекциях. Очевидно, вектор-столбцы

$$\boldsymbol{e}_1 = \bigl(1, 0, \ldots, 0\bigr)^T, \quad \boldsymbol{e}_2 = \bigl(0, 1, \ldots, 0\bigr)^T, \quad \ldots, \; \boldsymbol{e}_n = \bigl(0, 0, \ldots, 1\bigr)^T$$

образуют базис в $\mathbb{K}^n$ (этот базис называется стандартным); компоненты x^k произвольного вектор-столбца $\boldsymbol{x} = (x^1, x^2, \ldots, x^n)^T$ являются его координатами в стандартном базисе.

8.19. Пример. Пусть $\mathbb{K}_1$ — подполе поля $\mathbb{K}$. Тогда арифметическое пространство $\mathbb{K}^n$ является также и векторным пространством над $\mathbb{K}_1$. Наиболее интересным примером такого типа является пространство $\mathbb{C}^n(\mathbb{R})$, состоящее из комплекснозначных столбцов высоты n, в котором разрешено умножение столбца на вещественный числа. Стандартный базис в пространстве $\mathbb{C}^n(\mathbb{R})$ состоит из вектор-столбцов

$$\begin{aligned} \boldsymbol{e}_1 &= \bigl(1, 0, \ldots, 0\bigr)^T, & \boldsymbol{e}_2 &= \bigl(0, 1, \ldots, 0\bigr)^T, & &\ldots, & \boldsymbol{e}_n &= \bigl(0, 0, \ldots, 1\bigr)^T, \\ \boldsymbol{e}_{n+1} &= \bigl(i, 0, \ldots, 0\bigr)^T, & \boldsymbol{e}_{n+2} &= \bigl(0, i, \ldots, 0\bigr)^T, & &\ldots, & \boldsymbol{e}_{2n} &= \bigl(0, 0, \ldots, i\bigr)^T. \end{aligned}$$

Действительно, любой вектор-столбец с комплексными компонентами может быть представлен в виде следующей линейной комбинации столбцов $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{2n}$ с вещественными коэффициентами:

$$\begin{pmatrix} x^1 + iy^1 \\ x^2 + iy^2 \\ \vdots \\ x^n + iy^n \end{pmatrix} = x^1\boldsymbol{e}_1 + x^2\boldsymbol{e}_2 + \cdots + x^n\boldsymbol{e}_n + y^1\boldsymbol{e}_{n+1} + y^2\boldsymbol{e}_{n+2} + \cdots + y^n\boldsymbol{e}_{2n}.$$

Отсюда вытекает, что $\dim \mathbb{C}^n(\mathbb{R}) = 2n$.

8.20. Пример. Множество $\mathbb{K}^{m\times n}$ всех матриц размера $m \times n$ с элементами из поля $\mathbb{K}$ является векторным пространством над полем $\mathbb{K}$. *Стандартным базисом* этого пространства называют семейство матриц

$$\begin{array}{c@{}c} & \begin{array}{ccccc} & & \overset{k}{\downarrow} & & \end{array} \\ \begin{array}{c} \\ \\ j\rightarrow \\ \\ \\ \end{array} & \begin{pmatrix} 0 & \ldots & 0 & \ldots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \ldots & 1 & \ldots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \ldots & 0 & \ldots & 0 \end{pmatrix} \end{array} = \boldsymbol{e}_{jk}, \qquad \begin{array}{l} j = 1, \ldots, m, \\ k = 1, \ldots, n. \end{array}$$

Матрица $\boldsymbol{e}_{jk}$ имеет единственный ненулевой элемент, равный 1, расположенный на пересечении j-й строки и k-го столбца. Очевидно, координатами произвольной матрицы в этом базисе являются её элементы и $\dim \mathbb{K}^{m\times n} = mn$.

8.21. Пример. Множества направленных отрезков на прямой, на плоскости и в пространстве, обозначаемые соответственно V_1, V_2 и V_3, являются векторными пространствами над полем $\mathbb{R}$ размерности 1, 2 и 3 соответственно. Сложение направленных отрезков осуществляется по правилу параллелограмма (или, что эквивалентно, по правилу треугольника), умножение направленного отрезка на число α растягивает его в $|\alpha|$ раз и при $\alpha < 0$ изменяет направление на противоположное. Базисные векторы обозначают обычно символами $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$; в школьном курсе геометрии предполагалось, что эти векторы имеют единичную длину и взаимно перпендикулярны.

8.22. Пример. Множество $\mathbb{K}[t]_n$ всех многочленов степени $\leqslant n$ от переменной t, коэффициенты которых принадлежат полю $\mathbb{K}$, является векторным пространством над полем $\mathbb{K}$. Каждый такой многочлен имеет вид

$$f(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n, \quad a_k \in \mathbb{K}, k = 0, 1, \dots, n.$$

Коэффициенты многочлена можно рассматривать как координаты соответствующего вектора в *стандартном базисе*

$$\boldsymbol{e}_0 = 1, \quad \boldsymbol{e}_1 = t, \quad \boldsymbol{e}_2 = t^2, \quad \dots, \quad \boldsymbol{e}_n = t^n$$

(здесь удобно начинать нумерацию векторов базиса с нуля: тогда номер базисного вектора совпадает со степенью соответствующего многочлена) и $\dim \mathbb{K}[t]_n = n + 1$.

8.23. Пример. Множество $T(n)$ функций вида

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{n} a_k \cos kt + b_k \sin kt, \quad a_0, a_k, b_k \in \mathbb{R},\ k = 1, \dots, n,$$

называемых *тригонометрическими многочленами* порядка не выше n, является векторным пространством над полем $\mathbb{R}$. *Стандартным базисом* этого пространства называется семейство функций

$$\boldsymbol{e}_0 = \frac{1}{2}, \quad \begin{matrix} \boldsymbol{e}_1 = \cos t, & \boldsymbol{e}_2 = \cos 2t, & \dots, & \boldsymbol{e}_n = \cos nt, \\ \boldsymbol{e}_{-1} = \sin t, & \boldsymbol{e}_{-2} = \sin 2t, & \dots, & \boldsymbol{e}_{-n} = \sin nt; \end{matrix}$$

координатами вектора $\boldsymbol{x} = f(t)$ пространства $T(n)$ являются коэффициенты тригонометрического многочлена $f(t)$. Нумерация векторов базиса здесь, как и в предыдущем примере, начинается не с единицы, что, впрочем, оправдывается соображениями удобства. Ясно, что $\dim T(n) = 2n + 1$.

8.24. Пример. Аналогичное векторное пространство над полем $\mathbb{C}$ состоит из функций вида

$$f(t) = \sum_{k=-n}^{n} c_k e^{ikt}, \quad \text{где } i^2 = -1,$$

которые также называют тригонометрическими многочленами. Обозначим это пространство $T(\mathbb{C}; n)$. Стандартный базис этого пространства состоит из функций

$$\boldsymbol{e}_k = e^{ikt}, \quad k = -n, -n+1, \ldots, -1, 0, 1, \ldots, n-1, n;$$

очевидно, $\dim T(\mathbb{C}; n) = 2n+1$.

8.25. Пример. Наконец, рассмотрим необычный пример. Рассмотрим векторное пространство $\mathcal{V} = (\mathbb{R}_+, \mathbb{R}, \boxplus, \boxdot)$, определённое следующим образом:

(a) векторами являются положительные вещественные числа; положительное вещественное число x, рассматриваемое как вектор нашего пространства, будем обозначать $\boldsymbol{x}$;
(b) скалярами являются все вещественные числа,
(c) сложение векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ определяется как умножение соответствующих положительных чисел:

$$\boldsymbol{x} \boxplus \boldsymbol{y} \overset{\text{def}}{=\!=} x \cdot y;$$

(d) умножение вектора $\boldsymbol{x}$ на скаляр α определяется как возведение положительного числа x в степень α:

$$\alpha \boxdot \boldsymbol{x} \overset{\text{def}}{=\!=} x^{\alpha}.$$

Читателю предлагается самостоятельно проверить выполнение аксиом **V1**–**V8**, а также указать какой-либо базис этого пространства и найти его размерность.

3. Подпространства

A. Определение и основные свойства. Пусть $(\mathcal{V}, \mathbb{K}, +, \cdot)$ — векторное пространство над полем $\mathbb{K}$.

8.26. Определение. Множество $\mathcal{P}$ векторов пространства $\mathcal{V}$ называется *подпространством*, если оно замкнуто относительно операций $+$ и $\cdot$, т.е.

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{P},\ \forall \alpha, \beta \in \mathbb{K}: \quad \alpha\boldsymbol{x} + \beta\boldsymbol{y} \in \mathcal{P}.$$

Ясно, что множества $\{\mathbf{0}\}$ и $\mathcal{V}$ являются подпространствами в $\mathcal{V}$; они называются *тривиальными* подпространствами.

Тот факт, что $\mathcal{P}$ является подпространством в $\mathcal{V}$, будем записывать в виде $\mathcal{P} \Subset \mathcal{V}$, оставляя символ $\subset$ для обозначения подмножества.

Любая линейная оболочка является подпространством; доказывается это точно так же, как и в случае арифметического пространства $\mathbb{K}^n$ (см. теорему 5.8, с. 101).

Очевидно, если $\mathcal{P} \Subset \mathcal{V}$, то $(\mathcal{P}, \mathbb{K}, +, \cdot)$ само является векторным пространством, поэтому можно говорить о базисах и размерности $\mathcal{P}$.

B. Пополнение базиса.

8.27. Теорема (теорема о пополнении базиса)**.** *Пусть $\mathcal{P}$ — подпространство в векторном пространстве $\mathcal{V}$, причём $\dim\mathcal{P} = p < \dim\mathcal{V} = n$, $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p$ — базис в $\mathcal{P}$. Тогда найдутся такие векторы $\boldsymbol{e}_{p+1}, \ldots, \boldsymbol{e}_n \in \mathcal{V}\backslash\mathcal{P}$, что семейство векторов*

$$\boldsymbol{e}_1,\ \ldots,\ \boldsymbol{e}_p,\ \boldsymbol{e}_{p+1},\ \ldots,\ \boldsymbol{e}_n$$

является базисом в $\mathcal{V}$.

Доказательство. Так как $p < n$, то найдётся такой вектор $\boldsymbol{e}_{p+1} \in \mathcal{V}$, что векторы $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p, \boldsymbol{e}_{p+1}$ линейно независимы; при этом $\boldsymbol{e}_{p+1} \notin \mathcal{P}$, так как в противном случае получили бы $\dim \mathcal{P} > p$.

Если $p + 1 = n$, пополнение базиса завершено. Если $p + 1 < n$, продолжаем процесс. Процесс завершается через конечное число шагов, поскольку пространство $\mathcal{V}$ конечномерно. $\square$

C. Пересечение и сумма подпространств. Пусть $\mathcal{P}$ и $\mathcal{Q}$ — два подпространства в векторном пространстве $\mathcal{V}$.

8.28. Предложение. *Если $\mathcal{P} \Subset \mathcal{V}$ и $\mathcal{Q} \Subset \mathcal{V}$, то $\mathcal{P} \cap \mathcal{Q} \Subset \mathcal{V}$, т.е. пересечение двух подпространств векторного пространства также является подпространством.*

Доказательство состоит в проверке требований определения 8.26: нужно установить, что

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{P} \cap \mathcal{Q},\ \forall \alpha, \beta \in \mathbb{K} : \alpha\boldsymbol{x} + \beta\boldsymbol{y} \in \mathcal{P} \cap \mathcal{Q}.$$

Имеем:

$$\boldsymbol{x}, \boldsymbol{y} \in \mathcal{P} \cap \mathcal{Q} \iff \begin{cases} \boldsymbol{x}, \boldsymbol{y} \in \mathcal{P}, \\ \boldsymbol{x}, \boldsymbol{y} \in \mathcal{Q} \end{cases} \iff$$

$$\iff \forall \alpha, \beta \in \mathbb{K} : \begin{cases} \alpha\boldsymbol{x} + \beta\boldsymbol{y} \in \mathcal{P}, \\ \alpha\boldsymbol{x} + \beta\boldsymbol{y} \in \mathcal{Q} \end{cases} \iff \alpha\boldsymbol{x} + \beta\boldsymbol{y} \in \mathcal{P} \cap \mathcal{Q},$$

что и требовалось доказать. $\square$

Объединение подпространств, в отличие от пересечения, подпространством может не являться. Читателю рекомендуется убедиться в этом самостоятельно, рассмотрев подпространства $\mathcal{P} = L((1, 0)^T) \Subset \mathbb{R}^2$ и $\mathcal{Q} = L((0, 1)^T) \Subset \mathbb{R}^2$.

8.29. Определение. *Суммой* подпространств $\mathcal{P}, \mathcal{Q} \Subset \mathcal{V}$ называется линейная оболочка их объединения, т.е. множество

$$\mathcal{P} + \mathcal{Q} = \Big\{ \boldsymbol{x} + \boldsymbol{y} \ \Big| \ \boldsymbol{x} \in \mathcal{P},\ \boldsymbol{y} \in \mathcal{Q} \Big\}.$$

Очевидно, будучи линейной оболочкой, сумма подпространств также является подпространством.

8.30. Теорема. *Если $\mathcal{P}, \mathcal{Q} \Subset \mathcal{V}$, то*

$$\dim(\mathcal{P} + \mathcal{Q}) = \dim \mathcal{P} + \dim \mathcal{Q} - \dim(\mathcal{P} \cap \mathcal{Q}). \tag{8.4}$$

Доказательство. Введём обозначения

$$\dim \mathcal{P} = p, \quad \dim \mathcal{Q} = q, \quad \dim(\mathcal{P} \cap \mathcal{Q}) = r.$$

Пусть векторы $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_r$ образуют базис в $\mathcal{P} \cap \mathcal{Q}$. Пользуясь теоремой 8.27, достроим базис в $\mathcal{P} \cap \mathcal{Q}$ до базисов в $\mathcal{P}$ и $\mathcal{Q}$ следующим образом:

$$\underbrace{\overbrace{\boldsymbol{e}_1,\ \ldots,\ \boldsymbol{e}_r}^{r\ \text{векторов}},\ \overbrace{\boldsymbol{f}_1,\ \ldots,\ \boldsymbol{f}_{p-r}}^{p-r\ \text{векторов}}}_{p\ \text{векторов}} \text{ — базис в } \mathcal{P},$$

$$\underbrace{\overbrace{\boldsymbol{e}_1,\ \ldots,\ \boldsymbol{e}_r}^{r\ \text{векторов}},\ \overbrace{\boldsymbol{g}_1,\ \ldots,\ \boldsymbol{g}_{q-r}}^{q-r\ \text{векторов}}}_{q\ \text{векторов}} \text{ — базис в } \mathcal{Q}.$$

Семейство векторов

$$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_r,\ \boldsymbol{f}_1, \ldots, \boldsymbol{f}_{p-r},\ \boldsymbol{g}_1, \ldots, \boldsymbol{g}_{q-r}$$

линейно независимо, поскольку каждый из них не может быть линейно выражен через предыдущие (см. утверждение 7 теоремы 8.9), а их количество равно $r + (p-r) + (q-r) = p + q - r$; это число и есть размерность подпространства $\mathcal{P} + \mathcal{Q}$. □

D. Примеры. Для каждого из приведённых ниже примеров подпространств самостоятельно найдите какой-либо базис и размерность.

8.31. Пример. Для пространств $\mathcal{V}_1$, $\mathcal{V}_2$, $\mathcal{V}_3$ геометрических векторов на прямой, на плоскости и в пространстве, имеем $\mathcal{V}_1 \Subset \mathcal{V}_2 \Subset \mathcal{V}_3$.

8.32. Пример. $\mathbb{R}^n(\mathbb{R}) \Subset \mathbb{C}^n(\mathbb{R})$.

8.33. Пример. Подмножество в $\mathbb{K}^n(\mathbb{K})$, состоящее из столбцов, сумма элементов которых равна нулю, является подпространством в $\mathbb{K}^n(\mathbb{K})$.

8.34. Пример. В арифметическом пространстве $\mathbb{K}^n$ для любого $m \leqslant n$ множество всех вектор-столбцов вида $\left(x^1, \ldots, x^m, 0, \ldots, 0\right)^T$ образует подпространство.

8.35. Пример. Множество решений любой однородной системы, содержащей n неизвестных, является подпространством в арифметическом пространстве $\mathbb{K}^n$ (см. теорему 5.13, с. 103).

8.36. Пример. В векторном пространстве $\mathbb{K}^{n\times n}(\mathbb{K})$ квадратных матриц порядка n подпространствами являются следующие подмножества:

(a) подмножество симметричных матриц

$$S\mathbb{K}^{n\times n} = \left\{A \in \mathbb{K}^{n\times n} \;\middle|\; A^T = A\right\}$$

(как обычно, ^{T} означает транспонирование);

(b) подмножество кососимметричных матриц

$$A\mathbb{K}^{n\times n} = \left\{A \in \mathbb{K}^{n\times n} \;\middle|\; A^T = -A\right\}.$$

8.37. Пример. Подмножество, состоящее из матриц с нулевым следом:

$$\mathcal{P} = \left\{ A \in \mathbb{K}^{n\times n} \;\middle|\; \operatorname{tr} A = 0 \right\}.$$

(Напомним, что следом $\operatorname{tr} A$ квадратной матрицы A называется сумма её диагональных элементов.)

8.38. Пример. В векторном пространстве $\mathbb{K}[t]_n$ многочленов степени не выше n подпространствами являются множества

$$S\mathbb{K}[t]_n = \left\{ f(t) \in \mathbb{K}[t]_n \;\middle|\; f(-t) = f(t) \right\},$$

$$A\mathbb{K}[t]_n = \left\{ f(t) \in \mathbb{K}[t]_n \;\middle|\; f(-t) = -f(t) \right\},$$

состоящие из чётных и нечётных функций.

4. Изоморфизмы векторных пространств

Пусть $(\mathcal{V}, \mathbb{K}, +, \cdot)$ и $(\mathcal{W}, \mathbb{K}, \boxplus, \boxdot)$ — два векторных пространства над одним и тем же полем $\mathbb{K}$. Нулевые векторы этих пространств обозначим через $\mathbf{0}_{\mathcal{V}}$ и $\mathbf{0}_{\mathcal{W}}$ соответственно.

8.39. Определение. Взаимно однозначное отображение $\boldsymbol{A} : \mathcal{V} \to \mathcal{W}$ называется *изоморфизмом*, если оно обладает *свойством линейности*:

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V},\ \forall \alpha, \beta \in \mathbb{K}: \quad \boldsymbol{A}(\alpha \cdot \boldsymbol{x} + \beta \cdot \boldsymbol{y}) = \alpha \boxdot \boldsymbol{A}(\boldsymbol{x}) \boxplus \beta \boxdot \boldsymbol{A}(\boldsymbol{y}).$$

8.40. Теорема. *Пусть векторные пространства $\mathcal{V}$ и $\mathcal{W}$ изоморфны и $f : \mathcal{V} \to \mathcal{W}$ — изоморфизм.*

1. *Любой ненулевой вектор $\boldsymbol{x} \in \mathcal{V}$ при отображении f переходит в ненулевой вектор*:
$$\forall \boldsymbol{x} \in \mathcal{V},\ \boldsymbol{x} \neq \mathbf{0}_{\mathcal{V}}: \quad f(\boldsymbol{x}) \neq \mathbf{0}_{\mathcal{W}}.$$
2. *Если $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p \in \mathcal{V}$ — линейно независимые векторы, то векторы $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_p) \in \mathcal{W}$ также линейно независимы.*
3. *Если $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p \in \mathcal{V}$ — линейно зависимые векторы, причём нетривиальная линейная комбинация этих векторов, равная нулевому вектору $\mathbf{0}_{\mathcal{V}}$, имеет коэффициенты $\alpha^1, \ldots, \alpha^p$, то векторы $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_p) \in \mathcal{W}$ также линейно зависимы, причём нетривиальная линейная комбинация этих векторов, равная нулевому вектору $\mathbf{0}_{\mathcal{W}}$, имеет те же коэффициенты.*

Доказательство. 1. Пусть $\boldsymbol{x} \in \mathcal{V}$, $\boldsymbol{x} \neq \mathbf{0}_{\mathcal{V}}$. Предположим, что $f(\boldsymbol{x}) = \mathbf{0}_{\mathcal{W}}$. Тогда

$$f(\boldsymbol{x}) = \mathbf{0}_{\mathcal{W}} = 0 \boxdot \boldsymbol{y} = 0 \boxdot f(\boldsymbol{z}) = f(0 \cdot \boldsymbol{z}) = f(\mathbf{0}_{\mathcal{V}}).$$

Таким образом, в силу взаимной однозначности (биективности) отображения f, получаем $\boldsymbol{x} = \mathbf{0}_{\mathcal{V}}$; противоречие.

2. Пусть $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p \in \mathcal{V}$ — линейно независимые векторы. Предположим, что векторы $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_p) \in \mathcal{W}$ линейно зависимы, т.е. существуют такие коэффициенты $\beta^1, \ldots, \beta^p \in \mathbb{K}$, не все равные нулю, что

$$\beta^1 \boxdot f(\boldsymbol{x}_1) \boxplus \cdots \boxplus \beta^p \boxdot f(\boldsymbol{x}_p) = \mathbf{0}_{\mathcal{W}}.$$

Тогда

$$\beta^1 \boxdot f(\boldsymbol{x}_1) \boxplus \cdots \boxplus \beta^p f(\boldsymbol{x}_p) = \mathbf{0}_{\mathcal{W}} = f\Big(\beta^1 \cdot \boldsymbol{x}_1 + \cdots + \beta^p \cdot \boldsymbol{x}_p\Big),$$

откуда

$$\beta^1 \cdot \boldsymbol{x}_1 + \cdots + \beta^p \cdot \boldsymbol{x}_p = \mathbf{0}_{\mathcal{V}},$$

т.е. векторы $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ линейно зависимы; противоречие.

3. По условию

$$\alpha^1 \cdot \boldsymbol{x}_1 + \cdots + \alpha^p \cdot \boldsymbol{x}_p = \mathbf{0}_{\mathcal{V}}.$$

Тогда

$$\mathbf{0}_{\mathcal{W}} = f(\mathbf{0}_{\mathcal{V}}) = f\Big(\alpha^1 \cdot \boldsymbol{x}_1 + \cdots + \alpha^p \cdot \boldsymbol{x}_p\Big) = \alpha^1 \boxdot f(\boldsymbol{x}_1) \boxplus \cdots \boxplus \alpha^p \boxdot f(\boldsymbol{x}_p),$$

т.е. векторы $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_p)$ линейно зависимы, причём коэффициентами нетривиальной линейной комбинации, обращающейся в нулевой вектор $\mathbf{0}_{\mathcal{W}}$, являются скаляры $\alpha^1, \ldots, \alpha^p$. □

8.41. Предложение. *Отношение изоморфности векторных пространств обладает всеми свойствами отношения эквивалентности*:

(i) $\mathcal{V} \simeq \mathcal{V}$;

(ii) $\mathcal{V} \simeq \mathcal{W} \Rightarrow \mathcal{W} \simeq \mathcal{V}$;

(iii) *если* $\mathcal{V} \simeq \mathcal{W}$ *и* $\mathcal{W} \simeq \mathcal{U}$, *то* $\mathcal{V} \simeq \mathcal{U}$.

8.42. Замечание. Мы сознательно избегаем говорить, что отношение изоморфности *является* отношением эквивалентности: дело в том, что понятие «множество всех векторных пространств» в рамках «наивной» теории множеств противоречиво (ср. с парадоксом Расселла, см. добавление **??**).

Доказательство. (i) $\mathcal{V} \simeq \mathcal{V}$; изоморфизмом служит тождественное отображение $\mathrm{id} : \mathcal{V} \to \mathcal{V}$, т.е. отображение, которое каждому вектору $\boldsymbol{x} \in \mathcal{V}$ ставит в соответствие сам этот вектор: $\mathrm{id}(\boldsymbol{x}) = \boldsymbol{x}$.

(ii) Пусть $\mathcal{V} \simeq \mathcal{W}$ и $f : \mathcal{V} \to \mathcal{W}$ — изоморфизм. Благодаря взаимной однозначности (биективности) отображения f существует обратное отображение $f^{-1} : \mathcal{W} \to \mathcal{V}$ (см. теорему II.25 в добавлении 1, с. 331). Докажем, что отображение f^{-1} линейно, т.е. для любых $\boldsymbol{x}', \boldsymbol{y}' \in \mathcal{W}$ и любых $\alpha, \beta \in \mathbb{K}$ выполняется соотношение

$$f^{-1}\Big(\alpha \boldsymbol{x}' + \beta \boldsymbol{y}'\Big) = \alpha f^{-1}(\boldsymbol{x}') + \beta f^{-1}(\boldsymbol{y}').$$

Действительно, если $\boldsymbol{x}' = f(\boldsymbol{x})$ и $\boldsymbol{y}' = f(\boldsymbol{y})$, что эквивалентно соотношениям $\boldsymbol{x} = f^{-1}(\boldsymbol{x}')$ и $\boldsymbol{y} = f^{-1}(\boldsymbol{y}')$, то

$$\alpha \boldsymbol{x} + \beta \boldsymbol{y} = f^{-1}\Big(f(\alpha \boldsymbol{x} + \beta \boldsymbol{y})\Big) = f^{-1}\Big(\alpha f(\boldsymbol{x}) + \beta f(\boldsymbol{y})\Big) = f^{-1}\Big(\alpha \boldsymbol{x}' + \beta \boldsymbol{y}'\Big).$$

С другой стороны,

$$\alpha \boldsymbol{x} + \beta \boldsymbol{y} = \alpha f^{-1}(\boldsymbol{x}') + \beta f^{-1}(\boldsymbol{y}').$$

Из последних двух соотношений вытекает требуемое.

(iii) Если $f\colon \mathcal{V} \to \mathcal{W}$ и $g\colon W \to \mathcal{U}$ — изоморфизмы, то их композиция является биективным отображением (см. утверждение 2 теоремы **??** добавления 1, с. **??**). Докажем, что композиция отображений $g \circ f\colon \mathcal{V} \to \mathcal{U}$ является линейным отображением. Имеем:

$$(g \circ f)(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = g\Big(f(\alpha \boldsymbol{x} + \beta \boldsymbol{y})\Big) = g\Big(\alpha f(\boldsymbol{x}) + \beta f(\boldsymbol{y})\Big) = \\ = \alpha g\big(f(\boldsymbol{x})\big) + \beta g\big(f(\boldsymbol{y})\big) = \alpha (g \circ f)(\boldsymbol{x}) + \beta (g \circ f)(\boldsymbol{y}),$$

что и требовалось. □

8.43. Предложение. *Пусть $\mathcal{V}(\mathbb{K})$ — векторное пространство над полем $\mathbb{K}$, $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$ — базис в $\mathcal{V}$. Отображение $f : \mathcal{V} \to \mathbb{K}^n$, ставящее в соответствие каждому вектору $\boldsymbol{x} \in \mathcal{V}$ столбец X его координат в базисе $\boldsymbol{E}$, является изоморфизмом векторных пространств $\mathcal{V}$ и $\mathbb{K}^n$, $\mathcal{V} \simeq \mathbb{K}^n$. Этот изоморфизм называется координатным изоморфизмом, порождённым базисом $\boldsymbol{E}$.*

Доказательство. Биективность отображения вытекает из единственности разложения по базису (теорема 8.15). Проверим линейность. Если X и Y — столбцы координат векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ в базисе $\boldsymbol{E}$, то

$$\alpha \boldsymbol{x} + \beta \boldsymbol{y} = \alpha \boldsymbol{E} X + \beta \boldsymbol{E} Y = \boldsymbol{E}(\alpha X + \beta Y),$$

т.е. $\alpha X + \beta Y$ — столбец координат вектора $\alpha \boldsymbol{x} + \beta \boldsymbol{y}$. □

8.44. Пример. Примером координатного изоморфизма векторных пространств является отображение $V_3 \to \mathbb{R}^3$, которое ставит в соответствие каждому геометрическому вектору $\boldsymbol{x} \in V_3$ столбец его координат $X \in \mathbb{R}^3$ в некотором заранее зафиксированном базисе.

8.45. Следствие. *Все векторные пространства одинаковой размерности над одним и тем же полем изоморфны.*

Доказательство. Если $\mathcal{V}(\mathbb{K})$ и $\mathcal{W}(\mathbb{K})$ — векторные пространства над полем $\mathbb{K}$ одинаковой размерности n, то в силу предложения 8.43 $\mathcal{V} \simeq \mathbb{K}^n$ и $\mathcal{W} \simeq \mathbb{K}^n$. Поскольку $\simeq$ — отношение эквивалентности (см. теорему 8.41), получаем требуемое. □

8.46. Замечание. Если векторные пространства изоморфны, то изоморфизм между ними определён неоднозначно; более того, таких изоморфизмов имеется бесконечно много. В частности имеется много изоморфизмов пространства $\mathcal{V}$ с самим собой; такие изоморфизмы называются *автоморфизмами.*

8.47. Замечание. С алгебраической точки зрения изоморфные между собой пространства устроены одинаково; различие проявляется лишь в конкретной природе элементов этих пространств, что несущественно для исследования их алгебраических свойств.

5. Преобразование базисов и координат

А. Преобразование базисов. Предположим, что в векторном пространстве $(\mathcal{V}, \mathbb{K})$ заданы два базиса:

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

Будем называть базис $\boldsymbol{E}$ старым, а базис $\boldsymbol{E}'$ — новым.

8.48. Замечание. Обратите внимание, что штрихи, с помощью которых мы различаем векторы старого и нового базисов, ставятся не у буквы, обозначающей вектор, а у индекса, нумерующего вектор. Такая система обозначений позволит не только значительно сэкономить количество нужных букв, но и практически автоматически, не задумываясь писать правильные формулы, следя лишь за надлежащим расположением штрихов (у верхних или у нижних индексов). При этом нужно иметь в виду, что символы j и j' рассматриваются как *разные* буквы, т.е., например, в выражении $c_j^{j'}$ суммирование не производится (повторяющихся индексов нет), а в выражении $c_j^{j'} a^j$ производится суммирование по индексу j, который встречается в этом выражении дважды: один раз внизу у $c_j^{j'}$, а второй раз наверху у a^j. Более того, можно даже считать, что «штрихованные» индексы пробегают множество значений, состоящее из «штрихованных» натуральных чисел: $j' = 1', 2', \ldots, n'$.

Векторы нового базиса $\boldsymbol{E}'$ можно разложить по старому базису:

$$\begin{aligned} \boldsymbol{e}_{1'} &= c_{1'}^1 \boldsymbol{e}_1 + c_{1'}^2 \boldsymbol{e}_2 + \cdots + c_{1'}^n \boldsymbol{e}_n, \\ \boldsymbol{e}_{2'} &= c_{2'}^1 \boldsymbol{e}_1 + c_{2'}^2 \boldsymbol{e}_2 + \cdots + c_{2'}^n \boldsymbol{e}_n, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ \boldsymbol{e}_{n'} &= c_{n'}^1 \boldsymbol{e}_1 + c_{n'}^2 \boldsymbol{e}_2 + \cdots + c_{n'}^n \boldsymbol{e}_n. \end{aligned} \tag{8.5}$$

Используя правило суммирования Эйнштейна, запишем эти разложения компактно:

$$\boldsymbol{e}_{j'} = c_{j'}^j \boldsymbol{e}_j, \quad j = 1, 2, \ldots, n, \quad j' = 1', 2', \ldots, n'. \tag{8.6}$$

Составим матрицу, столбцы которой состоят из координат новых базисных векторов относительно старого базиса:

$$C = \begin{pmatrix} c_{1'}^1 & c_{2'}^1 & \ldots & c_{n'}^1 \\ c_{1'}^2 & c_{2'}^2 & \ldots & c_{n'}^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_{1'}^n & c_{2'}^n & \ldots & c_{n'}^n \end{pmatrix}.$$

Эта матрица называется *матрицей перехода от базиса $\boldsymbol{E}$ к базису $\boldsymbol{E}'$.*

8.49. Замечание. Предостерегая читателя от широко распространённой ошибки, автор считает своим долгом ещё раз подчеркнуть, что координаты новых базисных векторов относительно старого базиса образуют *столбцы* матрицы перехода. Так, если векторы нового базиса выражаются через векторы старого базиса двумерного векторного пространства по формулам

$$\begin{cases} \boldsymbol{e}_{1'} = 2\boldsymbol{e}_1 + 3\boldsymbol{e}_2, \\ \boldsymbol{e}_{2'} = 4\boldsymbol{e}_1 + 5\boldsymbol{e}_2, \end{cases}$$

то матрица перехода имеет вид

$$C = \begin{pmatrix} 2 & 4 \\ 3 & 5 \end{pmatrix},$$

т.е. *заполнение матрицы перехода производится «по столбцам»*.

Формулы (8.6) можно переписать в матричной форме

$$\boldsymbol{E}' = \boldsymbol{E}C. \tag{8.7}$$

Будем называть эту формулу, а также формулы (8.5) и (8.6) *формулами преобразования базиса*.

Линейная независимость векторов базиса влечёт линейную независимость столбцов матрицы перехода C и, следовательно, её невырожденность и обратимость. Будем обозначать элементы обратной матрицы C^{-1} через $c_j^{j'}$. Таким образом, элементы матриц C и C^{-1} обозначаются одной и той же буквой, различие состоит в положении штриха: у элементов матрицы C штрихом снабжён нижний индекс, а у элементов матрицы C^{-1} — верхний индекс.

Умножая обе части равенства (8.7) справа на обратную матрицу C^{-1}, получим $\boldsymbol{E}'C^{-1} = \boldsymbol{E}CC^{-1}$, откуда получаем *формулу обратного преобразования базиса*

$$\boldsymbol{E} = \boldsymbol{E}'C^{-1}. \tag{8.8}$$

Приведём другой вывод этой формулы. Умножим обе части формулы (8.6) на элемент обратной матрицы перехода $c_k^{j'}$ и просуммируем полученные выражения по j':

$$c_k^{j'}\boldsymbol{e}_{j'} = \underbrace{c_k^{j'}c_{j'}^{j}}_{=\delta_k^j}\boldsymbol{e}_j.$$

Подчёркнутое выражение равно символу Кронекера, поскольку оно выражает элемент произведения матриц C и C^{-1}, которое равно единичной матрице. Итак,

$$c_k^{j'}\boldsymbol{e}_{j'} = \delta_k^j\boldsymbol{e}_j = \boldsymbol{e}_k.$$

Заменяя[1] обозначение индекса суммирования j' в левой части на k', получаем формулу обратного преобразования базиса:

$$\boldsymbol{e}_k = c_k^{k'}\boldsymbol{e}_{k'}, \quad k = 1, 2, \ldots, n, \quad k' = 1', 2', \ldots, n'. \tag{8.9}$$

[1]Подчеркнём, что новое обозначение индекса суммирования должно содержать штрих, поскольку положение штриха важна для различения элементов прямой и обратной матриц перехода!

В. Преобразование координат. Теперь получим формулы, связывающие координаты произвольного вектора $\boldsymbol{x} \in \mathcal{V}$ относительно старого и нового базисов. Пусть

$$\boldsymbol{x} = x^j \boldsymbol{e}_j = x^{j'} \boldsymbol{e}_{j'}$$

— разложения вектора $\boldsymbol{x}$ по старому и новому базисам. Подставляя в эту формулу выражения (8.6), получим

$$x^j \boldsymbol{e}_j = x^{j'} \boldsymbol{e}_{j'} = x^{j'} c^j_{j'} \boldsymbol{e}_j.$$

В левой и правой частях этой формулы содержатся разложения вектора $\boldsymbol{x}$ по базису $\boldsymbol{E}$; из единственности разложения заключаем, что

$$x^j = c^j_{j'} x^{j'}, \quad j = 1, 2, \ldots, n, \quad j' = 1', 2', \ldots, n' \tag{8.10}$$

(по j' производится суммирование). Умножим обе части полученных соотношений на элемент матрицы перехода $c^{k'}_j$ и просуммируем полученные соотношения по j:

$$c^{k'}_j x^j = \underbrace{c^{k'}_j c^j_{j'}}_{=\delta^{k'}_{j'}} x^{j'} = \delta^{k'}_{j'} x^{j'} = x^{k'}.$$

Изменив в левой части обозначение индекса суммирования с j на k, получаем

$$x^{k'} = c^{k'}_k x^k, \quad k = 1, 2, \ldots, n, \quad k' = 1', 2', \ldots, n'. \tag{8.11}$$

Приведём вывод полученных формул в матричном виде. Введём столбцы координат $X = \left(x^1, \ldots, x^n\right)^T$ и $X' = \left(x^{1'}, \ldots, x^{n'}\right)^T$ вектора $\boldsymbol{x}$ в базисах $\boldsymbol{E}$ и $\boldsymbol{E}'$. Разложения вектора $\boldsymbol{x}$ по старому и новому базисам имеют вид

$$\boldsymbol{x} = \boldsymbol{E}X = \boldsymbol{E}'X'.$$

Поставляя вместо $\boldsymbol{E}'$ выражение (8.7), находим

$$\boldsymbol{E}X = \boldsymbol{E}CX',$$

откуда в силу единственности разложения по базису получаем

$$X = CX'. \tag{8.12}$$

Умножив обе части этой формулы слева на C^{-1}, приходим к соотношению

$$X' = C^{-1}X. \tag{8.13}$$

Запишем все полученные формулы вместе. Формулы перехода от старого базиса к новому:

$$\begin{aligned} \boldsymbol{e}_{j'} &= c^j_{j'} \boldsymbol{e}_j, \qquad & \boldsymbol{E}' &= \boldsymbol{E}C, \\ x^{j'} &= c^{j'}_j x^j, \qquad & X' &= C^{-1}X. \end{aligned}$$

Формулы обратного перехода:

$$\boldsymbol{e}_j = c^{j'}_j \boldsymbol{e}_{j'}, \qquad \boldsymbol{E} = \boldsymbol{E}C^{-1},$$

$$x^j = c^j_{j'} x^{j'}, \qquad X = CX'.$$

8.50. Замечание. Обратите внимание и запомните, что в формулах преобразования векторов базиса и формулах преобразования координат фигурируют взаимно обратные матрицы!

8.51. Замечание. В случае, когда требуется записать матрицу перехода C, зная формулы $X = CX'$ преобразования координат, нужно понимать, что при перемножении матриц C и X' происходит умножение строк матрицы C на столбец X', т.е. в этом случае *матрица парахода заполняется «по строкам»*. Например, из формул преобразования координат

$$\begin{cases} x^{1'} = 2x^1 + 3x^{x^2}, \\ x^{2'} = 4x^1 + 5x^{x^2} \end{cases}$$

следует, что матрица перехода имеет вид

$$C = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix}.$$

Сравните эту ситуацию с рассмотренной в замечании 8.49.

C. Полная линейная группа. Матрицей перехода от одного базиса n-мерного векторного пространства $(\mathcal{V}, \mathbb{K})$ к другому может служить любая невырожденная матрица. Легко видеть, что все такие матрицы образуют группу относительно операции умножения.[1] Эта группа называется *полной линейной группой* и обозначается $GL(n, \mathbb{K})$ или просто $GL(n)$.

6. Ориентация вещественного векторного пространства

Понятие ориентации является абстрактным обобщением наглядных представлений о направлении на прямой (вправо или влево), направлении вращения на плоскости (по часовой стрелке или против) и о винтовом движении в пространстве. Это понятие можно ввести лишь для *вещественных* векторных пространств.

A. Одноимённые и разноимённые базисы. Пусть $(\mathcal{V}, \mathbb{R})$ — векторное пространство над полем вещественных чисел. Рассмотрим в нём два базиса $\boldsymbol{E} = \big(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\big)$ и $\boldsymbol{E}' = \big(\boldsymbol{e}_{1'}, \dots, \boldsymbol{e}_{n'}\big)$.

8.52. Определение. Базисы $\boldsymbol{E}$ и $\boldsymbol{E}'$ называются *одноимёнными* (обозначение $\boldsymbol{E}' \simeq \boldsymbol{E}$), если матрица перехода от $\boldsymbol{E}$ к $\boldsymbol{E}'$ имеет положительный определитель, и *разноимёнными* в противном случае.

8.53. Предложение. *Отношение одноимённости на множестве всех базисов векторного пространства $\mathcal{V}$ является отношением эквивалентности.*

[1] Если это не ясно, нужно прочитать определение группы (определение 6.5, с. 129) и свойства невырожденных матриц (теорема 6.21, с. 6.21).

Доказательство. 1. Рефлексивность очевидна: $\boldsymbol{E} \simeq \boldsymbol{E}$, поскольку $\boldsymbol{E} = \boldsymbol{E}\mathbb{1}$.

2. Симметричность. Пусть $\boldsymbol{E}' \simeq \boldsymbol{E}$, т.е. существует такая матрица C, что $\boldsymbol{E}' = \boldsymbol{E}C$. Так как $\det C \neq 0$, существует обратная матрица C^{-1} и поэтому

$$\boldsymbol{E}'C^{-1} = \boldsymbol{E}CC^{-1} \iff \boldsymbol{E} = \boldsymbol{E}'C^{-1} \iff \boldsymbol{E} \simeq \boldsymbol{E}'.$$

3. Транзитивность. Если $\boldsymbol{E}' \simeq \boldsymbol{E}$, т.е. $\boldsymbol{E}' = \boldsymbol{E}C$, и $\boldsymbol{E}'' \simeq \boldsymbol{E}'$, т.е. $\boldsymbol{E}'' = \boldsymbol{E}'D$, то

$$\boldsymbol{E}'' = \boldsymbol{E}'D = (\boldsymbol{E}C)D = \boldsymbol{E}(CD) \iff \boldsymbol{E} \simeq \boldsymbol{G},$$

что и требовалось доказать. □

Таким образом, множество всех базисов в пространстве $\mathcal{V}$ разбивается на два класса эквивалентности следующим образом. Пусть $\boldsymbol{E}$ — некоторый базис. К одному классу эквивалентности относятся все базисы, одноимённые с $\boldsymbol{E}$ (и при этом одноимённые между собой), к другому — разноимённые с $\boldsymbol{E}$ (и при этом также одноимённые между собой).

8.54. 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.

В случае пространств размерностей 1, 2 и 3 абстрактное понятие ориентации имеет наглядную интерпретацию, которая обсуждалась в лекции III (см. с. 51).

Отметим, что термины «положительная ориентация» и «отрицательная ориентация» абстрактны, 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. Вопрос о том, существуют ли чисто физические процессы, позволяющие выбирать ориентацию пространства, т.е. процессы, не инвариантные относительно зеркального отражения, был решён положительно в 1957 г. с помощью эксперимента, установившего несохранение чётности в слабых взаимодействиях.

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.

В. Деформируемость базисов. Рассмотрим функцию

$$\boldsymbol{x}\colon [0,1] \to \mathcal{V}, \quad t \mapsto \boldsymbol{x}(t), \tag{8.14}$$

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 $\boldsymbol{E} = \bigl(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\bigr)$, то при каждом значении $t \in [0,1]$ вектор $\boldsymbol{x}(t)$

можно разложить по этому базису, в результате чего получим n числовых функций $x^1(t), \ldots, x^n(t)$, которые являются координатами вектор-функции $t \mapsto \boldsymbol{a}(t)$:

$$\boldsymbol{x}(t) = x^1(t)\boldsymbol{e}_1 + \cdots + x^n(t)\boldsymbol{e}_n. \tag{8.15}$$

8.55. Определение. Вектор-функция (8.14) называется *непрерывной*, если все её координаты (8.15) 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.)

8.56. Предложение. *Определение* 8.55 *корректно, т.е. не зависит от выбора базиса.*

Доказательство. Пусть $\boldsymbol{E}'$ — новый базис, C — матрица перехода от базиса $\boldsymbol{E}$ к базису $\boldsymbol{E}'$. Новые координаты $x^{k'}(t)$ вектор-функции $t \mapsto \boldsymbol{x}(t)$ являются линейными комбинациями старых координат $x^k(t)$ с постоянными коэффициентами (см. формулу (8.11)) и потому согласно известной теореме анализа являются непрерывными функциями. □

8.57. Определение. Будем говорить, что базис $\boldsymbol{E}' = \big(\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}\big)$ векторного пространства $\mathcal{V}$ получен из базиса $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$ *непрерывной деформацией*, если существует такое семейство непрерывных вектор-функций $\boldsymbol{A}(t) = \big(\boldsymbol{a}_1(t), \ldots, \boldsymbol{a}_n(t)\big)$, что

(i) при каждом $t \in [0, 1]$ семейство $\boldsymbol{A}(t)$ является базисом,
(ii) $\boldsymbol{A}(0) = \boldsymbol{E}$, $\boldsymbol{A}(1) = \boldsymbol{E}'$.

Семейство $\boldsymbol{A}(t)$ называется (непрерывной) деформацией базиса $\boldsymbol{E}$ в базис $\boldsymbol{E}'$.

Разумеется, матрица перехода $C(t)$ от базиса $\boldsymbol{E}$ к базису $\boldsymbol{A}(t)$ также зависит от t; очевидно, $C(0) = \mathbb{1}$, а $C(1)$ — матрица перехода от $\boldsymbol{E}$ к $\boldsymbol{E}'$. Будем говорить, что *матрица* $C(t)$ *осуществляет непрерывную деформацию единичной матрицы в матрицу* $C(1)$.

Таким образом, на множестве всех базисов векторного пространства возникает отношение деформируемости: $\boldsymbol{E}' \sim \boldsymbol{E}$, если базис $\boldsymbol{E}'$ получен непрерывной деформацией базиса $\boldsymbol{E}$.

8.58. Предложение. *Отношение деформируемости базисов является отношением эквивалентности.*

Доказательство. Для доказательства утверждения необходимо проверить выполнение свойств отношения эквивалентности (см. определение **??**, с. **??**).

1. Рефлексивность: для любого базиса $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ вектор-функции $\boldsymbol{a}_k(t) \equiv \boldsymbol{e}_k$, $k = 1, \ldots, n$, определяют, очевидно, деформацию этого базиса в себя.

2. Симметричность: для любой деформации $\boldsymbol{A}(t)$ базиса $\boldsymbol{E}$ в базис $\boldsymbol{E}'$ семейство $\boldsymbol{A}(1 - t)$ является деформацией базиса $\boldsymbol{E}'$ в базис $\boldsymbol{E}$.

3. Транзитивность. Если $\boldsymbol{A}(t)$ — деформация базиса $\boldsymbol{E}$ в базис $\boldsymbol{E}'$ и $\boldsymbol{B}(t)$ — деформация базиса $\boldsymbol{E}'$ в базис $\boldsymbol{E}''$, то формула

$$\boldsymbol{C}(t) = \begin{cases} \boldsymbol{A}(2t), & 0 \leqslant t \leqslant \frac{1}{2}, \\ \boldsymbol{B}(2t-1), & \frac{1}{2} \leqslant t \leqslant 1 \end{cases}$$

определяет деформацию базиса $\boldsymbol{E}$ в базис $\boldsymbol{E}''$. □

C. Деформации одноимённых базисов.

8.59. Теорема. *Два базиса одноимённы тогда и только тогда, когда они деформируемы друг в друга.*

Доказательство. Докажем, что деформируемые друг в друга базисы одноимённы. 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 анализа известно, что непрерывная функция, не обращающаяся в нуль на отрезке, сохраняет постоянный знак. Поэтому $\Delta(1) > 0$ и, следовательно, базис $\boldsymbol{E}' = \boldsymbol{A}(1)$ одноимёнен с базисом $\boldsymbol{E}$.

Обратное утверждение доказывается несколько сложнее. Предположим, что базисы $\boldsymbol{E}$ и $\boldsymbol{E}'$ одноимённы, т.е. в равенстве $\boldsymbol{E}' = \boldsymbol{E}C$ матрица C имеет положительный определитель.

Приведём матрицу C к упрощённому виду при помощи алгоритма Гаусса—Жордана; результатом будет единичная матрица, поскольку столбцы матрицы C линейно независимы. Как известно (см. теорему 7.1, с. 144), каждое элементарное преобразование строк матрицы эквивалентно её умножению справа на некоторую невырожденную матрицу, называемую матрицей элементарного преобразования. Таким образом, приведение матрицы C к упрощённому виду равносильно нахождению таких элементарных преобразований (матриц) $R_1, \ldots, R_p$, что справедливо равенство

$$R_p \ldots R_2 R_1 C = \mathbb{1}.$$

Из этого равенства легко получается соотношение

$$C = R_1^{-1} R_2^{-1} \ldots R_p^{-1}.$$

Таким образом, матрица C разлагается в произведение элементарных преобразований строк $R_1^{-1}, \ldots, R_p^{-1}$ (ясно, что преобразование, обратное к элементарному преобразованию строк, само является элементарным преобразованием строк). Напомним эти преобразования:

(1) перестановка двух строк местами;
(2) умножение любой строки на ненулевое число λ;
(3) прибавление к любой строке другой строки, умноженной на произвольное число.

Очевидно, определитель матрицы преобразования типа (1) равен -1 (ибо перестановка строк матрицы приводит к изменению знака определителя), определитель матрицы преобразования типа (2) равен λ, а определитель матрицы преобразования типа (3) равен 1 (такое преобразование строк матрицы не меняет её определителя).

Рассмотрим следующие преобразования, каждое из которых не меняет определитель:

(a) перестановка двух строк местами с одновременным изменением знака одной из переставляемых строк;
(b) умножение любой строки на ненулевое число λ с одновременным умножением какой-либо другой строки на обратное число $1/\lambda$;
(c) прибавление к любой строке другой строки, умноженной на произвольное число.

Убедимся, что с помощью этих преобразований матрицу C можно привести к виду

$$\begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & \det C \end{pmatrix}. \tag{8.16}$$

Действительно, каждый шаг алгоритма Гаусса—Жордана состоит в следующем:

(i) выбирается ведущая строка и при необходимости поднимается на нужное место; одновременно изменяется знак у опускаемой строки (преобразование (a));
(ii) ведущий элемент обращается в единицу при помощи преобразования (b): умножаем ведущую строку на подходящее число λ и одновременно умножаем последнюю строку матрицы на $1/\lambda$;
(iii) все элементы ведущего столбца обращаются в нуль (кроме ведущего элемента) при помощи преобразования (c).

Поскольку при каждом из преобразований (a)–(c) определитель матрицы не изменяется, в результате получим диагональную матрицу, все диагональные элементы которой суть единицы, кроме последнего, равного определителю исходной матрицы.

Матрицу (8.16) привести к единичной можно, разделив последнюю строку на положительное число $\det C$; назовём эту операцию преобразованием (d).

Каждое из преобразований (a)–(d), рассматриваемое как элементарное преобразование базиса, может быть осуществлено непрерывной деформацией. Убедимся в этом.

Матрица преобразования (a), имеющая вид

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & -1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \tag{8.17}$$

получается в результате следующей непрерывной деформации единичной матрицы:

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos\dfrac{\pi t}{2} & \cdots & -\sin\dfrac{\pi t}{2} & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \sin\dfrac{\pi t}{2} & \cdots & \cos\dfrac{\pi t}{2} & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix};$$

действительно, при $t = 0$ получаем единичную матрицу, а при $t = 1$ — матрицу (8.17).

Матрица преобразования (b) имеет вид

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \lambda & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & \lambda^{-1} & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix};$$

в случае, когда $\lambda > 0$, она получается в результате следующей непрерывной деформации единичной матрицы:

$$\Lambda(t) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & (\lambda-1)t+1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & (\lambda^{-1}-1)t+1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

Действительно,

$$\det \Lambda(t) = \Big((\lambda-1)t+1\Big)\Big((\lambda^{-1}-1)t+1\Big).$$

Выражения в скобках обращаются в нуль при

$$t_1 = \frac{1}{1-\lambda}, \quad t_2 = \frac{\lambda}{\lambda-1} = 1 - t_1,$$

так что получаем $t_1 > 1$, $t_2 < 0$ при $0 < \lambda < 1$ и $t_1 < 0$, $t_2 > 1$ при $\lambda > 1$. Таким образом, при $t \in [0,1]$ имеем $\det \Lambda(t) \neq 0$.

В случае, когда $\lambda < 0$, определитель $\det \Lambda(t)$ дважды обращается в нуль на отрезке $t \in [0,1]$, поэтому матрица $\Lambda(t)$ в качестве невырожденной непрерывной деформации не годится. Для $\lambda = -1$ используем непрерывную деформацию

$$\tilde{\Lambda}(t) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos \pi t & \cdots & -\sin \pi t & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \sin \pi t & \cdots & \cos \pi t & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix};$$

очевидно, $\det \tilde{\Lambda}(t) = 1$. Для других отрицательных значений λ воспользуемся очевидным фактом $\lambda = (-1) \cdot |\lambda|$.

Матрица преобразования (c) и соответствующая деформация единичной матрицы имеют вид

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \lambda & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & t\lambda & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

Наконец, для преобразования (d) имеем

$$\begin{pmatrix} 1 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & \Delta \end{pmatrix}, \quad D(t) = \begin{pmatrix} 1 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & (\Delta - 1)t + 1 \end{pmatrix}.$$

Проверим невырожденность матрицы $D(t)$ при $t \in [0, 1]$. Действительно, $\det D(t) = (\Delta - 1)t + 1$. При $\Delta = 1$ этот определитель тождественно равен 1, а при $\Delta \neq 1$ он обращается в нуль при $t_0 = 1/(1 - \Delta)$. Получаем $t_0 < 0$ в случае $\Delta > 1$ и $t_0 > 1$ в случае $0 < \Delta < 1$, т.е. $D(t) \neq 0$ при всех $t \in [0, 1]$.

Итак, мы построили непрерывную деформацию базиса $\boldsymbol{E}$ в одноимённый базис $\boldsymbol{E}'$, что и требовалось. $\square$

ГЛАВА 9

Аффинное пространство

Чтобы построить полноценную геометрию, одних векторов недостаточно: необходимы ещё точки. Пространство, состоящее из точек, и называется аффинным (или точечным) пространством. Определение аффинного пространства получается при помощи аксиоматизации построения вектора по двум точкам.

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. Тем не менее такое исследование является содержательным и позволяет решать многие задачи.

1. Определение аффинного пространства

A. Определение и простейшие свойства.

9.1. Определение. *Аффинное пространство* над полем $\mathbb{K}$ — это упорядоченная тройка $(\mathcal{A}, \mathcal{V}, \boxplus)$, состоящая из следующих элементов:

(a) множества $\mathcal{A}$, элементы которого называются *точками* и обозначаются прописными латинскими буквами $A, B, C, \ldots$,
(b) векторного пространства $\mathcal{V} = (\mathcal{V}, \mathbb{K}, +, \cdot)$ над полем $\mathbb{K}$ с линейными операциями над векторами $+$ и $\cdot$, которое называется *ассоциированным векторным пространством*,
(c) бинарной операции

$$\boxplus\colon\ \mathcal{A} \times \mathcal{V} \to \mathcal{A}, \quad (A, \boldsymbol{a}) \mapsto A \boxplus \boldsymbol{a},$$

называемой *операцией сдвига точки на вектор.*

Для элементов тройки $(\mathcal{A}, \mathcal{V}, \boxplus)$ должен быть выполнен следующий комплекс условий — *аксиом аффинного пространства*:

A1. для всех $A \in \mathcal{A}$ и всех $\boldsymbol{b}, \boldsymbol{c} \in \mathcal{V}$ имеем

$$(A \boxplus \boldsymbol{b}) \boxplus \boldsymbol{c} = A \boxplus (\boldsymbol{b} + \boldsymbol{c});$$

A2. для всех $A \in \mathcal{A}$ имеем $A \boxplus \mathbf{0} = A$;
A3. для любых двух точек $A, B \in \mathcal{A}$ существует единственный вектор $\boldsymbol{c} \in \mathcal{V}$, обладающий свойством $B = A \boxplus \boldsymbol{c}$.

В большинстве случаев для обозначения аффинного пространства достаточно одного символа $\mathcal{A}$. В ситуациях, когда требуется указать ассоциированное векторное пространство и/или поле скаляров, будем писать $(\mathcal{A}, \mathcal{V}, \mathbb{K})$ или $(\mathcal{A}, \mathcal{V})$.

Пример 9.1. Пусть $\mathcal{V}$ — произвольное векторное пространство. Определим аффинное пространство, положив $\mathcal{A} = \mathcal{V}$ и взяв в качестве операции сдвига $\boxplus$ операцию сложения $+$ векторов в $\mathcal{V}$. Проверим выполнение аксиом аффинного пространства:

A1: для любой «точки» $A = \boldsymbol{a} \in \mathcal{A}$ и любых векторов $\boldsymbol{b}, \boldsymbol{c} \in \mathcal{V}$ в силу ассоциативности сложения векторов (аксиома **V2**) имеем

$$(A \boxplus \boldsymbol{b}) \boxplus \boldsymbol{c} = (\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c} = \boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c}) = A \boxplus (\boldsymbol{b} + \boldsymbol{c});$$

A2: для любой «точки» $A = \boldsymbol{a} \in \mathcal{A}$ имеем

$$A \boxplus \mathbf{0} = \boldsymbol{a} + \mathbf{0} = \boldsymbol{a} = A;$$

A3: для любых двух «точек» $A = \boldsymbol{a} \in \mathcal{A}$ и $B = \boldsymbol{b} \in \mathcal{A}$ положим по определению $\boldsymbol{c} = \boldsymbol{b} - \boldsymbol{a} \equiv \boldsymbol{b} + (-\boldsymbol{a})$; тогда

$$B = \boldsymbol{b} = \boldsymbol{a} + \boldsymbol{c} = A \boxplus \boldsymbol{c}.$$

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.

Этот пример типичен: всякое аффинное пространство изоморфно такому. Кроме того, этот пример показывает, что обозначение операции сдвига $\boxplus$ тем же значком $+$, что и сложение векторов в ассоциированном векторном пространстве $\mathcal{V}$, не приводит к недоразумениям; мы также будем следовать этому соглашению.

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 $\dim \mathcal{A}$.

Аксиомы **A1**–**A3** образуют *вторую группу аксиом Вейля* векторно-точечной аксиоматики геометрии.

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 неопределяемые 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 $B = A + \boldsymbol{a}$.

Эти отношения должны удовлетворять восьми «векторным» аксиомам **V1–V8** и трём «аффинным» аксиомам **A1–A3**.

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.

B. Правила вычислений с точками и векторами. Согласно аксиоме **A3**, для любых точек $A, B \in \mathcal{A}$ существует единственный вектор $\boldsymbol{c} \in \mathcal{V}$, для которого $B = A + \boldsymbol{c}$. Этот вектор удобно обозначать $B - A$ или $\overrightarrow{AB}$ и называть соответствующую операцию «вычитанием точек» или «построением вектора с начальной точкой A и конечной точкой B».

9.2. Предложение. *Операция «вычитания точек»*

$$\mathcal{A} \times \mathcal{A} \to \mathcal{V}, \quad (A, B) \mapsto \overrightarrow{AB} \equiv B - A$$

обладает следующими свойствами:

(i) *для всех точек $A, B, C \in \mathcal{A}$ имеем*

$$\overrightarrow{AC} = \overrightarrow{AB} + \overrightarrow{BC} \iff (C - A) = (B - A) + (C - B)$$

(*правило треугольника*);

(ii) *для любой точки $A \in \mathcal{A}$ имеем $\overrightarrow{AA} = \mathbf{0}$ или, что то же самое, $A - A = \mathbf{0}$;*

(iii) *для всех $A, B \in \mathcal{A}$ и всех $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{V}$ справедливо соотношение*

$$(B + \boldsymbol{b}) - (A + \boldsymbol{a}) = (B - A) + (\boldsymbol{b} - \boldsymbol{a})$$

или, эквивалентно,

$$\overrightarrow{A'B'} = \overrightarrow{AB} + (\boldsymbol{b} - \boldsymbol{a}),$$

где $A' = A + \boldsymbol{a}$, $B' = B + \boldsymbol{b}$.

Доказательство. (i) Действительно, пусть $B = A + \boldsymbol{a}$, $C = B + \boldsymbol{b}$. Тогда

$$C = B + \boldsymbol{b} = (A + \boldsymbol{a}) + \boldsymbol{b} = A + (\boldsymbol{a} + \boldsymbol{b}) \quad \Rightarrow$$
$$\Rightarrow \quad C - A = \boldsymbol{a} + \boldsymbol{b} = (B - A) + (C - B).$$

(iii) Достаточно проверить, что

$$(A + \boldsymbol{a}) + (B - A) + (\boldsymbol{b} - \boldsymbol{a}) = B + \boldsymbol{b} \iff A + (B - A) = B,$$

но это — определение $B - A$. □

Итак, в аффинном пространстве и ассоциированном векторном пространстве можно выполнять следующие операции:

(i) сложение векторов: $+\colon\ \mathcal{V} \times \mathcal{V} \to \mathcal{V}$,

(ii) сдвиг точки на вектор: $+\colon\ \mathcal{A} \times \mathcal{V} \to \mathcal{V}$,

(iii) вычитание точек, или построение вектора с данными начальной и конечной точками: $-\colon\ \mathcal{A} \times \mathcal{A} \to \mathcal{V}$.

Заметим, что сумма точек $A + B$, вообще говоря, смысла не имеет, равно как и выражение αA, где $\alpha \in \mathbb{K}$, за исключением случая, когда $\mathcal{A} = \mathcal{V}$.

Пример 9.2. Следуя примеру 9.1, мы можем рассматривать арифметическое пространство $\mathbb{K}^n$ как аффинное пространство с ассоциированным векторным пространством $\mathbb{K}^n$; в таком качестве будем его обозначать $\mathbb{K}^n_{\text{афф}}$. Если $A = (a^1, \ldots, a^n)^T$ и $B = (b^1, \ldots, b^n)^T$ — две точки из аффинного пространства $\mathbb{K}^n_{\text{афф}}$, то вектор $\overrightarrow{AB}$ из векторного пространства $\mathbb{K}^n$ определяется как

$$\overrightarrow{AB} = \begin{pmatrix} b^1 - a^1 \\ \vdots \\ b^n - a^n \end{pmatrix} = B - A.$$

Мы получили хорошо известное школьное правило: координаты вектора равны разностям соответствующих координат его конца и начала.

C. Аффинная система координат.

9.3. Определение. *Система аффинных координат*, или *репе́р*, в аффинном пространстве $(\mathcal{A}, \mathcal{V}, \mathbb{K})$ — это пара, состоящая из точки $O \in \mathcal{A}$ (*начала координат*) и базиса $\boldsymbol{E} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ ассоциированного векторного пространства $\mathcal{V}$; обозначение $O\boldsymbol{E}$ или $O\boldsymbol{e}_1 \ldots \boldsymbol{e}_n$. *Координатами* точки $A \in \mathcal{A}$ в этой системе координат являются числа $x^j \in \mathbb{K}$, однозначно определяемые условием

$$A = O + x^j \boldsymbol{e}_j \iff \overrightarrow{OA} = x^j \boldsymbol{e}_j.$$

Вектор $\overrightarrow{OA}$ называется *радиус-вектором* точки A относительно начала координат O.

На рис. 9.1 изображены системы аффинных координат на прямой, на плоскости и в пространстве.

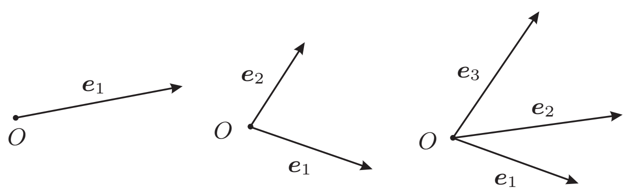


Рис. 9.1. Аффинные системы координат на прямой, на плоскости и в пространстве

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.

D. Изоморфизм аффинных пространств. Рассмотрим два аффинных пространства $(\mathcal{A}, \mathcal{V}, \mathbb{K})$ и $(\mathcal{B}, \mathcal{W}, \mathbb{K})$.

9.4. Определение. Пара отображений (F, f), где

$$F\colon \mathcal{A} \to \mathcal{B}, \quad f\colon \mathcal{V} \to \mathcal{W},$$

называется *изоморфизмом аффинных пространств* $(\mathcal{A}, \mathcal{V})$ и $(\mathcal{B}, \mathcal{W})$, если

(i) отображение F взаимно однозначно;
(ii) отображение f является изоморфизмом векторных пространств $\mathcal{V}$ и $\mathcal{W}$;
(iii) для любых точек $A, B \in \mathcal{A}$ выполняется соотношение

$$\overrightarrow{AB} = F\big(\overrightarrow{AB}\big).$$

Аффинные пространства $(\mathcal{A}, \mathcal{V}, \mathbb{K})$ и $(\mathcal{B}, \mathcal{W}, \mathbb{K})$ называются *изоморфными* (обозначение $\mathcal{A} \simeq \mathcal{B}$), если существует изоморфизм $F\colon \mathcal{A} \to \mathcal{B}$, $f\colon \mathcal{V} \to \mathcal{W}$.

Ясно, что отношение изоморфности аффинных пространств является отношением эквивалентности (ср. с предложением 8.41).

9.5. Предложение. *Каждое n-мерное аффинное пространство* $(\mathcal{A}, \mathcal{V}, \mathbb{K})$ *изоморфно аффинному пространству* $(\mathbb{K}^n_{\text{афф}}, \mathbb{K}^n, \mathbb{K}$.

Доказательство аналогично доказательству предложения 8.43. Зафиксируем в $\mathcal{A}$ какую-либо аффинную систему координат (репер) $\boldsymbol{OE}$, где $O \in \mathcal{A}$ — произвольная точка, $\boldsymbol{E}$ — произвольный базис в $\mathcal{V}$. Пара отображений (F, f), где $F\colon \mathcal{A} \to \mathbb{K}^n_{\text{афф}}$ — отображение, ставящее в соответствие каждой точке $A \in \mathcal{A}$ её аффинные координаты относительно репера $\boldsymbol{OE}$, а $f\colon \mathcal{V} \to \mathbb{K}^n$ — координатный изоморфизм, порождённый базисом $\boldsymbol{E}$, является аффинным изоморфизмом. $\square$

9.6. Следствие. *Все аффинные пространства одинаковой размерности над одним и тем же полем изоморфны.*

В первую очередь нас интересуют аффинные пространства размерности $\leqslant 3$, однако дальнейшее изложение проводится для общего случая: в рамках аксиоматики Вейля все геометрические конструкции совершенно идентичны в пространствах любой размерности. Одномерное аффинное пространство называется *прямой*, двумерное — *плоскостью*, трёхмерное — *пространством*.

2. Плоскости в аффинном пространстве

A. Мотивировка определений. Прежде чем ввести понятие плоскости в многомерном аффинном пространстве, проведём рассмотрение на наглядном интуитивном уровне, опираясь на школьные знания.

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. 9.2). 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 $\overrightarrow{M_0M} = t\boldsymbol{a}$.

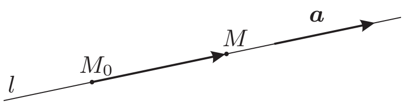


Рис. 9.2.

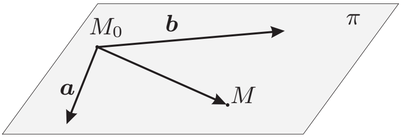


Рис. 9.3.

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. 9.3). 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 u и v, что $\overrightarrow{M_0M} = u\boldsymbol{a}+v\boldsymbol{b}$.

При аксиоматическом построении геометрии на базе системы аксиом Вейля мы положим эти свойства в основу определения, которое будет сформулировано сразу для многомерного случая.

B. Определение плоскости.

9.7. 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 $\pi(M_0, \mathcal{P})$:

$$\pi(M_0, \mathcal{P}) \stackrel{\text{def}}{=\!=} \left\{ M \in \mathcal{A} \;\middle|\; \overrightarrow{M_0M} \in \mathcal{P} \right\}.$$

Подпространство $\mathcal{P}$ называется *направляющим подпространством* плоскости π, а точка M_0 — *опорной точкой*. Число $n - p$ называется *коразмерностью* плоскости π и обозначается $\operatorname{codim} \pi$.

9.8. Определение. Плоскости размерности 1 называются прямыми. Плоскости коразмерности 1 называются гиперплоскостями.

Согласно определению 9.7 точки пространства $\mathcal{A}$ представляют собой 0-плоскости.

9.9. Предложение. *Каждая p-плоскость в аффинном пространстве сама является аффинным пространством размерности p.*

Доказательство. Требуется доказать, что множество $\pi(M_0, \mathcal{P})$ замкнуто относительно операции сдвига $+\colon \pi \times \mathcal{P} \to \pi$ на векторы $\boldsymbol{a} \in \mathcal{P}$, т.е. что для любой точки $A \in \pi$ и любого вектора $\boldsymbol{a} \in \mathcal{P}$ точка $B = A + \boldsymbol{a}$ снова лежит в π, т.е. $\overrightarrow{M_0B} \in \mathcal{P}$. Действительно, условие $A \in \pi$ означает, что $\overrightarrow{M_0A} \in \mathcal{P}$. Тогда

$$\overrightarrow{M_0B} = \underbrace{\overrightarrow{M_0A}}_{\in\mathcal{P}} + \underbrace{\overrightarrow{AB}}_{=\boldsymbol{a}\in\mathcal{P}} \in \mathcal{P}.$$

Подпространство $\mathcal{P} \Subset \mathcal{V}$ является ассоциированным векторным пространством аффинного пространства π, поэтому $\dim\pi = p$. □

9.10. 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.*

Доказательство. Условие $N_0 \in \pi(M_0,\mathcal{P})$ означает, что $\overrightarrow{M_0N_0} \in \mathcal{P}$. Требуется доказать совпадение множеств

$$\pi(M_0,\mathcal{P}) = \left\{M \in \mathcal{A} \;\middle|\; \overrightarrow{M_0M} \in \mathcal{P}\right\}, \quad \pi(N_0,\mathcal{P}) = \left\{M \in \mathcal{A} \;\middle|\; \overrightarrow{N_0M} \in \mathcal{P}\right\}.$$

Пусть $M \in \pi(M_0,\mathcal{P})$, 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. $M \in \pi(N_0,\mathcal{P})$. Обратно, если $M \in \pi(N_0,\mathcal{P})$, т.е. $\overrightarrow{N_0M} \in \mathcal{P}$, то

$$\overrightarrow{M_0M} = \underbrace{\overrightarrow{M_0N_0}}_{\in\mathcal{P}} + \underbrace{\overrightarrow{N_0M}}_{\in\mathcal{P}} \in \mathcal{P},$$

т.е. $M \in \pi(M_0,\mathcal{P})$. □

C. Прямые и отрезки. Одномерные плоскости называются прямыми.

Из предложения 9.10 легко получить следующее утверждение, которое в рамках аксиоматики Гильберта представляет собой одну из аксиом.

9.11. Предложение. *Через любые две различные точки M_0 и M_1 аффинного пространства проходит единственная прямая*; *её обозначают (M_0M_1).*

Доказательство. Если взять в качестве опорной точкой этой прямой точку M_0, а в качестве направляющего вектора — вектор $\overrightarrow{M_0M_1}$, то любая точка M этой прямой определяется условием $\overrightarrow{M_0M} = t\overrightarrow{M_0M_1}$. В частности, сами точки M_0 и M_1 получаются при $t = 0$ и $t = 1$ соответственно. □

Все предыдущие рассмотрения справедливы для любого поля скаляров $\mathbb{K}$. В вещественном случае, т.е. $\mathbb{K} = \mathbb{R}$, можно ввести следующие определения.

9.12. 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 $0 < t < 1$.

9.13. 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 $[M_0M_1]$:

$$[M_0M_1] \stackrel{\text{def}}{=\!=} \left\{ M \;\middle|\; \overrightarrow{M_0M} = t\overrightarrow{M_0M_1},\ 0 \leqslant t \leqslant 1 \right\}.$$

D. Пересекающиеся плоскости.

9.14. Предложение. *Если две плоскости пересекаются, то их пересечением является плоскость.*

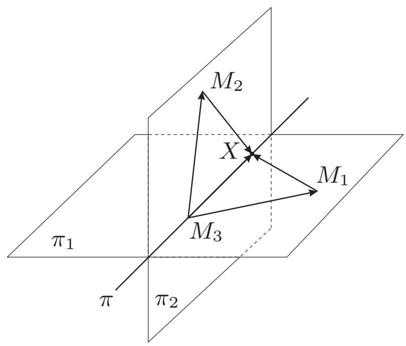


Рис. 9.4. Пересечение плоскостей

Доказательство. Рассмотрим в аффинном пространстве 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

$$X \in \pi_1 \iff \overrightarrow{M_1X} \in \mathcal{P}_1, \qquad X \in \pi_2 \iff \overrightarrow{M_2X} \in \mathcal{P}_2.$$

Далее,

$$\left.\begin{aligned} \overrightarrow{M_3X} &= \underbrace{\overrightarrow{M_3M_1}}_{\in\mathcal{P}_1} + \underbrace{\overrightarrow{M_1X}}_{\in\mathcal{P}_1} \in \mathcal{P}_1 \\ \overrightarrow{M_3X} &= \underbrace{\overrightarrow{M_3M_2}}_{\in\mathcal{P}_2} + \underbrace{\overrightarrow{M_2X}}_{\in\mathcal{P}_2} \in \mathcal{P}_2 \end{aligned}\right\} \quad \Rightarrow \quad \overrightarrow{M_3X} \in \mathcal{P}_1 \cap \mathcal{P}_2,$$

что и требовалось. □

E. Параллельные плоскости.

9.15. 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 $\mathcal{P}_1 \Subset \mathcal{P}_2$.

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).

9.16. 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 $\pi_2 = \pi_2(M_2, \mathcal{P}_2)$:

$$A \in \pi_1 \iff \overrightarrow{M_1A} \in \mathcal{P}_1, \qquad A \in \pi_2 \iff \overrightarrow{M_2A} \in \mathcal{P}_2.$$

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

$$\overrightarrow{AX} = \underbrace{\overrightarrow{M_1X}}_{\in \mathcal{P}_1} - \underbrace{\overrightarrow{M_1A}}_{\in \mathcal{P}_1} \in \mathcal{P}_1 \Subset \mathcal{P}_2 \quad \Rightarrow \quad \overrightarrow{AX} \in \mathcal{P}_2.$$

Далее,

$$\overrightarrow{M_2X} = \underbrace{\overrightarrow{M_2A}}_{\in \mathcal{P}_2} + \underbrace{\overrightarrow{AX}}_{\in \mathcal{P}_2} \in \mathcal{P}_2,$$

т.е. $X \in \pi_2$. □

F. Скрещивающиеся плоскости. Скрещивающиеся плоскости — это плоскости, которые не пересекаются и не параллельны.

9.17. Определение. Рассмотрим в аффинном пространстве $(\mathcal{A}, \mathcal{V})$ две непараллельные плоскости $\pi_1(M_1, \mathcal{P}_1)$ и $\pi_2(M_2, \mathcal{P}_2)$ размерностей p_1 и p_1 соответственно, причём $p_1 \leqslant p_2$. Непараллельность означает, что $\mathcal{P}_1 \not\Subset \mathcal{P}_2$. Если эти плоскости не имеют общих точек, будем говорить, что они *скрещиваются вдоль подпространства* $\mathcal{P}_1 \cap \mathcal{P}_2$.

Так как $\mathcal{P}_1 \not\Subset \mathcal{P}_2$, то $\dim(\mathcal{P}_1 \cap \mathcal{P}_2) < p_1$; поскольку для любых подпространств выполняется соотношение

$$\dim(\mathcal{P}_1 + \mathcal{P}_2) = \dim \mathcal{P}_1 + \dim \mathcal{P}_2 - \dim(\mathcal{P}_1 \cap \mathcal{P}_2),$$

получаем

$$\dim(\mathcal{P}_1 \cap \mathcal{P}_2) = p_1 + p_2 - \dim(\mathcal{P}_1 + \mathcal{P}_2) \geqslant p_1 + p_2 - n.$$

Итак, в n-мерном аффинном пространстве две плоскости размерностей p_1 и p_2, $p_1 \leqslant p_2$, могут скрещиваться вдоль подпространства, размерность которого заключена в пределах

$$\max\big(p_1 + p_2 - n,\ 0\big) \leqslant \dim(\mathcal{P}_1 \cap \mathcal{P}_2) \leqslant p_1 - 1.$$

В двумерном аффинном пространстве (на плоскости) не бывает скрещивающихся 1-плоскостей, т.е. прямых (подробнее о взаимном расположении прямых на плоскости см. в разделе 4B).

В трёхмерном аффинном пространстве существуют 1-плоскости, скрещивающиеся вдоль нульмерных направлений — обычные скрещивающиеся прямые.

G. Параметрическое уравнение плоскости. 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 $\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p\right)$ (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. определение 9.7) в виде

$$\boldsymbol{r} = \boldsymbol{r}_0 + t^1\boldsymbol{a}_1 + \cdots + t^p\boldsymbol{a}_p. \tag{9.1}$$

При изменении параметров $t^1, \ldots, t^p$ в поле $\mathbb{K}$ точка M с радиус-вектором $\boldsymbol{r}$, задаваемым формулой (9.1), пробегает всю рассматриваемую p-плоскость, поэтому равенство (9.1) называется *векторным параметрическим уравнением плоскости.*

Если ввести в аффинном пространстве аффинную систему координат $O\boldsymbol{e}_1 \ldots \boldsymbol{e}_n$, то векторное параметрическое уравнение плоскости (9.1) примет вид

$$X = X_0 + t^1 A_1 + \cdots + t^p A_p, \tag{9.2}$$

где X, X_0 — столбцы координат точек M и M_0 (или, эквивалентно, радиус-векторов $\boldsymbol{r}$ и $\boldsymbol{r}_0$), $A_1, \ldots, A_p$ — столбцы координат направляющих векторов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p$:

$$X = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix}, \quad X_0 = \begin{pmatrix} x_0^1 \\ x_0^2 \\ \vdots \\ x_0^n \end{pmatrix}, \quad A_1 = \begin{pmatrix} a_1^1 \\ a_1^2 \\ \vdots \\ a_1^n \end{pmatrix}, \quad \ldots, \quad A_p = \begin{pmatrix} a_p^1 \\ a_p^2 \\ \vdots \\ a_p^n \end{pmatrix}.$$

В развёрнутой координатной форме уравнение (9.2) имеет вид системы

$$\begin{cases} x^1 = x_0^1 + t^1 a_1^1 + \cdots + t^k a_k^1, \\ x^2 = x_0^2 + t^1 a_1^2 + \cdots + t^k a_k^2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x^n = x_0^n + t^1 a_1^n + \cdots + t^k a_k^n, \end{cases} \qquad t^j \in \mathbb{K},\ j = 1, 2, \ldots, p.$$

В частности, уравнение прямой (одномерной плоскости) можно записать в следующих формах:

(i) векторное параметрическое уравнение

$$\boldsymbol{r} = \boldsymbol{r}_0 + t\boldsymbol{a}, \tag{9.3}$$

где $\boldsymbol{r}_0$ — радиус-вектор опорной точки M_0, $\boldsymbol{a}$ — направляющий вектор;

(ii) система параметрических уравнений

$$\begin{cases} x^1 = x_0^1 + ta^1, \\ x^2 = x_0^2 + ta^2, \\ \dots\dots\dots\dots \\ x^n = x_0^n + ta^n, \end{cases} \tag{9.4}$$

где $(x^1, \dots, x^n)^T$ — координаты вектора $\boldsymbol{r}$, $(x_0^1, \dots, x_0^n)^T$ — координаты вектора $\boldsymbol{r}_0$, $(a^1, \dots, a^n)^T$ — координаты вектора $\boldsymbol{a}$.

H. Задание плоскости системой уравнений. Структура параметрического уравнения плоскости (9.1) или (9.2) в точности совпадает со структурой общего решения 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.

9.18. Теорема. *В n-мерном аффинном пространстве $\mathcal{A}$ в любых аффинных координатах всякая p-плоскость может быть задана неоднородной системой линейных уравнений ранга*[1] $r = n - p$.

Замечание. 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 (9.1) или (9.2). 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 $r = n - p$ (см.

[1]Под рангом системы подразумевается ранг её матрицы, основной или расширенной, которые для совместной системы

п. D лекции 7, с. D), 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

$$A(X - X_0) = 0 \iff AX = B, \quad \text{где } B = AX_0.$$

Поскольку X — столбец координат произвольной точки плоскости π в системе координат $O\boldsymbol{E}$, получаем, что найденная система $AX = B$ и определяет плоскость π. □

9.19. Замечание. Таким образом, аналогично подпространствам в пространстве $\mathbb{K}^n$ (и, следовательно, используя изоморфизмы, в любом векторном пространстве; см. замечание 5.14, с. 104), плоскости в аффинных пространствах можно задавать двумя взаимно двойственными способами:

(i) задавая опорную точку и базис направляющего подпространства;
(ii) задавая неоднородную систему линейных уравнений.

I. Гиперплоскости. Гиперплоскости — это плоскости коразмерности 1, поэтому они могут быть заданы неоднородной системой ранга 1, т.е. одним уравнением вида

$$A_1x^1 + A_2x^2 + \cdots + A_nx^n = B. \tag{9.5}$$

J. Полупространства. В этом разделе мы рассматриваем *вещественное* аффинное пространство $\mathcal{A}(\mathbb{R})$.

Рассмотрим некоторую гиперплоскость π в пространстве $\mathcal{A}$.

9.20. Определение. 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.

Найдём аналитическое условие того, что данная гиперплоскость π разделяет данные точки. Пусть в некоторой аффинной системе координат гиперплоскость π задана уравнением (9.5), а точки M_1 и M_2 имеют координаты $M_1(x_1^1, \ldots, x_1^n)$ и $M_2(x_2^1, \ldots, x_2^n)$. Для сокращения записи введём обозначение

$$F(M) \equiv F(x^1, \ldots, x^n) = A_1x^1 + A_2x^2 + \cdots + A_nx^n - B,$$

подразумевая, что точка M имеет координаты $(x^1, \ldots, x^n)$.

9.21. Теорема. *Точки M_1 и M_2, не принадлежащие гиперплоскости* (9.5)*, 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 $F(M) = 0$:

$$0 = F(M) = F\Big(M_1 + t(M_2 - M_1)\Big) = F(M_1) + t\Big(F(M_2) - F(M_1)\Big)$$

$$t_0 = \frac{F(M_1)}{F(M_1) - F(M_2)}, \tag{9.6}$$

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 (9.5).)

По определению, точки M_1 и M_2 разделены гиперплоскостью (9.5), если число (9.6) существует и лежит в промежутке $[0;1]$, т.е.

$$0 \leqslant \frac{F(M_1)}{F(M_1) - F(M_2)} \leqslant 1 \iff \begin{cases} \dfrac{F(M_1)}{F(M_1) - F(M_2)} \geqslant 0, \\ \dfrac{F(M_2)}{F(M_1) - F(M_2)} \leqslant 0. \end{cases}$$

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 $F(M_1)F(M_2) \leqslant 0$. □

9.22. Замечание. 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_-. В частности, прямая разбивает плоскость на две полуплоскости.

3. Ориентация в аффинной геометрии

9.23. Определение. Аффинное пространство называется *ориентированным*, если ориентировано его ассоциированное векторное пространство (см. определение **??**).

Плоскости в аффинном пространстве сами являются аффинными пространствами меньшей размерности, поэтому возникает вопрос о согласовании их ориентаций.

Без дополнительных данных ориентация аффинного пространства $(\mathcal{A}, \mathbb{R})$ и ориентация лежащей в нём гиперплоскости π никак друг с другом не связаны. Связь между этими ориентациями устанавливается при помощи выбора одной из двух сторон гиперплоскости π, т.е. одного из двух полупространств, на которые гиперплоскость разбивает пространство (см. п. J, с. 198).

Пусть $\boldsymbol{n}$ — произвольный вектор, не лежащий в направляющем подпространстве $\mathcal{P}$ гиперплоскости π. Выберем на π произвольную точку M_0 и отложим от неё вектор $\boldsymbol{n}$: $N_0 = M_0 + \boldsymbol{n}$ (т.е. $\overrightarrow{M_0N_0} = \boldsymbol{n}$). Так как $\boldsymbol{n} \notin \boldsymbol{P}$, то $N_0 \notin \pi$, т.е. N_0 лежит в одном из двух полупространств, на которые гиперплоскость π разбивает пространство $\mathcal{A}$, т.е. определяет одну из двух сторон гиперплоскости π. Легко доказать, что эта сторона не зависит от выбора точки M_0, т.е. определяется исключительно вектором $\boldsymbol{n}$, про который говорят, что он направлен *в эту сторону* (или внутрь соответствующего подпространства). Действительно, пусть $M_1 \in \pi$ — другая точка гиперплоскости π и $N_1 = M_1 + \boldsymbol{n}$; тогда

$$\overrightarrow{N_0N_1} = \underbrace{\overrightarrow{N_0M_0}}_{=\boldsymbol{n}} + \overrightarrow{M_0M_1} + \underbrace{\overrightarrow{M_1N_1}}_{=-\boldsymbol{n}} = \overrightarrow{M_0M_1} \in \mathcal{P},$$

т.е. прямая (N_0N_1) параллельна гиперплоскости π, так что точки N_0 и N_1 лежат по одну сторону гиперплоскости π.

9.24. Предложение. *Пусть $\boldsymbol{a} \in \mathcal{P}$ — произвольный вектор из направляющего подпространства гиперплоскости π, $\boldsymbol{n}$ — вектор, определяющий одну из сторон этой гиперплоскости. Вектор $\boldsymbol{n}' = \boldsymbol{a} + \alpha\boldsymbol{n}$ определяет ту же сторону гиперплоскости π тогда и только тогда, когда $\alpha > 0$.*

Доказательство. Для доказательства нужно убедиться, что точки N_0 и N_1, полученные сдвигом на векторы $\boldsymbol{n}$ и $\boldsymbol{n}'$ соответственно из произвольных точек $M_0 \in \pi$ и $M_1 \in \pi$, тогда и только тогда лежат в одном полупространстве, когда $\alpha > 0$.

Выше было доказано, что указание полупространства не зависит от выбора точки M_0, поэтому отложим вектор $\boldsymbol{n}$ от произвольной точки M_0, т.е. $N_0 = M_0 + \boldsymbol{n}$, а вектор $\boldsymbol{n}' = \boldsymbol{a} + \alpha\boldsymbol{n}$ — от точки $M_1 = M_0 - \boldsymbol{a}$, т.е.

$$\begin{aligned} N_1 &= M_1 + \boldsymbol{n}' = \\ &= (M_0 - \boldsymbol{a}) + (\boldsymbol{a} + \alpha\boldsymbol{n}) = \\ &= M_0 + \alpha\boldsymbol{n}. \end{aligned}$$

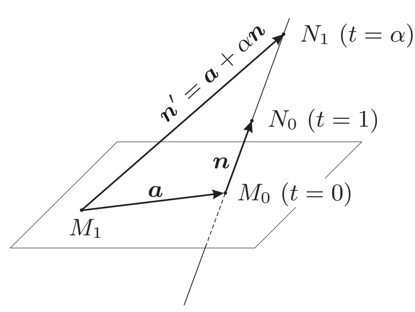


Рис. 9.5. Согласование ориентаций

Ясно, что точка N_1 лежит на прямой (M_0N_0), уравнение которой имеет вид $M = M_0 + t\boldsymbol{n}$: ей соответствует значение параметра $t = \alpha$.

Точкам M_0 и N_0 соответствуют значения параметра $t = 0$ и $t = 1$. Поэтому точки N_0 и N_1 лежат по одну сторону гиперплоскости π в том и только том случае, когда $\alpha > 0$. □

Пусть теперь задана ориентация гиперплоскости π, т.е. в направляющем подпространстве $\mathcal{P}$ этой гиперплоскости указан базис $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-1}\big)$, называемый положительным. Выберем одну из сторон этой гиперплоскости, указав вектор $\boldsymbol{n} \notin \mathcal{P}$. Тогда в пространстве $\mathcal{A}$

возникает базис $\boldsymbol{F} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-1}, \boldsymbol{n}\big)$. Ориентация пространства, задаваемая этим базисом, не зависит от выбора векторов $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-1}, \boldsymbol{n}$, а определяется лишь ориентациями, которые задают эти векторы. Действительно, пусть базис $\boldsymbol{E}' = \big(\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_{n-1}\big)$ задаёт ту же ориентацию гиперплоскости π, что и базис $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-1}\big)$, т.е. матрица перехода C от базиса $\boldsymbol{E}$ к базису $\boldsymbol{E}'$ ($\boldsymbol{E}' = \boldsymbol{E}C$) имеет положительный определитель, и пусть вектор $\boldsymbol{n}'$ задаёт ту же сторону гиперплоскости π, что и $\boldsymbol{n}$, т.е. $\boldsymbol{n}' = \boldsymbol{a} + \alpha\boldsymbol{n}$, где $\boldsymbol{a}$ — некоторый вектор из $\mathcal{P}$ и $\alpha > 0$. Найдём определитель матрицы перехода от базиса $\big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-1}, \boldsymbol{n}\big)$ пространства $\mathcal{A}$ к базису $\big(\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_{n-1}, \boldsymbol{n}'\big)$:

$$\det \left(\begin{array}{c|c} C & * \\ \hline 0 & \alpha \end{array}\right) = \det C \cdot \alpha > 0.$$

Итак, *ориентация гиперплоскости π и выбор её стороны однозначно определяют ориентацию пространства $\mathcal{A}$.* Легко доказать следующее утверждение (читателю предлагается сделать это самостоятельно).

9.25. Предложение. *Любые два из трёх следующих объектов однозначно определяют третий:*

(a) *ориентация гиперплоскости π;*
(b) *сторона гиперплоскости π;*
(c) *ориентация пространства $\mathcal{A}$.*

Например, если задана ориентация пространства $\mathcal{A}$ и с помощью некоторого вектора $\boldsymbol{n}$ указана сторона гиперплоскости π, то соответствующая ориентация гиперплоскости задаётся базисом $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-1}\big)$ в её направляющем подпространстве, обладающим тем свойством, что базис $\boldsymbol{F} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-1}, \boldsymbol{n}\big)$ в пространстве $\mathcal{A}$ положительно ориентирован.

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 $\boldsymbol{i}, \boldsymbol{j}$.

A. Различные виды уравнений прямой на плоскости. Векторное параметрическое уравнение прямой имеет вид

$$\boldsymbol{r} = \boldsymbol{r}_0 + t\boldsymbol{a}. \tag{9.3}$$

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 уравнение (9.3) в виде

$$\begin{cases} x = x_0 + ta_x, \\ y = y_0 + ta_y. \end{cases} \tag{9.7}$$

Полученная система называется *системой параметрических уравнений прямой*. Для краткости часто говорят «*параметрическое уравнение прямой*», помня, что на самом деле уравнений два. При решении задач параметрическое уравнение прямой — один из наиболее удобных типов уравнений.

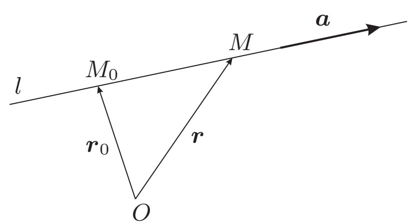


Рис. 9.6.

Попытаемся исключить из параметрического уравнения параметр t. Если обе координаты a_x и a_y направляющего вектора $\boldsymbol{a}$ рассматриваемой прямой отличны от нуля, то очевидным образом получаем

$$\frac{x - x_0}{a_x} = \frac{y - y_0}{a_y}. \tag{9.8}$$

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:

$$\frac{x - x_0}{a_x} = \frac{y - y_0}{a_y} \iff \frac{x - x_0}{a_x} = \frac{y - y_0}{a_y} = t \iff \begin{cases} x = x_0 + ta_x, \\ y = y_0 + ta_y. \end{cases}$$

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 нужно превратить в параметрическое по рецепту, указанному выше:

$$\frac{x - x_0}{0} = \frac{y - y_0}{a_y} \iff \frac{x - x_0}{0} = \frac{y - y_0}{a_y} = t \quad \Rightarrow \begin{cases} x = x_0, \\ y = y_0 + ta_y. \end{cases}$$

Можно поступить и иначе, преобразовав уравнение следующим образом:

$$\frac{x - x_0}{0} = \frac{y - y_0}{a_y} \iff a_y(x - x_0) = 0(y - y_0) \iff \begin{cases} x = x_0, \\ y\text{ — любой.} \end{cases}$$

Итак, каноническое уравнение прямой допускается записывать и в случае, когда одна из координат направляющего вектора прямой равна нулю.

Условие коллинеарности векторов $\overrightarrow{M_0M}$ и $\boldsymbol{a}$ может быть записано как равенство нулю определителя второго порядка, составленного из координат этих векторов:

$$\begin{vmatrix} x - x_0 & y - y_0 \\ a_x & a_y \end{vmatrix} = 0. \tag{9.9}$$

Это уравнение также называют *каноническим уравнением* прямой на плоскости.

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 $\boldsymbol{r}_2$.

На плоскости (размерность 2) прямая (размерность 1) является гиперплоскостью, поэтому (см. п. I) её можно задать уравнением вида

$$Ax + By = D, \tag{9.10}$$

называемым *общим уравнением прямой.* Общее уравнение легко получается из канонического:

$$\frac{x - x_0}{a_x} = \frac{y - y_0}{a_y} \quad \Longleftrightarrow \quad a_y(x - x_0) = a_x(y - y_0);$$

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 общее уравнение прямой (9.10). Ясно, что коэффициенты A и B одновременно не могут быть равны нулю: в противном случае уравнение не будет содержать переменных.

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 координатных осях[1]).

B. Взаимное расположение двух прямых на плоскости. Пусть на плоскости, т.е. в двумерном аффинном пространстве $\mathcal{A}$, две прямые заданы 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

$$\boldsymbol{r}_1 + t\boldsymbol{a}_1 = \boldsymbol{r}_2 + s\boldsymbol{a}_2 \iff t\boldsymbol{a}_1 - s\boldsymbol{a}_2 = \boldsymbol{r}_2 - \boldsymbol{r}_1.$$

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.

Если векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ линейно зависимы (коллинеарны), то рассматриваемые прямые параллельны. Если при этом $\boldsymbol{r}_2 - \boldsymbol{r}_1 \in L(\boldsymbol{a}_1) = L(\boldsymbol{a}_2)$, то эти параллельные прямые совпадают, поскольку имеют общую точку; 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. параллельные прямые различны.

Итак, доказано следующее утверждение.

9.26. Предложение (взаимное расположение двух прямых на плоскости). *Рассмотрим две прямые $\boldsymbol{r} = \boldsymbol{r}_1 + t\boldsymbol{a}_1$ и $\boldsymbol{r} = \boldsymbol{r}_2 + s\boldsymbol{r}_2$ на плоскости (т.е. в двумерном аффинном пространстве). Введём следующие матрицы, состоящие из координат векторов $\boldsymbol{r}_1(x_1, y_1)$, $\boldsymbol{a}(a_x, a_y)$, $\boldsymbol{r}_2(x_2, y_2)$, $\boldsymbol{a}_2(b_x, b_y)$ относительно произвольной аффинной системы координат:*

$$A = \begin{pmatrix} a_x & b_x \\ a_y & b_y \end{pmatrix}, \quad \tilde{A} = \begin{pmatrix} a_x & b_x & x_2 - x_1 \\ a_y & b_y & y_2 - y_1 \end{pmatrix}.$$

Возможны следующие варианты взаимного расположения этих прямых на плоскости:

(a) *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), что аналитически выражается условием* $\operatorname{rk} A = 2$*;*

(b) *прямые могут быть параллельными и 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; аналитически это выражается условиями* $\operatorname{rk} A = 1$ *и* $\operatorname{rk} \tilde{A} = 2$;

[1] Координатными осями называют прямые, проходящие через начало координат и параллельные базисным векторам.

(c) *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; аналитически это выражается условием* $\operatorname{rk} A = \operatorname{rk} \tilde{A} = 1$.

Поучительно провести аналогичный анализ в случае, когда прямые заданы как гиперплоскости общими уравнениями $A_1x + B_1y = D_1$ и $A_2x + B_2y = D_2$. Применяя к системе, состоящей из этих двух уравнений, теорему Кронекера—Капелли и результаты п. A, получим следующее утверждение.

9.27. Предложение. *Пусть на аффинной плоскости заданы две прямые*

$$A_1x + B_1y = D_1, \quad A_2x + B_2y = D_2.$$

Обозначим через r и R соответственно ранги основной и расширенной матриц этой системы:

$$r = \operatorname{rk}\begin{pmatrix} A_1 & B_1 \\ A_2 & B_2 \end{pmatrix}, \qquad R = \operatorname{rk}\begin{pmatrix} A_1 & B_1 & D_1 \\ A_2 & B_2 & D_2 \end{pmatrix}.$$

(a) *Данные прямые пересекаются в единственной точке тогда и только тогда, когда $r = R = 2$. В этом случае координаты точки пересечения (x_0, y_0) выражаются формулами Крамера:*

$$x_0 = \frac{\begin{vmatrix} D_1 & B_1 \\ D_2 & B_2 \end{vmatrix}}{\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}}, \quad y_0 = \frac{\begin{vmatrix} A_1 & D_1 \\ A_2 & D_2 \end{vmatrix}}{\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}}.$$

(b) *Прямые не имеют общих точек тогда и только тогда, когда $r = 1$ и $R = 2$.*

(c) *Прямые совпадают тогда и только тогда, когда $r = R = 1$.*

C. Пучок прямых. *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 аналитической геометрии часто 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.

9.28. Предложение. *Пусть $A_1x + B_1y = D_1D_1$ и $A_2x + B_2y = D_2$ — уравнения двух прямых, пересекающихся в точке $C(x_0, y_0)$. Прямая принадлежит пучку $\Pi(C)$ тогда и только тогда, когда она задаётся уравнением*

$$\alpha(A_1x + B_1y) + \beta(A_2x + B_2y) = \alpha D_1 + \beta D_2, \tag{9.11}$$

где α и β — числа, не равные одновременно нулю.

Доказательство. Достаточность. Рассмотрим уравнение (9.11). 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 (9.11) 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 (9.11). Следовательно, при любых α и β, не равных одновременно нулю, уравнение (9.11) 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 (9.11) с подходящими α и β.

Подставляя в (9.11) координаты (x_1, y_1) точки M_1, получим

$$\alpha(A_1x_1 + B_1y_1) + \beta(A_2x_1 + B_2y_1) = \alpha D_1 + \beta D_2 \iff$$
$$\iff \alpha\bigl[A_1x_1 + B_1y_1 - D_1\bigr] + \beta\bigl[A_2x_1 + B_2y_1 - D_2\bigr] = 0.$$

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, видим, что получающееся уравнение описывает прямую, проходящую через точку M_1. □

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

$$\text{собственный пучок} \;\mapsto\; \text{его центр}$$

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 были параллельны. 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

В этом разделе мы также используем терминологию элементарной геометрии: трёхмерное аффинное пространство называем пространством, 1-плоскости — прямыми, 2-плоскости (гиперплоскости) — плоскостями. 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).

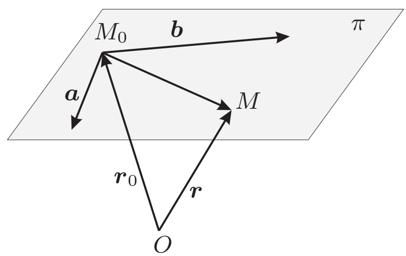


Рис. 9.7.

А. Уравнения плоскостей.
Плоскость в пространстве может быть задана следующими способами.

1. *Векторное параметрическое уравнение* (см. рис. 9.7)

$$\boldsymbol{r} = \boldsymbol{r}_0 + t\boldsymbol{a} + s\boldsymbol{b}, \tag{9.12}$$

где $\boldsymbol{r}_0$ — радиус-вектор опорной точки плоскости, $\boldsymbol{a}$ и $\boldsymbol{b}$ — линейно независимые направляющие векторы. В координатной записи уравнение (9.12) превращается в систему параметрических уравнений

$$\begin{cases} x = x_0 + ta_x + sb_x, \\ y = y_0 + ta_y + sb_y, \\ z = z_0 + ta_z + sb_z. \end{cases} \tag{9.13}$$

2. *Общее уравнение*:

$$Ax + By + Cz = D. \tag{9.14}$$

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, очевидно, следующий: рассматриваемая 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.

Чтобы перейти от общего уравнения к параметрическому, нужно найти общее решение неоднородной системы (9.14). Переход от параметрического задания 9.13 к уравнению (9.14) удобнее всего осуществить следующим образом. Уравнение (9.12) выражает факт линейной зависимости векторов $\boldsymbol{r}-\boldsymbol{r}_0$, $\boldsymbol{a}$ и $\boldsymbol{b}$, который может быть записан как равенство нулю определителя, составленного из координат этих векторов:

$$\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix} = 0. \tag{9.15}$$

Раскрывая определитель, получим уравнение вида (9.14); само же уравнение (9.15) будем по аналогии с уравнением 9.9 называть *каноническим уравнением плоскости.*

B. Свойства плоскостей в пространстве.

9.29. Предложение. *Существует единственная плоскость, проходящая через любые три точки M_1, M_2, M_3, не лежащие на одной прямой.*

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, очевидно, (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.

9.30. 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

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ a_x & a_y & a_z \end{vmatrix} = 0.$$

□

C. Уравнения прямых в пространстве. Векторное параметрическое уравнение прямой $\boldsymbol{r} = \boldsymbol{r}_0 + t\boldsymbol{a}$ в координатах имеет вид

$$\begin{cases} x = x_0 + ta_x, \\ y = y_0 + ta_y, \\ z = z_0 + ta_z \end{cases}$$

и называется системой *параметрических уравнений прямой*.

Исключение параметра t из параметрических уравнений приводит к *каноническому уравнению 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. Нули в знаменателях трактуются так же, как в планиметрическом случае (см. с. 202). В условиях задач прямые часто задаются каноническими уравнениями, однако при решении удобно преобразовать каноническое уравнение в параметрическое следующим образом:

$$\frac{x - x_0}{a_x} = \frac{y - y_0}{a_y} = \frac{z - z_0}{a_x} = t \quad \iff \quad \begin{cases} x = x_0 + ta_x, \\ y = y_0 + ta_y, \\ z = z_0 + ta_z. \end{cases}$$

Кроме того, прямая может быть 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; при этом частное решение системы представляет некоторую опорную точку прямой, а фундаментальное семейство решений соответствующей однородной системы, состоящее из одного вектора, — направляющий вектор.

D. Взаимное расположение плоскостей и прямых в пространстве. Доказательства следующих утверждений настолько подготовлены предыдущим, что не нуждаются в изложении.

Взаимное расположение двух плоскостей в пространстве. Рассмотрим две плоскости в пространстве, заданные уравнениями

$$A_1x + B_1y + C_1z = D_1, \quad A_2x + B_2y + C_2z = D_2.$$

Обозначим через r и R соответственно ранги следующих матриц, которые являются основной и расширенной матрицами неоднородной системы, состоящей из уравнений данных плоскостей:

$$r = \operatorname{rk}\begin{pmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{pmatrix}, \quad R = \operatorname{rk}\begin{pmatrix} A_1 & B_1 & C_1 & D_1 \\ A_2 & B_2 & C_2 & D_2 \end{pmatrix}.$$

В зависимости от соотношения чисел r и R возможны следующие ситуации (см. рис. 9.8).

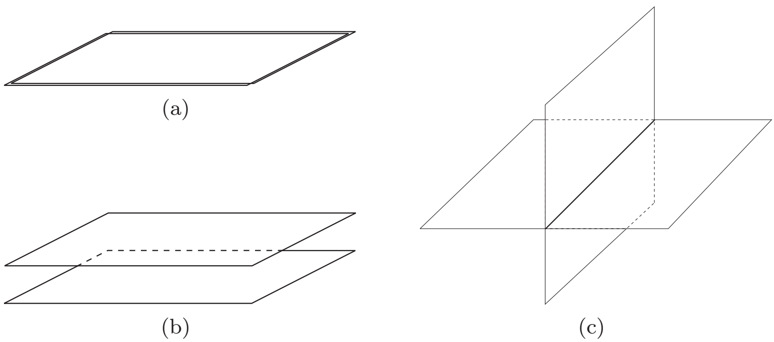


Рис. 9.8. Взаимное расположение двух плоскостей в пространстве

(a) $r = 1$, $R = 1$. Система совместна, множество её решений двумерно: плоскости параллельны и совпадают.
(b) $r = 1$, $R = 2$. Система несовместна: плоскости параллельны и не имеют общих точек.
(c) $r = 2$, $R = 2$. Система совместна, множество её решений одномерно: плоскости пересекаются, причём пересечением является прямая (ср. с предложением 9.14).

Взаимное расположение двух прямых в пространстве. Рассмотрим две прямые

$$\boldsymbol{r} = \boldsymbol{p} + t\boldsymbol{a}, \quad \boldsymbol{r} = \boldsymbol{q} - s\boldsymbol{b}$$

(для удобства в дальнейшем направляющий вектор второй прямой выбран в виде $-\boldsymbol{b}$). Общие точки прямых определяются из условия

$$\boldsymbol{p} + t\boldsymbol{a} = \boldsymbol{q} - s\boldsymbol{b} \iff t\boldsymbol{a} + s\boldsymbol{b} = \boldsymbol{q} - \boldsymbol{p}.$$

В произвольной аффинной системе координат это векторное уравнение можно записать в виде неоднородной системы линейных уравнений

$$t\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + s\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} q_1 - p_1 \\ q_2 - p_2 \\ q_3 - p_3 \end{pmatrix}.$$

Обозначим через r и R ранги основной и расширенной матриц этой системы:

$$r = \operatorname{rk}\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{pmatrix}, \quad R = \operatorname{rk}\begin{pmatrix} a_1 & b_1 & q_1 - p_1 \\ a_2 & b_2 & q_2 - p_2 \\ a_3 & b_3 & q_3 - p_3 \end{pmatrix}.$$

В зависимости от соотношения чисел r и R возможны следующие ситуации (см. рис. 9.9):

(a) $r = R = 1$. Система совместна, множество её решений одномерно: прямые совпадают.

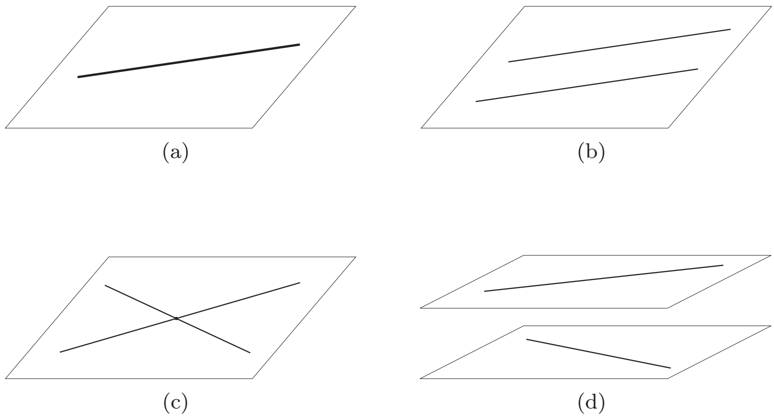


Рис. 9.9. Взаимное расположение двух прямых в пространстве

(b) $r = 1$, $R = 2$. Система несовместна, однако столбцы основной матрицы системы линейно зависимы, т.е. направляющие векторы прямых коллинеарны: прямые параллельны и не имеют общих точек. Эти прямые лежат в одной плоскости: в качестве её опорной точки можно взять произвольную точку любой из прямых, а в качестве направляющих векторов — направляющий вектор одной из прямых и вектор $\boldsymbol{q} - \boldsymbol{p}$, соединяющий опорные точки прямых.

(c) $r = R = 2$. Система совместна и её решение единственно: прямые пересекаются. Пересекающиеся прямые также лежат в одной плоскости: её опорной точкой может служить произвольная точка каждой из прямых, а направляющими векторами — направляющие векторы прямых.

(d) $r = 2$, $R = 3$. Система несовместна, столбцы основной матрицы системы линейно независимы, т.е. направляющие векторы прямых неколлинеарны: прямые скрещиваются. Скрещивающиеся прямые содержатся в параллельных плоскостях, направляющие векторы которых суть направляющие векторы этих прямых.

Взаимное расположение прямой и плоскости в пространстве. Пусть плоскость задана уравнением

$$Ax + By + Cz = D,$$

а прямая — системой уравнений

$$\begin{cases} A_1x + B_1y + C_1z = D_1, \\ A_2x + B_2y + C_2z = D_2, \end{cases}$$

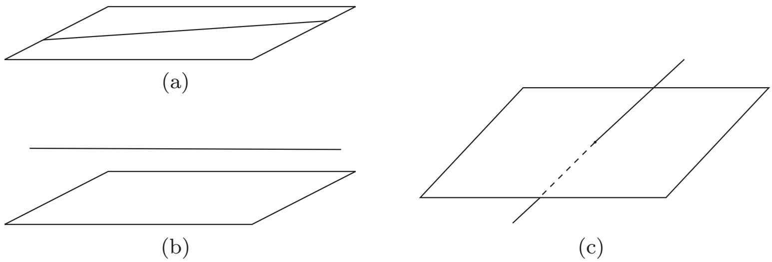


Рис. 9.10. Взаимное расположение прямой и плоскости в пространстве

ранг которой[1] равен двум. Анализ взаимного расположения прямой и плоскости сводится к исследованию системы

$$\begin{cases} A_1x + B_1y + C_1z = D_1, \\ A_2x + B_2y + C_2z = D_2, \\ \quad Ax + By + Cz = D. \end{cases}$$

Обозначим ранги основной и расширенной матриц этой системы через r и R соответственно; при этом $r \geqslant 2$.

Возможны следующие ситуации.

(a) $r = R = 2$. Система совместна и множество её решений одномерно: прямая параллельна плоскости и лежит в ней.

(b) $r = 2$, $R = 3$. Система несовместна: прямая и плоскость параллельны и не имеют общих точек.

(c) $r = R = 3$. Система совместна и имеет единственное решение: прямая пересекает плоскость в единственной точке.

Аналогичный результат можно получить, если задать плоскость уравнением $Ax + By + Cz = D$, а прямую — параметрическими уравнениями

$$\begin{cases} x = x_0 + ta_x, \\ y = y_0 + ta_y, \\ z = z_0 + ta_z. \end{cases} \tag{9.16}$$

Подставляя (9.16) в уравнение плоскости, получим

$$A(x_0 + ta_x) + B(y_0 + ta_y) + C(z_0 + ta_z) = D \iff \\ \iff t(Aa_x + Ba_y + Ca_z) = D - (Ax_0 + By_0 + Cz_0).$$

В случае $Aa_x + Ba_y + Ca_z \neq 0$ это уравнение имеет единственное решение

$$t_0 = \frac{D - (Ax_0 + By_0 + Cz_0)}{Aa_x + Ba_y + Ca_z},$$

[1] Поскольку система совместна, ранги её основной и расширенной матриц совпадают; это число мы и называем рангом системы.

подставив которое в (9.16), найдём координаты единственной точки пересечения прямой и плоскости.

Если же $Aa_x + Ba_y + Ca_z = 0$, то возможны следующие случаи:

(a) 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);

(b) 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.

Докажем, что в обоих этих случаях прямая и плоскость параллельны. Действительно, однородное уравнение $Ax + By + Cz = 0$ описывает направляющее подпространство плоскости, а соотношение $Aa_x + Ba_y + Ca_z = 0$ означает, что вектор с координатами (a_x, a_y, a_z) принадлежит этому подпространству, что и означает параллельность прямой и плоскости (см. определение 9.15).

E. Пучок плоскостей. *Пучком плоскостей* в пространстве называется множество всех плоскостей, проходящих через некоторую прямую l, называемую *осью* пучка. Пучок с центром l будем обозначать $\Pi(l)$.

9.31. 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

$$\alpha\big(A_1x + B_1y + C_1z\big) + \beta\big(A_2x + B_2y + C_2z\big) = \alpha D_1 + \beta D_2, \tag{9.17}$$

где α и β — числа, не равные одновременно нулю.

Доказательство вполне аналогично доказательству предложения 9.28.

Кроме того, здесь уместно было бы привести рассуждения, приводящие к понятию проективного пространства, однако ограничение объёма не позволяет нам этого сделать.

ГЛАВА 10

Евклидово пространство

ЕВКЛИДОВО пространство — это векторное пространство, в котором имеется операция скалярного умножения векторов, позволяющая ввести понятия длины (нормы) векторов и угла между векторами. Евклидовым является пространство, доступное нашему чувственному восприятию, что объясняет огромную роль, которую играет это понятие во всех разделах математики и физики.

1. Скалярное произведение векторов

А. Мотивировка определения. В разделе 5 гл. III было сформулировано элементарно-геометрическое определение скалярного произведения векторов как произведения длин этих векторов на косинус угла между ними. Важность этого понятия обусловлена не только той ролью, которое оно играет в геометрии, но и многочисленными его применениями в физике: например, как хорошо известно читателю из школьного курса, работа постоянной силы $\boldsymbol{F}$ при прямолинейном перемещении материальной точки на вектор $\boldsymbol{s}$ равна скалярному произведению этих векторов $(\boldsymbol{F}, \boldsymbol{s})$.

При изложении геометрии на основе векторно-точечной аксиоматики скалярное произведение является тем основным средством, которое позволяет ввести в геометрию измерение длин, углов, площадей и т. п. Вспоминая известные из элементарной геометрии (см. раздел 5 гл. III) факты, касающиеся этого понятия, и обращая точку зрения, т.е. принимая свойства скалярного произведения, установленные в рамках элементарно-геометрического подхода, за аксиомы, мы придём к общему определению скалярного произведения, пригодному не только для геометрических целей, но и используемое повсеместно в математическом обиходе.

Напомним, что в рамках элементарной геометрии скалярное произведение $(\boldsymbol{a}, \boldsymbol{b})$ векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ определялось как число, равное произведению длин векторов на косинус угла между ними:

$$(\boldsymbol{a}, \boldsymbol{b}) = |\boldsymbol{a}| \cdot |\boldsymbol{b}| \cdot \cos\varphi.$$

Из этого определения в лекции III были выведены основные свойства скалярного произведения (см. теорему 2.18).

B. Определение скалярного произведения и евклидова пространства. В аксиоматической теории мы обращаем полученные результаты и принимаем свойства скалярного произведения, установленные в теореме 2.18, за аксиомы.

Итак, пусть $(\mathcal{V}, \mathbb{R})$ — произвольное векторное пространство над полем *вещественных* чисел.[1]

10.1. Определение. *Скалярным умножением* на векторном пространстве $\mathcal{V}$ называется бинарная операция на $\mathcal{V}$
$$g\colon \mathcal{V} \times \mathcal{V} \to \mathbb{R}, \quad (\boldsymbol{x}, \boldsymbol{y}) \mapsto g(\boldsymbol{x}, \boldsymbol{y}),$$
обладающая следующими свойствами:

СП1: линейность по каждому аргументу:
$$g(\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2, \boldsymbol{y}) = \alpha_1 g(\boldsymbol{x}_1, \boldsymbol{y}) + \alpha_2 g(\boldsymbol{x}_2, \boldsymbol{y}),$$
$$g(\boldsymbol{x}, \beta_1 \boldsymbol{y}_1 + \beta_2 \boldsymbol{y}_2) = \beta_1 g(\boldsymbol{x}, \boldsymbol{y}_1) + \beta_2 g(\boldsymbol{x}, \boldsymbol{y}_2);$$

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 $(\boldsymbol{x}, \boldsymbol{y})$.

Перечисленные свойства **СП1–СП3** называются *аксиомами скалярного умножения* и образуют *третью группу аксиом Вейля.*

10.2. Определение. *Евклидовым векторным пространством* называется упорядоченная пара $(\mathcal{V}, g)$, где $\mathcal{V}$ — вещественное векторное пространство, g — скалярное умножение на нём.

Евклидовым точечным пространством называется упорядоченная пара $(\mathcal{A}, g)$, где $\mathcal{A}$ — вещественное аффинное пространство, g — скалярное умножение на его ассоциированном векторном пространстве.

Как векторные, так и точечные евклидовы пространства будем обозначать символом $\mathcal{E}$; тип пространства (векторное или точечное) будет ясен из контекста.

10.3. Определение. *Длиной вектора* $\boldsymbol{x}$ в евклидовом векторном пространстве называется число
$$|\boldsymbol{x}| = \sqrt{(\boldsymbol{x}, \boldsymbol{x})}. \tag{10.1}$$
Расстоянием между точками A *и* B в евклидовом точечном пространстве называется число
$$|AB| = \bigl|\overrightarrow{AB}\bigr| = \sqrt{\bigl(\overrightarrow{AB}, \overrightarrow{AB}\bigr)}. \tag{10.2}$$

Аксиома **СП3** обеспечивает, 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.

[1] Понятие скалярного произведения может быть введено и для векторных пространств над полем комплексных чисел; это будет следано в следующем семестре.

10.4. Замечание. В векторных пространствах, элементами которых являются функции (см. примеры 8.22–8.24, с. 169), 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 $\|\boldsymbol{x}\|$.

10.5. Определение. *Углом между векторами $\boldsymbol{x}$ и $\boldsymbol{y}$* в евклидовом векторном пространстве называется число

$$\varphi = \arccos \frac{(\boldsymbol{x}, \boldsymbol{y})}{|\boldsymbol{x}| \cdot |\boldsymbol{y}|}. \tag{10.3}$$

Углом между прямыми, имеющими направляющие векторы $\boldsymbol{a}$ и $\boldsymbol{b}$, в евклидовом точечном пространстве называется число

$$\varphi = \arccos \frac{|(\boldsymbol{a}, \boldsymbol{b})|}{|\boldsymbol{a}| \cdot |\boldsymbol{b}|}. \tag{10.4}$$

10.6. Замечание. Обратите внимание на различие формул (10.3) и (10.4): угол *между векторами* принимает значения $\varphi \in [0, \pi]$, а угол *между прямыми* — значения $\varphi \in [0, \pi/2]$.

Для формул (10.3) и (10.4) требует доказательства тот факт, что выражение под знаком арккосинуса не превышает по модулю единицы. Этот факт является следствием следующего важного утверждения.

10.7. Теорема (неравенство Коши—Буняковского—Шварца)**.** *Для любых $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{E}$ имеет место неравенство*

$$(\boldsymbol{x}, \boldsymbol{y})^2 \leqslant (\boldsymbol{x}, \boldsymbol{x}) \cdot (\boldsymbol{y}, \boldsymbol{y}).$$

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:

$$D = 4(\boldsymbol{x}, \boldsymbol{y}) - 4(\boldsymbol{x}, \boldsymbol{x}) \cdot (\boldsymbol{y}, \boldsymbol{y}),$$

откуда вытекает требуемое неравенство. $\square$

10.8. Предложение.

1. *Для любых $\boldsymbol{x} \in \mathcal{E}$ и $\alpha \in \mathbb{R}$ имеет место формула*[1]

$$|\alpha \boldsymbol{x}| = |\alpha| \cdot |\boldsymbol{x}|.$$

2. *Для любых $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{E}$ справедливы неравенства треугольника*

$$\big||\boldsymbol{x}| - |\boldsymbol{y}|\big| \leqslant |\boldsymbol{x} + \boldsymbol{y}| \leqslant |\boldsymbol{x}| + |\boldsymbol{y}|.$$

[1] Обратите внимание, что $|\alpha|$ означает модуль вещественного числа, а $|\boldsymbol{x}|$ — длину (норму) вектора в смысле определения 10.3.

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:

$$|\boldsymbol{x}+\boldsymbol{y}|^2 = (\boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y}) = |\boldsymbol{x}|^2 + 2(\boldsymbol{x}, \boldsymbol{y}) + |\boldsymbol{y}|^2 \leqslant \\ \leqslant |\boldsymbol{x}|^2 + 2|\boldsymbol{x}||\boldsymbol{y}| + |\boldsymbol{y}|^2 = \big(|\boldsymbol{x}| + |\boldsymbol{y}|\big)^2,$$

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 $|\boldsymbol{x}+\boldsymbol{y}| \geqslant \big||\boldsymbol{x}| - |\boldsymbol{y}|\big|$. □

C. Примеры евклидовых векторных пространств. В следующих примерах проверьте выполнение аксиом **СП1–СП3** самостоятельно.

Пример 10.1. Векторное пространство $\mathbb{R}^n(\mathbb{R})$ становится евклидовым, если для векторов (столбцов)

$$\boldsymbol{x} = X = \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix}, \quad \boldsymbol{y} = Y = \begin{pmatrix} y^1 \\ \vdots \\ y^n \end{pmatrix}$$

определить скалярное произведение по формуле

$$(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j=1}^{n} x^j y^j = X^T Y.$$

Пример 10.2. В векторном пространстве $\mathbb{R}^{m \times n}$ можно ввести скалярное произведение по формуле

$$(X, Y) = \operatorname{tr}(X^T Y).$$

Пример 10.3. В пространстве многочленов $\mathbb{R}[t]_n$ степени не выше n можно ввести скалярное произведение векторов (многочленов)

$$\boldsymbol{x} = x(t) = a_0 + a_1 t + \dots + a_n t^n, \quad \boldsymbol{y} = y(t) = b_0 + b_1 t + \dots + b_n t^n$$

следующим образом:

$$(\boldsymbol{x}, \boldsymbol{y}) = \int\limits_{\alpha}^{\beta} x(t) y(t) dt,$$

где α, β — произвольные вещественные числа

D. Выражение скалярного произведения в координатах. Матрица Грама. Пусть в евклидовом векторном пространстве $\mathcal{E}$ задан некоторый базис $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$. Получим выражение для скалярного произведения произвольных векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ через их координаты относительно этого базиса.

Разложения векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ по базису $\boldsymbol{E}$ имеют вид

$$\boldsymbol{x} = x^j \boldsymbol{e}_j, \quad \boldsymbol{y} = y^k \boldsymbol{e}_k, \quad j,k = 1, \ldots, n$$

(подразумевается суммирование по j и по k). Пользуясь свойством линейности скалярного умножения по каждому из аргументов, получаем

$$(\boldsymbol{x}, \boldsymbol{y}) = \Big(x^j \boldsymbol{e}_j, y^k \boldsymbol{e}_k\Big) = x^j y^k (\boldsymbol{e}_j, \boldsymbol{e}_k) = g_{jk} x^j y^k,$$

где введено обозначение

$$g_{jk} = (\boldsymbol{e}_j, \boldsymbol{e}_k)$$

для скалярных произведений векторов базиса; полученные числа g_{jk}, называемые *метрическими коэффициентами* рассматриваемого базиса, образуют квадратную матрицу $G = (g_{jk})$ — *матрицу Грама* базиса $\boldsymbol{E}$.

Таким образом, зная метрические коэффициенты, т.е. попарные скалярные произведения базисных векторов, можно вычислить скалярное произведение произвольных векторов $\boldsymbol{x}$ и $\boldsymbol{y}$:

$$(\boldsymbol{x}, \boldsymbol{y}) = g_{jk} x^j x^k. \tag{10.5}$$

В матричной форме эта формула имеет вид

$$(\boldsymbol{x}, \boldsymbol{y}) = X^T G Y, \tag{10.6}$$

где X и Y — столбцы координат векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ в базисе $\boldsymbol{E}$, а G — матрица Грама этого базиса.

Из аксиомы **СП2** вытекает симметричность матрицы Грама:

$$g_{jk} = (\boldsymbol{e}_j, \boldsymbol{e}_k) = (\boldsymbol{e}_k, \boldsymbol{e}_j) = g_{kj}.$$

Выяснение условий, при которых квадратная симметричная матрица обеспечивает положительную определённость выражения 10.6, т.е. может служить матрицей Грама[1], отложим до следующего семестра; сейчас мы ограничимся лишь рассмотрением простейшего случая базисов, матрица Грама которых равна единичной матрице.

[1]Разумеется не любая симметричная матрица может быть использована в качестве матрицы Грама, т.е. задавать скалярное произведение по формуле (10.6): например, при попытке использовать матрицу $\left(\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right)$ в качестве матрицы Грама обнаруживается, что «скалярный квадрат» вектора $\boldsymbol{a} = (1, -1)^T$ оказывается отрицательным:

$$(\boldsymbol{a}, \boldsymbol{a}) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}^T \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -2,$$

что противоречит аксиоме **СП3**.

2. Ортонормированные базисы

А. Ортогональные векторы. Пусть $\mathcal{E}$ — евклидово векторное пространство.

10.9. Определение. Векторы $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{E}$ называются *ортогональными*, если $(\boldsymbol{x}, \boldsymbol{y}) = 0$. Обозначение $\boldsymbol{x} \perp \boldsymbol{y}$.

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 $\mathcal{P}$:

$$\boldsymbol{x} \perp \mathcal{P} \quad \Longleftrightarrow \quad \forall \boldsymbol{y} \in \mathcal{P}: \ \boldsymbol{x} \perp \boldsymbol{y}.$$

10.10. Теорема.

1. $\boldsymbol{x} \perp \boldsymbol{x}$ *тогда и только тогда, когда* $\boldsymbol{x} = \boldsymbol{0}$.
2. *Если* $\boldsymbol{x} \in \mathcal{P}$ *и* $\boldsymbol{x} \perp \mathcal{P}$, *то* $\boldsymbol{x} = \boldsymbol{0}$.
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. *Если* $\boldsymbol{x} \perp \boldsymbol{y}$, *то*
$$|\boldsymbol{x} + \boldsymbol{y}|^2 = |\boldsymbol{x}|^2 + |\boldsymbol{y}|^2$$
(*теорема Пифагора*).
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 **СП1**.

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

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

$$(\boldsymbol{y}, \boldsymbol{x}) = \big(\boldsymbol{y}, \ \alpha^j \boldsymbol{x}_j\big) = \alpha^j \underbrace{(\boldsymbol{y}, \boldsymbol{x}_j)}_{=0} = 0 \iff \boldsymbol{y} \perp L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p).$$

4. Если $\boldsymbol{x} \perp \boldsymbol{y}$, то

$$|\boldsymbol{x} + \boldsymbol{y}|^2 = (\boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{x}) + 2 \underbrace{(\boldsymbol{x}, \boldsymbol{y})}_{=0} + (\boldsymbol{y}, \boldsymbol{y}) = |\boldsymbol{x}|^2 + |\boldsymbol{y}|^2.$$

5. Рассмотрим линейную комбинацию векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$, равную нулевому вектору:

$$\alpha^1 \boldsymbol{x}_1 + \cdots + \alpha^j \boldsymbol{x}_j + \cdots + \alpha^p \boldsymbol{x}_p = \boldsymbol{0}.$$

Умножая скалярно это равенство на вектор $\boldsymbol{x}_j$, получаем

$$\alpha^1 \underbrace{(\boldsymbol{x}_1, \boldsymbol{x}_j)}_{=0} + \cdots + \alpha^s \underbrace{(\boldsymbol{x}_j, \boldsymbol{x}_j)}_{=1} + \cdots + \alpha^p \underbrace{(\boldsymbol{x}_p, \boldsymbol{x}_j)}_{=0} = 0,$$

т.е. $\alpha^j = 0$, что означает линейную независимость векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$. $\square$

10.11. Определение. *Ортогональной проекцией* вектора $\boldsymbol{x} \in \mathcal{E}$ на подпространство $\mathcal{P} \Subset \mathcal{E}$ (см. рис. 10.1) называется такой вектор $\boldsymbol{x}' \in \mathcal{P}$, что $\boldsymbol{x} - \boldsymbol{x}' \perp \mathcal{P}$; обозначение $\boldsymbol{x}' = \operatorname{Pr}^{\perp}_{\mathcal{P}} \boldsymbol{x}$ (если ясно, что речь идёт об ортогональной проекции, используется сокращённое обозначение $\operatorname{Pr}_{\mathcal{P}} \boldsymbol{x}$).

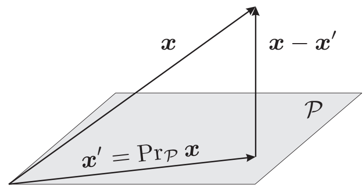


Рис. 10.1. Ортогональная проекция вектора на подпространство

B. Ортонормированный базис.

10.12. Определение. Семейство векторов $\big(\boldsymbol{i}_1, \dots, \boldsymbol{i}_p\big)$ евклидова векторного пространства называется *ортонормированным*, если

(i) длина каждого вектора равна 1,

(ii) любые два различных вектора ортогональны,

т.е. если

$$(\boldsymbol{i}_j, \boldsymbol{i}_k) = \delta_{jk} = \begin{cases} 1, & \text{если } j = k, \\ 0, & \text{если } j \neq k. \end{cases} \tag{10.7}$$

10.13. Определение. *Ортонормированным базисом* называется ортонормированное семейство векторов $\boldsymbol{I} = \big(\boldsymbol{i}_1, \dots, \boldsymbol{i}_n\big)$ евклидова векторного пространства, являющееся базисом.

Ортонормированные базисы очень удобны для вычислений.

10.14. Теорема. *Пусть $\big(\boldsymbol{i}_1, \dots, \boldsymbol{i}_p\big)$ — ортонормированное семейство векторов в евклидовом пространстве $\mathcal{E}$, $p \leqslant \dim \mathcal{E}$. Ортогональная проекция вектора $\boldsymbol{x} \in \mathcal{E}$ на подпространство $\mathcal{P} = L(\boldsymbol{i}_1, \dots, \boldsymbol{i}_p)$ равна*

$$\operatorname{Pr}^{\perp}_{\mathcal{P}} \boldsymbol{x} = (\boldsymbol{x}, \boldsymbol{i}_1)\boldsymbol{i}_1 + \dots + (\boldsymbol{x}, \boldsymbol{i}_p)\boldsymbol{i}_p = \sum_{j=1}^{p} (\boldsymbol{x}, \boldsymbol{i}_j)\boldsymbol{i}_j. \tag{10.8}$$

Доказательство. Достаточно показать, что вектор

$$\boldsymbol{y} = \boldsymbol{x} - \sum_{j=1}^{p} (\boldsymbol{x}, \boldsymbol{i}_j)\boldsymbol{i}_j \tag{10.9}$$

ортогонален каждому из векторов $\boldsymbol{i}_k$, $k = 1, \dots, p$. Действительно,

$$(\boldsymbol{y}, \boldsymbol{i}_k) = (\boldsymbol{x}, \boldsymbol{i}_k) - \sum_{j=1}^{p} (\boldsymbol{x}, \boldsymbol{i}_j) \underbrace{(\boldsymbol{i}_j, \boldsymbol{i}_k)}_{=\delta_{jk}} = (\boldsymbol{x}, \boldsymbol{i}_k) - (\boldsymbol{x}, \boldsymbol{i}_k) = 0,$$

что и требовалось. $\square$

10.15. Теорема. *Пусть $\boldsymbol{I} = \big(\boldsymbol{i}_1, \dots, \boldsymbol{i}_n\big)$ — произвольный ортонормированный базис в евклидовом пространстве $\mathcal{V}$.*

1. *Скалярное произведение двух векторов* $\boldsymbol{x}$ *и* $\boldsymbol{y}$ *выражается через их координаты относительно базиса* $\boldsymbol{I}$ *формулой*[1]
$$(\boldsymbol{x}, \boldsymbol{y}) = \delta_{jk} x^j y^k = \sum_{j=1}^{n} x^j y^j = X^T Y. \tag{10.10}$$

2. *Координаты произвольного вектора* $\boldsymbol{x}$ *относительно ортонормированного базиса могут быть вычислены по формулам*
$$x^k = (\boldsymbol{x}, \boldsymbol{i}_k), \tag{10.11}$$
называемым формулами Гиббса.

Доказательство. Утверждение 1 вытекает из того очевидного факта, что матрица Грама ортонормированного базиса является единичной матрицей: $G = \mathbb{1}$.

2. Рассмотрим разложение вектора $\boldsymbol{x}$ по ортонормированному базису $\boldsymbol{I}$:
$$\boldsymbol{x} = x^j \boldsymbol{i}_j$$
(суммирование по $j = 1, \ldots, n = \dim \mathcal{V}$). Умножив скалярно обе части этого разложения на $\boldsymbol{i}_k$, получим
$$(\boldsymbol{x}, \boldsymbol{i}_k) = (x^j \boldsymbol{i}_j, \boldsymbol{i}_k) = x^j (\boldsymbol{i}_j, \boldsymbol{i}_k) = x^j \delta_{jk} = x^k,$$
т.е. формулу (10.11). □

10.16. Замечание. Обратите внимание на положение индексов: 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.

C. Существование ортонормированных семейств.

10.17. Предложение. *В любом евклидовом пространстве* $\mathcal{E}$ *существуют ортонормированные семейства.*

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 членов ($p < n$, где $n = \dim \mathcal{E}$). Пусть $\bigl(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_p\bigr)$ — одно из таких семейств и $\boldsymbol{x} \notin L(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_p) \equiv \mathcal{P}$ — произвольный вектор; такой вектор существует, так как $\dim L(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_p) = p < n$ (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. □

[1] Здесь X и Y — столбцы координат векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ в базисе $\boldsymbol{I}$.

10.18. Замечание. Пошаговый метод построения ортонормированного семейства векторов, использованный в доказательстве предложения 10.17, называется методом ортогонализации Грама—Шмидта; более подробно мы обсудим его в следующем семестре.

D. Изоморфизм евклидовых пространств. Пусть $\mathcal{E}$ и $\mathcal{F}$ — два евклидовых векторных пространства с операциями скалярного умножения, которые мы обозначим $(\cdot,\cdot)_{\mathcal{E}}$ и $(\cdot,\cdot)_{\mathcal{F}}$.

10.19. Определение. Отображение $f : \mathcal{E} \to \mathcal{F}$ называется *изоморфизмом евклидовых векторных пространств*, если оно является изоморфизмом векторных пространств и обладает свойством

$$(\boldsymbol{x}, \boldsymbol{y})_E = \big(f(\boldsymbol{x}), f(\boldsymbol{y})\big)_F$$

для любых $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{E}$.

Пространства $\mathcal{E}$ и $\mathcal{F}$ называются *изоморфными*, если существует хотя бы один изоморфизм $f : \mathcal{E} \to \mathcal{F}$.

10.20. Теорема. *Любые два евклидовых векторных пространства одинаковой размерности изоморфны.*

Доказательство. Зафиксируем в $\mathcal{E}$ и в $\mathcal{F}$ ортонормированные базисы $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$ и $\boldsymbol{F} = \big(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n\big)$ соответственно. Отображение

$$f : \mathcal{E} \to \mathcal{F}, \quad \boldsymbol{x} = x^j \boldsymbol{e}_j \mapsto f(\boldsymbol{x}) = x^j \boldsymbol{f}_j \in F$$

является изоморфизмом евклидовых векторных пространств, поскольку во всех ортонормированных базисах скалярное произведение выражается одной и той же формулой:

$$(\boldsymbol{x}, \boldsymbol{y})_E = \sum_{j=1} x^j y^j = (f(\boldsymbol{x}), f(\boldsymbol{y}))_F.$$

$\square$

Аналогичные определения и результаты имеют место и для евклидовых точечных пространств.

3. Двумерная евклидова геометрия

Евклидова геометрия — это раздел геометрии, изучающий евклидовы пространства. По сравнению с аффинной геометрией в евклидовой геометрии имеется одно дополнительное первоначальное понятие — скалярное произведение — и три дополнительных аксиомы **СП1–СП3**.

Ортогональная (прямоугольная) система координат $O\boldsymbol{E}$ в евклидовом точечном пространстве состоит из фиксированной точки O (начала отсчёта) и $\boldsymbol{E} = \big(\boldsymbol{e}_1, \boldsymbol{e}_2\big)$ — ортонормированного базиса ассоциированного векторного пространства.

Рассмотрим геометрию двумерного евклидова точечного пространства (евклидовой плоскости) $\mathcal{E}$. Считаем, что в $\mathcal{E}$ зафиксирована ортогональная система координат $O\boldsymbol{E}$; по традиции векторы ортонормированного базиса на плоскости обозначают $\boldsymbol{i}, \boldsymbol{j}$, а координаты — x, y.

A. Нормальное уравнение прямой на плоскости. Будем считать, что в евклидовой плоскости задана ортогональная системе координат $O\boldsymbol{ij}$.

Пусть в евклидовой плоскости заданы ненулевой вектор $\boldsymbol{n}(A,B)$ и точка $M_0(x_0,y_0)$.

10.21. Предложение. *Множество всех точек M плоскости, для которых вектор $\boldsymbol{n}$ ортогонален вектору $\overrightarrow{M_0M}$, является прямой, проходящей через точку M_0. Вектор $\boldsymbol{n}$ называется* нормальным вектором *этой прямой.*

Доказательство. Условие ортогональности $\boldsymbol{n}\perp\overrightarrow{M_0M}$, т.е. $\big(\boldsymbol{n},\overrightarrow{M_0M}\big)=0$, в координатах записывается в виде $A(x-x_0)+B(y-y_0)=0$, что совпадает с общим уравнением прямой. □

Таким образом, уравнение прямой на плоскости может быть записано в виде

$$(\boldsymbol{r}-\boldsymbol{r}_0,\boldsymbol{n})=0,\quad (\boldsymbol{r},\boldsymbol{n})=D \tag{10.12}$$

или в прямоугольных координатах

$$A(x-x_0)+B(y-y_0)=0,\quad Ax+By=D, \tag{10.13}$$

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.*

10.22. Замечание. Обратите внимание, что запись уравнения прямой в виде (10.12) возможно в любой системе координат, а в виде (10.13) — только в прямоугольной, ибо только в ортонормированном базисе скалярное произведение векторов имеет вид (10.10).

B. Некоторые метрические задачи. Геометрические задачи, связанные с необходимостью использования скалярного произведения, т.е. включающие понятия расстояния, угла, перпендикулярности называются метрическими задачами в отличие от аффинных задач, где скалярное произведение не требуется.

10.23. Теорема. *Пусть на евклидовой плоскости $\mathcal{E}$ 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*

$$\boldsymbol{r}_2=\boldsymbol{r}_1-\frac{(\boldsymbol{r}_1,\boldsymbol{n})-D}{(\boldsymbol{n},\boldsymbol{n})}\boldsymbol{n}. \tag{10.14}$$

2. *Расстояние от точки $M_1(\boldsymbol{r}_1)$ до прямой l выражается формулой*

$$d(M_1,l)=\frac{|(\boldsymbol{r}_1,\boldsymbol{n})-D|}{|\boldsymbol{n}|}. \tag{10.15}$$

3. *Радиус-вектор* $\boldsymbol{r}_3$ *точки* M_3*, симметричной*[1] *точке* $M_1(\boldsymbol{r}_1)$ *относительно прямой* l*, выражается формулой*

$$\boldsymbol{r}_3 = \boldsymbol{r}_1 - 2\frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{(\boldsymbol{n}, \boldsymbol{n})}\boldsymbol{n}. \tag{10.16}$$

Доказательство. 1. Очевидно (см. рис. 10.2), $\overrightarrow{M_1M_2} = \overrightarrow{OM_2} - \overrightarrow{OM_1} = \lambda\boldsymbol{n}$. Умножим обе части этого соотношения скалярно на вектор $\boldsymbol{n}$:

$$\underbrace{(\overrightarrow{OM_2}, \boldsymbol{n})}_{=D} - \underbrace{(\overrightarrow{OM_1}, \boldsymbol{n})}_{=(\boldsymbol{r}_1,\boldsymbol{n})} = \lambda(\boldsymbol{n}, \boldsymbol{n}),$$

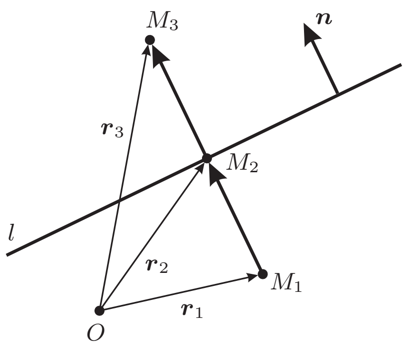


Рис. 10.2. К теореме 10.23

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:

$$\boldsymbol{r}_2 = \overrightarrow{OM_2} = \overrightarrow{OM_1} + \overrightarrow{M_1M_2} = \\ = \boldsymbol{r}_1 + \lambda\boldsymbol{n} = \\ = \boldsymbol{r}_1 - \frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{(\boldsymbol{n}, \boldsymbol{n})}\boldsymbol{n}.$$

2. Расстояние от точки M_1 до прямой l:

$$d(M_1, l) = d(M_1, M_2) = \left|\overrightarrow{M_1, M_2}\right| = \left|\frac{(\boldsymbol{r}_1, n) - D}{(\boldsymbol{n}, \boldsymbol{n})}\boldsymbol{n}\right| = \\ = \frac{|(\boldsymbol{r}_1, n) - D|}{(\boldsymbol{n}, \boldsymbol{n})}|\boldsymbol{n}| = \frac{|(\boldsymbol{r}_1, n) - D|}{|\boldsymbol{n}|}.$$

3. Для радиус-вектора $\boldsymbol{r}_3$ точки M_3, симметричной точке M_1 относительно прямой, имеем:

$$\boldsymbol{r}_3 = \overrightarrow{OM_3} = \overrightarrow{OM_1} + \overrightarrow{M_1M_3} = \overrightarrow{OM_1} + 2\overrightarrow{M_1M_2} = \\ = \boldsymbol{r}_1 + 2\lambda\boldsymbol{n} = \boldsymbol{r}_1 - 2\frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{(\boldsymbol{n}, \boldsymbol{n})}\boldsymbol{n}.$$

□

Формула (10.15) в прямоугольных координатах выглядит следующим образом:

$$d(M_1, l) = \frac{\left|Ax_1 + By_1 - D\right|}{\sqrt{A^2 + B^2}}, \tag{10.17}$$

где (x_1, y_1) — координаты точки M_1, (A, B) — координаты нормального вектора $\boldsymbol{n}$ прямой l.

---

[1] 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: $d(M_1, l) = d(M_3, l)$.

C. Нормированное уравнение прямой. Считаем, что система координат прямоугольная.

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 координатами нормального вектора являются его направляющие косинусы, и уравнение имеет вид

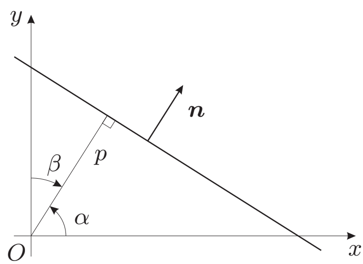


Рис. 10.3.

$$x \cos\alpha + y \cos\beta = p,$$

где α и β — углы, которые образует с осями координат вектор нормали $\boldsymbol{n}$. Так как, очевидно, $\beta = \pi/2 - \alpha$, уравнение может быть записано в виде

$$x \cos\alpha + y \sin\alpha = p. \tag{10.18}$$

Согласно (10.17), число p равно расстоянию от начала координат до прямой l. Конечно, в случае, когда прямая проходит через начало координат, $p = 0$.

Пусть $M_1(x_1, y_1)$ — некоторая точка плоскости. Расстояние от этой точки до прямой l, заданной нормированным уравнением (10.18), согласно теореме 10.23.2, равно

$$d(M_1, l) = \bigl|x_1 \cos\alpha + y_1 \sin\alpha - p\bigr|.$$

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; модуль этой величины, очевидно, равен расстоянию $d(M_1, l)$, а знак показывает, в какой из двух полуплоскостей, на которые разбивается плоскость прямой a, лежит точка M_1: если $\delta(M_1, a) < 0$, то точка M_1 и начало координат O лежат в одной полуплоскости, а если $\delta(M_1, a) > 0$ — то в разных (см. теорему 9.21, с. 198). Легко доказать, 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:

$$\delta(M_1, a) = \frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{|\boldsymbol{n}|} \cdot \operatorname{sgn} D,$$

где $\operatorname{sgn} D$ — знак числа D (см. с. 131).

4. Трехмерная евклидова геометрия

В трёхмерном евклидовом пространстве можно ввести операции векторного и смешанного умножения, для определения которых требуется только понятие скалярного произведения; вся необходимая информация об этих операциях была детально изложена в лекции III.

В пространствах размерности $\geqslant 3$ существуют аналоги векторного и смешанного произведений, однако соответствующие конструкции довольно сложны и целиком выходят за рамки нашего курса.

А. Плоскости в трехмерном евклидовом пространстве. Будем считать, что в трёхмерном евклидовом точечном пространстве $\mathcal{E}$ задана ортогональная системе координат $O\boldsymbol{ijk}$.

Пусть в $\mathcal{E}$ плоскости заданы ненулевой вектор $\boldsymbol{n}(A, B, C)$ и точка $M_0(x_0, y_0, z_0)$.

10.24. 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.*

Доказательство. Условие ортогональности $\boldsymbol{n} \perp \overrightarrow{M_0M}$, т.е. $\big(\boldsymbol{n}, \overrightarrow{M_0M}\big) = 0$, в координатах принимает вид $A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$, что совпадает с общим уравнением плоскости. □

Таким образом, уравнение плоскости в пространстве может быть записано в виде

$$(\boldsymbol{r} - \boldsymbol{r}_0, \boldsymbol{n}) = 0, \quad (\boldsymbol{r}, \boldsymbol{n}) = D, \tag{10.19}$$

не отличающемся от уравнения прямой на плоскости (10.19) (см. с. 226), что не удивительно, поскольку эти объекты являются гиперплоскостями в соответствующих пространствах. Координатная форма уравнений (10.19) в прямоугольных координатах:

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0, \quad Ax + By + Cz = D, \tag{10.20}$$

где $\boldsymbol{n}(A, B, C)$ — вектор нормали прямой. Формы уравнений (10.19) и (10.20) называют *нормальными.*

10.25. Замечание. Обратите внимание, что запись уравнения плоскости в виде (10.19) возможно в любой системе координат, а в виде (10.20) — только в прямоугольной, ибо только в ортонормированном базисе скалярное произведение векторов имеет вид (10.10) (см. с. 221).

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 (10.19) получаем

$$\big(\boldsymbol{r} - \boldsymbol{r}_0, [\boldsymbol{a}, \boldsymbol{b}]\big) = 0 \iff (\boldsymbol{r}, [\boldsymbol{a}, \boldsymbol{b}]) = (\boldsymbol{r}_0, \boldsymbol{a}, \boldsymbol{b}).$$

10.26. Теорема. *Пусть в евклидовом точечном пространстве $\mathcal{E}$ даны точка $M_1(\boldsymbol{r}_1)$ и плоскость π, заданная уравнением $(\boldsymbol{r}, \boldsymbol{n}) = D$* (см. рис. 10.4).

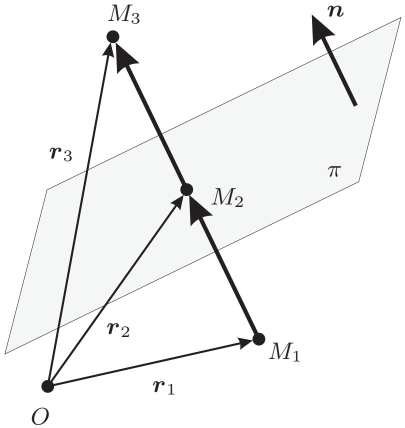


Рис. 10.4. К теореме 10.26

1. *Радиус-вектор $\boldsymbol{r}_2$ точки M_2, являющейся ортогональной проекцией точки $M_1(\boldsymbol{r}_1)$ на плоскость π, выражается формулой*
$$\boldsymbol{r}_2 = \boldsymbol{r}_1 - \frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{(\boldsymbol{n}, \boldsymbol{n})}\boldsymbol{n}. \qquad (10.21)$$
2. *Расстояние от точки $M_1(\boldsymbol{r}_1)$ до плоскости π выражается формулой*
$$d(M_1, \pi) = \frac{|(\boldsymbol{r}_1, \boldsymbol{n}) - D|}{|\boldsymbol{n}|}. \qquad (10.22)$$
3. *Радиус-вектор $\boldsymbol{r}_3$ точки M_3, симметричной точке $M_1(\boldsymbol{r}_1)$ относительно плоскости π, выражается формулой*
$$\boldsymbol{r}_3 = \boldsymbol{r}_1 - 2\frac{(\boldsymbol{r}_1, \boldsymbol{n}) - D}{(\boldsymbol{n}, \boldsymbol{n})}\boldsymbol{n}. \qquad (10.23)$$

Доказательство теоремы 10.26 дословно повторяет доказательство теоремы 10.23.

B. Уравнение прямой в пространстве. 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.*

Переход от уравнения Плюккера $[\boldsymbol{r}, \boldsymbol{a}] = \boldsymbol{b}$ к векторному параметрическому уравнению осуществляется следующим образом. Направляющий вектор прямой $[\boldsymbol{r}, \boldsymbol{a}] = \boldsymbol{b}$ можно выбрать равным $\boldsymbol{a}$. Найдем такую опорную точку $\boldsymbol{r}_0$ прямой, что её радиус-вектор ортогонален вектору $\boldsymbol{a}$. Умножим соотношение $[\boldsymbol{r}_0, \boldsymbol{a}] = \boldsymbol{b}$ прямой векторно на $\boldsymbol{a}$:
$$\big[\boldsymbol{a}, [\boldsymbol{r}_0, \boldsymbol{a}]\big] = [\boldsymbol{a}, \boldsymbol{b}].$$
Раскрывая двойное векторное произведение по формуле (2.11), находим
$$\boldsymbol{r}_0(\boldsymbol{a}, \boldsymbol{a}) - \boldsymbol{a}\underbrace{(\boldsymbol{a}, \boldsymbol{r}_0)}_{=0} = [\boldsymbol{a}, \boldsymbol{b}],$$
откуда
$$\boldsymbol{r}_0 = \frac{[\boldsymbol{a}, \boldsymbol{b}]}{(\boldsymbol{a}, \boldsymbol{a})}.$$

Таким образом, получаем параметрическое уравнение прямой

$$\boldsymbol{r} = \frac{[\boldsymbol{a}, \boldsymbol{b}]}{(\boldsymbol{a}, \boldsymbol{a})} + t\boldsymbol{a}.$$

10.27. Теорема. *Пусть в евклидовом точечной пространстве $\mathcal{E}$ даны точка $M_1(\boldsymbol{r}_1)$ и прямая l, заданная уравнением $\boldsymbol{r} = \boldsymbol{r}_0 + t\boldsymbol{a}$* (см. рис. 10.5).

1. *Радиус-вектор $\boldsymbol{r}_2$ точки M_2, являющейся ортогональной проекцией точки $M_1(\boldsymbol{r}_1)$ на прямую l, выражается формулой*
$$\boldsymbol{r}_2 = \boldsymbol{r}_0 + \frac{(\boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{a})}{(\boldsymbol{a}, \boldsymbol{a})}\boldsymbol{a}. \qquad (10.24)$$
2. *Расстояние от точки $M_1(\boldsymbol{r}_1)$ до прямой l выражается формулой*
$$d(M_1, l) = \frac{\big|[\boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{a}]\big|}{|\boldsymbol{a}|}. \qquad (10.25)$$
3. *Радиус-вектор $\boldsymbol{r}_3$ точки M_3, симметричной точке $M_1(\boldsymbol{r}_1)$ относительно прямой l, выражается формулой*
$$\boldsymbol{r}_3 = 2\boldsymbol{r}_0 - \boldsymbol{r}_1 + 2\frac{(\boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{a})}{(\boldsymbol{a}, \boldsymbol{a})}\boldsymbol{a}. \qquad (10.26)$$

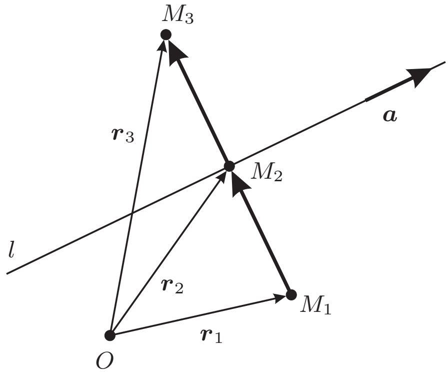


Рис. 10.5. К теореме 10.27

Доказательство. 1. Умножим обе части равенства

$$\overrightarrow{M_1M_2} = \overrightarrow{OM_2} - \overrightarrow{OM_1}$$

скалярно на вектор $\boldsymbol{a}$:

$$\underbrace{(\overrightarrow{M_1M_2}, \boldsymbol{a})}_{=0} - \underbrace{(\overrightarrow{OM_2}, \boldsymbol{n})}_{=(\boldsymbol{r}_0 + t\boldsymbol{a}, \boldsymbol{a})} - \underbrace{(\overrightarrow{OM_1}, \boldsymbol{n})}_{=(\boldsymbol{r}_1, \boldsymbol{a})} = (\boldsymbol{r}_0 - \boldsymbol{r}_1, \boldsymbol{a}) + t(\boldsymbol{a}, \boldsymbol{a}),$$

откуда

$$t_0 = \frac{(\boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{a})}{(\boldsymbol{a}, \boldsymbol{a})}$$

— значение параметра, отвечающее точке $M_2 \in l$. Поэтому для проекции M_2 точки M_1 на прямую l имеем:

$$\boldsymbol{r}_2 = \boldsymbol{r}_0 + t_0\boldsymbol{a} = \boldsymbol{r}_0 + \frac{(\boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{a})}{(\boldsymbol{a}, \boldsymbol{a})}\boldsymbol{a}.$$

3. Для точки M_3, симметричной точке M_1 относительно прямой l, имеем

$$\begin{aligned}\boldsymbol{r}_3 = \overrightarrow{OM_3} = \overrightarrow{OM_1} + 2\overrightarrow{M_1M_2} = \boldsymbol{r}_1 + 2(\boldsymbol{r}_2 - \boldsymbol{r}_1) &= \\ = 2\boldsymbol{r}_2 - \boldsymbol{r}_1 = 2\boldsymbol{r}_0 - \boldsymbol{r}_1 + 2&\frac{(\boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{a})}{(\boldsymbol{a}, \boldsymbol{a})}\boldsymbol{a}.\end{aligned}$$

2. 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. 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:

$$\left|[\boldsymbol{r}_0 - \boldsymbol{r}_1, \boldsymbol{a}]\right| = |\boldsymbol{a}| \cdot d(M_1, l),$$

откуда и вытекает формула (10.25).

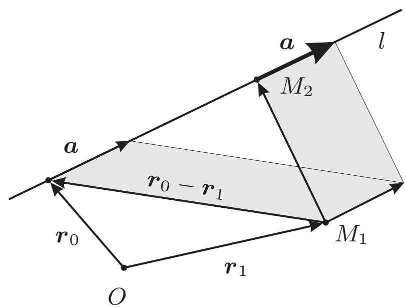


Рис. 10.6. Расстояние между скрещивающимися прямыми

□

ГЛАВА 11

Эллипс, гипербола, парабола на евклидовой плоскости

В ЭТОЙ лекции изучаются три важные линии на евклидовой плоскости — эллипс, гипербола и парабола, которые были известны ещё античным математикам. Каждая из этих трёх линий может быть получена как сечение кругового конуса некоторой плоскостью, и по этой причине они называются *кониками.* Коники, являясь 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 эллипса, гиперболы и параболы

A. Эллипс.

11.1. Определение (фокальное определение эллипса)**.** *Эллипсом* называется геометрическое место точек, сумма расстояний от каждой из которых до двух фиксированных точек (называемых *фокусами* эллипса) есть постоянная величина; требуется чтобы эта постоянная была больше расстояния между фокусами.

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 исключаем из рассмотрения оговоркой в конце предыдущего определения.

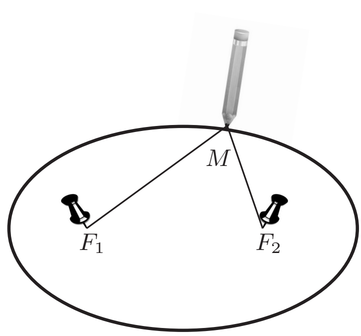


Рис. 11.1. Вычерчивание эллипса

Из определения 11.1 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 эллипс (см. рис. 11.1).

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 через b^2:

$$b^2 = a^2 - c^2.$$

Тогда последнее уравнение можно переписать в виде

$$b^2x^2 + a^2y^2 = a^2b^2$$

или

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \tag{11.1}$$

Это уравнение называется *каноническим уравнением эллипса.*

Мы установили, что координаты каждой точки эллипса удовлетворяют уравнению (11.1). Проверим обратное утверждение, а именно, что каждая точка $M(x, y)$, удовлетворяющая полученному уравнению, лежит на эллипсе, т.е. что для неё выполнено условие $r_1 + r_2 = 2a$. 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 (x, y) — решение уравнения; ясно, что $|x| \leqslant a$, $|y| \leqslant b$. Тогда

$$r_1 = |F_1M| = \sqrt{(x+c)^2 + y^2} = \sqrt{(x+c)^2 + b^2\left(1 - \frac{x^2}{a^2}\right)} =$$
$$= \sqrt{x^2 + 2xc + c^2 + b^2 - \frac{b^2}{a^2}x^2} = \sqrt{x^2\varepsilon^2 + 2x\varepsilon a + a^2} =$$
$$= \sqrt{(x\varepsilon + a)^2} = |x\varepsilon + a| = a + x\varepsilon$$

(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 с координатами (x, y) лежит на эллипсе.

Итак, доказано следующее утверждение.

11.2. Теорема (каноническое уравнение 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* (11.1).

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.

Из полученного уравнения 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

$$|x| \leqslant |a|, \quad |y| \leqslant |b|.$$

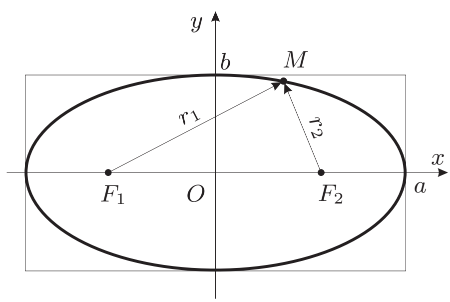


Рис. 11.2. Эллипс в канонической системе координат

Из выражения для длины левого фокального радиуса легко получаем

$$r_1 = a + x\varepsilon = \varepsilon\underbrace{\left(x + \frac{a}{\varepsilon}\right)}_{=d_1};$$

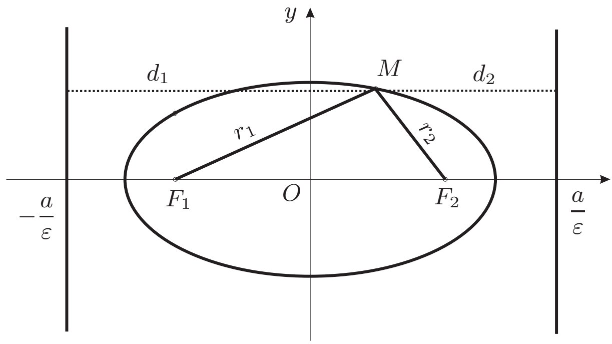


Рис. 11.3. Директрисы эллипса

выражение в скобках представляет собой расстояние d_1 от точки $M(x, y)$ до прямой $x = -\dfrac{a}{\varepsilon}$, называемой (левой) *директрисой* эллипса. Аналогичным образом из выражения $r_2 = a - x\varepsilon$ для длины правого фокального радиуса находим

$$r_2 = a - x\varepsilon = \varepsilon \Bigg(\underbrace{\frac{a}{\varepsilon} - x}_{=d_2} \Bigg);$$

выражение в скобках представляет собой расстояние d_2 от точки $M(x, y)$ до *правой директрисы* эллипса, задаваемой уравнением $x = -\dfrac{a}{\varepsilon}$. Таким образом, доказано следующее утверждение.

11.3. 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*:

$$\frac{r_1}{d_1} = \frac{r_2}{d_2} = \varepsilon < 1.$$

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.*

Отметим, что эллипс может быть определён не только фокальным определением, но также своим каноническим уравнением либо директориальным свойством (тогда фокальное определение превратится в теорему — фокальное свойство эллипса).

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 $\varepsilon=c/a<1$ — (числовой) эксцентриситет;
(xi) прямые $x=-\dfrac{a}{\varepsilon}$, $x=\dfrac{a}{\varepsilon}$ — директрисы.

B. Гипербола.

11.4. Определение (фокальное определение гиперболы)**.** *Гиперболой* называется геометрическое место точек, модуль разности расстояний от каждой из которых точки линии до двух фиксированных точек (называемых *фокусами* гиперболы) есть постоянная величина; требуется, чтобы эта постоянная была меньше расстояния между фокусами и отлична от нуля.

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 предыдущего определения.

На определении 11.4 основано вычерчивание гиперболы при помощи двух нитей, аналогичное построению эллипса, описанному выше. 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 и будем двигать карандаш на бумаге так, чтобы обе нити F_1MC и F_2MC были все время натянуты. Тогда остриё карандаша все время находится на одной ветви гиперболы с фокусами F_1 и F_2, так как

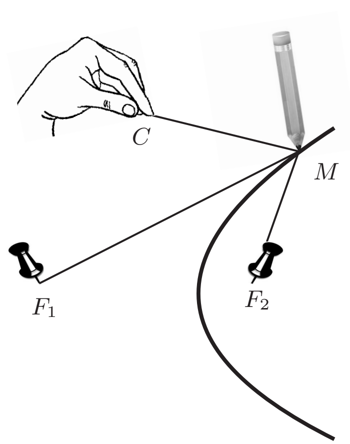


Рис. 11.4. Вычерчивание гиперболы

$$|F_1M|-|F_2M|=(|F_1M|+|MC|)-(|F_2M|+|MC|)=2a.$$

Если поменять местами булавки с привязанными к ним нитями, получится вторая ветвь гиперболы; вычерченные дуги гиперболы будут тем длиннее, чем длиннее нити.

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.*

Обратно, покажем, что любая точка (x, y), являющаяся 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

$$\begin{aligned} r_1^2 &= |F_1M|^2 = (x+c)^2 + y^2 = (x+c)^2 + b^2\left(\frac{x^2}{a^2} - 1\right) = \\ &= \left(\frac{b^2}{a^2} + 1\right)x^2 + 2xc + c^2 - b^2 = \frac{c^2}{a^2}x^2 + 2cx + a^2 = \\ &= \varepsilon^2 x^2 + 2\varepsilon^2 ax + a^2 = (\varepsilon x + a)^2. \end{aligned}$$

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.

11.5. Теорема (каноническое уравнение гиперболы)**.** *Линия на плоскости является гиперболой тогда и только тогда, когда существует такая прямоугольная декартова система координат* (*называемая канонической системой координат*), *в которой эта линия имеет уравнение*

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1,$$

где a и b — некоторые числа.

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 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

$$\sqrt{1 - \frac{a^2}{x^2}} = 1 - \frac{1}{2}\frac{a^2}{x^2} + o\left(\frac{1}{x2}\right), \quad x \to \infty,$$

получаем:

$$y = b\frac{x}{a}\left(1 - \frac{a^2}{2x^2} + o\left(\frac{1}{x2}\right)\right) = \frac{b}{a}x - \frac{ab}{2x} + o\left(\frac{1}{x}\right) = \frac{b}{a}x + o(1).$$

Таким образом, принимая во внимание симметричность гиперболы относительно осей координат, можем сделать вывод, что прямые

$$y = \pm\frac{b}{a}x \iff \frac{x}{a} \pm \frac{y}{b} = 0$$

являются *наклонными асимптотами* гиперболы.[1]

Всё вышесказанное позволяет сделать выводы о внешнем виде гиперболы (см. рис. 11.5)

[1]См. курс математического анализа.

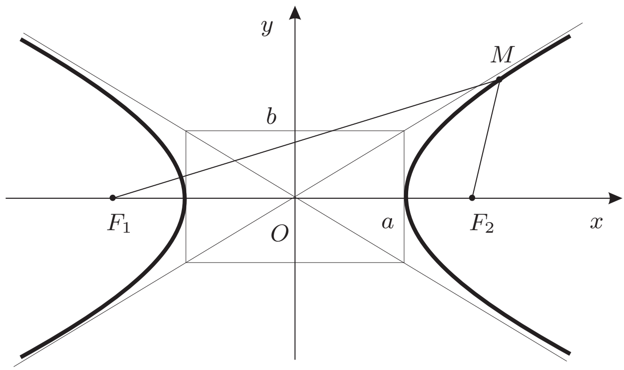


Рис. 11.5. Гипербола в канонической системе координат

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).

11.6. Теорема (директориальное свойство гиперболы, см. рис. 11.6). *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.

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).

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;

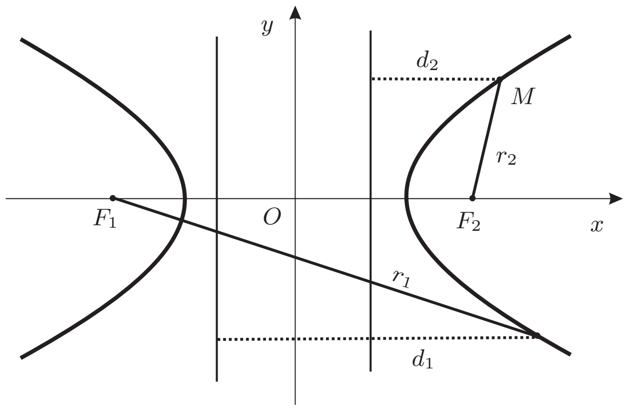


Рис. 11.6. Директрисы гиперболы

(viii) число b — мнимая полуось;

(ix) число $c = \sqrt{a^2 + b^2}$ — линейный эксцентриситет;

(x) число $\varepsilon = c/a > 1$ — (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.

Рассмотрим уравнение

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = -1.$$

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. 11.7).

Очевидно, взаимно сопряжённые гиперболы имеют общие асимптоты.

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.

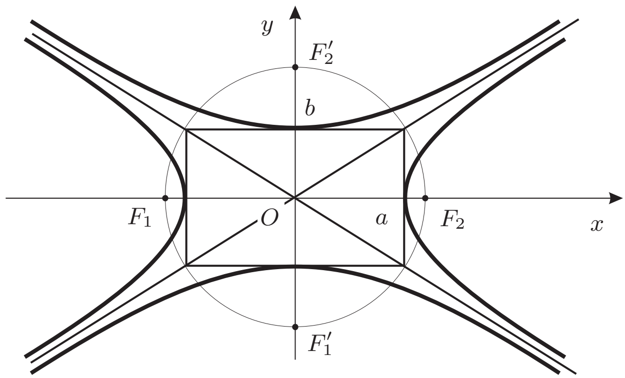


Рис. 11.7. Сопряжённые гиперболы

C. Парабола.

11.7. Определение (фокально-директориальное определение параболы)**.** *Параболой* называется геометрическое место точек, расстояние от каждой из которых до фиксированной точки (называемой фокусом) равно расстоянию до фиксированной прямой (называемой директрисой).

Для того чтобы нарисовать параболу, потребуется линейка, угольник, нить длиной, равной 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.

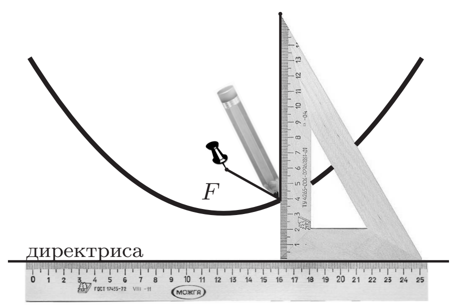


Рис. 11.8. Вычерчивание параболы

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

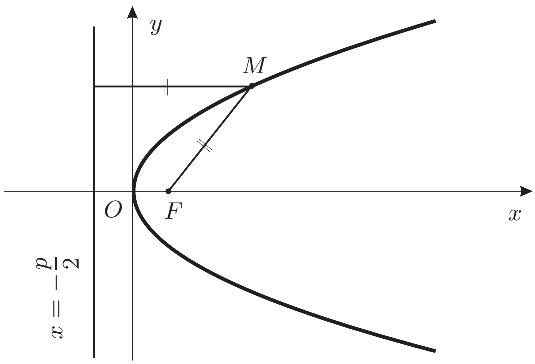


Рис. 11.9. Парабола в канонической системе координат

директрисы $x=-p/2$ 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*

$$y^2=2px.$$

Проверим, что любое решение полученного уравнения представляет точку параболы. Пусть (x,y) — решение уравнения; тогда

$$y^2=2px=\left(x+\frac{p}{2}\right)^2-\left(x-\frac{p}{2}\right)^2,$$

$$\left|x+\frac{p}{2}\right|=\sqrt{\left(x-\frac{p}{2}\right)^2+y^2},$$

т.е. точка (x,y) равноудалена от прямой $x=-p/2$ и точки $(p/2,0)$, а потому, согласно определению, лежит на параболе.

Итак, доказана следующая теорема.

11.8. Теорема (каноническое уравнение параболы)**.** *Линия на плоскости является параболой тогда и только тогда, когда существует такая прямоугольная декартова система координат (называемая канонической системой координат), в которой эта линия имеет уравнение*

$$y^2=2px,$$

где p — некоторое число.

Поскольку каноническое уравнение параболы эллипса не изменяется y на $-y$, парабола симметрична относительно оси Ox.

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. 11.10). 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 $x \to +\infty$).

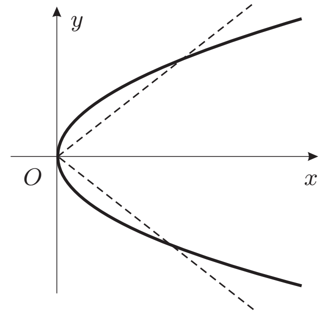


Рис. 11.10.

Эксцентриситет параболы считается равным единице по определению (попытайтесь это объяснить исходя из определений эксцентриситета эллипса и гиперболы).

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. *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

$$\frac{x - x_0}{l} = \frac{y - y_0}{m} \quad \Longleftrightarrow \quad x = x_0 + lt, \quad y = y_0 + mt,$$

т.е. единственная их общая точка. Найдём координаты l, m направляющего вектора касательной. Подставляя параметрические уравнения касательной в каноническое уравнение эллипса, получаем:

$$\underbrace{\frac{x_0^2}{a^2}+\frac{y_0^2}{b^2}}_{=1}+2t\left(\frac{x_0l}{a^2}+\frac{y_0m}{b^2}\right)+t^2\left(\frac{l^2}{a^2}+\frac{m^2}{b^2}\right)=1$$

или после преобразований

$$t^2\left(\frac{l^2}{a^2}+\frac{m^2}{b^2}\right)+2t\left(\frac{x_0l}{a^2}+\frac{y_0m}{b^2}\right)=0.$$

Это квадратное уравнение должно иметь один (двойной) корень, что возможно при выполнении условия

$$\frac{x_0l}{a^2}+\frac{y_0m}{b^2}=0,$$

так что можно положить

$$l=\frac{y_0}{b^2},\quad m=-\frac{x_0}{a^2}.$$

Каноническое уравнение касательной к эллипсу имеет вид

$$\frac{x-x_0}{y_0/b^2}=\frac{y-y_0}{-x_0/a^2}\iff\frac{x_0}{a^2}(x-x_0)+\frac{y_0}{b^2}(y-y_0)=0,$$

откуда, учитывая соотношение $x_0^2/a^2+y_0^2/b^2=1$, получаем

$$\frac{xx_0}{a^2}+\frac{yy_0}{b^2}=1.$$

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):

$$\frac{xx_0}{a^2}-\frac{yy_0}{b^2}=1$$

(проведите соответствующие выкладки самостоятельно).

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

$$yy_0=p(x+x_0),$$

где (x_0,y_0) — точка касания (докажите самостоятельно).

В. Оптические свойства эллипса, гиперболы и параболы. Оптические свойства эллипса, гиперболы и параболы лежат в основе многих технических приложений этих линий.

11.9. Теорема (оптическое свойство эллипса)**.** *Фокальные радиусы произвольной точки M_0 эллипса составляют равные углы с касательной к эллипсу, проведённой через точку M_0.*

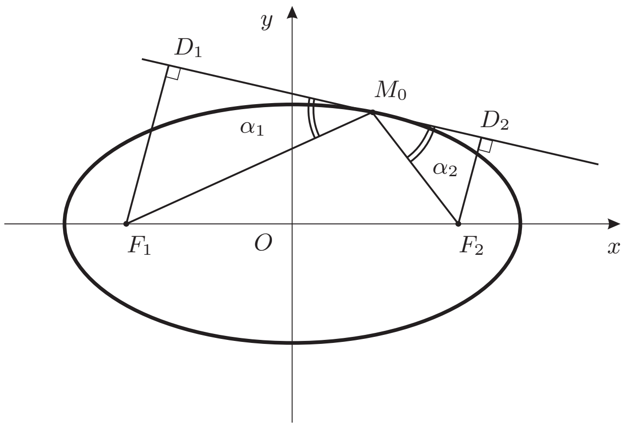


Рис. 11.11. Оптическое свойство эллипса

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

$$\frac{xx_0}{a^2} + \frac{yy_0}{b^2} = 1,$$

равно

$$|F_1D_1| = \frac{\left|\dfrac{(-c)\cdot x_0}{a^2} + \dfrac{0\cdot y_0}{b^2} - 1\right|}{\sqrt{\dfrac{x_0^2}{a^4} + \dfrac{y_0^2}{b^4}}} = \frac{\varepsilon x_0 + a}{a\sqrt{\dfrac{x_0^2}{a^4} + \dfrac{y_0^2}{b^4}}} = \frac{r_1}{a\sqrt{\dfrac{x_0^2}{a^4} + \dfrac{y_0^2}{b^4}}},$$

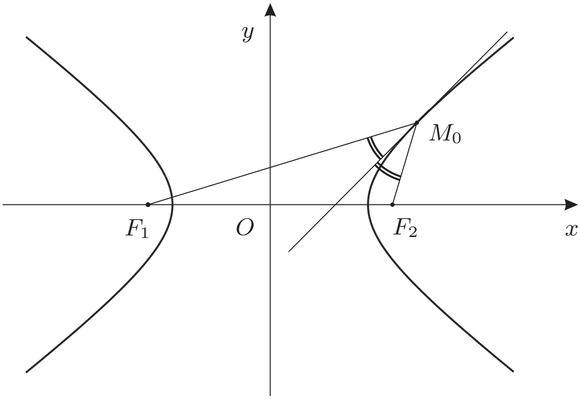
так что

$$\sin\alpha_1 = \frac{|F_1D_1|}{|F_1M_0|} = \frac{1}{a\sqrt{\dfrac{x_0^2}{a^4} + \dfrac{y_0^2}{b^4}}}.$$

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 $\sin\alpha_1 = \sin\alpha_2$ получаем $\alpha_1 = \alpha_2$, что и требовалось. $\square$



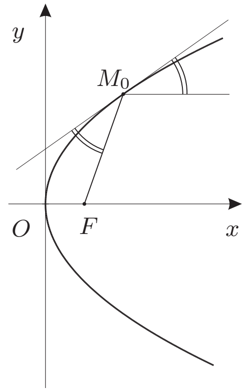


Рис. 11.12. Оптические свойства гиперболы и параболы эллипса

11.10. Теорема (оптическое свойство гиперболы)**.** *Фокальные радиусы произвольной точки M_0 гиперболы составляют равные углы с касательной к гиперболе, проведённой через точку M_0.*

11.11. 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.*

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.

Наиболее просто выводится полярное уравнение параболы. Поместим полюс в фокус параболы (см. рис. 11.13). Связь декартовых и полярных координат имеет вид

$$x - \frac{p}{2} = r\cos\varphi$$

(левая часть равенства — длина проекции вектора $\overrightarrow{FM}$ на ось Ox), а определение параболы даёт равенство

$$r = x + \frac{p}{2} \iff x = r - \frac{p}{2}$$

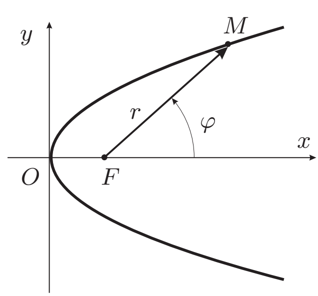


Рис. 11.13. Парабола в полярной системе координат

Таким образом,

$$r\cos\varphi = x - \frac{p}{2} = r - p \iff r = \frac{p}{1-\cos\varphi}.$$

Отметим, что при $\cos\varphi = 0$ (в частности, при $\varphi = \pi/2$) получаем $r = p$, т.е. p равно половине длины хорды параболы, проходящей через фокус перпендикулярно фокальной оси, что и объясняет её название — *фокальный параметр*.

Выведем полярное уравнение эллипса. Поместим полюс в *левый* фокус эллипса (см. рис. 11.14(a)). Запишем связь декартовых и полярных координат

$$x + c = r\cos\varphi$$

и выражение для левого фокального радиуса

$$r = \varepsilon x + a.$$

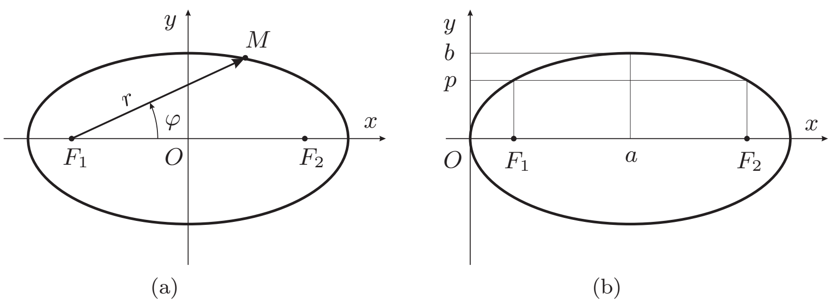


Рис. 11.14. Эллипс (a) в полярной системе координат, (b) в декартовой системе координат, отнесённой к левой вершине

Отсюда получаем

$$r = \varepsilon(r\cos\varphi - c) + a \iff r(1 - \varepsilon\cos\varphi) = a - \varepsilon c.$$

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:

$$p = a - \varepsilon c = a - \frac{c}{a}c = \frac{a^2 - c^2}{c} = \frac{b^2}{c}.$$

Итак, уравнение эллипса в полярной системе координат, ось которой совпадает с его фокальной осью, а полюс находится в левом фокусе, имеет вид

$$r = \frac{p}{1 - \varepsilon\cos\varphi}.$$

В случае гиперболы поместим полюс в *правый* фокус и ограничимся рассмотрением правой ветви гиперболы (см. рис. 11.15(a)). 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

$$r(1 - \varepsilon\cos\varphi) = \varepsilon c - a$$

или, 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.

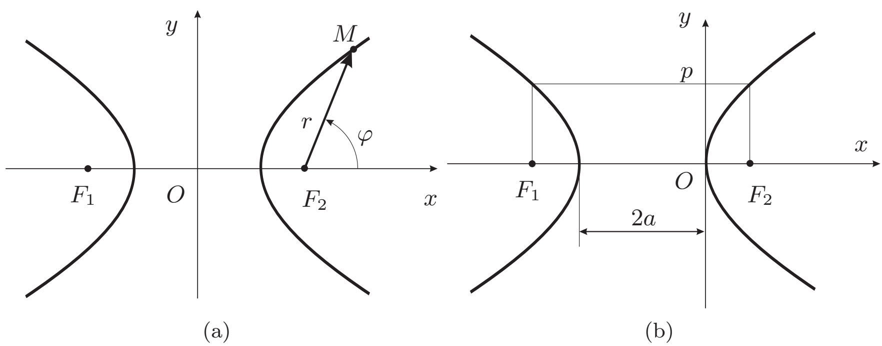


Рис. 11.15. Гипербола (a) в полярной системе координат, (b) в декартовой системе координат, отнесённой к правой вершине

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 при решении задачи Кеплера об обращении тел Солнечной системы вокруг Солнца.

В. 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.

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 (см. рис. 11.14(b)).

Формулы связи координат[1] имеют вид

$$x' = x - a, \quad y' = y.$$

Подставляя эти формулы в каноническое уравнение эллипса, получим

$$\frac{(x-a)^2}{a^2} + \frac{y^2}{b^2} = 1 \iff \frac{x^2}{a^2} - 2\frac{x}{a} + \frac{y^2}{b^2} = 0,$$

откуда

$$y^2 = 2\frac{b^2}{a}x - \frac{b^2}{a^2}x^2 \iff y^2 = 2px - \frac{p}{a}x^2,$$

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

$$y^2 = 2px + (\varepsilon^2 - 1)x^2.$$

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 (см. рис. 11.15(b)).

Подставляя формулы связи координат

$$x' = x + a, \quad y' = y$$

в каноническое уравнение гиперболы, получим

$$\frac{(x+a)^2}{a^2} - \frac{y^2}{b^2} = 1 \iff \frac{x^2}{a^2} + 2\frac{x}{a} - \frac{y^2}{b^2} = 0,$$

откуда

$$y^2 = 2\underbrace{\frac{b^2}{a}}_{=p}x + \frac{b^2}{a^2}x^2 \iff y^2 = 2px + \frac{p}{a}x^2.$$

Так как

$$\frac{b^2}{a^2} = \frac{c^2 - a^2}{a^2} = \varepsilon^2 - 1,$$

[1] Систематическому изучению преобразований координат посвящена лекция **??**.

последнее уравнение можно переписать в виде

$$y^2 = 2px + (\varepsilon^2 - 1)x^2.$$

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 $\varepsilon > 1$.

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.

С. Эллипс, гипербола и парабола как конические сечения. 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 образующих, впервые рассмотрел двуполостный конус, что позволило обнаружить вторую ветвь гиперболы, и предложил названия «эллипс», «гипербола» и «парабола» для конических сечений, или коник.

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.

Плоскость, проходящая через вершину конуса, может занимать относительно этого конуса следующие три положения (см. рис. 11.16):

[1] Т.е. имеющего такую же площадь.

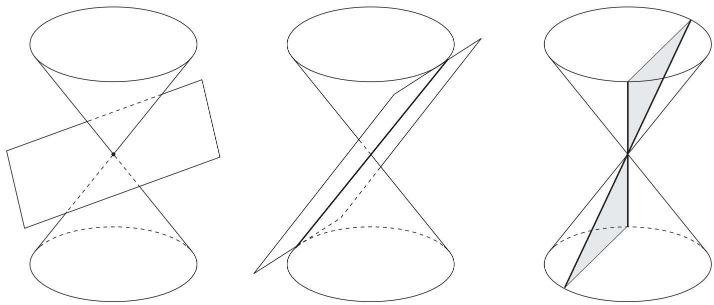


Рис. 11.16.

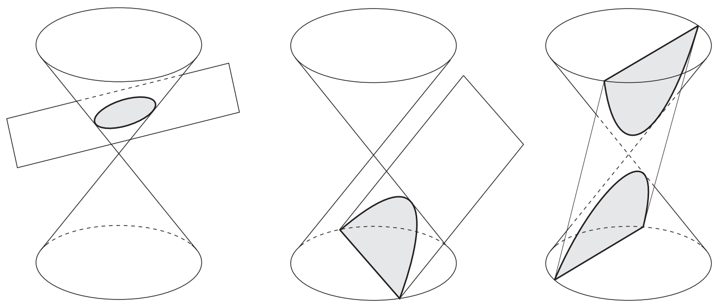


Рис. 11.17.

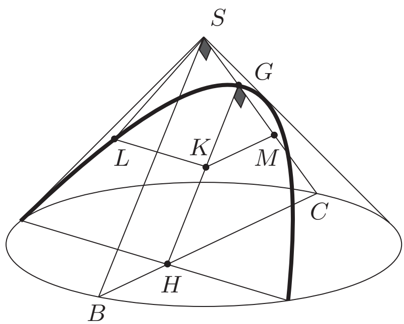
(i) иметь с конусом лишь одну общую точку — вершину конуса,
(ii) касаться конуса вдоль его образующей,
(iii) пересекать конус по двум его различным образующим.

Плоскость же, не проходящая через вершину конуса, может занимать относительно этого конуса следующие три положения (см. рис. 11.17):

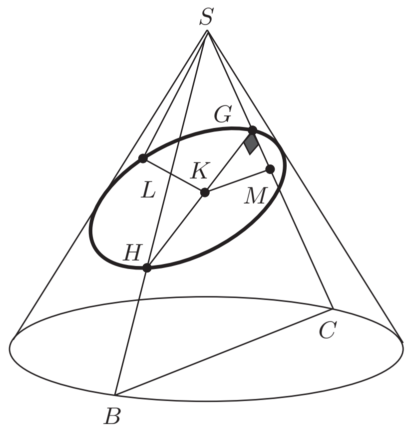
(i) пересекать все образующие конуса (в этом случае плоскость пересекает только одну полу конуса),
(ii) быть параллельной одной образующей конуса (в этом случае плоскость также пересекает только одну полу конуса),
(iii) быть параллельной двум различным образующим конуса (в этом случае плоскость пересекает обе полы конуса).

Изложим сначала 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 получаем кривые трёх типов:

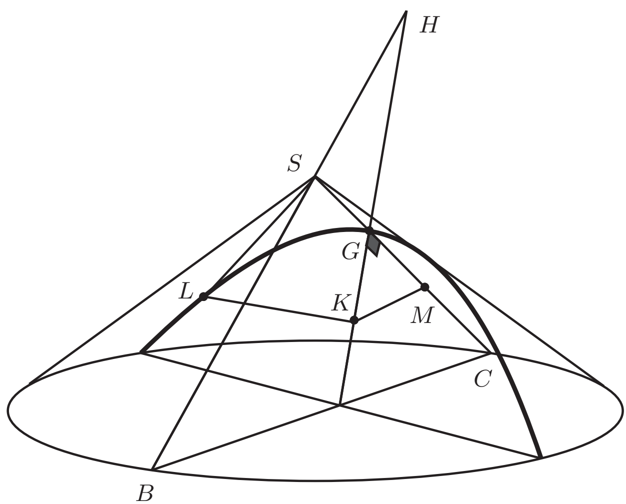
(i) эллипс, если $2\alpha < 90^\circ$;
(ii) гиперболу, если $2\alpha > 90^\circ$;
(iii) параболу, если $2\alpha = 90^\circ$.



(a) парабола



(b) эллипс



(c) гипербола

Рис. 11.18. Конические сечения

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 $|SM| = |SL|$.

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.

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 — вершины 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.

Аналогичные результаты получаются, если плоскость сечения проводить не перпендикулярно образующей конуса, а под произвольным углом. Чтобы продемонстрировать это, воспользуемся конструкцией, предложенной бельгийским математиком Ж. Данделеном (XIX в.)

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. 11.19).

Пусть M — произвольная точка на линии пересечения конуса с плоскостью π. Проведём через точку M образующую (SX) 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,

$$|MF_1| + |MF_2| = |MA_1| + |MA_2| = |A_1A_2| = \mathrm{const},$$

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.

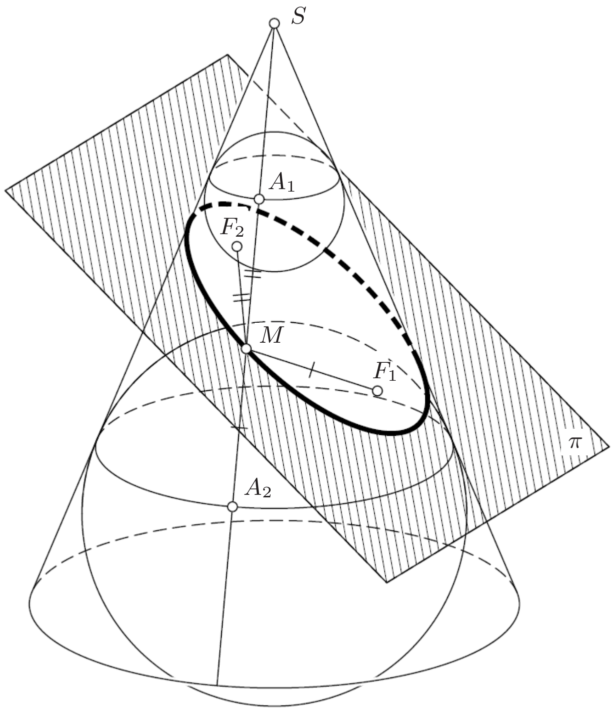


Рис. 11.19. Конструкция Данделена: эллипс

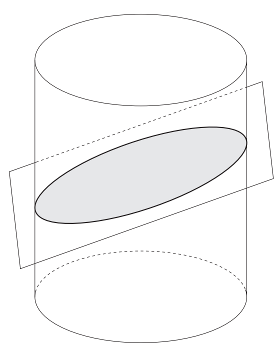


Рис. 11.20.

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.

Аналогично доказывается, что если секущая плоскость параллельна двум различным образующим конуса (т.е. пересекает обе его полы), то в сечении получается гипербола.

Рассмотрим случай, когда секущая плоскость π параллельна одной из образующей. 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. Возьмём произвольную точку M сечения конуса плоскостью π и найдём точку A пересечения образующей (SM) с

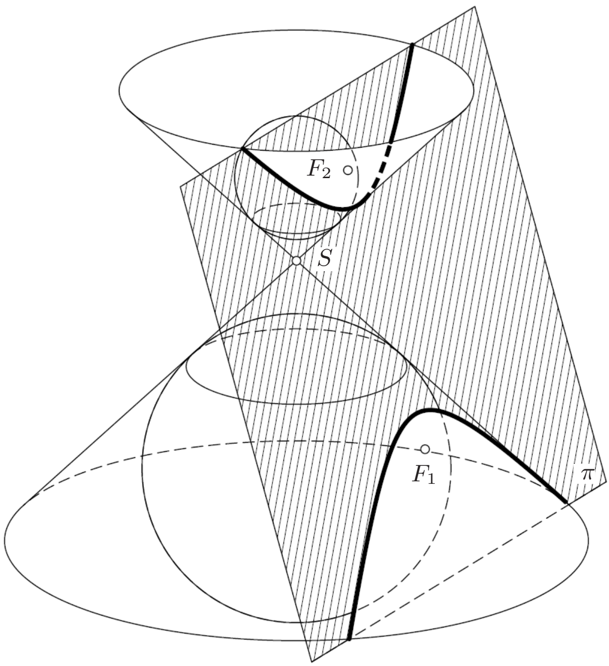


Рис. 11.21. Конструкция Данделена: гипербола

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.

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.

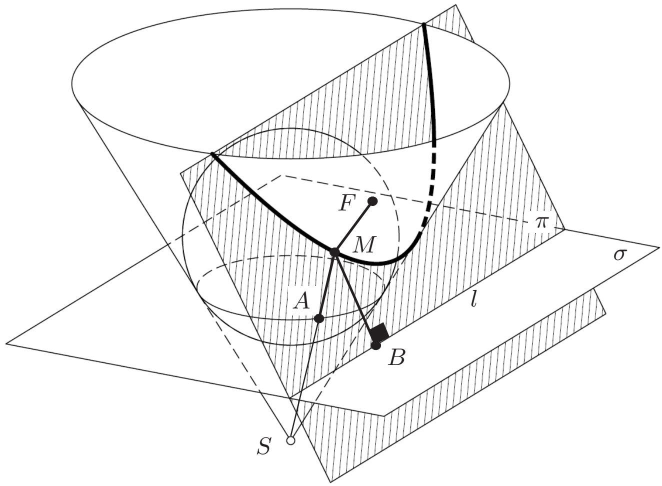


Рис. 11.22. Конструкция Данделена: парабола

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 12

Геометрические преобразования

1. Преобразование базисов и координат в аффинном пространстве

Пусть в аффинном пространстве $\mathcal{A}$ заданы две аффинные системы координат: $O\boldsymbol{E}$ («старая») и $O'\boldsymbol{E}'$ («новая»), где $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big)$ и $\boldsymbol{E}' = \big(\boldsymbol{e}_{1'}, \ldots, \boldsymbol{e}_{n'}\big)$ — «старый» и «новый» базисы в ассоциированном векторном пространстве $\mathcal{V}$, связанные формулами перехода

$$\boldsymbol{e}_{j'} = c^j_{j'}\boldsymbol{e}_j, \quad j = 1, \ldots, n, \quad j' = 1', \ldots, n', \tag{8.6}$$

или

$$\boldsymbol{E}' = \boldsymbol{E}C. \tag{8.7}$$

Разложим радиус-вектор $\boldsymbol{x}_0 = \overrightarrow{OO'}$ начала новой системы относительно старой по старому базису:

$$\overrightarrow{OO'} = x^j_0\boldsymbol{e}_j. \tag{12.1}$$

Введём столбцы координат

$$X_0 = \begin{pmatrix} x^1_0 \\ \vdots \\ x^n_0 \end{pmatrix}, \quad X = \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix}, \quad X' = \begin{pmatrix} x^{1'} \\ \vdots \\ x^{n'} \end{pmatrix}.$$

Очевидно,

$$\overrightarrow{OA} = \overrightarrow{OO'} + \overrightarrow{O'A}.$$

Радиус-векторы $\overrightarrow{OA}$ и $\overrightarrow{OO'}$ точек A и O' разложим по старому базису $\boldsymbol{E}$, а радиус-вектор $\overrightarrow{O'A}$ — по новому базису $\boldsymbol{E}'$. Тогда

$$\boldsymbol{E}X = \boldsymbol{E}X_0 + \boldsymbol{E}'X'.$$

Учитывая формулу (8.7) (см. с. 177) и пользуясь единственностью разложения по базису (см. теорему 8.15, с. 166), получаем

$$X = X_0 + CX'. \tag{12.2}$$

Отсюда легко получить и формулу обратного преобразования:

$$X' = C^{-1}(X - X_0) = C^{-1}X - C^{-1}X_0. \tag{12.3}$$

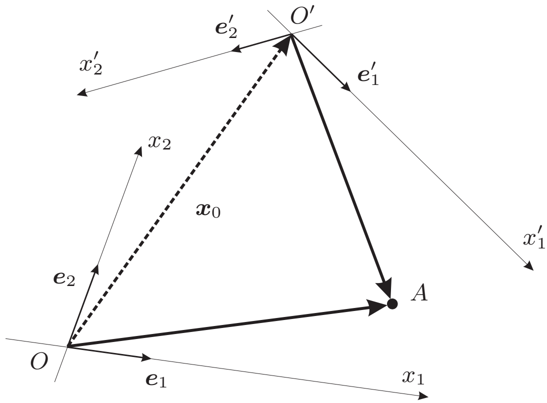


Рис. 12.1. Преобразование аффинного репера

Если положить здесь $X = O$, то получим координаты X_0' начала O системы $O\boldsymbol{E}$ в системе $O'\boldsymbol{E}'$:

$$X_0' = -C^{-1}X_0.$$

Для удобства часто[1] используются так называемые однородные координаты[2]: столбцы координат векторов и точек дополняются ещё одним элементом, равным 1, например,

$$Z = \begin{pmatrix} x^1 \\ \vdots \\ x^n \\ \hline 1 \end{pmatrix} = \left[\begin{array}{c} X \\ \hline 1 \end{array}\right], \quad Z' = \begin{pmatrix} x^{1'} \\ \vdots \\ x^{n'} \\ \hline 1 \end{pmatrix} = \left[\begin{array}{c} X' \\ \hline 1 \end{array}\right],$$

после чего, введя блочную матрицу

$$P = \left[\begin{array}{c|c} C & X_0 \\ \hline O & 1 \end{array}\right],$$

преобразования (12.2) и (12.3) можно записать в виде

$$Z = PZ', \quad Z' = QZ.$$

Действительно,

$$X = CX' + X_0 \iff PZ' = \left[\begin{array}{c|c} C & X_0 \\ \hline O & 1 \end{array}\right] \left[\begin{array}{c} X' \\ \hline 1 \end{array}\right] = \left[\begin{array}{c} CX' + X_0 \\ \hline 1 \end{array}\right]$$

и

[1] Особенно в компьютерной графике.

[2] Термин обязан своему происхождению проективной геометрии; об этом будет сказано подробнее в следующем семестре.

$$X' = C^{-1}X - C^{-1}X_0 \iff$$
$$\iff QZ = \left[\begin{array}{c|c} C^{-1} & -C^{-1}X_0 \\ \hline O & 1 \end{array}\right] \left[\begin{array}{c} X' \\ \hline 1 \end{array}\right] = \left[\begin{array}{c} C^{-1}X' - C^{-1}X_0 \\ \hline 1 \end{array}\right].$$

При помощи умножения легко проверить, что матрица Q является обратной к P:

$$QP = \left[\begin{array}{c|c} C^{-1} & -C^{-1}X_0 \\ \hline O & 1 \end{array}\right] \left[\begin{array}{c|c} C & X_0 \\ \hline O & 1 \end{array}\right] = \left[\begin{array}{c|c} \mathbb{1} & O \\ \hline O & 1 \end{array}\right].$$

Итак, формулы замены аффинных координат можно записать в матричном «однородном» виде:

$$Z = PZ', \quad Z' = P^{-1}Z, \tag{12.4}$$

где

$$P = \left[\begin{array}{c|c} C & X_0 \\ \hline O & 1 \end{array}\right], \quad P^{-1} \left[\begin{array}{c|c} C^{-1} & -C^{-1}X_0 \\ \hline O & 1 \end{array}\right]. \tag{12.5}$$

2. Преобразование базисов и координат в евклидовом пространстве

Формулы, полученные выше для преобразования базисов и координат в векторных (см. раздел 5 лекции 8, с. 176) и аффинных пространствах применимы и в случае евклидовых пространств. Однако в евклидовых пространствах преобразования базисов, переводящие один ортонормированный базис в другой, выделяются в специальный класс, который и будет предметом изучения в этом разделе.

А. Ортогональные преобразования базисов в евклидовом векторном пространстве. Пусть $\boldsymbol{I} = \big(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_n\big)$ и $\boldsymbol{I}' = \big(\boldsymbol{i}_{1'}, \ldots, \boldsymbol{i}_{n'}\big)$ — ортонормированные базисы в евклидовом *векторном* пространстве $\mathcal{E}$ («старый» и «новый»), связь между которыми устанавливается обычными формулами

$$\boldsymbol{i}_{k'} = c^k_{k'}\boldsymbol{i}_k, \quad \boldsymbol{I}' = \boldsymbol{I}C.$$

12.1. Определение. Матрица перехода C от одного ортонормированного базиса к другому называется *ортогональной матрицей.*

Установим основные свойства ортогональных матриц.

Так как оба базиса $\boldsymbol{I}$ и $\boldsymbol{I}'$ являются ортонормированными, скалярное произведение векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ выражается через их координаты по формуле (10.10):

$$(\boldsymbol{x}, \boldsymbol{y}) = X^TY = X'^TY',$$

где X, Y — столбцы координат векторов $\boldsymbol{x}$, $\boldsymbol{Y}$ в базисе $\boldsymbol{I}$, а X', Y' — в базисе $\boldsymbol{I}'$. Используя формулу (8.12) связи координат вектора в двух базисах, $X = CX'$, получаем тождество

$$X'^T Y' = X^T Y = (CX')^T (CY) = X'^T C^T C Y'.$$

Обе части полученного равенства

$$X'^T Y' = X'^T C^T C Y'$$

представляют собой многочлены от координат $x^{j'}$ и $y^{j'}$ векторов. Поскольку тождественное равенство многочленов имеет место тогда и только тогда, когда совпадают соответственные коэффициенты этих многочленов, приходим к соотношению

$$C^T C = \mathbb{1}.$$

Транспонируем обе части этого равенства:

$$(C^T C)^T = \mathbb{1}^T \quad \Rightarrow \quad CC^T = \mathbb{1}.$$

Умножим слева обе части на обратную матрицу C^{-1}:

$$C^{-1} C C^T = C^{-1} \mathbb{1} \quad \Rightarrow \quad C^T = C^{-1}.$$

Записав равенство $C^T C = \mathbb{1}$ в развёрнутом виде, получим

$$\sum_{j=1}^{n} c^j_{k'} c^j_{l'} = \delta_{k'l'}, \tag{12.6}$$

т.е. столбцы матрицы C составляют ортонормированное семейство (базис) пространства $\mathbb{R}^n$. Этот факт не удивителен, поскольку столбцы матрицы C состоят из координат ортонормированных векторов в ортонормированном базисе. Из равенства $CC^T = \mathbb{1}$ аналогично находим

$$\sum_{j'=1}^{n'} c^k_{j'} c^l_{j'} = \delta^{kl}. \tag{12.7}$$

Итак, доказана следующая теорема.

12.2. Теорема. *Матрица C является ортогональной тогда и только тогда, когда она обладает одним (а потому и любым другим) из следующих свойств:*

(1) $C^T C = \mathbb{1}$;
(2) $CC^T = \mathbb{1}$;
(3) $C^T = C^{-1}$;
(4) *ортонормированность столбцов, выраженная равенством* (12.6);
(5) *ортонормированность строк, выраженная равенством* (12.7).

Применяя теоремы об определителе произведения матриц и об определителе транспонированной матрицы (теорема 6.19, с. 141 и теорема 6.14, с. 135), из равенства $C^T C = \mathbb{1}$ получаем $\det C^T \cdot \det C = 1$, откуда

$$|\det C| = 1 \iff \det C = \pm 1.$$

Ортогональная матрица с определителем 1 называется *собственной*, а с определителем -1 — *несобственной*.

12.3. 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 (см. определение 6.5, с. 129). 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 $\det C^{-1} = 1$. □

12.4. Замечание. Обратите внимание, что несобственные ортогональные матрицы не образуют группу: множество таких матриц не содержит единичной матрицы, а произведение двух несобственных матриц является матрицей собственной.

B. Группы $O(1)$ и $SO(1)$. При $n = 1$ ортогональная матрица является числом $c^1_{1'}$, удовлетворяющим условию $(c^1_{1'})^2 = 1$. Таким образом,

$$O(1) = \{1,\ -1\}, \quad SO(1) = \{1\}.$$

C. Группы $O(2)$ и $SO(2)$. Найдём общий вид ортогональной матрицы порядка 2. Записав в развёрнутом виде условие $C^TC = \mathbb{1}$ для матрицы

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

получим уравнения

$$\begin{aligned} (c^1_{1'})^2 + (c^2_{1'})^2 &= 1, \\ c^1_{1'}c^1_{2'} + c^2_{1'}c^2_{2'} &= 0, \\ (c^1_{2'})^2 + (c^2_{2'})^2 &= 1. \end{aligned}$$

Первое и последнее уравнения означают, что найдутся такие углы α и β, что

$$c^1_{1'} = \cos\alpha, \qquad c^2_{1'} = \sin\alpha,$$
$$c^1_{2'} = \cos\beta, \qquad c^2_{2'} = \sin\beta,$$

а второе — что эти углы связаны соотношением

$$\cos\alpha\cos\beta + \sin\alpha\sin\beta = 0 \iff \cos(\alpha-\beta) = 0.$$

Таким образом, либо $\beta = \alpha + \pi/2$, и матрица C имеет вид

$$R = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}, \tag{12.8}$$

либо $\beta = \alpha - \pi/2$, и матрица C имеет вид

$$S = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}, \tag{12.9}$$

В первом случае матрица собственная, во втором — несобственная.

D. Поворот прямоугольной системы координат на плоскости. Преобразуем ортонормированный базис $(\boldsymbol{i}, \boldsymbol{j})$ на плоскости с помощью собственной ортогональной матрицы R, определённой формулой (12.8). Формулы преобразований для векторов базиса и координаты имеют следующий вид:

$$\text{переход } O\boldsymbol{ij} \to O\boldsymbol{i}'\boldsymbol{j}'\colon \quad \begin{cases} \boldsymbol{I}' = \boldsymbol{I}R, \\ X' = R^{-1}X = R^TX, \end{cases} \tag{12.10}$$

$$\text{переход } O\boldsymbol{i}'\boldsymbol{j}' \to O\boldsymbol{ij}\colon \quad \begin{cases} \boldsymbol{I} = \boldsymbol{I}'R^{-1} = \boldsymbol{E}'R^T, \\ X = RX', \end{cases} \tag{12.11}$$

где $X = (x, y)^T$, $X' = (x', y')^T$ — столбцы координат точки относительно систем координат $O\boldsymbol{ij}$ и $O\boldsymbol{i}'\boldsymbol{j}'$ соответственно. Такое преобразование есть не что иное, как поворот векторов базиса на угол α (см. рис. 12.2).

Преобразования векторов базиса и преобразование столбцов координат производятся взаимно обратными матрицами. Этот факт незаметен (более того, создаётся противоположное впечатление!), если формулы записать в координатном виде:

$$\begin{cases} \boldsymbol{i}' = \boldsymbol{i}\cos\alpha + \boldsymbol{j}\sin\alpha, \\ \boldsymbol{j}' = -\boldsymbol{i}\sin\alpha + \boldsymbol{j}\cos\alpha, \end{cases} \qquad \begin{cases} x' = x\cos\alpha + y\sin\alpha, \\ y' = -x\sin\alpha + y\cos\alpha \end{cases} \tag{12.12}$$

для перехода $O\boldsymbol{ij} \to O\boldsymbol{i}'\boldsymbol{j}'$ и

$$\begin{cases} \boldsymbol{i} = \boldsymbol{i}'\cos\alpha - \boldsymbol{j}'\sin\alpha, \\ \boldsymbol{j} = \boldsymbol{i}'\sin\alpha + \boldsymbol{j}'\cos\alpha, \end{cases} \qquad \begin{cases} x = x'\cos\alpha - y'\sin\alpha, \\ y = x'\sin\alpha + y'\cos\alpha. \end{cases} \tag{12.13}$$

для обратного перехода $O\boldsymbol{i}'\boldsymbol{j}' \to Oxy$. Конечно, всё дело здесь в порядке матричных сомножителей и ортогональности матрицы перехода.

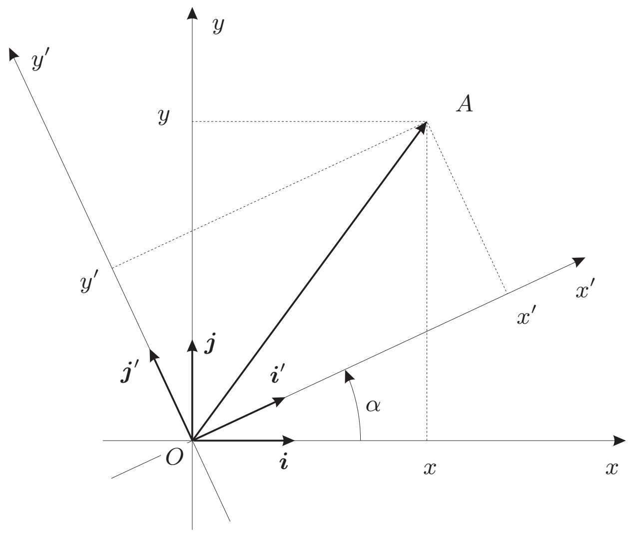


Рис. 12.2. Поворот прямоугольной декартовой системы координат

E. Ортогональные преобразования системы координат на плоскости. В случае, когда система координат $O'\boldsymbol{i}'\boldsymbol{j}'$ получена из системы $O\boldsymbol{i}\boldsymbol{j}$ поворотом осей на угол α и переносом начала координат на вектор $\overrightarrow{OO'}$, столбец координат которого относительно системы $O\boldsymbol{i}\boldsymbol{j}$ есть $\big(x_0, y_0\big)^T$, формулы преобразования имеют вид

$$\begin{cases} x = x'\cos\alpha - y'\sin\alpha + x_0, \\ y = x'\sin\alpha + y'\cos\alpha + y_0, \end{cases} \tag{12.14}$$

и

$$\begin{cases} x' = (x - x_0)\cos\alpha + (y - y_0)\sin\alpha, \\ y' = -(x - x_0)\sin\alpha + (y - y_0)\cos\alpha. \end{cases} \tag{12.15}$$

В матричных обозначениях

$$X = RX' + X_0, \tag{12.16}$$

$$X' = R^{-1}(X - X_0) = R^T(X - X_0). \tag{12.17}$$

Если ввести матрицы

$$Z = \left(\begin{array}{c} x \\ y \\ \hline 1 \end{array}\right) = \left[\begin{array}{c} X \\ \hline 1 \end{array}\right], \quad Z' = \left(\begin{array}{c} x' \\ y' \\ \hline 1 \end{array}\right) = \left[\begin{array}{c} X' \\ \hline 1 \end{array}\right],$$
$$P = \left(\begin{array}{cc|c} \cos\alpha & -\sin\alpha & x_0 \\ \sin\alpha & \cos\alpha & y_0 \\ \hline 0 & 0 & 1 \end{array}\right) = \left[\begin{array}{c|c} R & X_0 \\ \hline O & 1 \end{array}\right], \tag{12.18}$$

то преобразования координат запишутся в виде

$$Z = PZ', \quad Z' = P^{-1}Z, \tag{12.19}$$

причём обратная матрица P^{-1} имеет вид

$$P^{-1} = \left[\begin{array}{c|c} R^T & -R^T X_0 \\ \hline O & 1 \end{array}\right] = \left(\begin{array}{cc|c} \cos\alpha & \sin\alpha & x_0' \\ -\sin\alpha & \cos\alpha & y_0' \\ \hline 0 & 0 & 1 \end{array}\right) \tag{12.20}$$

(см. (12.5)), где

$$\begin{cases} x_0' = -x_0\cos\alpha - y_0\sin\alpha, \\ y_0' = x_0\sin\alpha - y_0\cos\alpha \end{cases} \iff \quad X_0' = -R^T X_0$$

— координаты точки O (начала системы координат $O\boldsymbol{ij}$) в системе координат $O'\boldsymbol{i}'\boldsymbol{j}'$.

Поворот координатных осей прямоугольной декартовой системы и перенос её начала, а также преобразования координат точек и уравнений линий называют *ортогональными преобразованиями.*

F. Ортогональные преобразования уравнения прямой на плоскости. 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

$$N^T X + d = 0, \tag{12.21}$$

где $N = \bigl(A, B\bigr)^T$ — столбец координат вектора нормали $\boldsymbol{n}$ прямой. При переходе к новой системе координат $O'\boldsymbol{i}'\boldsymbol{j}'$ по формулам (12.16) уравнение преобразуется следующим образом:

$$N^T\bigl(RX' + X_0\bigr) + d = 0 \iff (N^T R)X' + N^T X_0 + d = 0,$$

т.е. принимает вид, совпадающий с 12.21:

$$N'^T X' + d' = 0.$$

Матричные коэффициенты уравнения прямой преобразуются по формулам

$$N'^T = N^T R \iff N' = (N^T R)^T = R^T N = R^{-1}N, \quad d' = N^T X_0 + d.$$

(Обратите внимание на полное соответствие соотношения $N' = R^{-1}N$ формуле (12.10), по которой преобразуются координаты вектора при повороте системы координат!)

Если записать уравнение прямой в виде

$$MZ = 0,$$

[1] Напомним, что это утверждение верно только для прямоугольной системы координат (см. замечание 10.22, с. 223).

где

$$M = \begin{pmatrix} A & B & d \end{pmatrix}, \quad Z = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix},$$

а преобразование координат осуществить по формулам (12.19), получим

$$M(PZ') = 0 \iff (MP)Z' = 0,$$

т.е. матричный коэффициент M уравнения прямой преобразуется по формуле

$$M' = MP.$$

3. Аффинные преобразования

Формулу $X = CX' + X_0$ (см. (12.2)) 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

$$f : \mathcal{A} \to \mathcal{A}, \quad X' \mapsto X = X_0 + CX',$$

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.

А. Основные свойства аффинных преобразований. Пусть $(\mathcal{A}, \mathcal{V}, \mathbb{K})$ — аффинное пространство, 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 $\mathcal{V}$.

12.5. Определение. Отображение

$$f : \mathcal{A} \to \mathcal{A}, \quad X \mapsto X' = CX + X_0, \tag{12.22}$$

где C — невырожденная матрица, называется *аффинным преобразованием*[1] пространства $\mathcal{A}$. Аффинные преобразования можно записывать и с помощью однородных координат:

$$Z' = PZ, \quad P = \left[\begin{array}{c|c} C & X_0 \\ \hline O & 1 \end{array}\right].$$

Матрицы C и P называются соответственно *основной* и *расширенной матрицами преобразования*. Группа всех аффинных преобразований аффинного пространства $\mathcal{A}$ обозначается $\operatorname{Aff} \mathcal{A}$.

[1] Обратите внимание на различие в записи формул (12.22) и (12.2): сейчас точка X — прообраз, точка X' — образ, т.е. результат действия преобразования.

12.6. 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.*
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

$$X = A_0 + t^1A_1 + \cdots + t^pA_p,$$

где A_0 — опорная точка, $A_1, \ldots, A_p$ — линейно независимые направляющие векторы. После преобразования (12.22) получим

$$X' = C\Big(A_0 + t^1A_1 + \cdots + t^pA_p\Big) + X_0 \iff$$
$$\iff X' = (CA_0 + X_0) + t^1(CA_1) + \cdots + t^p(CA_p).$$

Поскольку векторы $CA_1, \ldots, CA_p$ линейно независимы (объясните это самостоятельно, взяв за образец доказательство утверждения 2 теоремы 8.40, с. 173), полученное уравнение определяет p-плоскость с опорной точкой $CA_0 + X_0$ и направляющими векторами $CA_1, \ldots, CA_p$.

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

$$\overrightarrow{X_1'X_3'} = X_3' - X_1' = (CX_3 + X_0) - (CX_1 + X_0) =$$
$$= C(X_3 - X_1) = C\overrightarrow{X_1X_3} = Ck\overrightarrow{X_1X_2} = kC(X_2 - X_1) =$$
$$= k\Big[(CX_2 + X_0) - (CX_1 + X_0)\Big] = k(X_2' - X_1') = k\overrightarrow{X_1'X_2'}.$$

Предложение доказано. □

B. Примеры аффинных преобразований плоскости.

1. *Параллельный перенос*, или *трансляция* на вектор X_0: $C = \mathbb{1}$, $X' = X + X_0$.

---

[1] Т.е. лежащие на одной прямой.

[2] Для неколлинеарных отрезков это утверждение не имеет места; приведите пример самостоятельно.

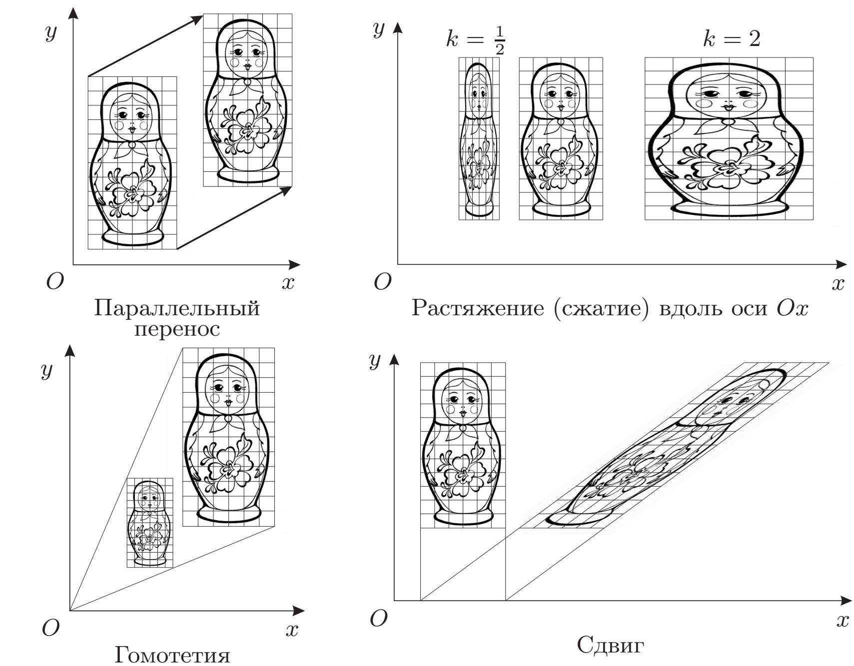


Рис. 12.3. Аффинные преобразования плоскости

2. *Гомотетия*, или *равномерное растяжение* (*сжатие*) с центром в начале координат[1]: $X' = kX$, где $k \neq 0$. Матрицей гомотетии является матрица вида $k\mathbb{1}$ (такие матрицы называются *скалярными*).
3. *Растяжение* (*сжатие*) *вдоль оси абсцисс*:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} kx \\ y \end{pmatrix}; \quad C = \begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix}.$$

   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*

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x + py \\ y \end{pmatrix}; \quad C = \begin{pmatrix} 1 & p \\ 0 & 1 \end{pmatrix}.$$

[1] Конечно, можно рассматривать гомотетию с центром, отличным от начала координат.

4. Ортогональные преобразования евклидовых пространств

A. Основные свойства ортогональных преобразований. *Ортогональное преобразование*, или *движение* — это аффинное преобразование, основная матрица C которого ортогональна:

$$X' = CX + X_0.$$

12.7. 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). $\square$

Определитель ортогональной матрицы равен 1 или -1. В первом случае движение сохраняет ориентацию плоскости и называется *собственным движением* или *движением первого рода*; во втором случае движение изменяет ориентацию и называется *несобственным движением* или *движением второго рода.*

Изучим более подробно движения евклидовой плоскости.

B. Движения плоскости. Каждое движение плоскости 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*:

$$\begin{cases} x' = x\cos\alpha - y\sin\alpha + x_0, \\ y' = x\sin\alpha + y\cos\alpha + y_0, \end{cases} \tag{12.23}$$

Во втором случае $\det S = -1$; такое движение *изменяет* ориентацию плоскости и называется *движением второго рода*, или *несобственным движением*:

$$\begin{cases} x' = x\cos\alpha + y\sin\alpha + x_0, \\ y' = x\sin\alpha - y\cos\alpha + y_0. \end{cases} \tag{12.24}$$

Классификацию движений удобно провести, используя понятие неподвижной точки. Пусть $f : \mathcal{E} \to \mathcal{E}$ — какое-либо преобразование плоскости. Точка M называется *неподвижной точкой* преобразования f, если $f(M) = M$.

С. Собственные движения плоскости. При $\alpha = 0$ (в этом случае $R = \mathbb{1}$) собственное движение (12.23) записывается формулами

$$\begin{cases} x' = x + x_0, \\ y' = y + y_0. \end{cases} \tag{12.25}$$

Если $x_0 = y_0 = 0$, то это тождественное преобразование; в противном случае указанное преобразование неподвижных точек не имеет.

12.8. Определение. Собственное движение плоскости, не имеющее неподвижных точек, называется *параллельным переносом*, или *трансляцией* (см. рис. 12.4(a)).

Ясно, что каждая трансляция $\boldsymbol{T}$ однозначно задаётся вектором (x_0, y_0) и может быть отождествлена с ним. Это объясняет еще одну возможную точку зрения на свободный вектор как на параллельный перенос.

Условие того, что точка M является неподвижной точкой собственного движения (12.23), 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,

$$\begin{cases} x(1-\cos\alpha) + \quad y\sin\alpha \quad = x_0, \\ \quad -x\sin\alpha \quad + y(1-\cos\alpha) = y_0. \end{cases} \tag{12.26}$$

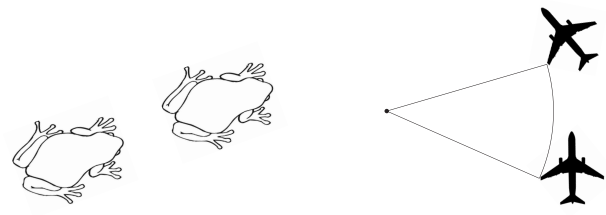
Определитель этой системы

$$\begin{vmatrix} 1-\cos\alpha & \sin\alpha \\ -\sin\alpha & 1-\cos\alpha \end{vmatrix} = (1-\cos\alpha)^2 + \sin^2\alpha = 2(1-\cos\alpha)$$

при $\alpha \neq 0$ отличен от нуля, так что система имеет единственное решение, а соответствующее собственное движение — единственную неподвижную точку.

12.9. Определение. Собственное движение плоскости, имеющее одну неподвижную точку, называется *вращением* или *поворотом*, а сама неподвижная точка — центром вращения.

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. 12.4(b)).



(a) (b)

Рис. 12.4. Собственные движения плоскости:
(a) параллельный перенос; (b) вращение

D. Несобственные движения плоскости. Обсудим теперь неподвижные точки несобственных движений плоскости (12.24); 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,

$$\begin{cases} x(1-\cos\alpha) - \quad y\sin\alpha \quad = x_0, \\ \quad -x\sin\alpha \quad + y(1+\cos\alpha) = y_0. \end{cases} \tag{12.27}$$

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).

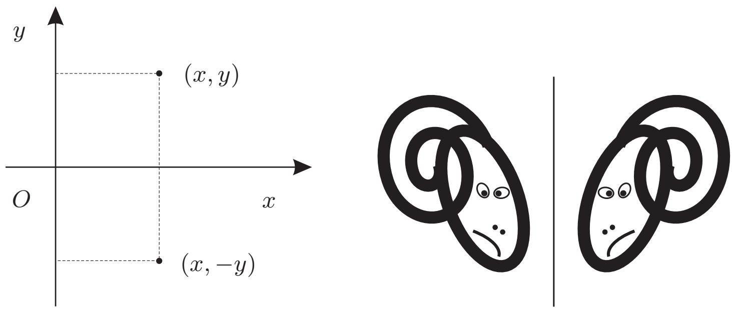
E. Осевая симметрия. Выясним геометрический смысл несобственного движения, неподвижные точки которого заполняют прямую.

12.10. Пример. Неподвижные точки несобственного движения вида (12.24) 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. 12.5(a)).

Докажем, что любое несобственное движение, имеющее прямую неподвижных точек, имеет аналогичный геометрический смысл.



(a) (b)

Рис. 12.5. Осевая симметрия

12.11. Предложение. *Пусть* $\boldsymbol{S}$ *— несобственное движение, для которого прямая* a *состоит из неподвижных точек. Справедливы следующие утверждения*:

(i) *для любой точки* X *и её образа* $X' = \boldsymbol{S}(X)$ *вектор* $\overrightarrow{XX'}$ *перпендикулярен прямой* a;
(ii) *середина отрезка* XX' *лежит на прямой* a;
(iii) *движение* $\boldsymbol{S}$ *инволютивно, т.е. его квадрат равен тождественному отображению*:

$$\boldsymbol{S} \circ \boldsymbol{S} = \mathrm{id} \iff \boldsymbol{S}^{-1} = \boldsymbol{S}.$$

Эти свойства проиллюстрированы на рис. 12.5.

Доказательство. В рассматриваемом случае система (12.27), 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 (12.27). Итак, для того чтобы рассматриваемое несобственное движение обладало прямой неподвижных точек, необходимо и достаточно выполнение соотношений

$$x_0 \sin\alpha + y_0(1 - \cos\alpha) = 0 \iff x_0(1 + \cos\alpha) + y_0 \sin\alpha = 0. \qquad (12.28)$$

Каждое из уравнение системы (12.27) при этом является уравнением прямой неподвижных точек. Таким образом, вектор нормали этой прямой можно взять, например, в виде $\boldsymbol{n} = \big(1 - \cos\alpha, -\sin\alpha\big)$; тогда направляющий вектор этой прямой есть $\boldsymbol{a} = \big(\sin\alpha, 1 - \cos\alpha\big)$.

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 (12.24):

$$\overrightarrow{XX'} = X' - X = (SX + X_0) - X = \begin{pmatrix} x\cos\alpha + y\sin\alpha + x_0 \\ x\sin\alpha - y\cos\alpha + y_0 \end{pmatrix} - \begin{pmatrix} x \\ y \end{pmatrix} = \\ = \begin{pmatrix} x(\cos\alpha - 1) + y\sin\alpha + x_0 \\ x\sin\alpha - y(\cos\alpha + 1) + y_0 \end{pmatrix}.$$

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 (12.28). Утверждение 1 доказано.

2. Доказательство того факта, что точка Y, являющаяся серединой отрезка XX', лежит на прямой неподвижных точек a, 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 a, т.е. в любое из уравнений (12.27). Принимая во внимание соотношения (12.28), обнаруживаем, что точка Y действительно лежит на прямой a:

$$\frac{1}{2}\Big(x(1+\cos\alpha)+y\sin\alpha+x_0\Big)(1-\cos\alpha)-\frac{1}{2}\Big(x\sin\alpha+y(1-\cos\alpha)+y_0\Big)\sin\alpha = x_0.$$

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 (12.28), находим

$$S^2 = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix} = \mathbb{1},$$

$$SX_0 + X_0 = \begin{pmatrix} x_0\cos\alpha + y_0\sin\alpha + x_0 \\ x_0\sin\alpha - y_0\cos\alpha + y_0 \end{pmatrix} = \begin{pmatrix} x_0(1 + \cos\alpha) + y_0\sin\alpha \\ x_0\sin\alpha + y_0(1 - \cos\alpha) \end{pmatrix} = O,$$

т.е. $\boldsymbol{S}\big(\boldsymbol{S}(X)\big) = X$, т.е. $\boldsymbol{S}^2 = \mathrm{id}$. □

Предложение 12.11 объясняет происхождение следующего термина.

12.12. 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.*

12.13. Замечание. Из предложения 12.11(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.

F. Скользящая симметрия.

12.14. Определение. Несобственное движение плоскости, не имеющее неподвижных точек, называется *скользящей симметрией*.

12.15. Предложение. *Каждая скользящая симметрия* $\boldsymbol{G}$ *является композицией осевой симметрии* $\boldsymbol{S}$ *и трансляции* $\boldsymbol{T}$ *вдоль её оси:* $\boldsymbol{G} = \boldsymbol{T} \circ \boldsymbol{S}$.

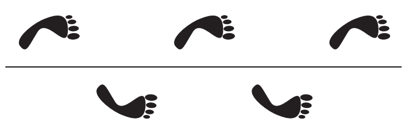


Рис. 12.6. Скользящая симметрия

Доказательство. Пусть $\boldsymbol{G}$ — скользящая симметрия, т.е. движение плоскости вида (12.24), не имеющее неподвижных точек; в этом случае система (12.27) несовместна. Покажем, что существует такой 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 $\boldsymbol{G} = \boldsymbol{T}^{-1} \circ \boldsymbol{G}$ 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

$$\left\{ \begin{aligned} x(1-\cos\alpha) - \quad y\sin\alpha \quad &= x_0 - p, \\ -x\sin\alpha \quad + y(1+\cos\alpha) &= y_0 - q. \end{aligned} \right. \tag{12.29}$$

Задача состоит в том, чтобы найти такие p и q, чтобы

(a) эта система определяла прямую (тогда преобразование $\boldsymbol{G}$ будет симметрией),

(b) полученная прямая была параллельна вектору $\boldsymbol{T}$.

Условие (a) будет выполнено, если уравнения системы (12.29) 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 (12.11), первое из этих равенств выполняется тождественно, а второе даёт уравнение для p и q:

$$\frac{-\sin\alpha}{1+\cos\alpha} = \frac{x_0 - p}{y_0 - q} \iff (x_0 - p)\cos\frac{\alpha}{2} + (y_0 - q)\sin\frac{\alpha}{2} = 0$$

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 (b) означает ортогональность вектора трансляции $\boldsymbol{T}$ и нормального вектора $\boldsymbol{n}$ прямой (12.29):

$$(\boldsymbol{n},\boldsymbol{T})=\begin{pmatrix}1-\cos\alpha\\ -\sin\alpha\end{pmatrix}^T\cdot\begin{pmatrix}p\\ q\end{pmatrix}=0 \iff p\sin\frac{\alpha}{2}-q\cos\frac{\alpha}{2}=0.$$

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. □

G. Композиции движений плоскости. Результаты, полученные выше, составляют содержание теоремы Шаля[1].

12.16. Теорема (теорема Шаля)**.** *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 $\boldsymbol{G}$*:

	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 $\boldsymbol{S}$

12.17. Замечание. Очевидно, что композиция двух собственных или двух несобственных движений является собственным движением, а композиция собственного и несобственного — несобственным движением.

[1] Мишель Шаль (фр. M. Chasles) — французский математик и механик (1793–1880).

12.18. 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 движением (см. замечание 12.17). Поскольку $\boldsymbol{S} \circ \boldsymbol{S} = \mathrm{id}$, имеем

$$\boldsymbol{F}_2 = (\boldsymbol{S} \circ \boldsymbol{S}) \circ \boldsymbol{F}_2 = \boldsymbol{S} \circ (\boldsymbol{S} \circ \boldsymbol{F}_2) = \boldsymbol{S} \circ \boldsymbol{F}_1,$$

что и требовалось. □

12.19. Замечание. 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.

12.20. 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. замечание 12.17), имеющим неподвижную точку O, так что по теореме Шаля представляет собой осевую симметрию. □

12.21. 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 12.13(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 $\boldsymbol{T} = \boldsymbol{S}_2\circ\boldsymbol{S}_1$.

Для доказательства утверждения 2(ii) заметим, что $\boldsymbol{R} = (\boldsymbol{R}\circ\boldsymbol{S})\circ\boldsymbol{S}$, а преобразование $\boldsymbol{R}\circ\boldsymbol{S}$ в силу предложения 12.20 является осевой симметрией. □

Из предложений 12.18 и 12.21 вытекает следующее утверждение.

12.22. Теорема. *Любое движение плоскости может быть представлено как композиция не более чем трёх осевых симметрий.*

5. Групповой подход в геометрии

А. Основные определения. 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* (так, наиболее интересными для нас будут аффинные и евклидовы пространствы), подмножества этих пространств — *геометрическими фигурами*, а слово *множество* будем использовать в остальных случаях.

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 $S(X)$.

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

$$g\circ f\colon X \to X, \quad x \mapsto g\big(f(x)\big).$$

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.25, с. 331).

Таким образом, множество $S(X)$ всех преобразований пространства X образует группу относительно операции композиции преобразований, которая называется *группой преобразований пространства* X. Читателю рекомендуется освежить в памяти содержание п. 6.5 раздела C главы 11 (см. сс. 129–132).

Пусть $(G, \circ)$ — группа, т.е. множество G с введённой на нём бинарной операцией $\circ$,

$$\circ : G \times G \to G, \quad (a, b) \mapsto a \circ b,$$

удовлетворяющей аксиомам группы **Г1–Г3** (см. с. 129).

12.23. Определение. Подмножество $H \subset G$ называется *подгруппой* группы G (обозначение[1] $H \Subset G$), если оно замкнуто относительно групповой операции $\circ$ и операции взятия обратного элемента, т.е.

(i) для всех $a, b \in H$ имеем $a \circ b \in H$,
(ii) для любого $a \in H$ имеем $a^{-1} \in H$.

Ясно, что каждая подгруппа H группы G сама является группой относительно той же групповой операции.

Пусть $H \Subset S(X)$ — некоторая подгруппа группы преобразований пространства X. Будем говорить, что геометрическая фигура $\Psi \subset X$ в пространстве X *эквивалентна* фигуре $\Phi \subset X$ относительно группы H, если в этой группе существует такое преобразование $h \in H$, что $\Psi = h(\Phi)$; обозначение $\Psi \overset{H}{\sim} \Phi$. Легко проверить, опираясь на свойство взаимной однозначности каждого преобразования из H, что отношение эквивалентности геометрических фигур в смысле приведённого определения действительно является отношением эквивалентности, т.е. обладает следующими свойствами:

(1) рефлексивность: $\Phi \overset{H}{\sim} \Phi$;
(2) симметричность: если $\Psi \overset{H}{\sim} \Phi$, то $\Phi \overset{H}{\sim} \Psi$;
(3) транзитивность: если $\Psi \overset{H}{\sim} \Phi$ и $\Xi \overset{H}{\sim} \Psi$, то $\Xi \overset{H}{\sim} \Phi$.

B. Эрлангенская программа Ф. Клейна. 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».

[1] Ср. с определением 8.26 и обозначением подпространства векторного пространства, с. 170.

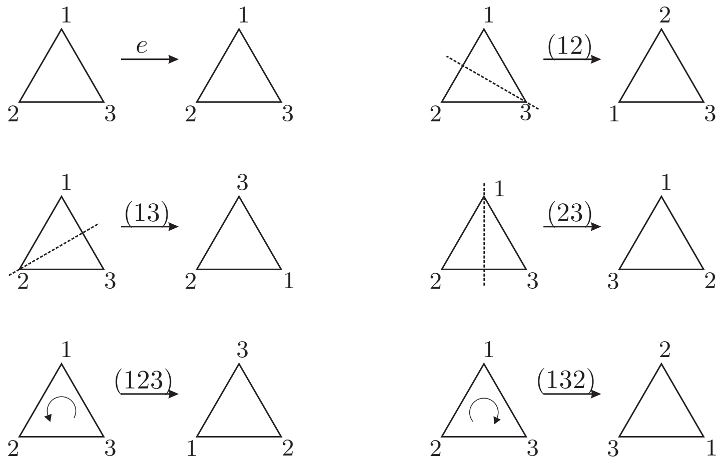


Рис. 12.7. Группа симметрий правильного треугольника

Такой подход к геометрии был впервые предложен Ф. Клейном[1] в 1872 г. в его работе, известной в истории математики как «Эрлангенская программа».

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 3), 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 4).

C. Группы симметрий. Группы являются также наиболее удобным и адекватным средством описания симметрий геометрических фигур.

Пусть Φ — некоторая геометрическая фигура (т.е. множество точек) на евклидовой плоскости $\mathcal{E}$. Преобразование плоскости $f : \mathcal{E} \to \mathcal{E}$ называется симметрией фигуры Φ, если $f(\Phi) = \Phi$. Например, отражение относительно прямой, содержащей диагональ ромба, является симметрией этого ромба.

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

[1]Феликс Клейн (нем. F. Klein) — немецкий математик (1849–1925).

фигуры Ф, обозначаемая символом Sym Ф, позволяет «измерить» степень симметричности фигуры: чем обширнее эта группа, тем фигура «более симметрична».

Пример 12.1. Отрезок $[AB]$ прямой обладает, помимо тождественного преобразования, единственной симметрией: симметрией отражения относительно собственного центра.

Пример 12.2. 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 (см. рис. 12.7).

ГЛАВА 13

Квадрики: канонизация и инварианты

1. Уравнение квадрики и его инварианты

А. Различные формы уравнения квадрики. В лекции 11 мы рассмотрели три важные линии на евклидовой плоскости — эллипс, гиперболу и параболу, и установили, что каждая из них задаётся в некоторой системе координат уравнением вида $F(x, y) = 0$, где $F(x, y)$ — многочлен второй степени от двух переменных:

(1) для эллипса $F(x, y) = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - 1$,

(2) для гиперболы $F(x, y) = \dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} - 1$,

(3) для параболы $F(x, y) = y^2 - 2px$.

Теперь выясним, какие ещё линии на плоскости можно задать уравнением вида $F(x, y) = 0$, где $F(x, y)$ — произвольный многочлен второй степени от двух переменных.

Итак, рассмотрим уравнение второй степени от двух переменных

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2b_1x + 2b_2y + c = 0; \tag{13.1}$$

многочлен в левой части этого уравнения обозначим $F(x, y)$. Пусть на евклидовой плоскости задана некоторая прямоугольная декартова система координат Oxy. Линия, точки которой (точнее, координаты точек) и только они являются решениями уравнения (13.1), называется *квадрикой*. При преобразовании системы координат уравнение квадрики, разумеется, изменяется; нашей окончательной целью является нахождение такой системы координат, в которой уравнение квадрики имеет по возможности наиболее простой вид.

Введём матрицы

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \text{где } a_{12} = a_{21}, \quad B = \begin{pmatrix} b_1 & b_2 \end{pmatrix}, X = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Тогда уравнение (13.1) можно записать в виде

$$X^T AX + 2BX + c = 0. \tag{13.2}$$

Введём также матрицы

$$D = \left(\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ \hline b_1 & b_2 & c \end{array}\right) = \left[\begin{array}{c|c} A & B^T \\ \hline B & c \end{array}\right] \qquad Z = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \left[\begin{array}{c} X \\ \hline 1 \end{array}\right];$$

тогда уравнение квадрики можно записать в виде

$$Z^T DZ = 0. \tag{13.3}$$

Отметим, что уравнение

$$W^T DW = 0, \quad \text{где} \quad W = \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

определяет в пространстве некоторый конус (конечно, не обязательно круговой). Действительно, начало координат удовлетворяет этому уравнению, 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 $W^T DW = 0$, а это и означает, что поверхность является конусом. В таком случае квадрику (13.3) можно рассматривать как сечение указанного конуса плоскостью $z = 1$.

B. Ортогональные преобразования уравнения квадрики. При преобразовании координат, описываемом формулами (**??**), уравнение квадрики $Z^T DZ = 0$ (см. (13.3)) 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

$$D' = P^T DP, \tag{13.4}$$

формально не отличающийся от (13.3), 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. Итак, при преобразовании координат матричные коэффициенты уравнения квадрики (13.2) преобразуются по формулам

$$A' = R^T AR, \quad B' = (X_0^T A + B)R, \quad c' = X_0^T AX_0 + 2BX_0 + c. \tag{13.5}$$

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.

C. Ортогональные инварианты уравнения квадрики.

13.1. 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 $\operatorname{tr}(UV) = \operatorname{tr}(VU)$, где U, V — произвольные матрицы, для которых определены оба произведения UV и VU (см. **??**), имеем
$$\operatorname{tr} A' = \operatorname{tr}\left(R^T AR\right) = \operatorname{tr}\left((R^T A)R\right) = \operatorname{tr}\left(R(R^T A)\right) = \operatorname{tr}\left((R^T R)A\right) = \operatorname{tr} A.$$
По теореме об определителе произведения матриц
$$\det A' = \det(R^T AR) = \det R^T \cdot \det A \cdot \det R = \det A,$$
поскольку $\det R = \det R^T = 1$. По той же теореме
$$\det D = \det(P^T DP) = \det P^T \cdot \det D \cdot \det P = \det D,$$
поскольку $\det P = \det P^T = 1$, в чём легко убедиться, разложив определитель матрицы P (см. (**??**)) по последней строке. $\square$

2. Канонизация уравнения квадрики при помощи ортогональных преобразований

Попытаемся упростить уравнение квадрики (13.1) (или (13.2), (13.3)), подбирая новую систему координат специальным образом.

A. Уничтожение слагаемого $2a_{12}xy$ при помощи поворота. Выберем угол поворота координатных 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:
$$A' = R^T AR = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$
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
$$A = (R^T)^{-1} A' R^{-1} = R(\lambda_1 \mathbb{1})R^{-1} = \lambda_1 RR^{-1} = \lambda_1 \mathbb{1},$$
т.е. матрица A также диагональна, т.е. $a_{12} = a_{21} = 0$, причём в любой системе координат.

Числа λ_1 и λ_2 могут быть найдены с помощью ортогональных инвариантов уравнения квадрики: поскольку
$$\operatorname{tr} A' = \lambda_1 + \lambda_2 = \operatorname{tr} A = S, \quad \det A' = \lambda_1 \cdot \lambda_2 = \det A = \delta,$$

по теореме Виета получаем, что λ_1 и λ_2 — корни квадратного уравнения

$$\lambda^2 - S\lambda + \delta = 0.$$

Умножая обе части соотношения $A' = R^T A R$ 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: $R = [R_1\ R_2]$. Соотношение $AR = RA'$ даёт

$$AR = RA' \iff A[R_1\ R_2] = [R_1\ R_2] \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = [\lambda_1 R_1\ \lambda_2 R_2],$$

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:

$$\det(A - \lambda\mathbb{1}) = \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_2 - \lambda \end{vmatrix} = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} =$$
$$= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} = \lambda^2 - S\lambda + \delta,$$

что в точности совпадает с найденным выше многочленом для определения λ_1 и λ_2.

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.

Коэффициенты полученного квадратичного многочлена, называемого характеристическим, являются ортогональными инвариантами уравнения квадрики, так что и сам он инвариантен, т.е. от выбора прямоугольной системы координат не зависит.

Дискриминант характеристического уравнения 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 оно всегда имеет вещественные корни.

Найдя корни λ_1 и λ_2 характеристического уравнения и решив однородные системы уравнений $(A - \lambda_1\mathbb{1})R_1 = 0$ и $(A - \lambda_2\mathbb{1})R_2 = 0$, 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 $\boldsymbol{i}, \boldsymbol{j}$.

Убедимся также, что получающиеся векторы $\boldsymbol{i}'$ и $\boldsymbol{j}'$ автоматически получаются ортогональными; для этого найдем их скалярное произведение $(\boldsymbol{i}', \boldsymbol{j}') = R_1^T R_2 = R_2^T R_1$. Умножая соотношения $AR_1 = \lambda_1 R_1$ и $AR_2 = \lambda_2 R_2$ слева на R_2^T и на R_1^T соответственно, получаем

$$R_2^T A R_1 = \lambda_1 R_2^T R_1, \quad R_1^T A R_2 = \lambda_2 R_1^T R_2.$$

Левые части полученных соотношений представляют собой матрицы порядка 1 (т.е. числа), причём они равны:

$$(R_2^T A R_1)^T = R_1^T A^T R_2 = R_1^T A R_2.$$

Поэтому равны и правые части:

$$\lambda_1 R_2^T R_1 = \lambda_2 R_1^T R_2 \iff (\lambda_1 - \lambda_2) R_1^T R_2 = 0.$$

Поскольку $\lambda_1 \neq \lambda_2$, заключаем отсюда, что $R_1^T R_2 = 0$, т.е. векторы $\boldsymbol{i}'$ и $\boldsymbol{j}'$ взаимно ортогональны.

Нетрудно найти выражение для угла поворота α:

$$AR_2 = \lambda_2 R_2 \iff \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix} = \begin{pmatrix} -\lambda_2 \sin\alpha \\ \lambda_2 \cos\alpha \end{pmatrix} \iff$$

$$\iff \begin{cases} -a_{11}\sin\alpha + a_{12}\cos\alpha = -\lambda_2 \sin\alpha, \\ -a_{21}\sin\alpha + a_{22}\cos\alpha = \quad \lambda_2\cos\alpha \end{cases}$$

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

$$\operatorname{ctg} 2\alpha = \frac{a_{11} - a_{22}}{2a_{12}}. \tag{13.6}$$

Итак, доказана следующая теорема.

13.2. Теорема. *Для уничтожения слагаемого $2a_{12}xy$ в уравнении квадрики* (13.1) *нужно перейти к новой системе координат, оси которой повёрнуты относительно осей исходной системы на угол α, определяемый соотношением* (13.6). *Другими словами, 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

$$\lambda^2 - S\lambda + \delta = 0 \iff \det(A - \lambda I) = 0. \tag{13.7}$$

B. Уничтожение линейных слагаемых $2b_1x+2b_2y$ при помощи переноса начала координат. Попытаемся теперь выбрать вектор переноса X_0 начала системы координат так, чтобы уравнение квадрики не содержало бы линейных слагаемых, т.е. $B' = O$ (O — нулевая матрица). Поскольку коэффициенты линейных членов уравнения квадрики преобразуются по формуле $B' = (X_0^T A + B)R$, для определения X_0 получаем уравнение

$$(X_0^T A + B)R = O \iff X_0^T A + B = O \iff (X_0^T A + B)^T = O$$

или, окончательно,

$$AX_0 = -B^T \tag{13.8}$$

(здесь использован факт симметричности матрицы A: $A^T = A$). Однако полученное уравнение не всегда разрешимо, так что уничтожение линейных слагаемых в уравнении квадрики возможно не во всех случаях.

Случай $\delta = \det A \neq 0$: центральные квадрики. Если $\det A \neq 0$, то уравнение $AX_0 = -B^T$ имеет единственное решение $X_0 = -A^{-1}B^T$, и матрица преобразования координат 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. 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

$$\lambda_1 x'^2 + \lambda_2 y'^2 + c' = 0, \tag{13.9}$$

где λ_1, λ_2 — корни характеристического уравнения (13.7), а c' — свободный член преобразованного уравнения, определяемый по формуле (13.5), в которой X_0 — решение уравнения (13.8). Матрица D' коэффициентов уравнения квадрики в полуканонической системе координат имеет вид

$$D' = \left(\begin{array}{cc|c} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ \hline 0 & 0 & c' \end{array}\right). \tag{13.10}$$

Коэффициент c' можно найти по формуле (13.5), однако если принять во внимание соотношение $AX_0 = -B^T$, вычисление можно упростить:

$$c' = X_0^T AX_0 + 2BX_0 + c = \underbrace{X_0^T(-B^T)}_{=-BX_0} + 2BX_0 + c = BX_0 + c. \tag{13.11}$$

Кроме того, поскольку определитель матрицы (13.10) равен

$$\det D' = \lambda_1\lambda_2 c' = \Delta,$$

легко получаем выражение для c' через ортогональные инварианты уравнения:

$$c' = \frac{\Delta}{\delta}. \tag{13.12}$$

Начало канонической системы координат $O'x'y'$ является центром симметрии линии, определяемой уравнением (13.9), а координатные оси $O'x'$ и $O'y'$ — её осями симметрии. Действительно, если точка с координатами (x', y') является решением уравнения (13.9), 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).

С помощью алгебраических преобразований полученное полуканоническое уравнение можно привести к каноническому виду.

В зависимости от знака $\det A$ выделяются следующие два подтипа.

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.

1. Пусть c' и $\lambda_{1,2}$ имеют разные знаки, т.е.

$$S \cdot \Delta < 0$$

(напомним, что $c' = \Delta/\delta$, а $\delta > 0$). Тогда из (13.9) 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 уравнение эллипса будет каноническим. Кроме того, полуоси эллипса можно выразить через ортогональные инварианты уравнения:

$$a = \sqrt{-\frac{\Delta}{\delta\lambda_1}}, \quad b = \sqrt{-\frac{\Delta}{\delta\lambda_2}}, \quad \lambda_1 < \lambda_2.$$

2. Пусть c' и $\lambda_{1,2}$ имеют один и тот же знак, т.е.

$$S \cdot \Delta > 0.$$

Тогда из (13.9) 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 и благодаря сходству с уравнением эллипса называется *уравнением мнимого эллипса*. «Полуоси» мнимого эллипса также можно выразить через ортогональные инварианты уравнения:

$$a = \sqrt{\frac{\Delta}{\delta\lambda_1}}, \quad b = \sqrt{\frac{\Delta}{\delta\lambda_2}}, \quad \lambda_1 < \lambda_2.$$

3. Наконец, в случае $c' = 0$, т.е. $\Delta = 0$, из (13.9) получаем уравнение
$$\lambda_1 x'^2 + \lambda_2 y'^2 = 0 \iff a^2 x'^2 + b^2 y'^2 = 0,$$
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:
$$(ax' + iby')(ax' - iby') = 0 \iff \left[\begin{aligned} &(ax' + iby') = 0, \\ &(ax' - iby') = 0, \end{aligned}\right.$$
что объясняет наименование рассматриваемой квадрики: пара мнимых прямых, пересекающихся в вещественной точке.

Гиперболический тип: $\delta = \det A < 0$**.** В этом случае числа λ_1 и λ_2 имеют разные знаки, и возможны следующие ситуации.

1. Если $c' \neq 0$ (т.е. $\Delta \neq 0$), то из полуканонического уравнения (13.9) 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 нумерация корней характеристического уравнения квадрики должна осуществляться 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.
2. Если $c' = 0$ (т.е. $\Delta = 0$), то полуканоническое уравнение (13.9) приводится к виду
$$a^2 x'^2 - b^2 y'^2 = 0, \quad \text{где} \quad a^2 = |\lambda_1|,\ b^2 = |\lambda_2|,$$
или, эквивалентно,
$$(ax' + by')(ax' - by') = 0 \iff \left[\begin{aligned} &(ax' + by') = 0, \\ &(ax' - by') = 0. \end{aligned}\right.$$
Последняя совокупность уравнений определяет пару пересекающихся прямых (очевидно, точкой пересечения является начало канонической системы координат).

Отметим, что полуканоническое уравнение получается из исходного 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 арифметических действий (например, деления обеих частей уравнения на свободный член уравнения и т. п.); относительно таких преобразований величины S, δ и Δ инвариантными быть не обязаны.

Случай $\delta = \det A = 0$: нецентральные квадрики. Условие $\det A = 0$ эквивалентно тому, что один из корней характеристического уравнения (13.7) 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 вид[1]

$$D' = \left(\begin{array}{cc|c} 0 & 0 & b_1' \\ 0 & S & b_2' \\ \hline b_1' & b_2' & c \end{array}\right), \quad \Delta = \det D' = -b_1'^2 S. \tag{13.13}$$

Если $\det A = 0$, то система уравнений (13.8) может либо оказаться несовместной, либо иметь бесконечно много решений.

Отметим следующий важный факт: любые переносы начала координат, выполняемые *после* поворота, приводящего матрицу коэффициентов квадрики к виду (13.13), не изменяют коэффициент b_1'. Действительно, такие переносы сохраняют вид матрицы (13.13), 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 сдвига начала координат нужно выбирать так, чтобы добиться, по возможности, исчезновения в матрице (13.13) элементов b_2' и c.

***Невырожденный параболический тип:* $\Delta \neq 0$.** Согласно теореме Кронекера—Капелли система (13.8) несовместна, если ранг матрицы A меньше ранга матрицы $\left[A \mid -B^T\right]$, или, что то же самое,

$$\operatorname{rk} A' < \operatorname{rk}\left[A' \mid -R^T B^T\right] \iff \operatorname{rk}\begin{pmatrix} 0 & 0 \\ 0 & S \end{pmatrix} < \operatorname{rk}\begin{pmatrix} 0 & 0 & -b_1' \\ 0 & S & -b_2' \end{pmatrix};$$

действительно, $\operatorname{rk} A = \operatorname{rk} A'$ и $\operatorname{rk}\left[A \mid -B^T\right] = \operatorname{rk}\left[A' \mid -R^T B^T\right]$, поскольку R — невырожденная матрица. 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:

$$S\left(y'^2 + 2\frac{b_2'}{S}y' + \frac{b_1'^2}{S^2}\right) - \frac{b_1'^2}{S} + 2b_1'\left(x' + \frac{c'}{2b_1'}\right) = 0,$$

так что

$$S\left(y' + \frac{b_2'}{S}\right)^2 + 2b_1'\left(x' + \frac{c'}{2b_1'} - \frac{b_1'^2}{2b_1'S}\right) = 0.$$

[1] Обратите внимание, что свободный член c при повороте не изменяется, см. формулы (13.5), с. 279.

Вводя новые переменные

$$\begin{cases} x'' = \pm\left(x' + \dfrac{c'}{2b'_1} - \dfrac{b'^2_1}{2b'_1 S}\right), \\ y'' = y' + \dfrac{b'_2}{S}, \end{cases} \tag{13.14}$$

получим полуканоническое уравнение

$$Sy''^2 + 2b'_1 x'', \tag{13.15}$$

а затем, выбирая надлежащим образом знак + или − в первой из формул (13.14), и каноническое уравнение параболы

$$y''^2 = 2px'', \tag{13.16}$$

где введено обозначение $p = \left|\dfrac{b'_1}{S}\right|$. Матрица коэффициентов полуканонического уравнения которого в системе координат $O''x''y''$ имеет вид

$$D'' = \left(\begin{array}{cc|c} 0 & 0 & b'_1 \\ 0 & S & 0 \\ \hline b'_1 & 0 & 0 \end{array}\right), \quad \det D'' = b'^2_1 S. \tag{13.17}$$

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 (13.14).

Из формул (13.14) следует, что координаты вершины параболы, т.е. начала канонической системы координат $O''x''y''$, отнесённые к повёрнутой системе $Ox'y'$, равны

$$\begin{cases} x'_0 = \mp\left(\dfrac{c'}{2b'_1} - \dfrac{b'^2_1}{2b'_1 S}\right), \\ y'_0 = -\dfrac{b'_2}{S}. \end{cases}$$

Для нахождения координат вершины в исходной системе Oxy нужно воспользоваться формулами (**??**):

$$X_0 = RX'_0.$$

Величина фокального параметра параболы в каноническом уравнении также может быть выражена через инварианты:

$$p = \left|\frac{b'_1}{S}\right| = \left|\frac{\sqrt{-\Delta/S}}{S}\right| = \sqrt{-\frac{\Delta}{S^3}}.$$

Вырожденный параболический тип: $\Delta = 0$**.** В случае, когда ранг матрицы A равен рангу матрицы $\left[A \mid -B^T\right]$, система (13.8) имеет бесконечно много решений. Выбрав любое из них, получаем следующую матрицу коэффициентов уравнения квадрики в преобразованной системе координат

$$D' = \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & S & 0 \\ \hline 0 & 0 & c' \end{array}\right), \tag{13.18}$$

т.е. само уравнение принимает полуканонический вид

$$Sy'^2 + c' = 0. \tag{13.19}$$

В зависимости от знаков чисел S и c' возможны следующие случаи.

1. Числа S и c' имеют разные знаки; тогда уравнение (13.19) 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 $y'' = -a$.
2. 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. геометрическая фигура, определяемая им, есть пустое множество. Если допускать комплексные решения, то по аналогии с предыдущим случаем можно назвать соответствующую квадрику парой мнимых параллельных прямых.
3. Наконец, если $c' = 0$, уравнение имеет вид
$$y'^2 = 0$$
и называется уравнением пары совпадающих прямых.

Для различения этих трёх квадрик, характеризуемых значениями инвариантов $\delta = \Delta = 0$, служит дополнительный инвариант K, теория которого изложена в следующем разделе.

Таким образом, доказана следующая теорема.

13.3. Теорема. *При помощи ортогональных преобразований координат* (*т.е. поворота осей координат и переноса начала координат*) *уравнение квадрики* (13.1) *может быть приведено к одному из девяти канонических типов, перечисленных в таблице на с.* 289.

3. Полуинвариант K

Выше было показано, 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.

Классификация квадрик

<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>

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. Докажем, что характеристический многочлен инвариантен относительно поворотов, т.е. ортогональных преобразований, описываемых матрицей

$$P = \left[\begin{array}{c|c} R & O \\ \hline O & 1 \end{array} \right].$$

Матрица P ортогональна, т.е. $P^{-1} = P^T$ (проверьте!), поэтому, принимая во внимание формулу (13.4) преобразования матрицы D коэффициентов уравнения квадрики, имеем

$$\begin{aligned} f_{D'}(\lambda) = \det(D' - \lambda \mathbb{1}) = \det\left(P^T D P - \lambda \mathbb{1}\right) = \\ = \det\left(P^{-1} D P - \lambda P^{-1} P\right) = \det\left(P^{-1}(D - \lambda \mathbb{1}) P\right) = \\ = \det P^{-1} \cdot \det(D - \lambda \mathbb{1}) \cdot \det P = \det(D - \lambda \mathbb{1}) = f_D(\lambda). \end{aligned}$$

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:

$$= \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ b_1 & b_2 & c - \lambda \end{vmatrix} - \begin{vmatrix} a_{11} & 0 & b_1 \\ a_{21} & \lambda & b_2 \\ b_1 & 0 & c - \lambda \end{vmatrix} - \lambda \begin{vmatrix} a_{22} - \lambda & b_2 \\ b_2 & c - \lambda \end{vmatrix} =$$

[1]Обратите внимание, что в этой формуле $\mathbb{1}$ в одном случае обозначает единичную матрицу порядка 3, а в другом — порядка 2.

в первом и третьем определителях снова применяем свойство линейности, второй раскладываем по второму столбцу:

$$= \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ b_1 & b_2 & c \end{vmatrix} - \begin{vmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ b_1 & b_2 & \lambda \end{vmatrix} - \lambda \begin{vmatrix} a_{11} & b_1 \\ b_1 & c-\lambda \end{vmatrix} -$$

$$- \lambda \left(\begin{vmatrix} a_{22} & b_2 \\ b_2 & c-\lambda \end{vmatrix} - \begin{vmatrix} \lambda & b_2 \\ 0 & c-\lambda \end{vmatrix} \right) =$$

$$= \Delta - \lambda \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} - \lambda \left(\begin{vmatrix} a_{11} & b_1 \\ b_1 & c \end{vmatrix} - \begin{vmatrix} a_{11} & 0 \\ b_1 & \lambda \end{vmatrix} \right) -$$

$$- \lambda \left(\begin{vmatrix} a_{22} & b_2 \\ b_2 & c \end{vmatrix} - \begin{vmatrix} a_{22} & 0 \\ b_2 & \lambda \end{vmatrix} - \begin{vmatrix} \lambda & b_2 \\ 0 & c \end{vmatrix} + \begin{vmatrix} \lambda & 0 \\ 0 & \lambda \end{vmatrix} \right) =$$

$$= \Delta - \left(\delta + \underbrace{\begin{vmatrix} a_{11} & b_1 \\ b_1 & c \end{vmatrix} + \begin{vmatrix} a_{22} & b_2 \\ b_2 & c \end{vmatrix}}_{=K} \right) \lambda + S\lambda^2 - \lambda^3.$$

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. Отметим, что имеется ещё одна величина, инвариантная относительно поворотов: свободный член c уравнения (см. (13.5)); c и K называются полуинвариантами (семиинвариантами).

Докажем, что K является инвариантом также и относительно сдвигов в случае $\delta = \Delta = 0$.

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
$$S(y'+y_0)+2b_2(y'+y_0)+c = 0 \iff Sy'^2+2(Sy_0+b_2)+(Sy_0^2+2b_2y_0+c) = 0.$$
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':
$$K' = \begin{vmatrix} S & Sy_0 + b_2 \\ Sy_0 + b_2 & Sy_0^2 + 2b_2 y_0 + c \end{vmatrix} = S\bigl(Sy_0^2+2b_2y_0+c\bigr)-\bigl(Sy_0+b_2\bigr)^2 = Sc-b_2^2 = K,$$
так что K — инвариант относительно сдвигов в случае $\delta = \Delta = 0$.

Завершим анализ вырожденного параболического случая. Уравнение квадрики допускает уничтожение слагаемого $b_2 y$ с помощью переноса, т.е. приводится к полуканоническому виду
$$Sy'^2 + c' = 0, \quad D' = \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & S & 0 \\ \hline 0 & 0 & c' \end{array}\right), \quad K = Sc',$$
откуда $c' = K/S$, а уравнение квадрики имеет вид
$$y'^2 + \frac{K}{S^2} = 0 \iff \begin{cases} y'^2 = a^2, & \text{где } a^2 = -\dfrac{K}{S^2}, \text{ если } K < 0, \\ y'^2 = -a^2, & \text{где } a^2 = \dfrac{K}{S^2}, \text{ если } K > 0, \\ y'^2 = 0, & \text{если } K = 0. \end{cases}$$

ГЛАВА 14

Квадрики в пространстве

1. Классификация квадрик. Канонические уравнения

Квадрика в трёхмерном пространстве, т.е. множество точек, координаты которых удовлетворяют уравнению

$$X^T AX + 2BX + c = 0, \tag{14.1}$$

где

$$X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad B = \begin{pmatrix} b_1, & b_2, & b_3 \end{pmatrix},$$

имеет в качестве графика некоторую поверхность.

Аналогично[1] тому, как это было сделано для линий второго порядка (см. гл. 14), каждое уравнение вида (14.1) можно привести к одному из семнадцати канонических типов, перечисленных ниже (некоторые из них изображены на рис. 14.1).

I. **Эллиптический тип.**

I.1. Эллипсоид (см. рис. 14.1(a)):

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1, \quad a > 0,\ b > 0,\ c > 0.$$

I.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.

I.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 состоит из единственной вещественной точки $O(0, 0, 0)$.

II. **Гиперболический тип.**

[1] В процессе канонизации уравнения используются некоторые факты линейной алгебры, которые будут изложены в следующем семестре.

II.1. Однополостный гиперболоид (см. рис. 14.1(b)):

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1, \quad a > 0,\ b > 0,\ c > 0.$$

II.2. Двуполостный гиперболоид (см. рис. 14.1(c)):

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1, \quad a > 0,\ b > 0,\ c > 0.$$

II.3. Конус (см. рис. 14.1(d)):

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0, \quad a > 0,\ b > 0,\ c > 0.$$

III. **Параболический тип.**

III.1. Эллиптический параболоид (см. рис. 14.1(e)):

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 2z, \quad a > 0,\ b > 0.$$

III.2. Гиперболический параболоид (см. рис. 14.1(f)):

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 2z, \quad a > 0,\ b > 0.$$

III.3. Эллиптический цилиндр (см. рис. 14.1(g)):

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad a > 0,\ b > 0.$$

Направляющей цилиндра является эллипс, образующие параллельны оси Oz.

III.4. 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.

III.5. 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).

III.6. Гиперболический цилиндр: (см. рис. 14.1(h)):

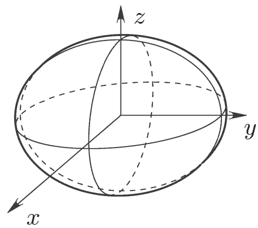
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \quad a > 0,\ b > 0.$$

Направляющей является гипербола, образующие параллельны оси Oz.

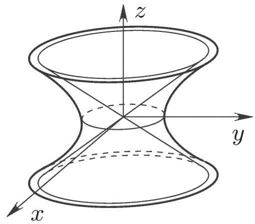
III.7. Пара пересекающихся плоскостей:

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

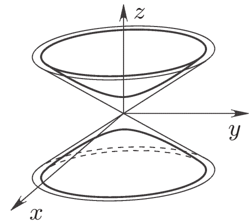
Линией пересечения плоскостей является ось Oz.



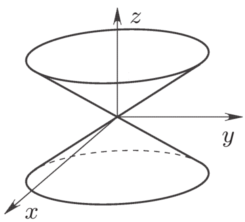
(a) Эллипсоид



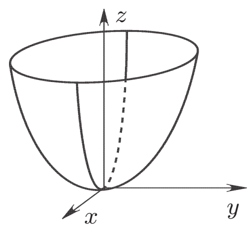
(b) Однополостный гиперболоид



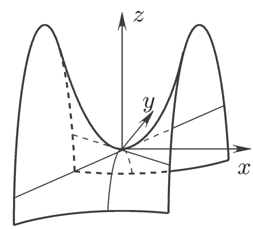
(c) Двуполостный гиперболоид



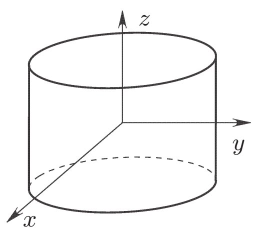
(d) Конус



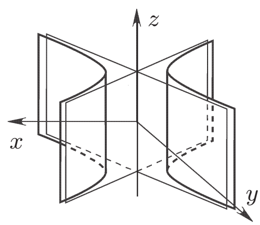
(e) Эллиптический параболоид



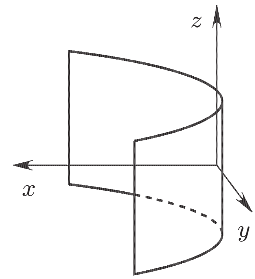
(f) Гиперболический параболоид



(g) Эллиптический цилиндр



(h) Гиперболический цилиндр



(i) Параболический цилиндр

Рис. 14.1. Основные типы поверхностей второго порядка

III.8. Параболический цилиндр: (см. рис. 14.1(i)):

$$y^2 = 2px,$$

где $p > 0$. Направляющей является парабола, образующие параллельны оси Oz.

III.9. Пара параллельных плоскостей

$$y^2 = a^2, \quad a > 0.$$

Плоскости параллельны плоскости Oxz.

III.10. Пара мнимых параллельных плоскостей:

$$y^2 = -a^2, \quad a > 0.$$

Это уравнение не имеет ни одного вещественного решения.

III.11. 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. 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 $2c$.

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 $a\sqrt{1 - \dfrac{h^2}{c^2}}$, $b\sqrt{1 - \dfrac{h^2}{c^2}}$, 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.

B. Двуполостный гиперболоид. Уравнение двуполостного гиперболоида имеет вид

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1 \iff -\frac{x^2}{a^2}-\frac{y^2}{b^2}+\frac{z^2}{c^2}=1,$$

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 $z\leqslant -c$.

Плоскость $z=h$ при $|h|<c$ не пересекает гиперболоид, при $|h|=c$ имеет единственную общую точку с гиперболоидом ($(0,0,c)$ при $h=c$ и $(0,0,-c)$ при $h=-c$) и при $|h|>c$ пересекает гиперболоид по эллипсу

$$\frac{x^2}{\left(a\sqrt{\dfrac{h^2}{c^2}-1}\right)^2}+\frac{y^2}{\left(b\sqrt{\dfrac{h^2}{c^2}-1}\right)^2}=1,$$

полуоси которого монотонно возрастают от 0 до $+\infty$, когда $|h|$ возрастает от c до $+\infty$.

Каждая плоскость $y=h$ пересекает гиперболоид по гиперболе

$$\frac{z^2}{\left(c\sqrt{1+\dfrac{h^2}{b^2}}\right)^2}-\frac{x^2}{\left(a\sqrt{1+\dfrac{h^2}{b^2}}\right)^2}=1,$$

полуоси которой монотонно возрастают (от c и a соответственно) до $+\infty$, когда $|h|$ возрастает от 0 до $+\infty$. Аналогично для сечений плоскостями $x=h$.

C. Однополостный гиперболоид. 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.

Каждая плоскость $z=h$ пересекает гиперболоид по эллипсу

$$\frac{x^2}{\left(a\sqrt{1+\dfrac{h^2}{c^2}}\right)^2}+\frac{y^2}{\left(b\sqrt{1+\dfrac{h^2}{c^2}}\right)^2}=1,$$

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.

Плоскость $y = h$ при $|h| < b$ пересекает гиперболоид по гиперболе

$$\frac{x^2}{\left(a\sqrt{1-\dfrac{h^2}{b^2}}\right)^2} - \frac{z^2}{\left(c\sqrt{1-\dfrac{h^2}{b^2}}\right)^2} = 1,$$

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

$$\frac{z^2}{\left(c\sqrt{\dfrac{h^2}{b^2}-1}\right)^2} - \frac{x^2}{\left(c\sqrt{\dfrac{h^2}{b^2}-1}\right)^2} = 1,$$

полуоси которой возрастают от 0 до $+\infty$, когда $|h|$ возрастает от b до $+\infty$. Аналогично для сечений плоскостями $x = h$.

D. Конус. 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.

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.

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 $O(0, 0, 0)$.

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.

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.

E. Эллиптический параболоид. 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.

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 $+\infty$.

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.

F. Гиперболический параболоид. 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.

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
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0.$$
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

Поверхность называется *l-кратно линейчатой* поверхностью, если через каждую её точку проходит ровно l различных прямых, называемых *прямолинейными образующими*.

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.

14.1. Предложение. *Однополостный гиперболоид является дважды линейчатой поверхностью.*

Доказательство. Пусть (x_0, y_0, z_0) — точка, лежащая на однополостном гиперболоиде

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1. \tag{14.2}$$

Рассмотрим прямую, проходящую через эту точку:

$$x = x_0 + lt, \quad y = y_0 + mt, \quad z = z_0 + nt. \tag{14.3}$$

Для краткости введём обозначения[1]

$$\begin{aligned} X &= \frac{x}{a}, & X_0 &= \frac{x_0}{a}, & L &= \frac{l}{a}, \\ Y &= \frac{y}{b}, & Y_0 &= \frac{y_0}{b}, & M &= \frac{m}{b}, \\ Z &= \frac{z}{c}, & Z_0 &= \frac{z_0}{c}, & N &= \frac{n}{c}. \end{aligned}$$

Уравнение гиперболоида (14.2) в новых обозначениях

$$X^2 + Y^2 - Z^2 = 1 \iff X^2 + Y^2 = 1 + Z^2, \tag{14.4}$$

а уравнение прямой (14.3) —

$$X = X_0 + Lt, \quad Y = Y_0 + Mt, \quad Z = Z_0 + Nt. \tag{14.5}$$

Подставляя (14.5) в (14.4), получим

$$\begin{gathered} (X_0 + Lt)^2 + (Y_0 + Mt)^2 - (Z_0 + Nt)^2 = 1 \iff \\ \iff \underbrace{(X_0^2 + Y_0^2 - Z_0^2)}_{=1} + 2t(X_0 L + Y_0 M - Z_0 N) + t^2(L^2 + M^2 - N^2) = 1 \\ \iff 2t(X_0 L + Y_0 M - Z_0 N) + t^2(L^2 + M^2 - N^2) = 0. \end{gathered}$$

[1] Это можно рассматривать как аффинное преобразование — растяжения/сжатия вдоль координатных осей.

Это уравнение выполняется тождественно (т. прямая (14.5) целиком лежит на гиперболоиде (14.4)) 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

$$X_0\cos\varphi + Y_0\sin\varphi = Z_0 \iff$$

$$\iff \sqrt{X_0^2+Y_0^2}\left(\underbrace{\frac{X_0}{\sqrt{X_0^2+Y_0^2}}}_{=\cos\theta}\cos\varphi + \underbrace{\frac{Y_0}{\sqrt{X_0^2+Y_0^2}}}_{=\sin\theta}\sin\varphi\right) = Z_0 \iff$$

$$\iff \cos(\varphi-\theta) = \frac{Z_0}{\sqrt{X_0^2+Y_0^2}} = \frac{Z_0}{\sqrt{1+Z_0^2}};$$

последняя дробь строго меньше единицы по модулю, поэтому тригонометрическое уравнение имеет ровно 2 решения на промежутке $[0, 2\pi)$:

$$\varphi - \theta = \pm\arccos\frac{Z_0}{\sqrt{1+Z_0^2}} \iff \varphi = \theta \pm \arccos\frac{Z_0}{\sqrt{1+Z_0^2}}.$$

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

$$L_\varepsilon = \cos\varphi = \frac{X_0Z_0 - \varepsilon Y_0}{1+Z_0^2}, \quad M_\varepsilon = \sin\varphi = \frac{Y_0Z_0 + \varepsilon X_0}{1+Z_0^2}, \quad \varepsilon = \pm 1.$$

Каждое из возможных значений ε определяет направляющий вектор прямолинейной образующей.

Таким образом, через точку (X_0, Y_0, Z_0) гиперболоида (14.4) проходят ровно две прямые, целиком лежащие на гиперболоиде:

$$X = X_0 + \frac{X_0 Z_0 - \varepsilon Y_0}{1 + Z_0^2} t, \quad Y = Y_0 + \frac{Y_0 Z_0 + \varepsilon X_0}{1 + Z_0^2} t, \quad Z = Z_0 + t, \qquad \varepsilon = \pm 1.$$

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 $t_0 = -Z_0$:

$$\begin{aligned} X_\varepsilon &= X_0 - \frac{X_0 Z_0 - \varepsilon Y_0}{1 + Z_0^2} Z_0 = \frac{X_0 + \varepsilon Y_0 Z_0}{1 + Z_0^2} = \varepsilon \frac{Y_0 Z_0 + \varepsilon X_0}{1 + Z_0^2} = \varepsilon M_\varepsilon, \\ Y_\varepsilon &= Y_0 - \frac{Y_0 Z_0 + \varepsilon X_0}{1 + Z_0^2} Z_0 = \frac{Y_0 - \varepsilon X_0 Z_0}{1 + Z_0^2} = -\varepsilon \frac{X_0 Z_0 - \varepsilon Y_0}{1 + Z_0^2} = -\varepsilon L_\varepsilon, \\ Z_\varepsilon &= 0, \qquad \varepsilon = \pm 1. \end{aligned}$$

Ясно, что эти две точки лежат на горловом эллипсе гиперболоида. Теперь можно записать уравнения прямолинейных образующих в виде

$$\begin{cases} X = X_\varepsilon - \varepsilon Y_\varepsilon t, \\ Y = Y_\varepsilon + \varepsilon X_\varepsilon t, \\ Z = t, \end{cases} \tag{14.6}$$

где $(X_\varepsilon, Y_\varepsilon)$ — точка горлового эллипса, $\varepsilon = \pm 1$.

Обратно, пусть (X_*, Y_*) — точка горлового эллипса однополостного гиперболоида (14.4); 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

$$\begin{aligned} X^2 + Y^2 - Z^2 &= (X_* - \varepsilon Y_* t)^2 + (Y_* + \varepsilon X_* t)^2 - t^2 = \\ &= \underbrace{(X_*^2 + Y_*^2)}_{=1} + 2t(X_\varepsilon Y_\varepsilon - X_\varepsilon Y_\varepsilon) + t^2 \underbrace{(X_*^2 + Y_*^2 - 1)}_{=0} = 1, \end{aligned}$$

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; к одному семейству относятся образующие, отвечающие $\varepsilon = 1$, к другому — отвечающие $\varepsilon = -1$. $\square$

Рассмотрим две прямолинейные образующие, проходящие через точки (X_1, Y_1) и (X_2, Y_2) горлового эллипса:

$$\begin{cases} X = X_1 - Y_1 t, \\ Y = Y_1 + X_1 t, \\ Z = t, \end{cases} \qquad \begin{cases} X = X_2 - \varepsilon Y_2 t, \\ Y = Y_2 + \varepsilon X_2 t, \\ Z = t; \end{cases} \tag{14.7}$$

при $\varepsilon = 1$ эти образующие принадлежат одному семейству, а при $\varepsilon = -1$ — разным.

Выясним вопрос о взаимном расположении этих прямых, используя метод, изложенный на с. 210. 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 найдём её определитель:

$$\begin{array}{ccc} \cdot(-1) & + & \\ \left|\begin{matrix} -Y_1 & -\varepsilon Y_2 & X_2 - X_1 \\ X_1 & \varepsilon X_2 & Y_2 - Y_1 \\ 1 & 1 & 0 \end{matrix}\right| & \end{array} = \begin{vmatrix} -Y_1 & Y_1 - \varepsilon Y_2 & X_2 - X_1 \\ X_1 & \varepsilon X_2 - X_1 & Y_2 - Y_1 \\ 1 & 0 & 0 \end{vmatrix} =$$

$$= \begin{vmatrix} Y_1 - \varepsilon Y_2 & X_2 - X_1 \\ \varepsilon X_2 - X_1 & Y_2 - Y_1 \end{vmatrix} = (Y_1 - \varepsilon Y_2)(Y_2 - Y_1) - (\varepsilon X_2 - X_1)(X_2 - X_1).$$

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 результатом, полученным на с. 210, 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.

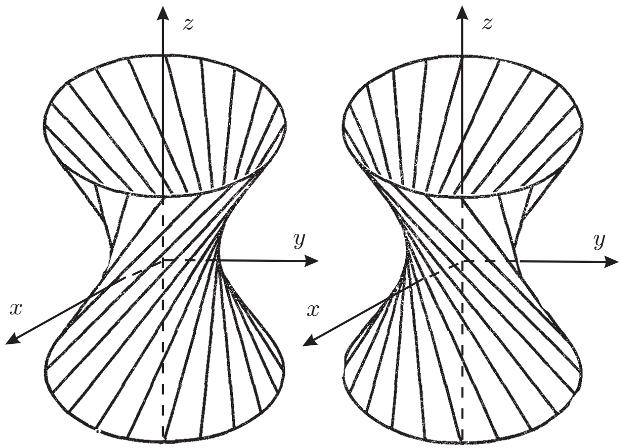


Рис. 14.2. Прямолинейные образующие однополостного гиперболоида

Рассмотрим три попарно различные одноименные прямолинейные образующие, проходящие через точки (X_1, Y_1), (X_2, Y_2), (X_3, Y_3) горлового эллипса. Рассмотрим 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, рассматриваемые три прямолинейные образующие не компланарны.

Доказана следующая теорема.

14.2. 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*
   (a) *параллельны, если проходят через диаметрально противоположные точки горлового эллипса,*
   (b) *и пересекаются в противном случае.*

При решении задач удобнее пользоваться другим методом нахождения прямолинейных образующих. Запишем уравнение однополостного гиперболоида в виде

$$\frac{x^2}{a^2}-\frac{z^2}{c^2}=1-\frac{y^2}{b^2} \iff \left(\frac{x}{a}-\frac{z}{c}\right)\left(\frac{x}{a}+\frac{z}{c}\right)=\left(1-\frac{y}{b}\right)\left(1+\frac{y}{b}\right).$$

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.

B. Гиперболический параболоид.

14.3. Предложение. *Гиперболический параболоид является дважды линейчатой поверхностью.*

Доказательство. Пусть точка (x_0, y_0, z_0) лежит на гиперболическом параболоиде

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=2z. \tag{14.8}$$

Рассмотрим прямую, проходящую через эту точку:

$$x=x_0+lt, \quad y=y_0+mt, \quad z=z_0+nt. \tag{14.9}$$

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 (14.10) и прямой (14.9) 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

$$(X_0 + Lt)^2 - (Y_0 + Mt)^2 = 2(Z_0 + Nt) \iff$$
$$\iff \underbrace{(X_0^2 - Y_0^2)}_{=2Z_0} + 2t(X_0L - Y_0M) + t^2(L^2 - M^2) = 2Z_0 + 2Nt \iff$$
$$\iff 2t(X_0L - Y_0M) + t^2(L^2 - M^2) = 2Nt.$$

Это уравнение выполняется тождественно, т.е. прямая целиком лежит на параболоиде, тогда и только тогда, когда

$$X_0L - Y_0M = N, \quad L^2 - M^2 = 0.$$

Поскольку направляющий вектор прямой определен лишь с точностью до ненулевого множителя, положим $L = 1$; тогда $M = \varepsilon$, где $\varepsilon = \pm 1$, $N = X_0 - \varepsilon Y_0$. Итак, через точку (X_0, Y_0, Z_0) параболоида проходят ровно две прямых

$$\begin{cases} X = X_0 + t, \\ Y = Y_0 + \varepsilon t, \\ Z = Z_0 + (X_0 - \varepsilon Y_0)t, \end{cases} \qquad \varepsilon = \pm 1.$$

□

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:

$$\begin{cases} X = X_1 + t, \\ Y = Y_1 + t, \\ Z = Z_1 + (X_1 - Y_1)t, \end{cases} \qquad \begin{cases} X = X_2 + t, \\ Y = Y_2 + t, \\ Z = Z_2 + (X_2 - Y_2)t, \end{cases}$$

---

[1] Как и в случае однополостного гиперболоида, это можно рассматривать как аффинное преобразование — растяжения/сжатия вдоль координатных осей.

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

$$r = \begin{cases} 1, & \text{если } X_1 - Y_1 = X_2 - Y_2 \iff X_1 - X_2 - Y_1 + Y_2 = 0, \\ 2 & \text{в противном случае.} \end{cases}$$

Для вычисления ранга второй матрицы найдём её определитель:

$$\begin{vmatrix} 1 & 1 & X_2 - X_1 \\ 1 & 1 & Y_2 - Y_1 \\ X_1 - Y_1 & X_2 - Y_2 & Z_2 - Z_1 \end{vmatrix} =$$

$$= \begin{vmatrix} 1 & 1 & X_2 - X_1 \\ 0 & 0 & Y_2 - Y_1 - X_2 + X_1 \\ X_1 - Y_1 & X_2 - Y_2 & Z_2 - Z_1 \end{vmatrix} =$$

$$= \begin{vmatrix} 1 & 0 & X_2 - X_1 \\ 0 & 0 & Y_2 - Y_1 - X_2 + X_1 \\ x_1 - Y_1 & X_2 - Y_2 - X_1 + Y_1 & Z_2 - Z_1 \end{vmatrix} = (Y_2 - Y_1 - X_2 + X_1)^2.$$

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:

$$\begin{vmatrix} 1 & X_2 - X_1 \\ X_1 - Y_1 & \frac{1}{2}\underbrace{(X_2 - Y_2)}_{X_1 - Y_1}(X_2 + Y_2) - \frac{1}{2}(X_1 - Y_1)(X_1 + Y_1) \end{vmatrix} =$$

$$= (X_1 - Y_1) \begin{vmatrix} 1 & X_2 - X_1 \\ 1 & \frac{1}{2}(X_2 + Y_2 - X_1 - Y_1) \end{vmatrix} =$$

$$= (X_1 - Y_1) \begin{vmatrix} 1 & X_2 - X_1 \\ 0 & \frac{1}{2}\underbrace{(-X_2 + Y_2 + X_1 - Y_1)}_{=0} \end{vmatrix} = 0.$$

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, рассматриваемые 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 $X - \varepsilon Y = 0$.

Доказана следующая теорема.

14.4. 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.*

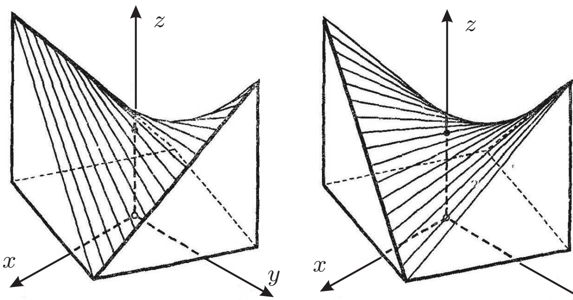


Рис. 14.3. Прямолинейные образующие гиперболического параболоида

Для практического нахождения уравнений прямолинейных образующих гиперболического параболоида используется следующий приём. Запишем уравнение параболоида (14.10) в виде

$$\left(\frac{x}{a}-\frac{y}{b}\right)\left(\frac{x}{a}+\frac{y}{b}\right)=2z. \tag{14.10}$$

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 (14.10), т.е. прямая целиком лежит на параболоиде.

ДОБАВЛЕНИЕ I

Язык логики в математике

1. Простейшие понятия математической логики

А. Логика высказываний. *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 является высказыванием (как и вообще любое определение), но фраза «если все стороны треугольника равны, то он *является* равносторонним» — высказывание, являющееся истинным в силу только что приведённого определения. 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, ложно, false) или истина (1, истинно, true). Для обозначения высказываний будем использовать прописные латинские буквы.

Из высказываний можно образовывать новые высказывания при помощи основных логических операций, к которым относятся:

(1) отрицание $\neg$,
(2) конъюнкция $\wedge$ (логический союз «*и*»),
(3) дизъюнкция $\vee$ (логический союз «*или*»),
(4) импликация $\Rightarrow$ (логический союз «если..., то...»; запись «$A \Rightarrow B$» означает «если A, то B» или «из A следует B»),
(5) эквивалентность $\Leftrightarrow$ (запись «$A \Leftrightarrow B$» означает «A и B эквивалентны», т.е. из A следует B, а из B следует A; в символической записи $(A \Rightarrow B) \wedge (B \Rightarrow A)$).

Логические операции удобно определять с помощью таблиц истинности.

I.1. 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.2. Определение. *Конъюнкция* — логическая операция, заключающаяся 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.3. Определение. *Дизъюнкция* — логическая операция, заключающаяся 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.4. Определение. *Импликация* — логическая операция, заключающаяся 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.*

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 повседневном языке 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...».

Приведем пример использования импликации с ложной посылкой в математике, приводящий к сокращению речи. Известен прием решения алгебраических иррациональных уравнений, 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 для 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.5. Определение. *Эквивалентность* — логическая операция, заключающаяся в соединении двух высказываний A и B в новое высказывание $A \iff B$ («A тогда и только тогда, когда B», «A эквивалентно B») по следующему правилу:

A	B	$A \iff B$
0	0	1
0	1	0
1	0	0
1	1	1

(I.7)

Из таблицы истинности видно, что высказывание $A \iff B$ истинно в том и только том случае, когда высказывания A и B оба истинны или оба ложны.

I.6. Теорема. *Справедливы следующие соотношения*:

(1) *коммутативность*:

$$A \vee B = B \vee A,$$
$$A \wedge B = B \wedge A;$$

(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) *законы поглощения*:

$$A \vee (A \wedge B) = A,$$
$$A \wedge (A \vee B) = A;$$

(5) *идемпотентность*:

$$A \vee A = A, \quad A \wedge A = A;$$

(6) *свойства констант*:

$$A \vee 1 = 1, \qquad A \vee 0 = A,$$
$$A \wedge 1 = A, \qquad A \wedge 0 = 0,$$

(7) *закон исключённого третьего*:

$$A \vee \neg A = 1, \quad A \wedge \neg A = 0;$$

(8) *закон двойного отрицания*:

$$\neg\neg A = A;$$

(9) *законы де Моргана*:

$$\neg(A \wedge B) = \neg A \vee \neg B;$$
$$\neg(A \vee B) = \neg A \wedge \neg B;$$

(10) *закон контрапозиции*:

$$(A \Rightarrow B) = (\neg B \Rightarrow \neg A);$$

(11) *соотношения*

$$(A \Rightarrow 1) = 1, \qquad (0 \Rightarrow A) = 1,$$
$$(1 \Rightarrow A) = A, \qquad (A \Rightarrow 0) = \neg A;$$

(12) *соотношения*

$$\neg(A \Rightarrow B) = (A \wedge \neg B); \quad (A \vee B) = (\neg A \Rightarrow B).$$

Доказательство. Каждое из приведенных соотношений может быть легко доказано при помощи таблиц истинности. Докажем, например, закон контрапозиции:

A	B	$A \Rightarrow B$	$\neg A$	$\neg B$	$\neg B \Rightarrow \neg A$
1	1	1	0	0	1
1	0	0	0	1	0
0	1	1	1	0	1
0	0	1	1	1	1

(здесь использованы таблицы истинности (I.1) и (I.4)). Сравнивая истинностные значения высказываний $A \Rightarrow B$ (третий столбец таблицы) и $\neg B \Rightarrow \neg A$ (последний столбец), видим, что они совпадают при любых истинностных значениях высказываний A и B. □

B. Логика предикатов. 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 какая-либо часть рассматриваемого предиката не определена, предикат считается ложным при этих значениях переменных.

С помощью логических операций $\neg$, $\wedge$, $\vee$, $\Rightarrow$ из данных предикатов можно строить более сложные предикаты.

Кванторы — это логические операции, которые по предикату $P(x)$ строят высказывание, дающее количественную характеристику области истинности предиката $P(x)$. Наиболее употребительны квантор общности $\forall$ и квантор существования $\exists$.

I.7. Определение. Пусть $P(x)$ — 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 x, истинно»[1] 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) — *навешиванием* на предикат P *квантора общности* по переменной x.

I.8. Определение. Пусть $P(x)$ — 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 x, 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) — *навешиванием* на предикат P *квантора существования* по переменной x.

В выражениях (I.8), (I.9) буква x является связанной переменной; от x эти выражения не зависят. Таким образом, навешивание квантора связывает переменную.

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 $\forall x$.

Пример I.1. $(\forall x)(x^2 \geqslant 0)$ и $(\exists x)(|x| \leqslant 0)$ — истинные высказывания; $(\forall x)(x^2 > 0)$ и $(\exists x)(x^2 + 1 = 0)$ — ложные высказывания.

I.9. Теорема. *Справедливы следующие аналоги законов де Моргана:*

$$\neg(\forall x \in X)P(x) \quad = \quad (\exists x \in X)\neg P(x), \tag{I.10}$$

$$\neg(\exists x \in X)P(x) \quad = \quad (\forall x \in X)\neg P(x), \tag{I.11}$$

Доказательство. Докажем соотношение (I.10); для этого требуется установить истинность двух импликаций «левая часть $\Rightarrow$ правая часть» и «правая часть $\Rightarrow$ левая часть».

[1] Разумеется, само высказывание, обозначенное выражением (I.8), может быть истинным или ложным.

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. $\square$

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.*

Если $P(x, y)$ — двухместный предикат, то его можно превратить в высказывание, навешивая 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

$$\begin{array}{rcl}
(\forall x)(\forall y)P(x,y) & = & (\forall y)(\forall x)P(x,y),\\
(\exists x)(\exists y)P(x,y) & = & (\exists y)(\exists x)P(x,y);
\end{array}$$

иными словами, одноименные кванторы можно переставлять. «Разноименные» кванторы можно переставлять только в одну сторону: утверждение

$$(\exists x)(\forall y)P(x,y) \implies (\forall y)(\exists x)P(x,y)$$

истинно, а утверждение

$$(\forall y)(\exists x)P(x,y) \implies (\exists x)(\forall y)P(x,y)$$

ложно. 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.

Таким образом, высказывания

$$(\exists x)(\forall y)P(x,y) \quad \text{и} \quad (\forall y)(\exists x)P(x,y)$$

не эквивалентны.

При написании высказывания, имеющего вид предиката, на который навешены кванторы 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).

2. Теоремы и доказательства

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 называют только достаточно важные утверждения; менее важные и технически несложные утверждения-теоремы обычно именуют леммами, предложениями, следствиями и прочими подобными терминами.

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.10. 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.*

Пример I.2. 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) четных чисел.

I.11. 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.*

Пример I.3. Для того чтобы число делилось на 2, достаточно, чтобы его последняя цифра была 0:

$$\underbrace{\text{последняя цифра } 0}_{\text{достаточное условие делимости на 2}} \quad\Longrightarrow\quad \text{число делится на 2.}$$

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* — 0*, то число делится на* 2» является признаком (одним из признаков) делимости на 2.

Достаточные условия стараются сделать возможно более широкими, т.е. охватывающими возможно большее число случаев, в которых интересующий нас факт все еще 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*: $A \iff B$. Критерии часто формулируются с помощью речевого оборота «*тогда и только тогда*».

Пример I.4. Примером критерия может служить следующее утверждение: «*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.12. Определение. Рассмотрим следующие четыре 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.

Как показывает следующая таблица истинности, взаимно обратные теоремы (I.13) и (I.14) 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	$A \Rightarrow B$	$B \Rightarrow A$
и	и	и	и
и	л	л	и
л	и	и	л
л	л	и	и

Взаимно противоположные теоремы (I.13) и (I.15) связаны аналогично.

Если для некоторой теоремы $A \Rightarrow B$ верна и обратная теорема $B \Rightarrow A$, то они могут быть объединены в критерий $A \iff B$.

Пример I.5. 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* $a^2 + b^2 = c^2$».

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».

ДОБАВЛЕНИЕ II

Множества, отношения, отображения

1. Множества и операции над ними

Множество — один из ключевых объектов математики, появляющийся во всех её областях. Никакого определения понятию множества не даётся; мы считаем это понятие первичным.

Общепризнано, что создателем теории множеств является Г. Кантор. 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 определенных недостатков; в частности, в 1903 г. Б. Расселл обнаружил следующий парадокс, демонстрирующий противоречивость наивной теории множеств.

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.

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, $\ldots$, X, Y, $\ldots$, элементы — строчными буквами a, b, $\ldots$, x, y, $\ldots$. Если элемент a принадлежит множеству A, пишем $a \in A$ $(A \ni a)$, а если не принадлежит — то $a \notin A$ $(A \not\ni A)$.

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, $[-2, 2] = \{x \mid x^2 \leqslant 4\}$.

В геометрии используется специфический термин — *геометрическое место точек*, означающий геометрическую фигуру как множество точек, обладающих некоторым свойством.

Для удобства вводится понятие множества, не содержащего никаких элементов; оно называется *пустым множеством* и обозначается $\varnothing$. Целесообразность введения понятия пустого множества обусловлена тем, что, задавая множество описанием (характеристическим свойством) его элементов, мы можем и не знать заранее, существует хотя бы один элемент, обладающий упомянутым свойством; например, мы не могли бы говорить о множестве корней произвольного уравнения, не убедившись предварительно, что данное уравнение имеет хотя бы один корень.

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 двухэлементным (или содержать другое количество элементов), 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 $b \in \{a\} \iff b = a$.

II.1. 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.

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).

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 $\{a, b\}$.

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.2. Определение. Множество B называется *подмножеством* множества A (запись $B \subseteq A$ или $A \supseteq B$), если каждый элемент B является также элементом A:

$$(B \subseteq A) \stackrel{\text{def}}{=\!=} (\forall x \in B \Rightarrow x \in A).$$

Отношение между множествами, выражаемое словами «B является подмножеством A», называется *отношением включения*. Очевидно, $A \subseteq A$.

Следует чётко различать знаки $\in$ и $\subseteq$ 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.

Принято считать, по определению, что $\varnothing \subseteq A$ 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; факт, что B является собственным подмножеством множества A, обозначается обычно $B \subset A$. Впрочем, символы $\subset$, $\supset$ часто используют как синонимы для $\subseteq$, $\supseteq$. Если требуется специально подчеркнуть, что $B \neq A$, используют запись $B \subsetneqq A$. Запись $B \not\subset A$ означает, что B не является подмножеством множества A.

II.3. 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:

$$A \cap B \stackrel{\text{def}}{=\!=} \{x \mid x \in A \wedge x \in B\}.$$

Пустое множество является подмножеством самого себя, но при этом само с собой не пересекается: $\varnothing \subseteq \varnothing$, $\varnothing \cap \varnothing = \varnothing$.

II.4. Определение. *Объединением* $A \cup B$ множеств A и B называется множество, состоящее из всех элементов, которые принадлежат либо множеству A, либо множеству B, либо обоим этим множествам:

$$A \cup B \stackrel{\text{def}}{=\!=} \{x \mid x \in A \vee x \in B\}.$$

II.5. Определение. *Разностью* $A \setminus B$ множеств A и B называется множество, состоящее из всех элементов, принадлежащих множеству A, но не принадлежащих множеству B:

$$A \setminus B \stackrel{\text{def}}{=\!=} \{x \mid x \in A \wedge x \notin B\}.$$

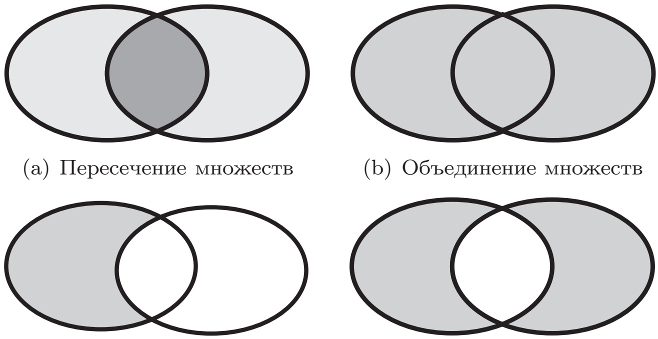
II.6. Определение. В некоторых разделах математики используется *симметрическая разность* $A \triangle B$ множеств A и B:

$$A \triangle B \stackrel{\text{def}}{=\!=} (A \cup B) \setminus (A \cap B).$$

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, разности множеств и их основные свойства.

Пример II.1. Пусть A и B — следующие подмножества числовой прямой: $A = [1; 3]$, $B = [2; 4]$. Тогда $A \cap B = [2; 3]$, $A \cup B = [1; 4]$, $A \setminus B = [1; 2)$, $B \setminus A = (3; 4]$, $A \triangle B = [1; 2) \cup (3; 4]$.

II.7. Определение. Будем говорить, что два множества A и B *находятся в общем положении*, если существуют такие элементы a, b, c, что $a \in A$, $a \notin B$, $b \in B$, $b \notin A$, $c \in A$, $c \in B$. Два множества в общем положении изображены на рис. II.2, где символ a, например, обозначает список всех элементов, содержащихся в множестве A, но не содержащихся в множестве B и т. д.



(a) Пересечение множеств

(b) Объединение множеств

(c) Разность множеств

(d) Симметрическая разность множеств

Рис. II.1. Диаграммы Эйлера, иллюстрирующие основные операции над множествами

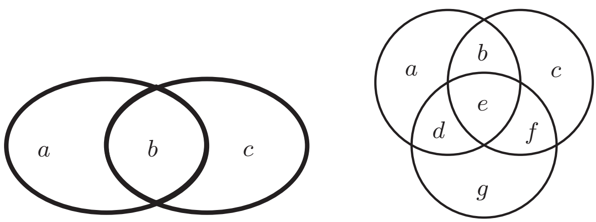


Рис. II.2. Множества в общем положении

Понятие трёх множеств в общем положении вводится аналогично; опираясь на диаграмму Эйлера, изображённую на рис. II.2, сформулируйте определение самостоятельно.

Множества, элементами которых являются также множества, называются обычно системами (совокупностями) множеств и обозначаются $\{A_\alpha \mid \alpha \in \mathfrak{A}\}$, где $\mathfrak{A}$ — множество, элементы которого «нумеруют» множества рассматриваемой системы; оно называется множеством индексов и может быть как конечным, так и бесконечным.

Пусть $\{A_\alpha \mid \alpha \in \mathfrak{A}\}$ — некоторая система множеств. Определим пересечение и объединение этих множеств:

$$\bigcap_{\alpha\in\mathfrak{A}} A_\alpha \stackrel{\text{def}}{=\!=} \{x \mid \forall\alpha \in \mathfrak{A} : x \in A_\alpha\},$$

$$\bigcup_{\alpha\in\mathfrak{A}} A_\alpha \stackrel{\text{def}}{=\!=} \{x \mid \exists\alpha \in \mathfrak{A} : x \in A_\alpha\}.$$

Используются также обозначения $\bigcap\limits_{k=m}^{n} A_k$, $\bigcup\limits_{k=m}^{n} A_k$ и им подобные.

Пример II.2. Пусть A_k — множество всех точек плоскости, лежащих внутри круга радиуса 2^k с центром в фиксированной точке, $n \in \mathbb{Z}$. Тогда $\bigcap\limits_{k=-\infty}^{\infty} A_k$ содержит единственную точку — общий центр рассматриваемых кругов, а $\bigcup\limits_{k=-\infty}^{\infty} A_k$ совпадает со всей плоскостью.

II.8. 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.9. Теорема. *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) *законы поглощения*:

$$(A \cap B) \cup B = B, \quad (A \cup B) \cap B = B;$$

(5) *законы дополнительности*:

$$A \cap A' = \varnothing, \quad A \cup A' = U;$$

(6) $A \setminus B = A \cap B'$;
(7) $\varnothing' = U$; $U' = \varnothing$;
(8) *законы де Моргана*:

$$(A \cap B)' = A' \cup B', \quad (A \cup B)' = A' \cap B'. \tag{II.1}$$

Доказательство. Докажем, например, соотношение

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C).$$

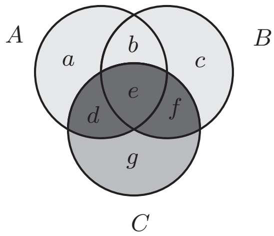


Рис. II.3.

Удобно воспользоваться техникой диаграмм Эйлера. Рассмотрим три множества A, B, C в общем положении (см. рис. II.3):

$$A = \{a, b, d, e\},\ B = \{b, c, e, f\},\ C = \{d, e, f, g\}.$$

Имеем:

$$A \cup B = \{a, b, c, d, e, f\},$$
$$(A \cup B) \cap C = \{d, e, f\},$$
$$A \cap C = \{d, e\}, \quad B \cap C = \{e, f\},$$
$$(A \cap C) \cup (B \cap C) = \{d, e, f\},$$

что и требовалось доказать.

Читателю рекомендуется доказать остальные соотношения самостоятельно. □

II.10. Определение. *Декартовым* (*прямым*) *произведением* $A \times B$ множеств A и B называется множество, состоящее из всех упорядоченных пар (a, b), где $a \in A$, $b \in B$:

$$A \times B = \{(a, b) \mid a \in A,\ b \in B\}.$$

Декартово произведение нескольких множеств $A_1, \ldots, A_n$ — это множество всех упорядоченных последовательностей, образованных элементами множеств $A_1, \ldots, A_n$:

$$A_1 \times \cdots \times A_n = \bigtimes_{i=1}^{n} A_i = \{(a_1, \ldots, a_n) \mid a_1 \in A_1, \ldots, a_n \in A_n\}.$$

2. Соответствия и отображения

A. Соответствия.

II.11. Определение. *Соответствием* R, заданным на упорядоченной паре множеств A и B, называется любое подмножество R декартова произведения этих множеств $A \times B$. 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: $a\ R\ b$.

Для соответствия $R \subseteq A \times B$ введём следующие термины:

(i) *область отправления* — множество A;

(ii) *область прибытия* — множество B;

(iii) *область определения* $\operatorname{Dom} R$ — множество всех *первых* элементов пар из R:

$$\operatorname{Dom} R = \{a \mid \exists b : (a, b) \in R\};$$

(iv) *область значений* $\operatorname{Im} R$ — множество всех *вторых* элементов пар из R:

$$\operatorname{Im} R = \{b \mid \exists a : (a, b) \in R\}.$$

В случае конечных множеств A и B соответствие $R \subseteq A \times B$ 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.4).

Выделяются следующие важные типы соответствий.

II.12. Определение. Соответствие $R \subseteq A \times B$ называется

(i) *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. 4(a));

(ii) *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. 4(b));

(iii) *функциональным*, если в нём нет пар с одинаковыми первыми и различными вторыми компонентами, т.е.

$$\forall (a_1, b_1), (a_2, b_2) \in R : b_1 \neq b_2 \implies a_1 \neq a_2$$

или, по контрапозиции,

$$\forall (a_1, b_1), (a_2, b_2) \in R : a_1 = a_2 \implies b_1 = b_2;$$

на графике из каждой точки области отправления исходит не более одной стрелки, т.е. запрещены ситуации вида $\bullet\begin{smallmatrix}\nearrow\\ \searrow\end{smallmatrix}$ (см. рис. 4(c));

(iv) *инъективным*, если в нём нет пар с различными первыми и одинаковыми вторыми компонентами, т.е.

$$\forall (a_1, b_1), (a_2, b_2) \in R : a_1 \neq a_2 \implies b_1 \neq b_2$$

или, по контрапозиции,

$$\forall (a_1, b_1), (a_2, b_2) \in R : b_1 = b_2 \implies a_1 = a_2;$$

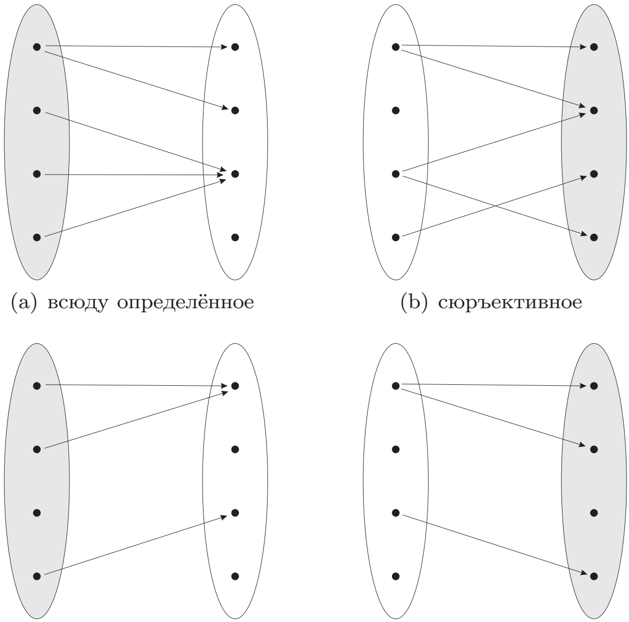
на графике в каждую точку области прибытия приходит не более одной стрелки, т.е. запрещены ситуации вида $\begin{smallmatrix}\searrow\\ \nearrow\end{smallmatrix}\bullet$ (см. рис. 4(d));

(v) *взаимно однозначным соответствием*, или *биективным соответствием*, или *биекцией* между множествами A и B если оно всюду определено, функционально, сюръективно и инъективно.

Замечание. 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).

Соответствия $R \subseteq A \times A$ называют *отношениями* на множестве A. Одним из наиболее важных типов отношений является отношение эквивалентности (см. раздел 1 главы III, с. 35).

B. 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.



(a) всюду определённое (b) сюръективное

(c) функциональное (d) инъективное

Рис. II.4. Типы соответствий (к определению II.12)

II.13. Определение. Пусть f — соответствие, определённое на множествах A и Y, т.е. подмножество декартова произведения $A \times Y$, обладающее свойством функциональности: в нём нет пар с одинаковыми первыми и разными вторыми компонентами, т.е.

$$\forall (a_1, y_1), (a_2, y_2) \in f \colon y_1 \neq y_2 \implies a_1 \neq a_2.$$

Пусть X — область определения соответствия f, т.е. множество

$$X = \operatorname{Dom} f \stackrel{\text{def}}{=\!=} \{x \in A \mid \exists y \colon (x, y) \in f\}.$$

Тогда упорядоченная тройка (f, X, Y) называется *отображением*, определённым ***на*** множестве X и принимающим значения ***в*** множестве Y.[1]

С понятием отображения связаны следующие термины и обозначения:

(a) 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 $x \xmapsto{f} y$; говорят также,

[1]Предлоги «на» и «в» в этом определении употребляются не произвольно, а играют самостоятельную роль, имеют терминологическое значение.

что отображение f переводит (преобразует, превращает; функция f отображает) элемент x в элемент y;[1]

(b) *образ множества* $C \subseteq A$ при отображении f — это множество
$$f(C) = \{f(c) \mid c \in C\} = \{y \in Y \mid \exists c \in C\colon (c, y) \in f\};$$

(c) X — *область определения* отображения f; говорят, что отображение f определено ***на*** X;

(d) A — *область отправления*[2] отображения f; говорят, что отображение f определено ***в*** A;

(e) Y — *область прибытия* отображения f; говорят, что отображение f принимает значения ***в*** множестве Y;

(f) $\operatorname{Im} f \stackrel{\text{def}}{=\!=} \{y \in Y \mid \exists x\colon y = f(x)\} \subseteq Y$ — *область значений* отображения f; говорят, что отображение f принимает значения ***на*** множестве $\operatorname{Im} f$;

(g) записи $f\colon X \to Y$, $X \xrightarrow{f} Y$ означают, что f — отображение с областью определения X и областью прибытия Y;

(h) отображение $f\colon X \to X$ часто называют (особенно в геометрии) *преобразованием* множества X;

(i) *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 (см. ниже п. D), которое может и не существовать;

(j) *прообраз подмножества* $B \subseteq Y$ — 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.14. Определение. Пусть $X \subset \tilde{X}$ и $Y \subseteq \tilde{Y}$. Отображение $f\colon X \to Y$ называется *сужением* отображения $g\colon \tilde{X} \to \tilde{Y}$ на подмножество X, если $f(x) = g(x)$ для всех $x \in X$; обозначение $f = g|_X$. В свою очередь, отображение $g\colon \tilde{X} \to \tilde{Y}$ называется *продолжением* отображения $f\colon X \to Y$ на множество $\tilde{X}$.

II.15. Определение. Отображение $f\colon X \to Y$ называется

(a) *сюръективным* (сюръекцией, отображением «на»), если $\operatorname{Im} f = Y$;

(b) *инъективным* (инъекцией), если из неравенства $x_1 \neq x_2$ следует $f(x_1) \neq f(x_2)$ или, по контрапозиции, из $f(x_1) = f(x_2)$ следует $x_1 = x_2$;

---

[1] 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.

[2] Понятие области отправления используется редко.

(c) *биективным* (взаимно однозначным, биекцией), если оно одновременно сюръективно и инъективно.

II.16. Определение. 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 $x \in X$.

Пример II.3. Отображения[1] $f\colon \mathbb{R} \to \mathbb{R}$, $g\colon \mathbb{R} \to \mathbb{R}_{\geqslant 0}$, $h\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, 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.17. Определение. В дальнейшем будут использоваться следующие отображения.

(a) Отображение $\mathrm{id}_X\colon X \to X$, определённое правилом $\mathrm{id}_X\colon x \mapsto x$, называется *тождественным*. Очевидно, оно биективно.
(b) Пусть $c \in X$. Отображение[2] $c\colon X \to X$, определённое правилом $c\colon x \mapsto c$, называется *постоянным отображением*, или *константой*.

II.18. Определение. Пусть X — произвольное множество. *Семейством*, или *упорядоченным набором* n элементов множества X называется произвольное отображение

$$f\colon \{1, 2, \ldots, n\} \to X, \quad k \mapsto f_k,$$

которое каждому натуральному числу $k \in \{1, 2, \ldots, n\}$ ставит в соответствие некоторый элемент f_k множества X, называемый *k-м членом семейства*. Таблица значений отображения f

k	1	2	$\ldots$	n
$f(k)$	f_1	f_2	$\ldots$	f_n

,

сокращённая до второй строки, т.е. $(f_1, f_2, \ldots, f_n)$, может служить адекватным обозначением семейства.

Таким образом, понятие *семейства* элементов отличается от понятия *множества* элементов тем, что, во-первых, элементы системы пронумерованы, и во-вторых, среди них могут быть равные.

Подсемейством семейства $f\colon \{1, 2, \ldots, n\} \to X$ называется сужение отображение f на какое-либо подмножество множества $\{1, 2, \ldots, n\}$. Например, подсемейством семейства $(f_1, f_2, f_3, f_4, f_5, f_6)$ является семейство (f_1, f_2, f_5) (отображение f сужено с множества $\{1, 2, 3, 4, 5, 6\}$ на множество $\{1, 2, 5\}$).

Аналогично можно ввести понятие бесконечной последовательности элементов множества.

[1] Читатель, конечно, легко установит смысл обозначений $\mathbb{R}_{\geqslant 0}$ и $\mathbb{R}_{>0}$.

[2] Автор надеется, что обозначение элемента множества и отображения одним и тем же символом не вызовет путаницы у читателя.

II.19. Определение. Пусть X — произвольное множество. *Последовательностью* элементов множества X называется произвольное отображение

$$f\colon \mathbb{N} \to X, \quad k \mapsto f_k,$$

которое ставит в соответствие каждому натуральному числу $k \in \mathbb{N}$ некоторый элемент f_k множества X, называемый *k-м членом последовательности. Подпоследовательностью* последовательности $f\colon \mathbb{N} \to X$ называется сужение отображения f на какое-либо (как правило, бесконечное) подмножества множества $\mathbb{N}$.

C. Композиция отображений.

II.20. Определение. *Композицией* (или *суперпозицией*) отображений $f\colon X \to Y$ и $g\colon Y \to Z$ называется отображение $g \circ f\colon X \to Z$, определённое условием

$$(\forall x \in X)\ (g \circ f)(x) = g\bigl(f(x)\bigr)$$

(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.21. Замечание. 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 преобразований $f, g\colon X \to X$, когда обе композиции $g \circ f$ и $f \circ g$ определены, коммутативность может не иметь места, как показывают примеры числовых функций $f\colon x \mapsto x^2$ и $g\colon x \mapsto \sin x$:

$$(g \circ f)(x) = g(f(x)) = \sin(x^2), \quad (f \circ g)(x) = f(g(x)) = (\sin x)^2.$$

II.22. Теорема. *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

$$(h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x))) = \\ = (h \circ g)(f(x)) = ((h \circ g) \circ f)(x). \quad \square$$

$\square$

D. Обратное отображение.

II.23. Определение. Пусть $f\colon X \to Y$ и $g\colon Y \to X$ — отображения. Если

$$g \circ f = \mathrm{id}_X \quad \text{и} \quad f \circ g = \mathrm{id}_Y, \tag{II.2}$$

то отображение g называется *обратным* к отображению f и обозначается f^{-1}. Таким образом,

$$f(x) = y \quad \Longleftrightarrow \quad f^{-1}(y) = x.$$

Отображения f и f^{-1} называют *взаимно обратными.*

II.24. Замечание. 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.25. 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*

$$(h \circ f)^{-1} = f^{-1} \circ h^{-1}. \tag{II.3}$$

4. *Если отображение $f\colon X \to Y$ биективно, то $f^{-1}\colon Y \to X$ также биективно, причём*

$$(f^{-1})^{-1} = f. \tag{II.4}$$

ДОБАВЛЕНИЕ 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. Теорема. *Отношение равномощности множеств обладает всеми свойствами отношения эквивалентности: оно рефлексивно, симметрично и транзитивно.*

III.3. Замечание. Мы сознательно избегаем говорить, что отношение равномощности *является* отношением эквивалентности, чтобы не пришлось указывать, на каком множестве задано это отношение; как показывает парадокс Расселла, о «множестве всех множеств» говорить нельзя, а конкретизировать класс всех множеств, которые рассматривать допустимо, 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. теорему **??** в лекции 1, с. **??**). $\square$

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*.

Таким образом, в случае бесконечного множества имеется взаимно однозначное соответствие между ним самим и его частью, а в случае конечного множества такое соответствие невозможно.

Пример III.1. 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 $n \mapsto 2n$.

2. Конечные множества. Комбинаторика

Комбинаторика — это раздел математики, в котором изучаются конечные множества и различные комбинации их элементов (комбинаторные конфигурации).

A. Мощность конечных множеств.

III.5. 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.6. Теорема. *Мощность декартова произведения конечных множеств A и B равна произведению мощностей этих множеств*:

$$|A \times B| = |A| \cdot |B|.$$

Для декартова произведения нескольких множеств имеет место аналогичное соотношение

$$\left| \bigtimes_{i=1}^{n} A_i \right| = \prod_{i=1}^{n} |A_i|.$$

Доказательство. Теорема III.6 становится очевидной, если элементы множества $A \times B$ записать в таблицу вида

	a_1	a_2	$\dots$	a_m
b_1	(a_1, b_1)	(a_2, b_1)	$\dots$	(a_m, b_1)
b_2	(a_1, b_2)	(a_2, b_2)	$\dots$	(a_m, b_2)
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
b_n	$(a_1, b_2n$	(a_2, b_n)	$\dots$	(a_m, b_n)

где $A = \{a_1, a_2, \dots\}$, $B = \{b_1, b_2, \dots\}$. $\square$

III.7. Теорема. *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 различных подмножеств множества $A = \{1, 2, \dots, n\}$. Закодируем каждое подмножество $X \subset A$ последовательностью Y (длины n), 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). □

Теорема III.6 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.8. Теорема. *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
$$|A \cup B| = |A \setminus B| + |B \setminus A| + |A \cap B| = (a - q) + (b - q) + q = a + b - q,$$
что и требовалось. □

Теорема III.8 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.

В дальнейшем нам будет удобно пользоваться следующим понятием.

III.9. Определение. *Факториалом* $n!$ натурального числа n называется произведение $1 \cdot 2 \cdot \dots \cdot n$. По определению полагают $0! = 1$. Факториал как функцию неотрицательного целого аргумента можно определить рекурсивно:

$$0! = 1, \quad (n + 1)! = n! \cdot (n + 1).$$

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, дающая приближённое значение $n!$ при больших значениях n:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n, \tag{III.1}$$

где $\pi \approx 3{,}14159\ldots$ — число π (отношение длины окружности к её диаметру), $e \approx 2{,}71828\ldots$ — число e (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:

$$a = 10! = 3\,628\,800, \quad b = \sqrt{2\pi \cdot 10} \cdot \left(\frac{10}{e}\right) \approx 3\,598\,696, \quad \frac{a - b}{a} \approx 0{,}0083.$$

B. Размещения. 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.10. 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).

III.11. Теорема. 1. *Число $\overline{A}_n^k$ повторных упорядоченных выборок объёма k из n-множества, т.е. число 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. число размещений без повторений из* n *по* k*, равно*

$$A_n^k = \frac{n!}{(n-k)!}. \tag{III.3}$$

В зарубежной литературе вместо $\overline{A}_n^k$, A_n^k используются обозначения $\overline{P}_k^n$, P_k^n соответственно (обратите внимание на положение индексов!).

Доказательство. 1. При образовании повторной 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. 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

$$A_n^k = \underbrace{n(n-1)\cdots(n-k+1)}_{k\text{ сомножителей}} =$$

$$= n(n-1)\cdots(n-k+1)\cdot\overbrace{\frac{(n-k)(n-k-1)\cdots 2\cdot 1}{(n-k)(n-k-1)\cdots 2\cdot 1}}^{\text{дробь равна 1}} =$$

$$= \frac{n(n-1)\cdots(n-k+1)(n-k)(n-k-1)\cdots 2\cdot 1}{(n-k)(n-k-1)\cdots 2\cdot 1} = \frac{n!}{(n-k)!}. \quad \square$$

$\square$

C. Перестановки. Размещения объёма n, составленные из элементов n-множества A, называются *перестановками* элементов множества A. Используя формулу (III.3) при $k = n$, находим число P_n перестановок элементов n-множества:

$$P_n = n!. \tag{III.4}$$

Более удобное определение перестановки множества A связано с понятием взаимно однозначного отображения.

III.12. Определение. *Перестановкой* множества A называется любое взаимно однозначное отображение $\sigma\colon A \to A$ множества A в себя. Перестановку можно задать таблицей значений отображения σ:

a_1	a_2	$\ldots$	a_n
$\sigma(a_1)$	$\sigma(a_2)$	$\ldots$	$\sigma(a_n)$

Множество всех перестановок n-множества обозначается S_n; более подробно оно рассматривается в п. **??** (см. с. **??**).

III.13. Определение. Пусть $A = \{a_1, \dots, a_n\}$ — n-множество A. *Перестановкой с повторениями* типа $(k_1, \dots, k_n)$ из элементов множества A называется упорядоченный набор, состоящий из $k = k_1 + \dots + k_n$ элементов множества A, в котором элемент a_1 встречается k_1 раз, …, элемент a_n встречается k_n раз.

III.14. Теорема. *Количество $P(k_1, \dots, k_n)$ перестановок с повторениями типа $(k_1, \dots, k_n)$ из элементов n-множества A равно*

$$P(k_1, \dots, k_n) = \frac{k!}{k_1! \dots k_n!}, \quad k = k_1 + \dots + k_n. \tag{III.5}$$

Доказательство. Каждую перестановку с повторениями типа $(k_1, \dots, k_n)$ из элементов n-множества A будем кодировать словом W длины $k = k_1 + \dots + k_n$ в алфавите $A = \{a_1, \dots, a_n\}$. Подсчитаем число перестановок букв, при которых слово W не меняется. Буквы a_1, имеющиеся в слове W в количестве k_1 штук, можно произвольно переставлять на k_1 местах, на которых они расположены; это можно сделать $k_1!$ способами. Аналогично буквы a_2, …, a_n можно переставить $k_2!$, …, $k_n!$ способами соответственно. По правилу произведения число перестановок букв, при которых слово W не меняется, равно $k_1! \cdots k_n!$. Общее число перестановок k букв равно $k!$. Следовательно, число слов, т.е. перестановок типа $(k_1, \dots, k_n)$, выражается формулой (III.5). $\square$

D. Сочетания и биномиальные коэффициенты.

III.15. 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.16. Теорема. *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]

$$C_n^k = \frac{n!}{k!(n-k)!}. \tag{III.6}$$

Числа C_k^n называются *биномиальными коэффициентами*; причина такого названия вскоре выяснится.

Доказательство. Каждое k-сочетание $\{a_1, \dots, a_k\}$ из элементов данного n-множества можно превратить в размещение $(a_{i_1}, \dots, a_{i_k})$, по-разному упорядочивая его элементы, $k!$ способами. Таким образом, каждому сочетанию отвечает $k!$ размещений, поэтому

$$C_n^k = \frac{A_n^k}{k!} = (\mathrm{III}.3) \frac{n!}{k!(n-k)!}. \quad \square$$

$\square$

[1]В зарубежной литературе вместо ${}^k_n$ используются обозначения C_k^n, ${}^nC_k$, $\binom{n}{k}$ (обратите внимание на положение индексов!).

III.17. Теорема. *Биномиальные коэффициенты C_n^k обладают следующими свойствами*:

$$1. \qquad \sum_{k=0}^{n} C_n^k = 2^n; \tag{III.7}$$

$$2. \qquad C_n^0 = C_n^n = 1, \quad C_n^1 = C_n^{n-1} = n; \tag{III.8}$$

$$3. \qquad C_n^k = \frac{\overbrace{n(n-1)\cdots(n-k+1)}^{k \text{ сомножителей}}}{\underbrace{k(k-1)\cdots 1}_{k \text{ сомножителей}}}; \tag{III.9}$$

$$4. \qquad C_n^{n-k} = C_n^k; \tag{III.10}$$

$$5. \qquad C_n^k + C_n^{k+1} = C_{n+1}^{k+1}. \tag{III.11}$$

Доказательство. Соотношение (III.7) 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.7 равно 2^n. Алгебраическое доказательство будет приведено ниже (см. с. 343)

Свойства (III.8), (III.9) вытекают непосредственно из формулы (III.6). Докажем (III.10):

$$C_n^{n-k} = \frac{n!}{(n-k)!(n-(n-k))!} = \frac{n!}{(n-k)!k!} = C_n^k.$$

Отметим, что соотношение (III.10) 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.10)).

Докажем (III.11) 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 кабинет министров состоит из $n+1$ членов, и требуется сформировать комиссию, состоящую из $k+1$ членов; это можно сделать C_{n+1}^{k+1} способами. Если в состав комиссии входит премьер-министр, то оставшиеся члены комиссии могут быть выбраны C_n^k способами; если же в состав комиссии премьер-министр не входит, то комиссия может быть сформирована C_n^{k+1} способами. По правилу сложения $C_{n+1}^{k+1} = C_n^k + C_n^{k+1}$. $\square$

На свойстве (III.11) основан простой способ вычисления чисел C_n^k, который можно проиллюстрировать следующей диаграммой:

$$\begin{tikzcd}[column sep=small]
 & & & & C_0^0 \arrow[dl] \arrow[dr] & & & & \\
 & & & C_1^0 \arrow[dl] \arrow[dr] & & C_1^1 \arrow[dl] \arrow[dr] & & & \\
 & & C_2^0 \arrow[dl] \arrow[dr] & & C_2^1 \arrow[dl] \arrow[dr] & & C_2^2 \arrow[dl] \arrow[dr] & & \\
 & C_3^0 \arrow[dl] \arrow[dr] & & C_3^1 \arrow[dl] \arrow[dr] & & C_3^2 \arrow[dl] \arrow[dr] & & C_3^3 \arrow[dl] \arrow[dr] & \\
C_4^1 & & C_4^1 & & C_4^2 & & C_4^3 & & C_4^4
\end{tikzcd}$$

Как указывает формула (III.11), 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] Отметим, что 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 девять строк (с нулевой по восьмую включительно) треугольника Паскаля имеют вид

$n=0$									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		
$n=7$		1		7		21		35		35		21		7		1	
$n=8$	1		8		28		56		70		56		28		8		1

[1]Блез Паскаль (Blaise Pascal, 1623–1662) — французский математик, физик, литератор и философ. Классик французской литературы, один из основателей математического анализа, теории вероятностей и проективной геометрии, создатель первых образцов счётной техники, автор основного закона гидростатики.

Приведём без доказательства ещё ряд замечательных свойств, которыми обладают биномиальные коэффициенты C_n^k и треугольник Паскаля.

1. Боковые стороны треугольника состоят из единиц, а вдоль линий, параллельных боковым сторонам треугольника, стоят:

(1) натуральные числа;
(2) *треугольные числа*, т.е. числа кругов, которые можно расположить в виде правильного треугольника:
(3) *тетраэдральные числа*, т.е. числа шаров, которые можно расположить в виде правильного тетраэдра;
(4) числа, являющиеся обобщениями треугольных и тетраэдральных чисел на многомерный случай.

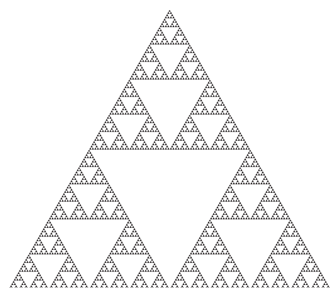
2. Если строки в треугольнике Паскаля выровнять по левому краю, то суммы чисел, расположенных вдоль наклонных линий, идущих слева направо и снизу вверх (см. диаграмму на рис. III.1), равны числам Фибоначчи, т.е. элементам последовательности, определённой рекуррентно соотношениями

$$F_0 = 1, \quad F_1 = 1, \quad F_n = F_{n-1} + F_{n-2}.$$

Первые десять чисел Фибоначчи:

$$1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ 55.$$

3. 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[1] — 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.



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

[1]Неформально говоря, фрактал — это сложная геометрическая фигура, обладающая свойством самоподобия, то есть 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.

[2]Вацлав Серпинский (Wacław Sierpiński, 1882–1969) — польский математик.

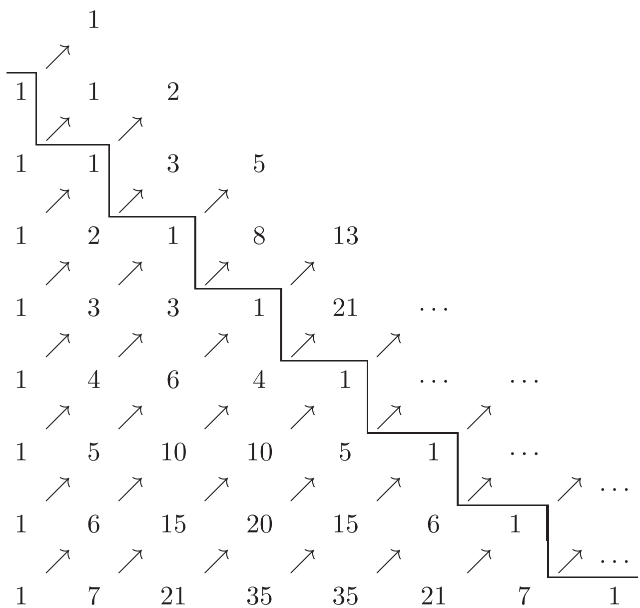


Рис. III.1. Треугольник Паскаля и числа Фибоначчи

T_2, состоящее из девяти равных правильных треугольников второго ранга. Продолжая этот процесс, получим бесконечную последовательность

$$T_0 \supset T_1 \supset \dots \supset T_n \supset \dots,$$

пересечение членов которой называется треугольником Серпинского.

E. Бином Ньютона. Известные из школьного курса формулы сокращенного 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.18. Теорема. *Для любых чисел a и b и любого натурального n справедлива следующая формула, называемая формулой бинома Ньютона:*

$$(a+b)^n = \sum_{k=0}^{n} C_n^k a^{n-k} b^k =$$

$$= a^n + C_n^1 a^{n-1} b + C_n^2 a^{n-2} b^2 + \dots + C_n^k a^{n-k} b^k + \\ + \dots + C_n^{n-1} a b^{n-1} + b^n. \quad \text{(III.12)}$$

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

$$(a+b)^n = \sum_{k=0}^{n} |B_k| a^{n-k} b^k = \sum_{k=0}^{n} C_n^k a^{n-k} b^k. \quad \square$$

$\square$

Индуктивное доказательство Докажем формулу Ньютона *методом математической индукции*. Установим сначала базу индукции, т.е. справедливость формулы (III.12) при $n = 1$:

$$(a+b)^1 = a + b = C_1^0 a + C_1^1 b.$$

Предположение индукции состоит в том, что формула (III.12) справедлива при некотором значении n. Наша задача — пользуясь формулой (III.12), вывести её справедливость для показателя степени $n+1$:

$$(a+b)^{n+1} = (a+b) \cdot (a+b)^n = (a+b) \left(\sum_{k=0}^{n} C_n^k \; a^{n-k} \; b^k \right) =$$

$$= a \left(\sum_{k=0}^{n} C_n^k \; a^{n-k} \; b^k \right) + b \left(\sum_{k=0}^{n} C_n^k \; a^{n-k} \; b^k \right) =$$

$$= \sum_{k=0}^{n} C_n^k \; a^{n-k+1} \; b^k + \sum_{k=0}^{n} C_n^k \; a^{n-k} \; b^{k+1} =$$

$$= a^{n+1} + \underbrace{\sum_{k=1}^{n} C_n^k \; a^{n-k+1} \; b^k}_{\substack{\text{Замена } k=p+1, \\ k=1,\dots,n, \\ p=0,\dots,n-1}} + \underbrace{\sum_{k=0}^{n-1} C_n^k \; a^{n-k} \; b^{k+1}}_{\text{Замена } k=p} + b^{n+1} =$$

$$= a^{n+1} + \underbrace{\sum_{p=0}^{n-1} C_n^{p+1}\ \underline{a^{n-p}\ b^{p+1}} + \sum_{p=0}^{n-1} C_n^{p}\ \underline{a^{n-p}\ b^{p+1}}}_{\text{подобные слагаемые}} + b^{n+1} =$$

$$= a^{n+1} + \sum_{p=0}^{n-1} \underbrace{\left(C_n^{p+1} + C_n^{p}\right)}_{=C_{n+1}^{p+1}}\ a^{n-p}\ b^{p+1} + b^{n+1} =$$

$$= C_{n+1}^{0}\ a^{n+1} + \underbrace{\sum_{p=0}^{n-1} C_{n+1}^{p+1}\ a^{(n+1)-(p+1)}\ b^{p+1}}_{\substack{\text{Замена } k=p+1,\\ p=0,\dots,n-1,\\ k=1,\dots,n}} + C_{n+1}^{n+1}\ b^{n+1} =$$

$$= C_{n+1}^{0}\ a^{n+1} + \sum_{k=1}^{n} C_{n+1}^{k}\ a^{(n+1)-k}\ b^{k} + C_{n+1}^{n+1}\ b^{n+1} =$$

$$= \sum_{k=0}^{n+1} C_{n+1}^{k}\ a^{(n+1)-k}\ b^{k}.$$

$\square$

Формула бинома Ньютона позволяет установить ещё несколько интересных свойств биномиальных коэффициентов. Например, взяв в формуле (III.12) $a = b = 1$, получим

$$(1+1)^n = \sum_{k=0}^{n} C_n^k\ 1^{n-k}\ 1^k,$$

откуда

$$\sum_{k=0}^{n} C_n^k = 2^n$$

(этот результат был получен комбинаторным способом в теореме III.17, формула (III.7)). Аналогично, взяв $a = 1$, $b = -1$, находим

$$\sum_{k=0}^{n} (-1)^k\, C_n^k = 0.$$

При $a = 1$, $b = 2$ получаем

$$\sum_{k=0}^{n} 2^k\, C_n^k = 3^n.$$

3. Бесконечные множества

Напомним (см. определение III.4, с. 332), 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.19. Определение. Множество называется *счётным*, если оно равномощно множеству $\mathbb{N}$ натуральных чисел. Неформально говоря, множество счётно, если его элементы можно перенумеровать.

III.20. 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 $B \subset A = \{a_1, a_2, \dots\}$. Удалим из последовательности $a_1, a_2, \dots$ те элементы, которые 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 $A_1 = \{a_{11}, a_{12}, \dots\}$, $A_2 = \{a_{21}, a_{22}, \dots\}$, …. Расположив элементы каждого из них слева направо в последовательность и записав последовательности одну под другой, получим бесконечную таблицу

$$
\begin{array}{ccccccccccc}
a_{11} & \rightarrow & a_{12} & & a_{13} & \rightarrow & a_{14} & & a_{15} & \rightarrow & \dots \\
 & \swarrow & & \nearrow & & \swarrow & & \nearrow & & \swarrow & \\
a_{21} & & a_{22} & & a_{23} & & a_{24} & & a_{25} & & \dots \\
\downarrow & \nearrow & & \swarrow & & \nearrow & & \swarrow & & \nearrow & \\
a_{31} & & a_{32} & & a_{33} & & a_{34} & & a_{35} & & \dots \\
 & \swarrow & & \nearrow & & \swarrow & & \nearrow & & & \\
a_{41} & & a_{42} & & a_{43} & & a_{44} & & a_{45} & & \dots \\
\downarrow & \nearrow & & \swarrow & & \nearrow & & & & & \\
a_{51} & & a_{52} & & a_{53} & & a_{54} & & a_{55} & & \dots \\
 & \swarrow & & \nearrow & & & & & & & \\
\dots & & \dots & & \dots & & \dots & & \dots & & \dots
\end{array}
$$

которую можно развернуть в последовательность

$$a_{11},\ a_{12},\ a_{21},\ a_{31},\ a_{22},\ a_{13},\ a_{14},\ a_{23},\ a_{32},\ a_{41},\ \dots .$$

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-то 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 потому также счётно. □

Пример III.2. Приведем важные примеры счётных множеств.

(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, а именно не более n для многочленов степени n (это утверждение называется «основная теорема алгебры», см. п. **??**).

III.21. Теорема. *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 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 $B \sqcup P$ и P оба счётны, между ними существует биекция, которая легко продолжить до биекции между $P \sqcup Q \sqcup B$ и $P \sqcup Q$ тождественным на Q отображением (т.е. каждый элемент множества Q соответствует сам себе). □

B. Несчётные множества.

III.22. Теорема. *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 вниз таблицу, строками которой являются наши последовательности:

$$\begin{aligned} \alpha_1 &= \underline{\alpha_{11}}\ \alpha_{12}\ \alpha_{13}\ \ldots \\ \alpha_2 &= \alpha_{21}\ \underline{\alpha_{22}}\ \alpha_{23}\ \ldots \\ \alpha_3 &= \alpha_{31}\ \alpha_{32}\ \underline{\alpha_{33}}\ \ldots \\ &\ \ \vdots \end{aligned}$$

(через α_{ij} обозначен j-й член i-й последовательности). Рассмотрим теперь последовательность, образованную стоящими на диагонали членами α_{11}, α_{22}, ...; ее i-й член есть 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 — противоречие. □

III.23. Теорема.

$$[0,1] \sim 2^{\mathbb{N}} \sim \mathcal{P}(\mathbb{N}).$$

Числовой отрезок $[0,1]$*, множество* $2^{\mathbb{N}}$ *и множество* $\mathcal{P}(\mathbb{N})$ *всех подмножеств натурального ряда равномощны и несчётны.*

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 $x_0 x_1 x_2 \ldots$ соответствует число, являющееся суммой бесконечной геометрической прогрессии

$$\frac{x_0}{2} + \frac{x_1}{2^2} + \frac{x_2}{2^3} + \ldots.$$

[1] Обозначение для этого множества объясняется следующим образом. Согласно определению II.13(**??**), символом Y^X обозначается 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 $\{0,1\}$); см. также определение натуральных чисел в п. **??**.

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, $[0,1] \sim \mathbb{S}$.

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. $\square$

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.

C. Теорема Кантора и сравнение мощностей.

III.24. Теорема (теорема Кантора)**.** *Никакое множество X не равномощно множеству 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 $z \in Z \iff z \notin f(z)$ по построению множества Z и $z \notin f(z) \iff z \notin Z$ по предположению $f(z) = Z$. Полученное противоречие показывает, что Z не соответствует ни одному элементу множества X, т.е. отображение f не биективно. $\square$

Определение равномощности уточняет интуитивную идею о множествах «одинакового размера». Дадим формальное определение ситуации, когда одно множество «больше» другого.

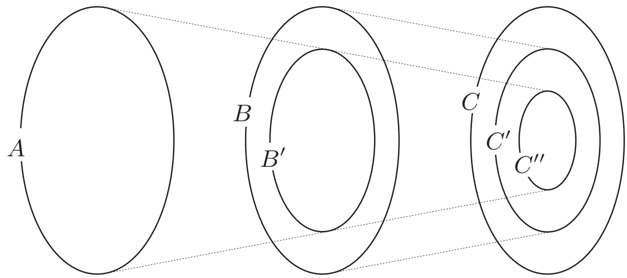
III.25. Определение. Будем говорить, что мощность множества A не больше мощности множества B (обозначение $A \preccurlyeq B$), если A равномощно некоторому подмножеству множества B (возможно, самому B).[1]

[1] Мы сознательно отказываемся от записи $|A| \leqslant |B|$, чтобы избежать ненужной аллюзии со сравнением чисел.

III.26. Теорема. *Отношение* $\preccurlyeq$ *обладает*[1] *всеми свойствами отношения частичного порядка*:

(1) *если* $A \sim B$, *то* $A \preccurlyeq B$; *в частности,* $A \preccurlyeq A$ (*рефлексивность*);
(2) *если* $A \preccurlyeq B$ *и* $B \preccurlyeq C$, *то* $A \preccurlyeq C$ (*транзитивность*);
(3) *если* $A \preccurlyeq B$ *и* $B \preccurlyeq A$, *то* $A \sim B$ (*антисимметричность*).
(4)

Доказательство. Утверждение (1) очевидно. Доказательство (2) можно проиллюстрировать следующей диаграммой:



Если A находится в биективном соответствии с $B' \subset B$, а B — в биективном соответствии с $C' \subset C$, то при втором соответствии множеству B' соответствует некоторое множество $C'' \subset C' \subset C$, так что $A \sim C$.

Антисимметричность отношения $\preccurlyeq$ составляет содержание теоремы Кантора—Бернштейна (см. далее). □

III.27. Теорема (теорема Кантора—Бернштейна)**.** *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.

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.

Вернёмся к вопросу о сравнении мощностей. Для данных множеств A и B теоретически имеются четыре возможности:

(1) A равномощно некоторой части B, а B равномощно некоторой части A; в этом случае множества равномощны;
(2) A равномощно некоторой части B, но B не равномощно никакой части A; в этом случае говорят, что A *имеет меньшую мощность, чем* B;
(3) B равномощно некоторой части A, но A не равномощно никакой части B; в этом случае говорят, что A *имеет большую мощность, чем* B;

[1] См. замечание III.3 на с. 332.

(4) ни A не равномощно никакой части B, ни B не равномощно никакой части A. Этот случай на самом деле невозможен, но доказательство этого факта, опирающееся на теорему Цермело, выходит за рамки данной книги.

D. Континуум-гипотеза. 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 $\mathbb{S}$ и $2^{\mathbb{N}}$). В силу того, что $\mathbb{R} \sim 2^{\mathbb{N}}$, мощность континуума обозначают также $2^{\aleph_0}$.

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 $\mathfrak{c} = \aleph_1$.

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. Гедель доказал, что континуум-гипотеза не может быть опровергнута в рамках аксиоматической теории Цермело—Френкеля, а в 1963 г. П. Коэн 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.