Ильин В.А., Ким Г.Д.

Линейная алгебра и аналитическая геометрия: Учебник. - М.: Изд-во Моск. ун-та, 1998. - 320 с. ISBN 5-211-03814-2.

Книга представляет собой учебник по объединенному курсу линейной алгебры и аналитической геометрии. Она написана на основе лекций, читавшихся ее авторами в Московском государственном университете им. М.В. Ломоносова. В изложении материала, вполне традиционного по своей тематике, авторы придерживаются тенденции объединения двух математических дисциплин в одну, добиваясь лаконичности геометрических доказательств и наглядности алгебраических абстракций. Наряду с традиционными темами книга содержит сведения из общей алгебры, элементы теории множеств. Особое внимание уделяется вычислительным аспектам алгебраических методов. Книга является теоретической поддержкой компьютерного учебника по линейной алгебре, созданного авторами и участвующего в учебном процессе факультета ВМиК МГУ.

Для студентов вузов, обучающихся по специальности "Прикладная математика".

ББК 22.12

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Линейная алгебра является широко используемым аппаратом для всех разделов математики и ее приложений. Особенно возросла ее роль в связи с развитием вычислительной техники и математики. Не будет большим преувеличением утверждать, что любое математическое приложение в вычислительной практике на том или ином этапе сводится к решению алгебраической задачи.

Логическая структура линейной алгебры исключительно проста, она основана на небольшом числе удобных в обращении понятий и аксиом. Однако абстрактный характер алгебраических понятий затушевывает это ее свойство и затрудняет первоначальный опыт изучения линейной алгебры. Объединение линейной алгебры и аналитической геометрии в один курс позволяет подчеркнуть геометрическую природу линейной алгебры и сделать ее объекты более наглядными. По существу, линейная алгебра и аналитическая геометрия настолько связаны, что между ними трудно провести четкую грань, "во многих случаях они отличаются друг от друга лишь языком: каждую из этих дисциплин можно понимать как перевод другой" (Ж. Дьедонне).

При написании этой книги мы придерживались традиции объединения (переплетения) линейной алгебры и аналитической геометрии, установившейся в системе преподавания на факультете вычислительной математики и кибернетики МГУ им. М.В. Ломоносова. При совместном изучении этих дисциплин геометрические представления своей наглядностью делают алгебраические понятия и факты более воспринимаемыми, помогают уяснить, а зачастую и предвидеть не всегда очевидные факты. В свою очередь алгебраический формализм позволяет проводить геометрические исследования более компактно.

Авторы использовали ряд методических приемов из учебников, написанных А.Г. Курошем [9], И.М. Гельфандом [3], Н.В. Ефимовым [4], Г.Е. Шиловым [10], В.В. Воеводиным [2], А.И. Кострикиным [8], В.А. Ильиным и Э.Г. Позняком [6].

Нам приятно подчеркнуть благотворное влияние на методические особенности предлагаемой книги идей первого лектора по данному курсу на факультете ВМиК МГУ В.В. Воеводина.

Авторы считают своим приятным долгом поблагодарить ректора МГУ В.А. Садовничего и декана факультета ВМиК МГУ Д.П. Костомарова, оказавших существенную поддержку изданию этой книги, и научного редактора книги Л.В. Крицкова за ценные замечания, способствовавшие ее улучшению.

В.А. Ильин, Г.Д. Ким

внедиагональные элементы равны нулю. Обозначение: $\mathrm{diag}(a_{11}, \ldots, a_{nn})$. Диагональная матрица, у которой все диагональные элементы равны между собой, называется *скалярной*. Скалярная матрица, у которой все диагональные элементы равны 1, называется *единичной (тождественной)* и обозначается символами I или E. Отметим, что для каждого порядка n существует своя единичная матрица. Число $\mathrm{tr}\,A = a_{11} + \ldots + a_{nn}$ называется *следом* матрицы $A = (a_{ij}) \in \mathbb{R}^{n\times n}$.

Матрица размера $1 \times n$ называется ***строчной матрицей***, или ***матрицей-строкой***, или ***вектор-строкой***. Матрица размера $m \times 1$ называется ***столбцовой матрицей***, или ***матрицей-столбцом***, или ***вектор-столбцом***.

Компактная форма записи матрицы. Пусть $A = (a_{ij}) \in \mathbb{R}^{m\times n}$. Обозначим i-ю строку и i-й столбец матрицы A символами a'_i и a_i:

$$a'_i = \begin{bmatrix} a_{i1} & a_{i2} & \ldots & a_{in} \end{bmatrix} \quad \text{и} \quad a_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix}.$$

В этих обозначениях матрица A может быть записана более компактно:

$$A = \begin{bmatrix} a'_1 \\ a'_2 \\ \vdots \\ a'_m \end{bmatrix} \quad \text{или} \quad A = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix}. \tag{1.1}$$

Такой формой записи мы нередко будем пользоваться, причем не только лаконичности ради.

Матрицы специального вида. Квадратная матрица $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ называется ***верхней (правой) треугольной***, если $a_{ij} = 0$ при $i > j$, и ***нижней (левой) треугольной***, если $a_{ij} = 0$ при $i < j$. Общий вид треугольных матриц:

$$R = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}; \; L = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}. \tag{1.2}$$

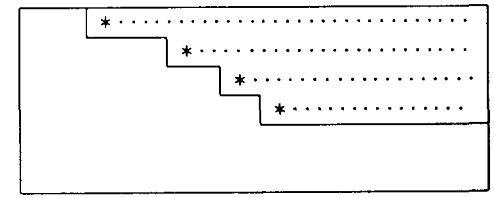
Матрица $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ называется ***верхней (правой) ступенчатой***, если она обладает следующими свойствами:

1) если i-я строка нулевая, то $(i+1)$-я строка также нулевая;

2) если первые ненулевые элементы i-й и $(i+1)$-й строк расположены в столбцах с номерами k_i и k_{i+1}, то $k_i < k_{i+1}$.

Эти свойства означают, что все нулевые строки являются последними и что все элементы, расположенные слева и под первым ненулевым элементом каждой строки, равны нулю. Происхождение названия

становится понятным из "рисунка" ступенчатой матрицы:



(здесь все неотмеченные элементы равны нулю, а заведомо ненулевые элементы помечены знаком $*$).

Если в определении верхней ступенчатой матрицы поменять ролями строки и столбцы, то получим определение ***нижней (левой) ступенчатой*** матрицы.

Ступенчатая матрица, у которой $k_i = i$, называется ***трапециевидной***. Общий вид трапециевидных матриц:

$$R = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1r} & a_{1,r+1} & \ldots & a_{1n} \\ 0 & a_{22} & \ldots & a_{2r} & a_{2,r+1} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & a_{rr} & a_{r,r+1} & \ldots & a_{rn} \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \end{bmatrix}, \tag{1.3}$$

$$L = \begin{bmatrix} a_{11} & 0 & \ldots & 0 & 0 & \ldots & 0 \\ a_{21} & a_{22} & \ldots & 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{r1} & a_{r2} & \ldots & a_{rr} & 0 & \ldots & 0 \\ a_{r+1,1} & a_{r+1,2} & \ldots & a_{r+1,r} & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mr} & 0 & \ldots & 0 \end{bmatrix},$$

где $a_{ii} \neq 0$ при $i = \overline{1, r}$.

До сих пор мы рассматривали матрицы, элементами которых служат числа. Так в основном будет и впредь. Одно из немногих в этом отношении исключений составляют матрицы, элементами которых являются также матрицы. Остановимся на этом более подробно.

Разобьем матрицу $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ системой горизонтальных и вертикальных линий на клетки (блоки). ***Клеточной (блочной) матрицей*** называется матрица, элементами которой служат эти клетки. Общий вид клеточной матрицы:

$$A = \left[\begin{array}{c|c|c|c} A_{11} & A_{12} & \ldots & A_{1k} \\ \hline A_{21} & A_{22} & \ldots & A_{2k} \\ \hline \ldots & \ldots & \ldots & \ldots \\ \hline A_{s1} & A_{s2} & \ldots & A_{sk} \end{array}\right],$$

где A_{ij} – клетка, расположенная в i-й клеточной строке и в j-м клеточном столбце. Квадратная клеточная матрица $A = (A_{ij})$ с квадратными клетками на главной диагонали называется *квазидиагональной*, если $A_{ij} = O$ при $i \neq j$, и *квазитреугольной*, если $A_{ij} = O$ при $i > j$ (или $i < j$).

Как будет видно из дальнейшего, блочные матрицы по многим характеристикам близки к числовым матрицам (см., например, с. 17, свойство 1). Это продуктивно используется в вычислительной математике для обработки большого объема информации.

§2. Операции над матрицами

Равенство матриц. Две матрицы $A = (a_{ij})$ и $B = (b_{ij})$ одинакового размера $m \times n$ называются *равными*, если $a_{ij} = b_{ij}$, $i = \overline{1,m}$, $j = \overline{1,n}$. Обозначение: $A = B$.

Линейные операции. *Суммой* матриц $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ и $B = (b_{ij}) \in \mathbb{R}^{m\times n}$ называется матрица $C = (c_{ij}) \in \mathbb{R}^{m\times n}$, элементы которой определены равенством

$$c_{ij} = a_{ij} + b_{ij}, \quad i = \overline{1,m},\ j = \overline{1,n}.$$

Обозначение: $C = A + B$.

Матрица $-A = (-a_{ij}) \in \mathbb{R}^{m\times n}$ называется *противоположной* к матрице $A = (a_{ij}) \in \mathbb{R}^{m\times n}$.

Теорема 2.1. *Операция сложения матриц обладает следующими свойствами:* $\forall A, B, C \in \mathbb{R}^{m\times n}$ *и* $O \in \mathbb{R}^{m\times n}$

1) $A + B = B + A$ *(сложение матриц коммутативно);*

2) $(A + B) + C = A + (B + C)$ *(сложение матриц ассоциативно);*

3) $A + O = O + A = A$;

4) $A + (-A) = -A + A = O$.

Доказательство. Эти свойства непосредственно вытекают из определения и доказываются по единой схеме. Докажем свойство 2. Матрицы $(A+B)+C$ и $A+(B+C)$ имеют одинаковый размер $m \times n$, при этом их элементы, расположенные в одинаковых позициях, равны, так как $\{(A + B) + C\}_{ij} = \{A + B\}_{ij} + \{C\}_{ij} = (a_{ij} + b_{ij}) + c_{ij} = a_{ij} + (b_{ij} + c_{ij}) = \{A\}_{ij} + \{B + C\}_{ij} = \{A + (B + C)\}_{ij}$. ■

Разностью матриц $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ и $B = (b_{ij}) \in \mathbb{R}^{m\times n}$ называется матрица $X = (x_{ij}) \in \mathbb{R}^{m\times n}$ такая, что $A = B + X$.

Обозначение: $X = A - B$. Очевидно, что для любых матриц $A, B \in \mathbb{R}^{m\times n}$ существует единственная разность $A - B$, при этом

$$A - B = A + (-B) = (a_{ij} - b_{ij}).$$

Произведением матрицы $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ *на число* $\alpha \in \mathbb{R}$ называется матрица $C = (c_{ij}) \in \mathbb{R}^{m\times n}$, элементы которой определены равенством

$$c_{ij} = \alpha a_{ij}, \quad i = \overline{1,m},\ j = \overline{1,n}.$$

Обозначение: $C = \alpha A$.

Теорема 2.2. *Операция умножения матрицы на число обладает следующими свойствами:* $\forall A, B \in \mathbb{R}^{m\times n}$, $\forall \alpha, \beta \in \mathbb{R}$

1) $1 \cdot A = A$;

2) $(\alpha\beta)A = \alpha(\beta A)$;

3) $\alpha(A + B) = \alpha A + \alpha B$ *(умножение матрицы на число дистрибутивно относительно сложения матриц);*

4) $(\alpha + \beta)A = \alpha A + \beta A$ *(умножение матрицы на число дистрибутивно относительно сложения чисел);*

5) $-A = (-1)A$.

Все эти свойства непосредственно вытекают из определения и проверяются по той же схеме, что и свойство 2 из теоремы 2.1. ■

Итак, на множестве $\mathbb{R}^{m\times n}$ всех вещественных матриц одинакового размера $m \times n$ любые две матрицы можно сложить и любую матрицу можно умножить на вещественное число. При этом результаты выполнения этих операций также будут матрицами из $\mathbb{R}^{m\times n}$. Обратим внимание на эту особенность множества $\mathbb{R}^{m\times n}$, так как в дальнейшем мы неоднократно будем сталкиваться с множествами, наделенными операциями сложения и умножения на число, которые обладают свойствами, сформулированными в теоремах 2.1 и 2.2.

Операции сложения матриц и умножения матрицы на число позволяют однозначно определить матрицу $\alpha_1 A_1 + \alpha_2 A_2 + \ldots + \alpha_k A_k = \sum_{i=1}^{k} \alpha_i A_i$, называемую *линейной комбинацией матриц* $A_1, A_2, \ldots, A_k \in \mathbb{R}^{m\times n}$ *с коэффициентами* $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}$.

Умножение матриц. *Произведением матриц* $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ и $B = (b_{ij}) \in \mathbb{R}^{n\times k}$ называется матрица $C = (c_{ij}) \in \mathbb{R}^{m\times k}$, элементы которой определены равенством

$$c_{ij} = \sum_{s=1}^{n} a_{is} b_{sj}, \quad i = \overline{1,m},\ j = \overline{1,k}.$$

Обозначение: $C = AB$.

Уже из определения следует, что произведение матриц зависит от порядка сомножителей; произведение AB определено лишь в том случае, когда размеры матриц A и B согласованы специальным образом: число столбцов левой матрицы должно совпадать с числом строк правой. Это означает, что оба произведения AB и BA определены тогда и только тогда, когда A и B имеют размеры $m \times n$ и $n \times m$ соответственно. При этом размеры матриц AB и BA совпадают лишь при $m = n$. Следовательно, равенство $AB = BA$ возможно лишь для квадратных матриц одинакового порядка. Однако и в этом случае произведение матриц, вообще говоря, может зависеть от порядка сомножителей:

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

Матрицы A и B, для которых $AB = BA$, называются *перестановочными* или *коммутирующими*. Очевидно, что на множестве $\mathbb{R}^{n\times n}$

1) нулевая и единичная матрицы перестановочны с любой другой матрицей;

2) любые две диагональные матрицы перестановочны.

Теорема 2.3. *Операция умножения матриц обладает следующими свойствами:*

1) $(AB)C = A(BC)$ *(свойство ассоциативности),*

2) $\alpha(AB) = (\alpha A)B = A(\alpha B)$, $\quad \forall \alpha \in \mathbb{R}$,

3) $A(B+C) = AB+BC$, $(A+B)C = AC+BC$ *(свойство дистрибутивности),*

выполненными для любых матриц A, B, C, *для которых левые части равенств имеют смысл.*

Доказательство. Эти свойства проверяются непосредственно. Остановимся подробнее на свойстве 1. Так как произведение $(AB)C$ имеет смысл, то размеры матриц A, B и C согласованы соответствующим образом. Пусть $A = (a_{ij}) \in \mathbb{R}^{m\times n}$, $B = (b_{ij}) \in \mathbb{R}^{n\times k}$, $C = (c_{ij}) \in \mathbb{R}^{k\times p}$. Тогда и произведение $A(BC)$ имеет смысл, при этом размеры обеих матриц $(AB)C$ и $A(BC)$ совпадают. Проверим совпадение элементов этих матриц, расположенных в одинаковых позициях. Обозначим $AB = U = (u_{ij})$, $BC = V = (v_{ij})$, $(AB)C = S = (s_{ij})$, $A(BC) = T = (t_{ij})$. Тогда для любых $i = \overline{1,m}$, $j = \overline{1,p}$

$$s_{ij} = \sum_{q=1}^{k} u_{iq}c_{qj} = \sum_{q=1}^{k} \left(\sum_{r=1}^{n} a_{ir}b_{rq}\right) c_{qj},$$

$$t_{ij} = \sum_{r=1}^{n} a_{ir}v_{rj} = \sum_{r=1}^{n} a_{ir} \left(\sum_{q=1}^{k} b_{rq}c_{qj}\right).$$

Числа c_{qj} в первой двойной сумме (a_{ir} – во второй) не зависят от индекса суммирования r (соответственно q), т.е. являются общими множителями слагаемых внутренней суммы. Следовательно, c_{qj} (соответственно a_{ir}) можно внести под знак суммирования, так что

$$s_{ij} = \sum_{q=1}^{k} \left(\sum_{r=1}^{n} a_{ir}b_{rq}c_{qj}\right) = \sum_{q=1}^{k}\sum_{r=1}^{n} a_{ir}b_{rq}c_{qj},$$

$$t_{ij} = \sum_{r=1}^{n} \left(\sum_{q=1}^{k} a_{ir}b_{rq}c_{qj}\right) = \sum_{r=1}^{n}\sum_{q=1}^{k} a_{ir}b_{rq}c_{qj}.$$

Отсюда и из свойств двойной суммы следует, что $s_{ij} = t_{ij}$. ∎

Транспонирование матрицы. Пусть $A = (a_{ij}) \in \mathbb{R}^{m\times n}$. Матрица $A^T = (a^t_{ij}) \in \mathbb{R}^{n\times m}$ называется *транспонированной* к матрице A, если

$$a^t_{ij} = a_{ji}, \quad i = \overline{1,n},\ j = \overline{1,m}.$$

Для обозначения транспонированной матрицы используются также символы A^t и A'.

Переход от матрицы A к A^T называется *транспонированием матрицы* A. Заметим, что при транспонировании матрицы A ее строки становятся столбцами A^T с теми же номерами, а столбцы – строками. Другими словами, транспонирование – это вращение матрицы в пространстве на 180° вокруг главной диагонали.

Теорема 2.4. *Операция транспонирования матриц обладает следующими свойствами:*

1) $(A+B)^T = A^T + B^T$,

2) $(\alpha A)^T = \alpha A^T$, $\quad \forall \alpha \in \mathbb{R}$,

3) $(AB)^T = B^T A^T$,

4) $(A^T)^T = A$,

выполненными для всех матриц A, B, *для которых имеют смысл левые части равенств.*

Доказательство. Проверим свойство 3, остальные непосредственно вытекают из определения. Положим $A = (a_{ij}) \in \mathbb{R}^{m\times n}$, $B = (b_{ij}) \in \mathbb{R}^{n\times k}$, произведение AB при этом имеет смысл. Нетрудно проверить, что произведение $B^T A^T$ также имеет смысл, при этом размеры матриц $(AB)^T$ и $B^T A^T$ совпадают. Элементы этих матриц, стоящие в одинаковых позициях, равны, так как

$$\{(AB)^T\}_{ij} = \{AB\}_{ji} = \sum_{s=1}^{n} a_{js}b_{si} = \sum_{s=1}^{n} b^t_{is}a^t_{sj} = \{B^T A^T\}_{ij}. \ ∎$$

Некоторые свойства операций. Приведем несколько полезных фактов, касающихся операций над матрицами. Доказательство этих свойств предоставляется читателю.

1. Операции сложения, умножения на число и умножения блочных матриц совершаются по тем же правилам, по которым они совершаются с обычными числовыми матрицами:

а) если блочные матрицы $A = (A_{ij})$, $B = (B_{ij})$ имеют одинаковый размер и одинаковым образом разбиты на клетки, то сумме матриц A и B при том же разбиении на клетки отвечает блочная матрица $C = (C_{ij})$ с элементами $C_{ij} = A_{ij} + B_{ij}$;

б) произведению αA отвечает блочная матрица $C = (C_{ij})$ с элементами $C_{ij} = \alpha A_{ij}$;

в) если $A = (A_{ij})$ и $B = (B_{ij})$ – блочные матрицы, у которых число столбцов блока A_{is} равно числу строк блока B_{sj} при любых i, s, j, то произведению AB соответствует блочная матрица $C = (C_{ij})$ с элементами

$$C_{ij} = \sum_{s} A_{is}B_{sj}.$$

2. Вектор-столбец $e_i = (0 \ldots 0\,1\,0 \ldots 0)^T \in \mathbb{R}^{n\times 1}$ и вектор-строку $e'_i = (0 \ldots 0\,1\,0 \ldots 0) \in \mathbb{R}^{1\times m}$ (у которых все компоненты равны 0, кроме i-й, равной 1) будем называть *i-м единичным столбцом* и *i-й единичной строкой*. Если $A = (a_{ij}) \in \mathbb{R}^{m\times n}$, то в обозначениях (1.1)

$$Ae_i = a_i, \qquad e'_i A = a'_i. \tag{2.1}$$

3. Если $A = (a_{ij}) \in \mathbb{R}^{m\times n}$, а $b' = (\alpha_1\,\alpha_2 \ldots \alpha_m)$ и $b = (\alpha_1\,\alpha_2 \ldots \alpha_n)^T$ – вектор-строка и вектор-столбец соответственно, то

$$Ab = \sum_{i=1}^{n} \alpha_i a_i, \qquad b'A = \sum_{i=1}^{m} \alpha_i a'_i. \tag{2.2}$$

4. Если $A = (a_{ij}) \in \mathbb{R}^{m\times n}$, $B = (b_{ij}) \in \mathbb{R}^{n\times k}$, то

$$AB = [\ Ab_1 \quad Ab_2 \quad \ldots \quad Ab_k\ ]\,, \qquad AB = \begin{bmatrix} a_1'B \\ a_2'B \\ \vdots \\ a_m'B \end{bmatrix}, \tag{2.3}$$

т.е. ***столбцы произведения** AB **суть линейные комбинации столбцов матрицы** A, **а строки – линейные комбинации строк** B.*

§3. Элементарные преобразования матриц

Приведение матрицы к ступенчатой форме. *Элементарными преобразованиями матрицы* называются преобразования следующих типов:

1) перестановка двух строк (столбцов) матрицы;

2) умножение строки (столбца) матрицы на число, отличное от нуля;

3) прибавление к одной строке (столбцу) матрицы другой ее строки (соответственно столбца), умноженной на любое число.

Теорема 3.1 (об основном процессе). *Произвольная ненулевая матрица конечным числом элементарных преобразований только строк первого и третьего типов может быть приведена к верхней ступенчатой форме.*

Доказательство. Пусть $A = (a_{ij}) \in \mathbb{R}^{m\times n}$, $A \neq O$. Процесс приведения этой матрицы к верхней ступенчатой форме состоит в общем случае из $k = \min(m,n)$ шагов. Иногда, как это будет видно ниже, он обрывается раньше, давая нужный результат.

Первый шаг. а) Так как $A \neq O$, то в ней должен быть хотя бы один ненулевой столбец. Пусть k_1 – номер *первого* из них. В k_1-м столбце существует хотя бы один ненулевой элемент a_{ik_1}. Если $a_{1k_1} \neq 0$, то переходим к п. "б". Если же $a_{1k_1} = 0$, то, поменяв местами 1-ю и i-ю строки (т.е. выполнив элементарное преобразование *строк* первого типа), получим матрицу

$$\begin{bmatrix} 0 & \ldots & 0 & a_{1k_1} & \ldots & a_{1n} \\ 0 & \ldots & 0 & a_{2k_1} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & a_{mk_1} & \ldots & a_{mn} \end{bmatrix},$$

в которой $a_{1k_1} \neq 0$. Хотя при таком переходе элементы матрицы A могли измениться, мы оставили прежние обозначения с тем, чтобы акцентировать внимание лишь на тактической стороне процесса.

б) Элемент a_{1k_1} назовем *ведущим (главным)* элементом первого шага. С его помощью аннулируем все расположенные под ним элементы k_1-го столбца. Для этого из всех строк, начиная со второй, вычтем первую строку, умноженную на a_{2k_1}/a_{1k_1}, a_{3k_1}/a_{1k_1}, $\ldots$,

a_{mk_1}/a_{1k_1} соответственно (т.е. выполним элементарные преобразования *строк* третьего типа).

После выполнения первого шага матрица A переходит в матрицу

$$\left[\begin{array}{cccc|ccc} 0 & \ldots & 0 & a_{1k_1} & a_{1,k_1+1} & \ldots & a_{1n} \\ 0 & \ldots & 0 & 0 & a_{2,k_1+1} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & 0 & a_{m,k_1+1} & \ldots & a_{mn} \end{array}\right],$$

в которой *первая* строка является *первой* строкой строящейся верхней ступенчатой матрицы. Если при этом все строки, начиная со второй, стали нулевыми, то весь процесс заканчивается, так как матрица уже приведена к верхней ступенчатой форме. Если же в этих строках есть хотя бы один ненулевой элемент, т.е. если матрица

$$A_1 = \begin{bmatrix} a_{2,k_1+1} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots \\ a_{m,k_1+1} & \ldots & a_{mn} \end{bmatrix} \neq O\,,$$

то переходим ко второму шагу.

Второй шаг. Второй шаг аналогичен первому. Он состоит в применении к матрице A_1 процедуры описанного выше первого шага. При этом можно считать, что выполняются элементарные преобразования строк всей матрицы A, так как нулевые элементы этих строк, расположенные в первых k_1 столбцах, при элементарных преобразованиях строк не изменяются. В результате *второго* шага уже и *вторая* строка матрицы A станет *второй* строкой строящейся верхней ступенчатой матрицы.

Переход к следующему шагу аналогичен уже известному переходу от первого шага ко второму. Повторяя описанные преобразования на следующих шагах, самое большее через $k = \min(m,n)$ шагов мы получим требуемый результат. Отметим, что ведущим элементом i-го шага является первый ненулевой элемент i-й строки, т.е. a_{ik_i}.

Описанный здесь процесс будем называть ***основным процессом*** приведения матрицы к ступенчатой форме. ■

Замечания к основному процессу. 1. Основной процесс с незначительной модификацией может быть использован и для приведения матрицы к верхней трапециевидной форме. Для этого нужно привлечь еще и элементарные преобразования столбцов: переставить в ступенчатой матрице столбцы так, чтобы k_i-й столбец оказался на i-м месте. Итак, произвольная ненулевая матрица элементарными преобразованиями строк и перестановками столбцов может быть приведена к верхней трапециевидной форме.

2. Квадратная матрица с помощью основного процесса приводится к треугольной форме.

3. Если в основном процессе поменять ролями строки и столбцы, то матрица A приведется к нижней ступенчатой (трапециевидной) форме.

4. В ручных вычислениях во избежание больших чисел целесообразно в основном процессе использовать элементарные преобразования строк второго типа: сокращать все элементы на общий множитель.

5. Во избежание дробных чисел в ручных вычислениях удобно также в качестве ведущего элемента выбирать элемент, равный 1. Если такого элемента нет, то, как правило, его можно получить, используя элементарные преобразования строк и перестановки столбцов.

6. Идеи основного процесса используются во многих компьютерных алгоритмах вычислительной алгебры. Выбор ведущего элемента здесь представляет собой особую проблему, так как от этого зависит точность вычислений. Исследование

этой проблемы выходит за рамки данной книги; отметим лишь, что ведущий элемент не должен быть "маленьким". Именно этим определяется многообразие алгоритмов, реализующих основной процесс, с различными стратегиями выбора ведущего элемента.

Матрицы элементарных преобразований. Элементарные преобразования матрицы просты и удобны в матричных исследованиях. Однако словесное описание выполняемых преобразований весьма утомительно как само по себе, так и для его восприятия. Этого можно избежать, если ввести некоторые матрицы специального вида.

Матрицами элементарных преобразований называются квадратные матрицы D_i, P_{ij}, L_{ij} вида

$$D_i = \begin{array}{cc} \left\| \begin{array}{ccccccc} 1 & & & \vdots & & & \\ & \ddots & & \vdots & & & \\ & & 1 & \vdots & & & \\ \cdots & \cdots & \cdots & \alpha & \cdots & \cdots & \cdots \\ & & & \vdots & 1 & & \\ & & & \vdots & & \ddots & \\ & & & \vdots & & & 1 \end{array} \right\| & \begin{array}{c} \\ \\ \\ i \\ \\ \\ \\ \end{array} \\ \begin{array}{ccccccc} & & & i & & & \end{array} & \end{array}, \quad \alpha \neq 0,$$

$$P_{ij} = \begin{array}{cc} \left\| \begin{array}{ccccccccc} 1 & & & \vdots & & \vdots & & & \\ & \ddots & & \vdots & & \vdots & & & \\ & & 1 & \vdots & & \vdots & & & \\ \cdots & \cdots & \cdots & 0 & \cdots & 1 & \cdots & \cdots & \cdots \\ & & & \vdots & \ddots & \vdots & & & \\ \cdots & \cdots & \cdots & 1 & \cdots & 0 & \cdots & \cdots & \cdots \\ & & & \vdots & & \vdots & 1 & & \\ & & & \vdots & & \vdots & & \ddots & \\ & & & \vdots & & \vdots & & & 1 \end{array} \right\| & \begin{array}{c} \\ \\ \\ i \\ \\ j \\ \\ \\ \\ \end{array} \\ \begin{array}{ccccccccc} & & & i & & j & & & \end{array} & \end{array},$$

$$L_{ij} = \begin{array}{cc} \left\| \begin{array}{ccccccccc} 1 & & & \vdots & & \vdots & & & \\ & \ddots & & \vdots & & \vdots & & & \\ & & 1 & \vdots & & \vdots & & & \\ \cdots & \cdots & \cdots & 1 & \cdots & 0 & \cdots & \cdots & \cdots \\ & & & \vdots & \ddots & \vdots & & & \\ \cdots & \cdots & \cdots & \beta & \cdots & 1 & \cdots & \cdots & \cdots \\ & & & \vdots & & \vdots & 1 & & \\ & & & \vdots & & \vdots & & \ddots & \\ & & & \vdots & & \vdots & & & 1 \end{array} \right\| & \begin{array}{c} \\ \\ \\ i \\ \\ j \\ \\ \\ \\ \end{array} \\ \begin{array}{ccccccccc} & & & i & & j & & & \end{array} & \end{array}, \quad \forall \beta \in \mathbb{R}, \qquad (3.1)$$

в которых все диагональные элементы, кроме указанных, равны 1, а все внедиагональные элементы, кроме указанных, равны 0.

Теорема 3.2. *Умножение матрицы A на матрицы элементарных преобразований P_{ij}, D_i, L_{ij} справа равносильно элементарным преобразованиям столбцов матрицы A первого, второго и третьего типов соответственно, а умножение слева на матрицы P_{ij}, D_i, L_{ij}^T – аналогичным элементарным преобразованиям строк.*

Доказательство. Докажем одно из сформулированных утверждений: умножение матрицы A на L_{ij} справа равносильно прибавлению к i-му столбцу матрицы A ее j-го столбца, умноженного на β. Действительно, пусть $l_1, l_2, \ldots, l_n$ – столбцы матрицы L_{ij}. Тогда, как следует из (2.3), (2.2) и (2.1),

$$AL_{ij} = [\ Al_1 \ \ \ldots \ \ Al_n\ ] = \{l_k = e_k \text{ при } k \neq i\} =$$
$$= [a_1 \ldots a_{i-1}\ Al_i\ a_{i+1} \ldots a_n] = \{Al_i = a_i + \beta a_j\} =$$
$$= [a_1 \ldots a_{i-1}\ (a_i + \beta a_j)\ a_{i+1} \ldots a_n]. \ \blacksquare$$

Итак, с помощью матриц элементарных преобразований все элементарные преобразования матрицы могут быть записаны весьма лаконично: $P_{ij}A$, D_iA, L_{ij}^TA, AP_{ij}, AD_i, AL_{ij}.

В свете доказанной теоремы можно по-иному сформулировать теорему 3.1: *для любой ненулевой матрицы A существуют матрицы элементарных преобразований $L_1, \ldots, L_k$ такие, что произведение $L_k \ldots L_1 A$ имеет верхнюю ступенчатую форму.*

§4. Определители

Перестановки. Упорядоченная совокупность чисел α_1, α_2, …, α_n, в которой

1) $\alpha_i \in \{1, 2, \ldots, n\}$, $\quad i = \overline{1, n}$;
2) $\alpha_i \neq \alpha_j$ при $i \neq j$,

называется ***перестановкой*** из чисел $1, 2, \ldots, n$.

Перестановка $1, 2, \ldots, n$ называется ***натуральной***.

Замечание. Аналогично рассматриваются перестановки из n произвольных символов: достаточно перенумеровать эти символы и иметь дело с их номерами $1, 2, \ldots, n$[1].

Преобразование перестановки, при котором два ее числа α_i и α_j с номерами $i \neq j$ меняются местами, называется ***транспозицией***.

Говорят, что два числа α_i и α_j в перестановке $\alpha_1, \alpha_2, \ldots, \alpha_n$ образуют ***инверсию (беспорядок)***, если большее из них предшествует меньшему, т.е. если $\alpha_i > \alpha_j$ при $i < j$, и ***порядок*** – в противном случае, т.е. если $\alpha_i < \alpha_j$ при $i < j$. Перестановка называется ***четной***, если общее число инверсий в ней четно, и ***нечетной*** – в противном случае. Общее число инверсий в перестановке $\alpha_1, \alpha_2, \ldots, \alpha_n$ обозначается символами $\sigma(\alpha_1, \alpha_2, \ldots, \alpha_n)$ или $\sigma(\alpha)$.

Теорема 4.1. *Число всевозможных перестановок из n чисел равно $n!$.*

Доказательство. Переберем все перестановки из n чисел. В качестве α_1 можно взять любое из этих чисел. Это дает n возможностей. В каждой из них α_1 уже выбрано и в качестве α_2 можно выбрать любое из $n-1$ оставшихся чисел. Это означает, что число различных способов выбрать α_1 и α_2 равно $n(n-1)$. Продолжая эти рассуждения, получим, что число различных способов выбрать $\alpha_1, \alpha_2, \ldots, \alpha_n$ равно $n(n-1) \cdot \ldots \cdot 2 \cdot 1 = n!$. ■

Теорема 4.2. *Каждая транспозиция меняет четность перестановки.*

Доказательство. 1. Пусть в перестановке $\alpha_1, \alpha_2, \ldots, \alpha_n$ меняются местами два соседних числа α_i и α_{i+1}, т.е.

[1] В § 8 будет дано общее определение перестановки n-го порядка.

$$\ldots,\alpha_i,\alpha_{i+1},\ldots \quad\longmapsto\quad \ldots,\alpha_{i+1},\alpha_i,\ldots$$

(здесь многоточия заменяют числа, которые не затрагивались при транспозиции). Очевидно, что в обеих перестановках числа, оставшиеся на местах, составляют одни и те же инверсии друг с другом и с α_i, α_{i+1}. Если числа α_i и α_{i+1} раньше составляли инверсию, то в новой перестановке она пропадает; если же они не составляли инверсию, то теперь появится одна новая инверсия. Таким образом, общее число инверсий в новой перестановке отличается от старой на единицу, т.е. четность при такой транспозиции меняется.

2. Пусть теперь между переставляемыми числами α_i и α_j расположено k чисел, т.е.

$$\ldots,\alpha_i,\alpha_{i+1},\ldots,\alpha_{i+k},\alpha_j\ldots \quad\longmapsto\quad \ldots,\alpha_j,\alpha_{i+1},\ldots,\alpha_{i+k},\alpha_i\ldots$$

Новую перестановку можно получить из старой, последовательно меняя местами соседние числа: α_i поменять местами $(k+1)$ раз с соседними числами $\alpha_{i+1},\alpha_{i+2},\ldots,\alpha_{i+k},\alpha_j$, а затем α_j поменять местами k раз с числами $\alpha_{i+k},\ldots,\alpha_{i+2}$, α_{i+1}. При этом четность перестановки изменится $2k+1$ раз. Следовательно, и при такой транспозиции четность перестановки меняется. ■

Теорема 4.3. *Все $n!$ перестановок из n чисел могут быть упорядочены так, чтобы каждая последующая отличалась от предыдущей на одну транспозицию, причем начинать это упорядочение можно с любой перестановки.*

Доказательство. Проведем индукцию по n. Для $n=2$ утверждение теоремы легко проверить: $(1,2),(2,1)$ и $(2,1),(1,2)$.

Пусть утверждение теоремы верно для $n-1$ чисел. Докажем его для n чисел. Пусть первая перестановка имеет вид $\alpha_1,\alpha_2,\ldots,\alpha_n$.

а) Сначала упорядочим все перестановки, начинающиеся с α_1. Таких перестановок $(n-1)!$, и по индуктивному предположению они могут быть упорядочены нужным образом, начиная с перестановки $\alpha_1,\alpha_2,\ldots,\alpha_n$, так как это сводится к упорядочению перестановок из $n-1$ чисел, начиная с $\alpha_2,\alpha_3,\ldots,\alpha_n$.

б) Далее в последней перестановке из этого списка сделаем одну транспозицию, поменяв местами числа α_1 и α_2. И снова упорядочим все перестановки, начинающиеся с α_2, и т.д. ■

Следствие 1. *При $n\geq 2$ число четных перестановок равно числу нечетных.* Действительно, после упорядочения в списке всех перестановок четные и нечетные перестановки будут чередоваться, а так как $n!$ четно при $n\geq 2$, то количества четных и нечетных перестановок совпадают и равны $n!/2$.

Следствие 2. *От каждой перестановки из n чисел можно перейти к любой другой перестановке из этих же чисел при помощи конечного числа транспозиций.*

Теорема 4.4. *Если $\alpha_1,\alpha_2,\ldots,\alpha_n$ – перестановка из первых n натуральных чисел с числом инверсий s, то после преобразования ее в натуральную перестановку индексные номера $1,2,\ldots,n$ образуют новую перестановку с тем же числом инверсий s.*

Доказательство. Рассмотрим в перестановке

$$\alpha_1,\alpha_2,\ldots,\alpha_i,\ldots,\alpha_j,\ldots,\alpha_n \tag{4.1}$$

любые ее два числа α_i и α_j.

Числа α_i и α_j образуют либо инверсию ($\alpha_i>\alpha_j$, $i<j$), либо порядок ($\alpha_i<\alpha_j$, $i<j$). После преобразования перестановки (4.1) в натуральную числа α_i и α_j будут располагаться следующим образом:

$$1,2,\ldots,\alpha_j,\ldots,\alpha_i,\ldots,\alpha_n \quad \text{в случае инверсии,}$$
$$1,2,\ldots,\alpha_i,\ldots,\alpha_j,\ldots,\alpha_n \quad \text{в случае порядка,}$$

причем в обоих случаях $i<j$. Это означает, что числа α_i и α_j в перестановке (4.1) и их индексы i и j в перестановке индексных номеров одновременно образуют либо инверсию, либо порядок. Следовательно, обе эти перестановки имеют одинаковое число инверсий s. ■

Построение определителя n-го порядка. Пусть $A=(a_{ij})$ – квадратная матрица n-го порядка. Рассмотрим произведение элементов этой матрицы, взятых по одному из каждой строки и каждого столбца:

$$a_{1\alpha_1}a_{2\alpha_2}\ldots a_{n\alpha_n}. \tag{4.2}$$

Заметим, что в этом произведении сомножители упорядочены в порядке возрастания номеров строк, при этом номера столбцов α_1, α_2, …, α_n образуют перестановку из чисел $1,2,\ldots,n$, так как $\alpha_i\in\{1,2,\ldots,n\}$ и $\alpha_i\neq\alpha_j$ при $i\neq j$. Произведений вида (4.2) в матрице A столько, сколько существует перестановок $\alpha_1,\alpha_2,\ldots,\alpha_n$ из n чисел, т.е. $n!$.

Определителем квадратной матрицы $A=(a_{ij})$ *n-го порядка* называется сумма всевозможных произведений $a_{1\alpha_1}a_{2\alpha_2}\ldots a_{n\alpha_n}$ элементов матрицы, взятых по одному из каждой строки и каждого столбца, причем если сомножители в этом произведении упорядочены в порядке возрастания номеров строк, то оно берется со знаком $(-1)^{\sigma(\alpha_1,\alpha_2,\ldots,\alpha_n)}$.

Для обозначения определителя приняты символы $|A|$, $\det A$. Итак,

$$\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n}\\ a_{21} & a_{22} & \ldots & a_{2n}\\ \ldots & \ldots & \ldots & \ldots\\ a_{n1} & a_{n2} & \ldots & a_{nn}\end{vmatrix} = \sum_{\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_n)} (-1)^{\sigma(\alpha)}a_{1\alpha_1}a_{2\alpha_2}\ldots a_{n\alpha_n}, \tag{4.3}$$

где суммирование ведется по всевозможным перестановкам $(\alpha_1,\alpha_2,\ldots,\alpha_n)$ из чисел $1,2,\ldots,n$.

Каждое произведение в сумме (4.3) называется *членом определителя*, а число $(-1)^{\sigma(\alpha)}$ – его *знаком*.

Из свойств перестановки следует, что число всевозможных членов определителя n-го порядка равно $n!$ и что при $n\geq 2$ число положительных членов равно числу отрицательных и равно $n!/2$.

Замечание 1. Свойства 4 и 5 часто объединяют, называя их *свойством линейности определителя* относительно строк и столбцов.

Свойство 6. *При перестановке местами двух строк (столбцов) матрицы ее определитель меняет знак.*

Доказательство. Пусть в матрице $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ переставляются i-й и j-й столбцы и B – результат этой перестановки:

$$A = [a_1 \dots a_i \dots a_j \dots a_n], \quad B = [a_1 \dots a_j \dots a_i \dots a_n].$$

Очевидно, что определители матриц A и B состоят из одних и тех же членов. Сравним их знаки. Члену $a_{\alpha_1 1} \dots a_{\alpha_i i} \dots a_{\alpha_j j} \dots a_{\alpha_n n}$ в $|A|$ соответствует перестановка $\alpha_1, \dots, \alpha_i, \dots, \alpha_j, \dots, \alpha_n$, а в $|B|$ – перестановка $\alpha_1, \dots, \alpha_j, \dots, \alpha_i, \dots, \alpha_n$. Эти перестановки отличаются друг от друга одной транспозицией, т.е. имеют разную четность. Отсюда следует, что все члены $|A|$ входят в $|B|$ с обратным знаком. Это означает, что $|A| = -|B|$. ■

Замечание 2. Отметим, что свойство 6 относится к случаю, когда переставляются строки (столбцы) с разными номерами.

Свойство 7. *Определитель матрицы, имеющей две одинаковые строки (столбца), равен нулю.*

Утверждение вытекает из свойства 6: достаточно в матрице поменять местами одинаковые строки, тогда $|A| = -|A| = 0$. ■

Свойство 8. *Если одна строка (столбец) матрицы является линейной комбинацией других ее строк (столбцов), то определитель матрицы равен нулю.*

Утверждение вытекает из свойства линейности определителя и свойства 7. ■

Свойство 9. *Если к какой-либо строке (столбцу) матрицы прибавить линейную комбинацию других ее строк (столбцов), то ее определитель не изменится.*

Утверждение вытекает из свойств 5 и 8. ■

Миноры и алгебраические дополнения. Пусть $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ и $k \in \mathbb{N}, 1 \le k \le \min(m,n)$. Выберем в матрице A произвольные k строк и k столбцов с номерами $i_1 < i_2 < \dots < i_k$ и $j_1 < j_2 < \dots < j_k$ соответственно. Элементы матрицы A, стоящие на пересечении выбранных строк и столбцов, образуют квадратную матрицу k-го порядка. Определитель этой матрицы называется *минором* k-го порядка матрицы A, расположенным в строках с номерами $i_1, i_2, \dots, i_k$ и столбцах с номерами $j_1, j_2, \dots, j_k$. Для обозначения миноров приняты символы $M^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}$, $M\binom{i_1 i_2 \dots i_k}{j_1 j_2 \dots j_k}$, M_k, M. Итак,

$$M^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k} = \begin{vmatrix} a_{i_1 j_1} & \dots & a_{i_1 j_k} \\ \dots & \dots & \dots \\ a_{i_k j_1} & \dots & a_{i_k j_k} \end{vmatrix}.$$

Пусть теперь $A = (a_{ij})$ – квадратная матрица n-го порядка и $M^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}$ – ее минор. Если вычеркнуть в матрице A строки и столбцы, в которых расположен заданный минор, то оставшиеся элемен-

ты матрицы A образуют квадратную матрицу $(n-k)$-го порядка. Определитель этой матрицы называется *дополнительным минором к минору* $M^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}$. Дополнительный минор обозначается символами $\overline{M}^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}$, $\overline{M}$, M^σ. Очевидно, что исходный минор является дополнительным к своему дополнительному минору. Дополнительный минор к минору $M^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}$, взятый со знаком $(-1)^{\sum_{p=1}^{k}(i_p + j_p)}$, называется *алгебраическим дополнением* к минору $M^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}$ и обозначается символом $A^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}$. Итак,

$$A^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k} = (-1)^{i_1+i_2+\dots+i_k+j_1+j_2+\dots+j_k} \overline{M}^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}.$$

Теорема 4.5 (теорема Лапласа). *Пусть* $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ *и* $k \in \mathbb{N}$, $1 \le k \le n-1$. *Пусть в матрице* A *выбраны произвольные* k *строк (или столбцов). Тогда определитель матрицы* A *равен сумме всевозможных произведений миноров* k*-го порядка, расположенных в выбранных строках (соответственно столбцах), на их алгебраические дополнения.*

Доказательство. Пусть выбраны строки с номерами $i_1 < i_2 < \dots < i_k$. Следует доказать, что

$$\det A = \sum_{(j_1, j_2, \dots, j_k)} M^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k} A^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}, \tag{4.6}$$

где суммирование ведется по всевозможным значениям $j_1, j_2, \dots, j_k$ $(1 \le j_1 < j_2 < \dots < j_k \le n)$.

Для доказательства рассмотрим подробнее правую часть требуемого равенства (4.6). Минор $M^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}$, как определитель k-го порядка, представляет собой сумму произведений k элементов матрицы A. Точно так же и алгебраическое дополнение $A^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}$ является суммой произведений $n-k$ элементов матрицы A. Значит, произведение $M^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k} A^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}$, а следовательно, и вся правая часть (4.6) представляет собой сумму произведений n элементов матрицы A. Обозначим эту сумму через S и покажем, что она совпадает с $\det A$ как с суммой (4.3) членов определителя с соответствующими знаками.

1. Сначала покажем, что произведение $M^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k} A^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}$ представляет собой некоторую сумму членов $\det A$, причем с теми же знаками, с какими они входят в $\det A$.

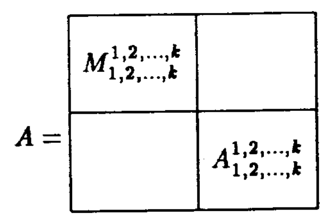


Рис. 1

а) В простейшем случае, когда минор $M^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}$ находится в левом верхнем углу матрицы A (рис. 1), дополнительный минор будет занимать правый нижний угол, при этом он будет совпадать с алгебраическим дополнением, так как $(-1)^{(1+2+\dots+k)+(1+2+\dots+k)} = 1$.

Возьмем произвольный член минора $M^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k}$: $(-1)^{\sigma(\alpha)} a_{1\alpha_1} a_{2\alpha_2} \cdot \ldots \cdot a_{k\alpha_k}$ и произвольный член дополнительного к нему минора:

$$(-1)^{\sigma(\beta)} a_{k+1,\beta_{k+1}} a_{k+2,\beta_{k+2}} \ldots a_{n\beta_n}.$$

Тогда произведение $M^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k} A^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k}$ есть сумма произведений вида

$$(-1)^{\sigma(\alpha)+\sigma(\beta)} a_{1\alpha_1} a_{2\alpha_2} \ldots a_{k\alpha_k} a_{k+1,\beta_{k+1}} \ldots a_{n\beta_n}.$$

В этом произведении все сомножители стоят в разных строках и разных столбцах матрицы A, следовательно, оно будет членом $\det A$. Знак этого члена равен $(-1)^{\sigma(\alpha_1,\ldots,\alpha_k,\beta_{k+1},\ldots,\beta_n)}$, но $\sigma(\alpha_1,\ldots,\alpha_k,\beta_{k+1},\ldots,\beta_n) = \sigma(\alpha) + \sigma(\beta)$, так как никакое α_i ни с каким β_j не образует инверсий: все $\alpha_i \le k$, а все $\beta_j > k$. Таким образом, произведение $M^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k} A^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k}$ представляет собой некоторую сумму членов $\det A$, причем с теми же знаками, с какими они входят в $\det A$.

б) Общий случай минора $M^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k}$ сводится к рассмотренному следующим образом. Будем переставлять i_1-ю строку матрицы A последовательно со всеми предыдущими до тех пор, пока она не займет место первой строки. Затем точно так же будем переставлять i_2-ю, ..., i_k-ю строки до тех пор, пока они не займут места второй, ..., k-й строк соответственно. Аналогично переставляются j_1-й, j_2-й, ..., j_k-й столбцы до тех пор, пока они не займут места первого, второго, ..., k-го столбцов. При этом всего будет выполнено

$$\begin{aligned}(i_1 - 1) + \ldots + (i_k - k) + (j_1 - 1) + \ldots + (j_k - k) &= \\ = (i_1 + \ldots + i_k) + (j_1 + \ldots + j_k) - 2(1 + \ldots + k)\end{aligned} \tag{4.7}$$

перестановок строк и столбцов матрицы A. В результате этих перестановок матрица A преобразуется в матрицу B, в которой рассматриваемый минор $M^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k}$ матрицы A займет левый верхний угол. Так как при указанных преобразованиях взаимное расположение строк и столбцов дополнительного минора не изменилось, то дополнительный минор к минору $M^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k}$ в матрице A остается дополнительным к нему и в матрице B. Из п. "а" следует, что произведение $M^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k} \overline{M}^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k}$ является суммой членов $\det B$ с теми же знаками, с какими они входят в $\det B$. Но, согласно свойству 6 определителя и (4.7), $\det A = (-1)^{(i_1+\ldots+i_k)+(j_1+\ldots+j_k)} \det B$. Следовательно, слагаемые произведения $(-1)^{(i_1+\ldots+i_k)+(j_1+\ldots+j_k)} M^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k} \overline{M}^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k}$ входят в $\det A$ со своими знаками.

2. Из доказанного в п. 1 следует, что вся сумма S представляет собой некоторую сумму членов $\det A$ со своими знаками.

3. Покажем теперь, что в сумму S входят все члены $\det A$. Пусть

$$a_{1\alpha_1} a_{2\alpha_2} \ldots a_{n\alpha_n} \tag{4.8}$$

– произвольный член $\det A$. В этом произведении соберем отдельно сомножители, расположенные в строках с номерами $i_1, i_2, \ldots, i_k$:

$$a_{i_1\alpha_{i_1}} a_{i_2\alpha_{i_2}} \ldots a_{i_k\alpha_{i_k}}. \tag{4.9}$$

Они расположены в различных столбцах с номерами $\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_k}$. Эти номера однозначно определяются заданием члена (4.8). Обозначим через M минор k-го порядка матрицы A, расположенный в строках с номерами $i_1, i_2, \ldots, i_k$ и столбцах с номерами $\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_k}$. Тогда произведение (4.9) будет членом этого минора, а произведение остальных сомножителей (4.8), не вошедших в (4.9), – членом дополнительного минора $\overline{M}$ к минору M. Итак, любой член определителя матрицы A может быть получен умножением определенного этим членом минора M на дополнительный к нему минор $\overline{M}$. Из доказанного в п. 1 следует, что при умножении минора M на его алгебраическое дополнение получится член (4.8) с его знаком.

4. Осталось доказать, что каждый член $\det A$ входит в сумму S ровно один раз и что других слагаемых в этой сумме нет. Для этого достаточно показать, что количество слагаемых в сумме S равно $n!$. Действительно, минор $M^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k}$ состоит из $k!$ слагаемых, алгебраическое дополнение к нему – из $(n-k)!$ слагаемых. Значит, произведение $M^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k} A^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k}$ состоит из $k!(n-k)!$ членов определителя матрицы A. Таких произведений в сумме S столько, сколько существует миноров k-го порядка в строках $i_1, i_2, \ldots, i_k$, т.е. столько, сколько существует способов выбрать k столбцов $j_1, j_2, \ldots, j_k$ из n столбцов матрицы A, или C_n^k. Таким образом, количество слагаемых в сумме S равно $C_n^k k!(n-k)! = n!$. ■

Разложение определителя по строке (столбцу). Если в теореме Лапласа выбрать $k = 1$ и строку (столбец) с номером i, то минорами первого порядка, расположенными в i-й строке (столбце), будут сами элементы a_{ij} (a_{ji}). Обозначив через A_{ij} алгебраическое дополнение к элементу a_{ij}, получим из теоремы Лапласа, что

$$\det A = \sum_{j=1}^{n} a_{ij} A_{ij} \quad (\text{или} \quad \det A = \sum_{j=1}^{n} a_{ji} A_{ji}). \tag{4.10}$$

Представление определителя (4.10) называется *разложением определителя по i-й строке* (соответственно *по i-му столбцу*). Итак, *определитель матрицы равен сумме всех произведений элементов произвольной ее строки (столбца) на свои алгебраические дополнения.*

Определитель квазитреугольной матрицы.

Теорема 4.6. *Определитель квазитреугольной матрицы равен произведению определителей диагональных клеток.*

Доказательство. Пусть A – верхняя квазитреугольная матрица m-го порядка (рис. 2). Применим теорему Лапласа к группе из k

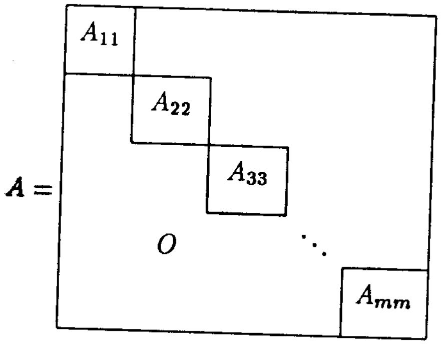


Рис. 2

столбцов матрицы A, образующих ее первый клеточный столбец.

В этих столбцах все миноры k-го порядка, кроме $|A_{11}|$, заведомо равны нулю, а алгебраическое дополнение к минору $|A_{11}|$ совпадает с дополнительным минором. Согласно теореме Лапласа отсюда имеем, что $|A| = |A_{11}| \cdot |\tilde{A}|$, где $\tilde{A}$ – тоже квазитреугольная матрица, но уже $(m-1)$-го порядка, начинающаяся с клетки A_{22}.

Поступая точно так же еще $(m-1)$ раз, получим, что

$$|A| = |A_{11}| \cdot |A_{22}| \cdot \ldots \cdot |A_{mm}|. \tag{4.11}$$

Аналогично получается равенство (4.11) и для нижней квазитреугольной матрицы. ■

Теорема 4.7. *Определитель произведения квадратных матриц равен произведению определителей матриц-сомножителей.*

Доказательство. Пусть $A = (a_{ij})$, $B = (b_{ij})$ – квадратные матрицы n-го порядка. Рассмотрим матрицу C вида

$$C = \left[\begin{array}{cccc|cccc} a_{11} & a_{12} & \ldots & a_{1n} & & & & \\ a_{21} & a_{22} & \ldots & a_{2n} & & & & \\ \ldots & \ldots & \ldots & \ldots & & O & & \\ a_{n1} & a_{n2} & \ldots & a_{nn} & & & & \\ \hline -1 & 0 & \ldots & 0 & b_{11} & b_{12} & \ldots & b_{1n} \\ 0 & -1 & \ldots & 0 & b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & -1 & b_{n1} & b_{n2} & \ldots & b_{nn} \end{array}\right].$$

С одной стороны, C – квазитреугольная матрица, поэтому согласно (4.11)

$$|C| = |A| \cdot |B|. \tag{4.12}$$

С другой стороны, не меняя определителя матрицы C, преобразуем ее так, чтобы клетка A стала нулевой. Сначала аннулируем первую строку матрицы A, для чего прибавим к первой строке матрицы C ее $(n+1)$-ю, $(n+2)$-ю, $\ldots, 2n$-ю строки, умноженные на $a_{11}, a_{12}, \ldots, a_{1n}$ соответственно. В результате этих преобразований $\det C$ остается без изменения, первые n элементов первой строки матрицы C станут нулевыми, а вторые n элементов заполнятся элементами первой строки матрицы AB. Теперь в новой матрице выполним аналогичные преобразования второй строки: ко второй строке матрицы C прибавим ее $(n+1)$-ю, $(n+2)$-ю, $\ldots, 2n$-ю строки,

умноженные на $a_{21}, a_{22}, \ldots, a_{2n}$ соответственно. После этих преобразований определитель матрицы C не изменится, первые n элементов второй строки матрицы C станут нулевыми, а вторые n элементов заполнятся элементами второй строки матрицы AB. Проделав аналогичные преобразования с третьей, ... , n-й строками матрицы C, получим матрицу

$$C_1 = \left[\begin{array}{c|c} O & AB \\ \hline -I & B \end{array}\right],$$

определитель которой равен определителю матрицы C. Переставив в матрице C_1 первый и $(n+1)$-й столбцы, второй и $(n+2)$-й столбцы, $\ldots, n$-й и $2n$-й столбцы, получим квазитреугольную матрицу

$$C_2 = \left[\begin{array}{c|c} AB & O \\ \hline B & -I \end{array}\right],$$

определитель которой отличается от определителя матрицы C_1 множителем $(-1)^n$. Таким образом, $|C| = |C_1| = (-1)^n|C_2| = |AB|$. Отсюда и из (4.12) получаем, что

$$|AB| = |A| \cdot |B|. \quad ■ \tag{4.13}$$

Вычисление определителя. Многие задачи линейной алгебры связаны с проблемой вычисления определителя.

Было бы весьма затруднительным вычислять определитель исходя из его определения, т.е. вычисляя непосредственно все $n!$ членов и определяя их знаки. Для этого нужно выполнить $n!n$ операций умножения (без учета менее трудоемкой операции сложения чисел). Так, для $n = 100$ число умножений превосходит 10^{153} – с этим не в состоянии справиться даже самые мощные современные компьютеры.

Теорема Лапласа позволяет упростить проблему вычисления определителя, сводя ее к вычислению определителей более низких порядков. Однако, как видно из (4.11), (4.12), этот подход существенно эффективен лишь для матриц с большим числом нулевых элементов, сгруппированных специальным образом.

Среди различных методов вычисления определителя особое место в приложениях занимает *метод Гаусса*.

Метод Гаусса применяется для решения широкого класса матричных задач. Идея этого метода проста и естественна. Она состоит в том, что:

- выделяется тип матрицы, для которой задача решается достаточно просто;
- выделяется тип преобразований, которые либо не изменяют решений задачи, либо изменяют их контролируемым образом;
- произвольная матрица выделенными преобразованиями приводится к выделенному типу, тем самым задача общего вида сводится к более простой.

В применении к задаче вычисления определителя эта схема выглядит следующим образом:

– определитель треугольной матрицы равен произведению диагональных элементов (свойство 1);

– элементарные преобразования матрицы либо не изменяют определителя (свойство 9), либо изменяют (свойства 4 и 6), но так, что эти изменения можно легко восстановить;

– произвольная квадратная матрица элементарными преобразованиями приводится к треугольной форме (теорема 3.1).

Метод Гаусса вычисления определителя состоит в приведении матрицы элементарными преобразованиями к треугольному виду, вычислении определителя получившейся треугольной матрицы и восстановлении исходного определителя, если использовались элементарные преобразования типов 1 и 2 (§3).

Замечание 3. В методе Гаусса вычисления определителя можно использовать элементарные преобразования как строк, так и столбцов.

Замечание 4. Для вычисления определителя n-го порядка по методу Гаусса требуется выполнить $\frac{n^3}{3} + O(n^2)$ операций умножения двух чисел. Теперь для $n = 100$ определитель может быть вычислен быстрее чем за одну секунду на компьютере с быстродействием 10^6 арифметических операций в секунду.

§5. Обратная матрица

Условие обратимости. Матрица A^{-1} называется *обратной к матрице* A, если $AA^{-1} = A^{-1}A = I$. Матрица A, для которой существует обратная матрица, называется *обратимой*.

Из определения следует, что обратимой может быть лишь квадратная матрица, так как равенство $AA^{-1} = A^{-1}A$ возможно лишь для квадратных матриц A и A^{-1} одинакового порядка. Но не каждая квадратная матрица обратима. Так, матрица $A = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$ при умножении справа на любую матрицу дает матрицу с нулевой первой строкой, т.е. ни для какой матрицы B произведение AB не может совпадать с единичной матрицей I. Выясним, какие свойства матрицы обеспечивают ее обратимость.

Квадратная матрица A называется *вырожденной (особенной)*, если $|A| = 0$, и *невырожденной (неособенной)*, если $|A| \neq 0$.

Пусть $A = (a_{ij}) \in \mathbb{R}^{n\times n}$. Матрица

$$\widehat{A} = \begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix}, \tag{5.1}$$

составленная из алгебраических дополнений A_{ij} к элементам a_{ij} матрицы A, называется *присоединенной (взаимной)* к матрице A.

Теорема 5.1 (о фальшивом разложении определителя). *Сумма произведений элементов одной строки (столбца) матрицы на алгебраические дополнения к элементам другой ее строки (соответственно столбца) равна нулю.*

Доказательство. Пусть $A = (a_{ij}) \in \mathbb{R}^{n\times n}$. Покажем, что для любых ее двух строк i, j, где $i \neq j$,

$$\sum_{s=1}^{n} a_{is}A_{js} = 0\,. \tag{5.2}$$

Рассмотрим вспомогательную матрицу B, которая отличается от A только j-й строкой: на месте j-й строки в B стоит i-я строка A.

С одной стороны, $\det B = 0$ (§4, свойство 7). С другой стороны, $\det B = \sum_{s=1}^{n} a_{is}A_{js}$, так как в разложении $\det B$ по j-й строке алгебраические дополнения к элементам j-й строки матрицы B получаются вычеркиванием j-й строки и поэтому совпадают с алгебраическими дополнениями A_{js} к элементам j-й строки матрицы A. Отсюда следует (5.2). Аналогично доказывается столбцовый вариант теоремы. ■

Доказанную теорему иногда называют теоремой о "чужих" алгебраических дополнениях. Напомним, что в этой терминологии умножение на "свои" алгебраические дополнения дает разложение (4.10) определителя по строке (столбцу соответственно).

Теорема 5.2 (критерий обратимости). *Матрица обратима тогда и только тогда, когда она не вырождена.*

Доказательство. Необходимость. Пусть матрица A обратима. Тогда существует обратная матрица A^{-1} такая, что $AA^{-1} = I$. Взяв определители от обеих частей этого равенства, согласно (4.13) получим, что $|A| \cdot |A^{-1}| = |I|$, т.е. $|A| \cdot |A^{-1}| = 1$. Следовательно, $|A| \neq 0$.

Достаточность. Пусть $|A| \neq 0$. Покажем, что матрица $\frac{1}{|A|}\widehat{A}$ является обратной к матрице A. В самом деле, из разложения определителя по строке (столбцу) и теоремы 5.1 имеем, что $A\widehat{A} = \widehat{A}A = |A|I$. Следовательно, $A\frac{1}{|A|}\widehat{A} = \frac{1}{|A|}\widehat{A}A = I$, т.е.

$$A^{-1} = \frac{1}{|A|}\widehat{A}. \quad ■ \tag{5.3}$$

Теорема 5.3 (о единственности обратной матрицы). *Если A – квадратная матрица и $AB = I$ (или $BA = I$), то $B = A^{-1}$.*

Доказательство. Из равенства $AB = I$ следует, что B – квадратная матрица и, согласно (4.13), A не вырождена. Следовательно, A обратима и для нее существует обратная матрица A^{-1}. Тогда $A^{-1} = A^{-1}(AB) = (A^{-1}A)B = IB = B$. Таким образом, $B = A^{-1}$. Аналогично рассматривается случай, когда $BA = I$. ■

Доказанная теорема устанавливает свойство единственности обратной матрицы, и, более того, из нее следует, что для квадратной матрицы A одного из равенств $AA^{-1} = I$ или $A^{-1}A = I$ достаточно, чтобы матрица A^{-1} была обратной к матрице A.

Некоторые свойства обратной матрицы.

1. $I^{-1} = I$, так как $I \cdot I = I$.
2. $|A^{-1}| = 1/|A|$, так как $|A| \cdot |A^{-1}| = 1$.
3. $(A^{-1})^{-1} = A$, так как $AA^{-1} = A^{-1}A = I$.
4. $(A^T)^{-1} = (A^{-1})^T$, так как $(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$.
5. $(AB)^{-1} = B^{-1}A^{-1}$, так как $(AB)(B^{-1}A^{-1}) = I$.

Вычисление обратной матрицы.

Соотношение (5.3) дает явный вид обратной матрицы. Оно полезно в теоретических исследованиях и совершенно неэффективно для практического вычисления (разве что для матриц второго порядка) вследствие большого объема требуемых вычислений. В самом деле, для получения обратной матрицы к матрице n-го порядка согласно (5.3) требуется вычислить n^2 определителей $(n-1)$-го порядка и один определитель n-го порядка. В вычислительной математике используются различные дополнительные приемы вычисления обратной матрицы, которые по объему вычислений равносильны вычислению всего лишь двух определителей n-го порядка. Опишем один из них.

Теорема 5.4. *Произвольная невырожденная матрица элементарными преобразованиями только строк (только столбцов) приводится к единичной матрице.*

Доказательство. Рассмотрим строчный вариант теоремы. Пусть $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ и $\det A \neq 0$. Применим к ней основной процесс (теорема 3.1). Так как A – квадратная матрица, то окончательная ступенчатая матрица будет треугольной. Ввиду невырожденности исходной матрицы она также будет невырожденной и ее диагональные элементы будут отличны от нуля. Поделив каждую *строку* на ее диагональный элемент, т.е. выполнив элементарные преобразования строк, получим треугольную матрицу вида

$$\begin{bmatrix} 1 & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & 1 & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

Если представить процесс приведения матрицы к верхней ступенчатой форме как преобразование матрицы "слева направо" (в таком порядке аннулируются столбцы), то теперь будем выполнять аналогичные преобразования "справа налево". На первом шаге с помощью последней строки аннулируем все наддиагональные элементы последнего столбца, вычитая из первых $(n-1)$ строк последнюю строку, умноженную на $a_{1n}, a_{2n}, \dots, a_{n-1,n}$ соответственно. На втором шаге из первых $(n-2)$ строк вычитаем $(n-1)$-ю строку, умноженную на $a_{1,n-1}, a_{2,n-1}, \dots, a_{n-2,n-1}$ соответственно. Выполняя аналогичные преобразования, через $(n-1)$ шагов получим единичную матрицу. Отметим, что на каждом шаге изменяются элементы только одного аннулируемого столбца. Если в доказательстве поменять ролями строки и столбцы, то получим утверждение столбцового варианта теоремы. ∎

Доказанная теорема может быть переформулирована в терминах матриц элементарных преобразований (теорема 3.2). Для строчного варианта: *существуют матрицы элементарных преобразований $L_1, L_2, \dots, L_k$ такие, что $L_kL_{k-1}\dots L_2L_1A = I$.* Отсюда в силу теоремы 5.3 следует, что $A^{-1} = L_kL_{k-1}\dots L_2L_1$ или $A^{-1} = L_kL_{k-1}\dots L_2L_1I$.

Это означает, что ***для получения обратной матрицы достаточно к строкам единичной матрицы I применить те преобразования, которые приводят матрицу A к единичной матрице.*** Для этого удобно составить расширенную матрицу $[A|I]$ и над строками этой матрицы выполнить те преобразования, которые матрицу A приводят к единичной; тогда на месте матрицы I окажется обратная матрица A^{-1}. Итак,

$$\begin{array}{|c|c|}\hline A & I \\ \hline\end{array} \xmapsto{\text{преобразования строк}} \begin{array}{|c|c|}\hline I & A^{-1} \\ \hline\end{array}. \qquad (5.4)$$

Аналогично в столбцовом варианте

$$\begin{array}{|c|}\hline A \\ \hline I \\ \hline\end{array} \xmapsto{\text{преобразования столбцов}} \begin{array}{|c|}\hline I \\ \hline A^{-1} \\ \hline\end{array}. \qquad (5.5)$$

Этот метод вычисления обратной матрицы называется ***методом Жордана*** или ***методом Гаусса–Жордана.***

Замечание. Если в расширенных матрицах (5.4) и (5.5) на место единичной матрицы I поставить матрицу B, то вместо матрицы A^{-1} получим в первом случае матрицу $A^{-1}B$, а во втором – BA^{-1}:

$$\begin{array}{|c|c|}\hline A & B \\ \hline\end{array} \xmapsto{\text{преобразования строк}} \begin{array}{|c|c|}\hline I & A^{-1}B \\ \hline\end{array}, \quad \begin{array}{|c|}\hline A \\ \hline B \\ \hline\end{array} \xmapsto{\text{преобразования столбцов}} \begin{array}{|c|}\hline I \\ \hline BA^{-1} \\ \hline\end{array},$$

так как $A^{-1}B = L_kL_{k-1}\dots L_2L_1B, \quad BA^{-1} = BL_1L_2\dots L_k$.

Глава II. Теоретико-множественные понятия

Здесь излагаются первоначальные теоретико-множественные понятия, которые будут использоваться в последующих главах.

§6. Множества

Под ***множеством*** в математике понимается совокупность объектов, называемых ***элементами*** множества.

Как правило, множество обозначается прописной буквой какого-либо алфавита, а его элементы – строчными буквами того же или другого алфавита. Для некоторых множеств приняты стандартные обозначения. Так, буквами $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ обозначают соответственно множества всех натуральных, целых, рациональных, вещественных чисел. Множества с конечным числом элементов могут быть описаны путем явного перечисления всех их элементов, элементы при этом заключаются в фигурные скобки. Например, $\{0,1,2\}$ – множество остатков от деления целых чисел на число 3.

Множество, не содержащее ни одного элемента, называется ***пустым множеством***. Мы будем обозначать его символом $\varnothing$.

Множество S называется ***подмножеством*** множества X, если имеет место импликация: $x \in S \Rightarrow x \in X$. Обозначение: $S \subset X$. Пустое множество по определению является подмножеством любого множества.

Для задания подмножества $S \subset X$ используется его характеристическое свойство, т.е. свойство, присущее только элементам из S. Например, запись $\{n = 2k \mid k \in \mathbb{Z}\}$ задает множество всех четных чисел.

Два множества X и Y называются ***равными***, если каждое из них является подмножеством другого, т.е.

$$X = Y \iff \begin{cases} X \subset Y, \\ Y \subset X. \end{cases} \tag{6.1}$$

Если $S \subset X$, причем $S \neq \varnothing$, $S \neq X$, то S называется ***собственным подмножеством*** множества X.

Объединением (***суммой*** или ***соединением***) множеств X и Y называется множество всех элементов, которые принадлежат хотя бы одному из множеств X или Y.

Обозначение: $X \cup Y$. Итак, $X \cup Y = \{x \mid x \in X \text{ или } x \in Y\}$.

Пересечением множеств X и Y называется множество всех элементов, одновременно принадлежащих как X, так и Y.

Обозначение: $X \cap Y$. Итак, $X \cap Y = \{x \mid x \in X, x \in Y\}$.

Разностью множеств X и Y называется множество всех элементов из X, которые не содержатся в Y.

Обозначение: $X \setminus Y$ или $X - Y$. Итак, $X \setminus Y = \{x \mid x \in X, x \notin Y\}$.

Если $Y \subset X$, то разность $X \setminus Y$ называется ***дополнением*** множества Y до множества X. Обозначение: $\overline{Y}$, $\mathrm{C}Y$ или $\mathrm{C}_X Y$.

Декартовым произведением множеств X и Y называется множество всевозможных упорядоченных пар (x,y), в которых $x \in X, y \in Y$.

Обозначение: $X \times Y$. Итак, $X \times Y = \{(x,y) \mid x \in X, y \in Y\}$.

Декартово произведение $X \times X$ называется ***декартовым квадратом*** множества X и обозначается символом X^2. Множество всех пар (x,x), где $x \in X$, называется ***диагональю*** декартова квадрата множества X.

Примеры. 1. Множество $\mathbb{Q}$ всех рациональных чисел можно рассматривать как декартово произведение $\mathbb{Z} \times \mathbb{N}$.

2. Множество декартовых координат всех точек плоскости представляет собой декартов квадрат $\mathbb{R}^2$.

3. Если упорядоченную пару (a,b) вещественных чисел изображать точкой плоскости с абсциссой a и ординатой b, то декартовы произведения $X \times Y$ и $Y \times X$ множеств, указанных на рис. 1, изображаются точками соответствующих прямоугольников.

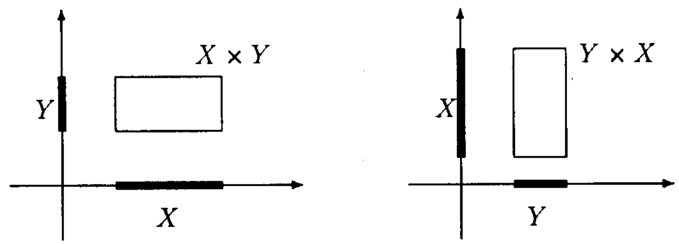


Рис. 1

§7. Эквивалентность

В математике, в логических рассуждениях, а также в обыденной жизни мы сталкиваемся с необходимостью сравнивания двух элементов множества: равенство (и неравенство) двух чисел, равенство матриц, равенство множеств, подобие треугольников, эквивалентность уравнений и т.д. Во всех этих случаях между *двумя* элементами множества определено некоторое отношение, и в каждом случае введено *правило*, по которому устанавливается, находятся ли заданные два элемента в этом отношении или нет. Несмотря на разнообразие этих отношений, оказывается, что с математической точки зрения они являются конкретными проявлениями одного и того же понятия. Введем его.

Бинарное отношение. Говорят, что на множестве X задано ***бинарное отношение*** $\mathcal{R}$, если указано непустое подмножество $\mathcal{R}$ декартова квадрата этого множества. Если при этом $(x,y) \in \mathcal{R}$, то говорят, что элементы x и y связаны отношением $\mathcal{R}$, и обозначают символом $x\mathcal{R}y$.

Равенство $x = y$ и неравенство $x < y$ действительных чисел, равенство матриц $A = B$, равенство множеств $X = Y$ и пр. – все это примеры бинарных отношений. В самом деле, рассмотрим один из них – отношение "$<$" на множестве $\mathbb{R}$ всех действительных чисел. Пара чисел $x, y \in \mathbb{R}$, для которой $x < y$, является элементом $\mathbb{R}^2$.

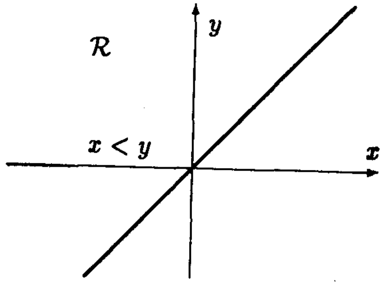


Рис. 1

Множество всех таких пар образует подмножество $\mathcal{R}$ декартова квадрата $\mathbb{R}^2$ – на рис. 1 они изображаются точками плоскости $\mathbb{R}^2$, которые лежат выше прямой $y = x$. Задание этого множества $\mathcal{R}$ определяет бинарное отношение на множестве действительных чисел, а отношение $x\mathcal{R}y$ (т.е. $(x, y) \in \mathcal{R}$) есть не что иное, как $x < y$.

Заметим, что бинарное отношение равенства вещественных чисел задается прямой $y = x$, т.е. диагональю декартова квадрата $\mathbb{R}^2$.

Бинарное отношение $\mathcal{R}$ на множестве X называется:

- *рефлексивным*, если $x\mathcal{R}x, \forall x \in X$;
- *симметричным*, если имеет место импликация $x\mathcal{R}y \Rightarrow y\mathcal{R}x$;
- *транзитивным*, если имеет место импликация $x\mathcal{R}y, y\mathcal{R}z \Rightarrow x\mathcal{R}z$.

Отношение эквивалентности. Важным классом бинарных отношений являются бинарные отношения, описывающие свойство "схожести", – отношения эквивалентности.

Бинарное отношение $\mathcal{E}$ на множестве X называется *отношением эквивалентности*, если оно рефлексивно, симметрично и транзитивно. Если пара элементов $x, y \in X$ связана отношением эквивалентности, то говорят, что x и y *эквивалентны*, и обозначают символом $x \sim y$. В конкретных случаях вместо этого символа могут быть использованы и другие, например $x = y$, $x \equiv y$.

Отношения эквивалентности играют важную роль в математических исследованиях. Назвав два элемента эквивалентными, мы игнорируем те различия между ними, которые не существенны в рассматриваемой задаче. В каждой конкретной задаче мы различаем или не различаем элементы лишь в отношении тех их свойств, которыми интересуемся в данный момент. Известное правило равенства обыкновенных дробей $(p_1/q_1 = p_2/q_2$, если $p_1 q_2 = q_1 p_2)$ является отношением эквивалентности. Назвав дроби $\frac{1}{2}$ и $\frac{2}{4}$ равными, мы игнорируем их несовпадение как дробей и считаем, что они определяют одно и то же рациональное число. Уже этот простейший пример подсказывает, что эквивалентные элементы множества целесообразно объединять в один объект.

Пусть на множестве X задано отношение эквивалентности $\mathcal{E}$ и $a \in X$. Множество всевозможных элементов $x \in X$, эквивалентных a, называется *классом эквивалентности, порожденным элементом a*. Обозначение: $\mathrm{cl}\,(a)$. Итак, $\mathrm{cl}\,(a) = \{x \in X | x \sim a\}$. Любой элемент класса эквивалентности называется *представителем класса*.

Теорема 7.1. *Класс эквивалентности порождается любым своим представителем, т.е. если $b \in \mathrm{cl}\,(a)$, то $\mathrm{cl}\,(b) = \mathrm{cl}\,(a)$.*

Для доказательства равенства этих множеств достаточно показать двустороннее вложение (6.1) их друг в друга. Действительно,

$$c \in \mathrm{cl}\,(b) \Rightarrow c \sim b, \text{ но } b \sim a \Rightarrow c \sim a \Rightarrow c \in \mathrm{cl}\,(a),$$
$$c \in \mathrm{cl}\,(a) \Rightarrow c \sim a, \text{ но } a \sim b \Rightarrow c \sim b \Rightarrow c \in \mathrm{cl}\,(b). \quad \blacksquare$$

Теорема 7.2. *Два класса эквивалентности либо не пересекаются, либо совпадают.*

Действительно, любые два класса $\mathrm{cl}\,(a)$ и $\mathrm{cl}\,(b)$ либо не пересекаются, либо имеют хотя бы один общий элемент. Но в последнем случае они совпадают, так как если $c \in \mathrm{cl}\,(a)$, $c \in \mathrm{cl}\,(b)$, то в силу теоремы 7.1 $\mathrm{cl}\,(a) = \mathrm{cl}\,(b)$.$\blacksquare$

Из теоремы 7.2 и очевидного факта, что любой элемент множества содержится в одном из классов эквивалентности ($a \in \mathrm{cl}\,(a)$), следует, что все множество разбивается на непересекающиеся классы эквивалентности. Множество всех классов эквивалентности называется *фактор-множеством множества X по отношению эквивалентности $\mathcal{E}$* и обозначается символом $X|\mathcal{E}$.

Как уже отмечалось выше, одни и те же элементы множества, эквивалентные при одном отношении эквивалентности, могут оказаться неэквивалентными при другом. Все зависит от задачи, а в разных задачах мы можем интересоваться различными свойствами одних и тех же элементов. Проиллюстрируем это на примерах различных отношений эквивалентности на множестве $\mathbb{Z}$ целых чисел.

Примеры. 1. Положим $m \sim n$, если $m = n$. Это тривиальный пример отношения эквивалентности, сводящегося к простому совпадению. Очевидно, что $\mathrm{cl}\,(m) = \{m\}$ и $\mathbb{Z}|\mathcal{E} = \mathbb{Z}$.

2. Положим $m \sim n$, если m и n имеют одинаковую четность. Нетрудно проверить, что это бинарное отношение является отношением эквивалентности. Очевидно, что при таком отношении эквивалентности все множество $\mathbb{Z}$ разбивается на два класса эквивалентности: $C_0 = \mathrm{cl}\,(0) = \{2k \,|\, k \in \mathbb{Z}\}$ – множество всех четных чисел и $C_1 = \mathrm{cl}\,(1) = \{2k + 1 \,|\, k \in \mathbb{Z}\}$ – множество всех нечетных чисел. Итак, $\mathbb{Z}|\mathcal{E} = \{C_0, C_1\}$.

Такое разбиение множества целых чисел на классы нам знакомо из обыденной жизни, например деление множества домов улицы на четную и нечетную сторону.

3. Пусть $p \in \mathbb{N}, p > 1$. Два целых числа m и n называются *сравнимыми по модулю p*, если при делении на p они дают одинаковые остатки, т.е. если $m - n = pk$, где $k \in \mathbb{Z}$.

Обозначение: $m \equiv n (\mathrm{mod}\, p)$. Итак,

$$m \equiv n (\mathrm{mod}\, p) \iff m - n = pk, k \in \mathbb{Z}. \tag{7.1}$$

Положим $m \sim n$, если $m \equiv n (\mathrm{mod}\, p)$. Нетрудно проверить, что бинарное отношение (7.1) является отношением эквивалентности. Найдем классы эквивалентности. Пусть m при делении на p дает в остатке r. Очевидно, $0 \le r \le p - 1$. Тогда $\mathrm{cl}\,(m) = \{n = pk + r \,|\, k \in \mathbb{Z}\}$ – множество всех целых чисел n, дающих при делении на p остаток r. Так как при делении на p возможно ровно p различных остатков: $0, 1, \dots, p-1$, то количество классов эквивалентности равно p. Обозначения: $C_r = \{n = pk + r \,|\, k \in \mathbb{Z}\}$, где $0 \leq r \leq p-1$, и $\mathbb{Z}_p$ – фактор-множество множества $\mathbb{Z}$ по отношению эквивалентности (7.1). В этих обозначениях $\mathbb{Z}_p = \{C_0, C_1, \dots, C_{p-1}\}$. Множество $\mathbb{Z}_p$ называют ***множеством классов вычетов по модулю p***.

§8. Отображения

Определения, простейшие свойства. Пусть X, Y – два множества. *Отображением* f множества X во множество Y называется закон, посредством которого произвольному элементу $x \in X$ ставится в соответствие однозначно определенный элемент $y \in Y$; при этом элемент y называется *образом* элемента x, а элемент x – *прообразом* элемента y. Символически отображение записывается в виде $f : X \to Y$ или $X \xrightarrow{f} Y$. Запись $y = f(x)$ или $x \mapsto y$ означает, что элемент x при отображении f переходит в элемент y.

Отображение множества X во множество Y называют также *преобразованием множества* X в Y, или *оператором*, действующим из множества X во множество Y, или *функцией*, определенной на X со значениями в Y. Все эти названия употребляются в одинаковом смысле, и их использование диктуется соображениями удобства или желанием подчеркнуть тот или иной аспект. В случае $X = Y$ говорят об *отображении в себя*.

Полным прообразом элемента $y \in Y$ называется множество[1]

$$f^{-1}(y) = \{x \in X | f(x) = y\}.$$

Образом отображения f называется множество

$$\operatorname{im} f = \{y = f(x) | x \in X\}.$$

Вместо символа $\operatorname{im} f$ используется также символ $f(X)$.

Отображение $f : X \to Y$ называется:

- *инъективным*, если из того, что $x_1 \neq x_2$, следует, что $f(x_1) \neq f(x_2)$, или, другими словами, если уравнение

$$f(x) = y \tag{8.1}$$

при любом $y \in Y$ имеет не более одного решения;

- *сюръективным* или *отображением на*, если $\operatorname{im} f = Y$, или, другими словами, если уравнение (8.1) при любом $y \in Y$ имеет хотя бы одно решение;

- *биективным* или *взаимно однозначным*, если оно и инъективно, и сюръективно, или, другими словами, если уравнение (8.1) при любом $y \in Y$ имеет, и притом единственное, решение.

[1] Символ $f^{-1}(y)$ не следует ассоциировать с обратным отображением (8.5), которое может и не существовать.

Примеры. 1. Пусть $a \in \mathbb{R}$ и $\mathbb{R}_+$ – множество всех неотрицательных действительных чисел. Соответствие $a \mapsto |a|$ определяет три различных отображения $f : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R} \mapsto \mathbb{R}_+$, $h : \mathbb{R}_+ \to \mathbb{R}_+$. Отображение f не инъективно и не сюръективно, g – не инъективно, но сюръективно, h – биективно.

2. Пусть $A \in \mathbb{R}^{n \times n}$. Соответствие $A \mapsto \det A$ определяет отображение $f : \mathbb{R}^{n \times n} \to \mathbb{R}$, не инъективное, но сюръективное.

3. Пусть $a, b \in \mathbb{R}$. Соответствие $(a, b) \mapsto a + b$ определяет отображение $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, не инъективное, но сюръективное. Таким образом, операцию сложения чисел можно рассматривать как отображение декартова квадрата множества $\mathbb{R}$ в $\mathbb{R}$. Этим обстоятельством мы воспользуемся в дальнейшем.

4. Пусть X – конечное множество, состоящее из n элементов. Перенумеруем эти элементы и будем считать, что $X = \{1, 2, \dots, n\}$. При биективном отображении f множества X на себя его элементы преобразуются следующим образом: $1 \mapsto \alpha_1$, $2 \mapsto \alpha_2$, $\dots, n \mapsto \alpha_n$, где $\alpha_i \in \{1, 2, \dots, n\}$ и $\alpha_i \neq \alpha_j$ при $i \neq j$. Этот закон преобразования элементов принято записывать в виде таблицы

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \alpha_1 & \alpha_2 & \dots & \alpha_n \end{pmatrix} \tag{8.2}$$

или просто в виде

$$\alpha_1, \alpha_2, \dots, \alpha_n, \tag{8.3}$$

где $\alpha_i \in \{1, 2, \dots, n\}$ и $\alpha_i \neq \alpha_j$ при $i \neq j$. Биективное отображение множества X на себя называют *подстановкой* этого *множества* (*подстановкой n-го порядка*) и обозначают символом (8.2) или *перестановкой* этого *множества* (*перестановкой n-го порядка*) и обозначают символом (8.3).

Два отображения $f : X \to Y$ и $g : X \to Y$ называются *равными*, если $f(x) = g(x)$, $\forall x \in X$. Обозначение: $f = g$.

Тождественным (единичным) отображением на множестве X называется отображение $e_X : X \to X$, которое переводит каждый элемент $x \in X$ в себя.

Произведение отображений. *Произведением (суперпозицией* или *композицией)* отображений $g : X \to Y$ и $f : Y \to Z$ называется отображение $fg : X \to Z$, определенное правилом

$$fg(x) = f(g(x)), \quad \forall x \in X. \tag{8.4}$$

Таким образом, произведение отображений есть последовательное выполнение отображений-сомножителей, причем если символ отображения рассматривать как рецепт для выполнения определенных действий, то символ произведения fg следует читать справа налево.

Заметим, что произведение отображений некоммутативно. Даже в случаях, когда оба произведения fg и gf имеют смысл, произведение, вообще говоря, зависит от порядка. В этом легко убедиться на примере, когда $X = Y = \mathbb{R}$, $f(x) = x + 5$, $g(x) = |x|$.

Произведение отображений обладает следующими свойствами.

1. $fe_X = f$; $e_Y f = f$ *для любого отображения* $f : X \to Y$. Проверка этого свойства предоставляется читателю.

2. *Произведение отображений ассоциативно, т.е. если* $h : X \to Y$, $g : Y \to Z$, $f : Z \to U$, *то* $f(gh) = (fg)h$.

Доказательство. В соответствии с определением равенства отображений нужно просто сравнить значения отображений $f(gh) : X \to U$ и $(fg)h : X \to U$ в произвольной "точке" $x \in X$. Согласно определению (8.4) произведения имеем

$$(f(gh))(x) = f((gh)(x)) = f(g(h(x))) = (fg)(h(x)) = ((fg)h)(x). \quad \blacksquare$$

3. *Произведение инъективных (сюръективных, биективных) отображений инъективно (соответственно сюръективно, биективно).*

Доказательство. Пусть $g : X \to Y$, $f : Y \to Z$ – инъективные отображения и пусть $x_1 \neq x_2$. Тогда из инъективности g следует, что $g(x_1) \neq g(x_2)$, а из инъективности f следует, что $f(g(x_1)) \neq f(g(x_2))$, т.е. $fg(x_1) \neq fg(x_2)$.

Пусть теперь f, g сюръективны. Тогда для любого $z \in Z$ в силу сюръективности f существует $y \in Y$ такой, что $z = f(y)$. Но для этого элемента y в силу сюръективности g существует элемент $x \in X$ такой, что $y = g(x)$. Таким образом, для любого элемента $z \in Z$ существует элемент $x \in X$ такой, что $z = f(g(x))$, т.е. $z = fg(x)$. Биективность вытекает из сюръективности и инъективности. ■

Обратное отображение. Пусть $f : X \to Y$. Отображение $f^{-1} : Y \to X$ называется *обратным* к отображению f, если

$$f^{-1}f = e_X, \quad ff^{-1} = e_Y. \tag{8.5}$$

Заметим, что из этого определения следует, что f – обратное отображение к отображению f^{-1}. Отображение, для которого существует обратное отображение, называется *обратимым*.

Если выполнено одно из равенств (8.5), то f^{-1} называется соответственно *левым* или *правым обратным*.

Теорема 8.1 (критерий обратимости). *Отображение обратимо тогда и только тогда, когда оно биективно.*

Лемма. *Если* $g : X \to Y$, $f : Y \to X$ *и* $fg = e_X$, *то* g *инъективно, а* f *сюръективно.*

Действительно, если g не инъективно, то существуют элементы $x_1, x_2 \in X$ такие, что $x_1 \neq x_2$, а $g(x_1) = g(x_2)$. Тогда $x_1 = e_X(x_1) = fg(x_1) = f(g(x_1)) = f(g(x_2)) = fg(x_2) = e_X(x_2) = x_2$. Следовательно, g инъективно. Далее, если x – произвольный элемент X, то $x = e_X(x) = fg(x) = f(g(x)) = f(y)$, где $y = g(x) \in Y$. Это доказывает сюръективность f. Лемма доказана.

Доказательство теоремы. Необходимость. Пусть f обратимо. Тогда из (8.5) и леммы следует, что f инъективно и сюръективно, т.е. f биективно.

Достаточность. Пусть f биективно, тогда для любого $y \in Y$ существует единственный прообраз $x \in X$. Построим отображение $g : Y \to X$, положив $g(y) = x$. Тогда для любого $y \in Y$ имеем $(fg)(y) = f(g(y)) = f(x) = y$, т.е. $fg = e_Y$, а для любого $x \in X$ имеем $(gf)(x) = g(f(x)) = g(y) = x$, т.е. $gf = e_X$. Таким образом, $g = f^{-1}$ и f обратимо. ■

Отметим еще два свойства обратимых отображений.

1. *Обратное отображение единственно*, так как если f_1^{-1}, f_2^{-1} – два обратных отображения к отображению $f : X \to Y$, то $f_1^{-1} = f_1^{-1}e_Y = f_1^{-1}(ff_2^{-1}) = (f_1^{-1}f)f_2^{-1} = e_X f_2^{-1} = f_2^{-1}$.

2. *Произведение обратимых отображений обратимо, при этом* $(fg)^{-1} = g^{-1}f^{-1}$. Действительно, произведение fg обратимо как произведение биективных отображений, при этом $(fg)(g^{-1}f^{-1}) = f(gg^{-1})f^{-1} = ff^{-1} = e_Y$, $(g^{-1}f^{-1})(fg) = g^{-1}(f^{-1}f)g = g^{-1}g = e_X$.

§9. Алгебраические законы

Внутренний закон композиции. ***Внутренним законом композиции (алгебраической операцией)*** на множестве X **называется** отображение

$$* : X \times X \to X,$$

т.е. закон, посредством которого любой упорядоченной паре элементов $a, b \in X$ ставится в соответствие однозначно определенный элемент $c \in X$. Тот факт, что $(a, b) \mapsto c$, записывается символически в виде $a * b = c$. В конкретных случаях вместо символа $*$ используют символы $+$, $-$, $\times$, : и др.

Примеры. 1. На множестве натуральных чисел $\mathbb{N}$ операции сложения и умножения чисел являются алгебраическими операциями, так как любые два натуральных числа можно сложить (и умножить), при этом результатом будет также натуральное число.

На этом же множестве операции вычитания и деления не являются алгебраическими операциями, так как результаты выполнения этих операций не всегда будут натуральными числами.

2. На множестве вещественных чисел $\mathbb{R}$ операции сложения, вычитания, умножения (но не деления) чисел будут алгебраическими операциями.

3. На множестве ненулевых вещественных чисел $\mathbb{R}\backslash\{0\}$ операции умножения и деления (но не сложения и вычитания) будут алгебраическими операциями.

4. На множестве $\mathbb{R}^{m\times n}$ матриц размера $m\times n$, где $m \neq n$, операции сложения и вычитания (но не умножения) матриц являются алгебраическими операциями.

5. На множестве $\mathbb{R}^{n\times n}$ квадратных матриц n-го порядка операции сложения, вычитания и умножения матриц – алгебраические операции.

Алгебраическая операция $*$ на множестве X называется:

- ***коммутативной***, если $a*b=b*a, \quad \forall a,b\in X$,
- ***ассоциативной***, если $(a*b)*c=a*(b*c), \quad \forall a,b,c\in X$.

Операции сложения и умножения чисел в $\mathbb{R}$ коммутативны и ассоциативны. Сложение матриц, как уже отмечалось в §2, также коммутативно и ассоциативно. Примером некоммутативной, но ассоциативной алгебраической операции могут служить операции умножения матриц в $\mathbb{R}^{n\times n}$ и суперпозиции отображений на множестве всех отображений множества X в себя.

Элемент $e\in X$ называется ***нейтральным элементом множества*** X ***относительно алгебраической операции*** $*$, если $\forall a\in X$:

$$a*e=e*a=a.$$

Примеры. 1. Очевидно, что число $0\in\mathbb{R}$ является нейтральным элементом относительно операции сложения чисел в $\mathbb{R}$, а матрица $O\in\mathbb{R}^{m\times n}$ – нейтральным элементом относительно сложения матриц в $\mathbb{R}^{m\times n}$. Операции вычитания чисел и матриц не обладают нейтральными элементами.

2. Также очевидно, что $1\in\mathbb{R}$ и $I\in\mathbb{R}^{n\times n}$ являются нейтральными элементами относительно операций умножения чисел в $\mathbb{R}$ и матриц в $\mathbb{R}^{n\times n}$ соответственно. Операция умножения на множестве всех четных чисел, являясь алгебраической операцией, не обладает нейтральным элементом.

Теорема 9.1. *Нейтральный элемент единствен.*

Доказательство. Действительно, если e_1 и e_2 – два нейтральных элемента, то $e_1*e_2=e_1$, так как e_2 – нейтральный элемент, и $e_1*e_2=e_2$, так как e_1 – нейтральный элемент. Значит, $e_1=e_2$. ∎

Пусть $*$ – алгебраическая операция на множестве X, обладающая нейтральным элементом e. Элемент x' называется ***симметричным элементом*** для элемента $x\in X$, если $x*x'=x'*x=e$.

Примеры. Очевидно, что любое число $a\in\mathbb{R}$ и любая матрица $A\in\mathbb{R}^{m\times n}$ обладают симметричным элементом относительно операции сложения – ими будут соответственно противоположное число $-a$ и противоположная матрица $-A$.

Что же касается умножения чисел в $\mathbb{R}$ и умножения матриц в $\mathbb{R}^{n\times n}$, то не каждый элемент обладает симметричным: только ненулевое число $a\in\mathbb{R}$ и невырожденная матрица $A\in\mathbb{R}^{n\times n}$ имеют симметричные элементы – ими будут соответственно обратное число a^{-1} и обратная матрица A^{-1} (§5).

Теорема 9.2. *Симметричный элемент относительно ассоциативной алгебраической операции единствен.*

Доказательство. Действительно, если x' и x'' – два симметричных элемента к элементу $x\in X$, то

$$x'=x'*e=x'*(x*x'')=(x'*x)*x''=e*x''=x''. \quad ∎$$

Замечание. В определении нейтрального и симметричного элементов мы считали, что $x*e=e*x=x$ и $x*x'=x'*x=e$. Но можно рассматривать только равенства $x*e=x$ и $x*x'=e$ (или $e*x=x$ и $x'*x=e$) и говорить о *правом нейтральном* и *правом симметричном* элементах (или о *левом нейтральном* и *левом симметричном*). Этим обобщением мы будем иногда пользоваться.

Говорят, что алгебраическая операция $*$ на множестве X обладает *обратной операцией*, если для любых двух элементов $a,b\in X$ уравнения $a*x=b$ и $y*a=b$ имеют единственное решение.

Наличие обратной операции означает существование двух алгебраических операций. Первая ставит в соответствие любой упорядоченной паре элементов $a,b\in X$ однозначно определенный элемент $x\in X$ такой, что $a*x=b$, и называется ***правой обратной операцией*** к операции $*$, а вторая – элемент $y\in X$ такой, что $y*a=b$, и называется ***левой обратной операцией***. Но если алгебраическая операция $*$ коммутативна, то $x=y$ и обе операции совпадают и определяют *обратную операцию* к операции $*$. Очевидно, вычитание чисел в $\mathbb{R}$ и вычитание матриц в $\mathbb{R}^{m\times n}$ – операции, обратные к операции сложения в $\mathbb{R}$ и $\mathbb{R}^{m\times n}$.

Пусть $*_1$ и $*_2$ – две алгебраические операции на множестве X. Алгебраическая операция $*_1$ называется ***дистрибутивной справа*** относительно алгебраической операции $*_2$, если

$$(a*_2 b)*_1 c=(a*_1 c)*_2(b*_1 c), \quad \forall a,b,c\in X;$$

дистрибутивной слева, если

$$a*_1(b*_2 c)=(a*_1 b)*_2(a*_1 c), \quad \forall a,b,c\in X;$$

и ***дистрибутивной***, если она дистрибутивна и справа и слева.

Пример. Умножение чисел в $\mathbb{R}$ (матриц в $\mathbb{R}^{n\times n}$) дистрибутивно относительно сложения в $\mathbb{R}$ ($\mathbb{R}^{n\times n}$ соответственно). Но сложение не дистрибутивно относительно умножения.

Пусть на множестве X задано отношение эквивалентности. Нами уже отмечалось в §7, что, назвав элементы множества эквивалентными, мы ожидаем, что в одной и той же ситуации они должны проявлять себя одинаково. В первую очередь это относится к алгебраическим операциям, в которых они участвуют. Алгебраическая операция $*$ на множестве X называется *согласованной* с отношением эквивалентности $\sim$ на этом множестве, если из того, что $a_1\sim a_2, b_1\sim b_2$, следует, что $a_1*b_1\sim a_2*b_2$, т.е. алгебраическая операция, примененная к эквивалентным элементам, дает эквивалентные результаты. Это заставляет нас при введении тех или иных алгебраических операций проверять выполнение этого требования. Мы не будем останавливаться на этом, а предлагаем читателю самому убедиться в справедливости данного свойства у всех рассматриваемых операций.

Заметим, что в наших рассуждениях, говоря о внутренних законах композиции, мы всюду использовали термин "алгебраическая операция". И это не случайно. Как правило, термин "внутренний закон композиции" используют в тех случаях, когда наряду с внутренним законом композиции рассматривается и внешний закон композиции.

Внешний закон композиции. Пусть X и P – два множества. *Внешним законом композиции на множестве* X называется отображение

$$(\cdot):\quad P\times X\to X,$$

т.е. закон, посредством которого любому элементу $\alpha\in P$ и любому элементу $x\in X$ ставится в соответствие однозначно определенный элемент $c\in X$. Тот факт, что $(\alpha,x)\mapsto c$, обозначается символом

$c = \alpha x$.

Умножение матрицы $A \in \mathbb{R}^{m \times n}$ на число $\alpha \in \mathbb{R}$ является внешним законом композиции на множестве $\mathbb{R}^{m \times n}$. Умножение чисел в $\mathbb{R}$ можно рассматривать и как внутренний закон композиции, и как внешний.

Внешний закон композиции на множестве X называется *дистрибутивным относительно внутреннего закона композиции* $*$ в X, если

$$\alpha(a * b) = \alpha a * \alpha b, \ \forall \alpha \in P, \forall a, b \in X.$$

Внешний закон композиции на множестве X называется *дистрибутивным относительно внутреннего закона композиции* $*'$ в P, если

$$(\alpha *' \beta)a = \alpha a *' \beta a, \ \forall \alpha, \beta \in P, \forall a \in X.$$

Умножение матрицы $A \in \mathbb{R}^{m \times n}$ на число $\alpha \in \mathbb{R}$ дистрибутивно как относительно сложения матриц, так и относительно сложения чисел, ибо $\alpha(A + B) = \alpha A + \alpha B$, $(\alpha + \beta)A = \alpha A + \beta A$.

Глава III. Геометрические векторы

В этой главе предполагаются известными из средней школы основные геометрические понятия, такие, как точка, прямая, отрезок, луч, длина отрезка, направление на прямой и т.д.

С математической точки зрения логически безупречным методом введения этих понятий является аксиоматический метод, получивший завершение в трудах Гильберта. Существуют различные версии аксиом евклидовой геометрии. Мы оставим в стороне эти исследования, ограничившись лишь ссылкой на литературу [4, 5] и формулировками основных фактов, на которые мы будем опираться в дальнейшем.

Отрезок $[AB]$ – это множество, состоящее из точек A и B и всех точек прямой (AB), которые лежат между точками A и B. Каждому отрезку $[AB]$ можно поставить в соответствие неотрицательное действительное число $|AB|$, называемое *длиной отрезка* $[AB]$ и удовлетворяющее аксиомам:

1) аксиома тождества:
$|AB| \geq 0, \forall A, B$,
$|AB| = 0$ тогда и только тогда, когда $A = B$;

2) аксиома симметрии:
$|AB| = |BA|, \forall A, B$;

3) неравенство треугольника:
$|AB| \leq |AC| + |CB|, \forall A, B, C$,
$|AB| = |AC| + |CB|$ тогда и только тогда, когда точка C лежит на отрезке $[AB]$.

Длина отрезка $[AB]$ называется также *расстоянием* между точками A и B и обозначается символом $\rho(A, B)$.

Любая точка A, лежащая на прямой l, разбивает эту прямую на два *луча* l_+ и l_- с началами в точке A. Эти лучи называются *дополнительными* друг к другу. Точка A принадлежит обоим лучам. Две точки $B \neq A$ и $C \neq A$ прямой l принадлежат одному лучу тогда и только тогда, когда отрезок $[BC]$ не содержит точки A, и принадлежат дополнительным лучам, если точка A является внутренней точкой отрезка $[BC]$. Луч с началом в точке A, на котором лежит точка $B \neq A$, обозначается символом $[AB)$.

Два луча, лежащие на одной прямой, называются *одинаково направленными* (*сонаправленными*), если их пересечение есть луч, и *противоположно направленными*, если их пересечение не является лучом.

Всякая прямая l, лежащая в плоскости P, разбивает эту плоскость на две полуплоскости P_+ и P_-, про которые говорят, что они *определяются прямой* l. Прямая l принадлежит каждой из этих полуплоскостей. Две точки $B \notin l$ и $C \notin l$ плоскости P лежат в одной полуплоскости, определяемой прямой l, тогда и только тогда, когда $[BC] \cap l = \varnothing$.

Два луча $[AB)$ и $[CD)$, лежащие на параллельных несовпадающих прямых, принадлежат некоторой плоскости P. Лучи $[AB)$ и $[CD)$ называются *одинаково направленными* (*сонаправленными*), если они лежат в одной полуплоскости, определяемой прямой AC, и *противоположно направленными*, если они лежат в разных полуплоскостях. Обозначение: $[AB) \uparrow\uparrow [CD)$ и $[AB) \uparrow\downarrow [CD)$ соответственно. Свойство сонаправленности лучей транзитивно.

Пусть P – некоторая плоскость и A, B, O – три ее различные точки, не лежащие на одной прямой. Прямая OA разбивает эту плоскость на две полуплоскости. Точка B расположена в одной из них. Обозначим эту полуплоскость P_+. Аналогично прямая OB разбивает плоскость P на две полуплоскости. Пусть P_- –

полуплоскость, в которой лежит точка A. *Выпуклым углом между лучами* $[OA)$ и $[OB)$ называется пересечение множеств P_+ и P_-. Обозначение: $\angle AOB$. Итак, $\angle AOB = P_+ \cap P_-$. Величина выпуклого угла $\angle AOB$ называется *углом между лучами* $[OA)$ и $[OB)$. Обозначение: $\widehat{AOB}$. Очевидно, $0 < \widehat{AOB} < \pi$. По определению угол между сонаправленными лучами равен 0, а между противоположно направленными равен π.

§10. Направленные отрезки

Упорядоченная пара точек (A, B) называется ***направленным отрезком*** с началом в точке A и концом в точке B. Обозначение: $\overrightarrow{AB}$.

Рис. 1

Направленный отрезок $\overrightarrow{AB}$ изображается стрелкой, идущей из его начала в его конец (рис. 1). Направленный отрезок $\overrightarrow{AB}$ называют также ***связанным вектором***, а точку A – ***точкой*** его ***приложения***. Если точки A и B различны, то направленный отрезок $\overrightarrow{AB}$ называется ***ненулевым***; если же точки A и B совпадают, то направленный отрезок $\overrightarrow{AB}$, точнее, $\overrightarrow{AA}$ называется ***нулевым*** и обозначается символом θ_A.

Направленный отрезок $\overrightarrow{AB}$ называется ***параллельным прямой*** l (***плоскости*** P), если либо он нулевой, либо прямая AB совпадает с прямой l (соответственно лежит в плоскости P), либо прямая AB параллельна прямой l (соответственно плоскости P).

Обозначение: $\overrightarrow{AB} \parallel l$, $\overrightarrow{AB} \parallel P$.

Направленные отрезки $\overrightarrow{A_1B_1}, \overrightarrow{A_2B_2}, \ldots, \overrightarrow{A_kB_k}$ называются ***коллинеарными*** (***компланарными***), если существует прямая (соответственно плоскость), которой параллелен каждый из этих отрезков.

Длиной направленного отрезка $\overrightarrow{AB}$ называется длина отрезка $[AB]$. Обозначение: $|\overrightarrow{AB}|$. Как следует из определения, длина нулевого и только нулевого направленного отрезка равна нулю.

Ненулевые направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{CD}$ называются ***одинаково направленными*** (***сонаправленными***), если лучи $[AB)$ и $[CD)$ имеют одинаковое направление, и ***противоположно направленными***, если лучи $[AB)$ и $[CD)$ имеют противоположные направления.

Обозначение: $\overrightarrow{AB} \uparrow\uparrow \overrightarrow{CD}$ и $\overrightarrow{AB} \uparrow\downarrow \overrightarrow{CD}$ соответственно.

Направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{CD}$ называются ***равными***, если середины отрезков $[AD]$ и $[BC]$ совпадают (рис. 2).

Обозначение: $\overrightarrow{AB} = \overrightarrow{CD}$. Как следует из определения, нулевой направленный отрезок равен любому другому нулевому и только нулевому направленному отрезку.

Из свойств параллелограмма (рис. 2) следует, что ненулевые направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{CD}$, не лежащие на одной прямой, равны тогда и только тогда, когда четырехугольник $ABDC$ – параллелограмм. Для равных ненулевых отрезков, лежащих на одной прямой, возможен один из четырех вариантов расположения, изображенных на рис. 2.

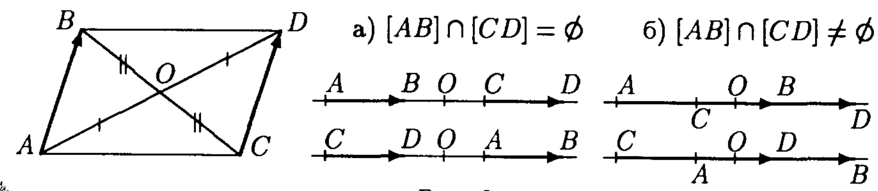


Рис. 2

Теорема 10.1. *Направленные отрезки $\overrightarrow{AB}$ и $\overrightarrow{CD}$ равны тогда и только тогда, когда они имеют:*

1) одинаковую длину: $|\overrightarrow{AB}| = |\overrightarrow{CD}|$

и, в случае $|\overrightarrow{AB}| \neq 0$,

2) одинаковое направление: $\overrightarrow{AB} \uparrow\uparrow \overrightarrow{CD}$.

Утверждение теоремы вытекает непосредственно из определения равенства направленных отрезков и свойств параллелограмма. ■

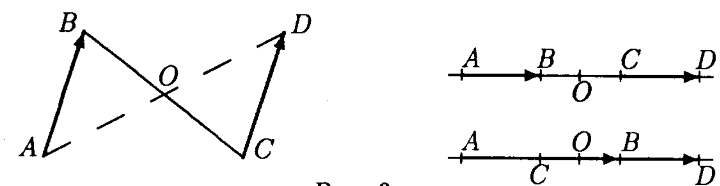


Рис. 3

Теорема 10.2. *Для любого направленного отрезка $\overrightarrow{AB}$ и любой точки C существует, и притом единственная, точка D такая, что $\overrightarrow{AB} = \overrightarrow{CD}$.*

Доказательство. Пусть (рис. 3) точка O – середина $[BC]$. Из определения равенства направленных отрезков следует, что точка D – точка на прямой AO, симметричная точке A относительно точки O. Такая точка определена однозначно. ■

Замечание. Теорему 10.2 формулируют и в других терминах: *направленный отрезок можно отложить от любой точки или направленный отрезок можно перенести в любую точку.*

Теорема 10.3. *Отношение равенства направленных отрезков является отношением эквивалентности на множестве всех направленных отрезков.*

В самом деле, отношение равенства направленных отрезков является бинарным отношением, которое обладает свойствами:

а) рефлексивности (направленный отрезок равен самому себе);

б) симметричности, так как справедливы импликации
$$|\overrightarrow{AB}| = |\overrightarrow{CD}| \Rightarrow |\overrightarrow{CD}| = |\overrightarrow{AB}|, \qquad \overrightarrow{AB} \uparrow\uparrow \overrightarrow{CD} \Rightarrow \overrightarrow{CD} \uparrow\uparrow \overrightarrow{AB};$$
в) транзитивности, так как равенство длин (т.е. чисел) и сонаправленность направленных отрезков (т.е. лучей) обладают свойством транзитивности. ■

Прямая l с заданным на ней направлением называется *осью*. *Величиной* направленного отрезка $\overrightarrow{AB}$ на оси l называется число
$$(\overrightarrow{AB}) = \begin{cases} |\overrightarrow{AB}|, & \overrightarrow{AB} \uparrow\uparrow l, \\ -|\overrightarrow{AB}|, & \overrightarrow{AB} \uparrow\downarrow l. \end{cases} \tag{10.1}$$

Из определения вытекают следующие факты.

1°. Нулевые направленные отрезки, и только они, имеют нулевую величину.

2°. $(\overrightarrow{AB}) = -(\overrightarrow{BA})$.

Лемма Шаля. *При любом расположении точек A, B и C на прямой имеет место равенство*
$$(\overrightarrow{AB}) = (\overrightarrow{AC}) + (\overrightarrow{CB}).$$

Доказательство. Если какие-либо две точки совпадают (например, $A = B$), то утверждение очевидно, так как $(\overrightarrow{AA}) = (\overrightarrow{AC}) + (\overrightarrow{CA})$. Пусть A, B, C – различные точки. Тогда возможны только три варианта расположения этих точек:

1) точка C между точками A и B (рис. 4);

Рис. 4

2) точка A между точками B и C;

3) точка B между точками A и C.

В первом случае, как видно из рис. 4, $|\overrightarrow{AB}| = |\overrightarrow{AC}| + |\overrightarrow{CB}|$ и $\overrightarrow{AB} \uparrow\uparrow \overrightarrow{AC} \uparrow\uparrow \overrightarrow{CB}$, откуда следует, что $(\overrightarrow{AB}) = (\overrightarrow{AC}) + (\overrightarrow{CB})$. Отсюда во втором случае имеем $(\overrightarrow{BC}) = (\overrightarrow{BA}) + (\overrightarrow{AC})$ или $(\overrightarrow{AB}) = (\overrightarrow{AC}) + (\overrightarrow{CB})$. Третий случай рассматривается аналогично второму. ■

§11. Свободный вектор

Определение и терминология. Известно (теорема 10.3), что отношение равенства направленных отрезков является отношением эквивалентности на множестве направленных отрезков. Оно разбивает это множество на непересекающиеся классы эквивалентности (§7).

Класс эквивалентности направленных отрезков называется *свободным вектором* или просто *вектором*. Векторы обозначают строчными латинскими буквами $\mathbf{a}$, $\mathbf{b}$. Итак, вектор $\mathbf{a} = \mathrm{cl}(\overrightarrow{AB})$ состоит из всех направленных отрезков, равных $\overrightarrow{AB}$. Так как класс эквивалентности (§7) порождается любым своим представителем, то вектор $\mathbf{a} = \mathrm{cl}(\overrightarrow{AB})$ можно задавать любым направленным отрезком $\overrightarrow{CD} = \overrightarrow{AB}$, т.е. $\mathbf{a} = \mathrm{cl}(\overrightarrow{CD})$. Если вместо направленного отрезка $\overrightarrow{AB}$ используется направленный отрезок $\overrightarrow{CD} = \overrightarrow{AB}$, то говорят, что вектор $\mathbf{a}$ отложен от точки C. Символ $\mathbf{a} = \mathrm{cl}(\overrightarrow{AB})$ используется применительно к геометрическому вектору $\mathbf{a}$ только в тех ситуациях, когда подчеркивается отношение этого вектора к классу эквивалентности. Обычно вместо символа $\mathbf{a} = \mathrm{cl}(\overrightarrow{AB})$ используется символ $\mathbf{a} = \overrightarrow{AB}$, который в зависимости от контекста читается как "вектор $\mathbf{a}$, порожденный направленным отрезком $\overrightarrow{AB}$" или "вектор $\mathbf{a}$, отложенный от точки A".

Длиной вектора $\mathbf{a}$ (*величиной вектора* $\mathbf{a}$ *на оси*) называется длина (соответственно величина) порождающего его направленного отрезка; векторы $\mathbf{a}_1$, $\mathbf{a}_2$, ..., $\mathbf{a}_k$ называются *коллинеарными* (*компланарными*), если коллинеарны (соответственно компланарны) порождающие их направленные отрезки; векторы $\mathbf{a}$ и $\mathbf{b}$ называются *одинаково направленными* (*противоположно направленными*), если одинаково (соответственно противоположно) направлены порождающие их направленные отрезки. Очевидно, что эти определения корректны несмотря на произвол в выборе направленных отрезков.

Линейные операции над векторами.

Сложение векторов. Сумма векторов $\mathbf{a}$ и $\mathbf{b}$ определяется следующим образом. Отложим вектор $\mathbf{a}$ от произвольной точки A, пусть B – конец этого вектора, т.е. $\mathbf{a} = \overrightarrow{AB}$. Затем отложим вектор $\mathbf{b}$ от точки B, пусть $\mathbf{b} = \overrightarrow{BC}$. *Суммой* $\mathbf{a} + \mathbf{b}$ *векторов* $\mathbf{a}$ и $\mathbf{b}$ называется вектор, порожденный направленным отрезком $\overrightarrow{AC}$ (рис. 1).

Рис. 1

Это правило сложения векторов называется *правилом треугольника*. Очевидно, что этот же вектор $\mathbf{a} + \mathbf{b}$ для неколлинеарных векторов $\mathbf{a}$ и $\mathbf{b}$ может быть получен (рис. 2) как диагональ параллелограмма, остроенного на векторах $\mathbf{a}$ и $\mathbf{b}$. Это правило сложения векторов азывается ***правилом параллелограмма.***

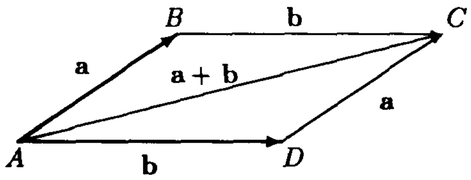


Рис. 2

Теорема 11.1. *Операция сложения векторов обладает сле-ющими свойствами:*

1) $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$, $\forall \mathbf{a},\mathbf{b}$ *(свойство коммутативности);*

2) $(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})$, $\forall \mathbf{a},\mathbf{b},\mathbf{c}$ *(свойство ассоциативности);*

3) существует такой вектор $\mathbf{0}$, *называемый нулевым вектором, по* $\mathbf{a}+\mathbf{0}=\mathbf{0}+\mathbf{a}=\mathbf{a}$, $\forall \mathbf{a}$ *(свойство существования нейтрального емента);*

4) для любого вектора $\mathbf{a}$ *существует такой вектор* $-\mathbf{a}$ *(называ-ый противоположным к вектору* $\mathbf{a}$*), что* $\mathbf{a}+(-\mathbf{a})=\mathbf{0}$ *(свойство ществования симметричного элемента).*

Доказательство. Коммутативность и ассоциативность сло-ения в случае неколлинеарных векторов $\mathbf{a}$, $\mathbf{b}$ и $\mathbf{c}$ проверяются не-средственным построением (рис. 3) векторов левой и правой ча-ей соответствующих равенств. Случай коллинеарности предлага-ся рассмотреть читателю.

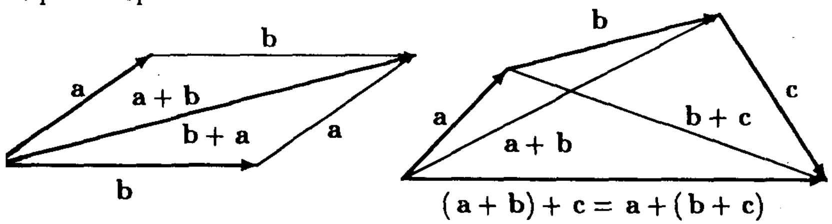


Рис. 3

Свойства 3 и 4 очевидны: нулевым вектором $\mathbf{0}$ будет класс эквива-нтности нулевых направленных отрезков, противоположным к век-ру $\mathbf{a}=\overrightarrow{AB}$ будет вектор $-\mathbf{a}=\overrightarrow{BA}$. ■

Разностью векторов $\mathbf{b}$ и $\mathbf{a}$ называется вектор $\mathbf{x}$ такой, что $\mathbf{a}+\mathbf{x}=$

Обозначение: $\mathbf{b}-\mathbf{a}$.

Теорема 11.2. *Для любых векторов* $\mathbf{a}$ *и* $\mathbf{b}$ *существует, и итом единственная, разность* $\mathbf{b}-\mathbf{a}$.

Доказательство. В качестве разности $\mathbf{b}-\mathbf{a}$ можно взять ктор $\mathbf{b}+(-\mathbf{a})$, так как $\mathbf{a}+(\mathbf{b}+(-\mathbf{a}))=\mathbf{a}+((-\mathbf{a})+\mathbf{b})=$

$(\mathbf{a}+(-\mathbf{a}))+\mathbf{b}=\mathbf{0}+\mathbf{b}=\mathbf{b}$. Эта разность единственна, так как если $\mathbf{c}$ – еще одна разность, то $\mathbf{c}=\mathbf{c}+\mathbf{0}=(\mathbf{c}+\mathbf{a})+(-\mathbf{a})=\mathbf{b}+(-\mathbf{a})$. ■

Замечание. Правило параллелограмма сложения неколлинеарных векторов $\mathbf{a}$ и $\mathbf{b}$ позволяет построить и разность $\mathbf{b}-\mathbf{a}$ как другую диагональ параллелограмма (рис. 4).

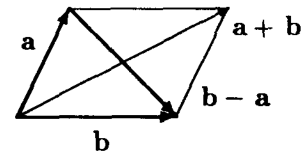


Рис. 4

Умножение вектора на число. *Произведением вектора* $\mathbf{a}$ *на вещественное число* α называется вектор $\mathbf{b}$, удовлетворяющий следующим условиям:

1) $|\mathbf{b}|=|\alpha|\cdot|\mathbf{a}|$
и, в случае $\mathbf{b}\neq\mathbf{0}$,

2) $\mathbf{b}\uparrow\uparrow\mathbf{a}$, если $\alpha>0$, и $\mathbf{b}\uparrow\downarrow\mathbf{a}$, если $\alpha<0$.

Обозначение: $\mathbf{b}=\alpha\mathbf{a}$. Очевидно, что $0\,\mathbf{a}=\alpha\,\mathbf{0}=\mathbf{0}$.

Теорема 11.3. *Операция умножения вектора на число обладает следующими свойствами: для любых векторов* $\mathbf{a}$, $\mathbf{b}$ *и чисел* $\alpha,\beta\in\mathbb{R}$

1) $1\cdot\mathbf{a}=\mathbf{a}$*;*

2) $(\alpha\beta)\,\mathbf{a}=\alpha(\beta\,\mathbf{a})$*;*

3) $(\alpha+\beta)\,\mathbf{a}=\alpha\,\mathbf{a}+\beta\,\mathbf{a}$ *(свойство дистрибутивности умножения на число относительно сложения чисел);*

4) $\alpha(\mathbf{a}+\mathbf{b})=\alpha\,\mathbf{a}+\alpha\,\mathbf{b}$ *(свойство дистрибутивности умножения на число относительно сложения векторов).*

Доказательство. Свойство 1 очевидно. Свойства 2 и 3 проверяются перебором различных вариантов знаков и абсолютных значений чисел α и β. Свойство 4 вытекает из подобия треугольников (рис. 5). Здесь следует отдельно рассмотреть случай, когда векторы $\mathbf{a}$ и $\mathbf{b}$ коллинеарны. ■

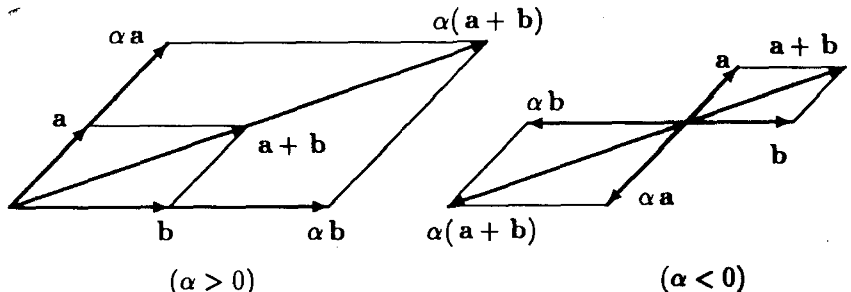


Рис. 5

§12. Векторы на прямой, на плоскости и в пространстве

До сих пор мы рассматривали свободные векторы в пространстве. Мы определили векторы и операции над ними, исходя из направленных отрезков, начала и концы которых – любые точки пространства. Свойства этих операций относились к произвольным векторам пространства.

Можно дать такое же определение вектора и операций над ними, оставаясь в пределах некоторой плоскости или даже прямой. При этом оказывается, что все факты, изложенные в § 10, 11, останутся, по существу, неизменными, незначительное изменение коснется лишь их формулировок: в них добавится уточнение, какое именно множество векторов рассматривается.

В самом деле, пусть V_1, V_2 и V_3 – множества всех векторов на прямой, на плоскости и в пространстве соответственно. Как следует из определения суммы векторов, если $\mathbf{a}, \mathbf{b} \in V_1$ (V_2 или V_3), то $\mathbf{a}+\mathbf{b} \in V_1$ (V_2 или V_3 соответственно), т.е. операция сложения векторов не выводит за пределы данной прямой (плоскости или пространства). Значит, операция сложения векторов является алгебраической операцией (или внутренним законом композиции) на каждом из множеств V_1, V_2 или V_3 (§9). Аналогично операция умножения вектора на действительное число является внешним законом композиции на каждом из этих множеств. Если к этому добавить еще, что каждое из них содержит нулевой вектор $\mathbf{0}$ и противоположный вектор $-\mathbf{a}$ к любому своему вектору $\mathbf{a}$, то теоремы 11.1–11.3 можно отнести к каждому из множеств V_1, V_2, V_3. Сформулируем этот итог в виде следующего утверждения.

Утверждение. *Множество V_n, где $n = 1,2,3$, всех векторов на прямой (на плоскости или в пространстве) наделено внутренним законом композиции (называемым сложением) и внешним законом композиции (называемым умножением на число), которые обладают следующими свойствами: для любых $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V_n$ и $\alpha, \beta \in \mathbb{R}$*

1) $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$;
2) $(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})$;
3) $\exists\, \mathbf{0} \in V_n : \mathbf{a}+\mathbf{0}=\mathbf{a},\ \forall \mathbf{a} \in V_n$;
4) $\forall \mathbf{a} \in V_n \ \ \exists\, (-\mathbf{a}) \in V_n : \mathbf{a}+(-\mathbf{a})=\mathbf{0}$;
5) $1 \cdot \mathbf{a}=\mathbf{a},\ \forall \mathbf{a} \in V_n$;
6) $(\alpha\beta)\mathbf{a}=\alpha(\beta\mathbf{a})$;
7) $(\alpha+\beta)\mathbf{a}=\alpha\mathbf{a}+\beta\mathbf{a}$;
8) $\alpha(\mathbf{a}+\mathbf{b})=\alpha\mathbf{a}+\alpha\mathbf{b}$.

Отметим, что, прежде чем сделать этот вывод, мы проверили лишь факт наличия законов композиции и справедливость свойств 3 и 4. Очевидно, остальные свойства не нуждаются в проверке.

К свойствам 1–8 остается добавить, что операция сложения векторов на каждом из множеств V_1, V_2, V_3 обладает обратной операцией,

называемой вычитанием.

Замечание. Свойства 1–8 говорят о том, что множества V_1, V_2, V_3, будучи, вообще говоря, различными, обладают общими свойствами. Такими же свойствами обладает и совсем непохожее на них множество $\mathbb{R}^{m\times n}$ матриц размера $m\times n$ (§2). Очевидно, это относится и к множеству $\mathbb{R}$ всех действительных чисел, если операцию умножения чисел рассматривать как внешний закон композиции. Уже эти примеры наводят на мысль о целесообразности общего взгляда на эти множества. В следующей главе мы увидим, что рассмотренные в примерах множества – это лишь частные проявления того круга формальных понятий, который составляет основу линейной алгебры и аналитической геометрии.

Глава IV. Введение в теорию линейных пространств

§13. Вещественное линейное пространство

Мы будем рассматривать множества, наделенные двумя законами композиции: внутренним и внешним (§9). Внутренний закон композиции будем называть сложением, а внешний – умножением на вещественное число. Согласно этой терминологии на этих множествах указаны два правила:

• правило, посредством которого любой упорядоченной паре элементов a, b множества ставится в соответствие однозначно определенный элемент $a+b$ из этого же множества, называемый суммой элементов a и b;

• правило, посредством которого любому вещественному числу α и любому элементу a множества ставится в соответствие однозначно определенный элемент αa этого же множества, называемый произведением элемента a на число α.

Непустое множество V называется ***вещественным линейным пространством*** [1], если на нем заданы два закона композиции:

внутренний закон композиции, подчиненный аксиомам

1) $a+b=b+a$, $\forall a,b\in V$ (аксиома коммутативности),

2) $(a+b)+c=a+(b+c)$, $\forall a,b,c\in V$ (аксиома ассоциативности),

3) $\exists\, \theta\in V: a+\theta=a$, $\forall a\in V$,

4) $\forall\, a\in V\ \ \exists\,(-a)\in V$: $a+(-a)=\theta$;

внешний закон композиции, подчиненный аксиомам

5) $1\cdot a=a$, $\forall a\in V$,

6) $(\alpha\beta)a=\alpha(\beta a)$, $\forall\alpha,\beta\in\mathbb{R},\forall a\in V$;

и если оба закона связаны между собой аксиомами

7) $(\alpha+\beta)a=\alpha a+\beta a$, $\forall\alpha,\beta\in\mathbb{R},\forall a\in V$ (аксиома дистрибутивности умножения на число относительно сложения чисел),

8) $\alpha(a+b)=\alpha a+\alpha b$, $\forall\alpha\in\mathbb{R},\forall a,b\in V$ (аксиома дистрибутивности умножения на число относительно сложения элементов V).

Линейное пространство называют также *векторным пространством.*

Отметим, что определение линейного пространства не содержит каких-либо конкретных описаний его элементов и операций над ними. Требуется только, чтобы эти операции были определены и подчинялись сформулированным выше восьми аксиомам, называемым *аксиомами линейного пространства.* Это означает,

[1] В гл. XII будут рассматриваться линейные пространства общего вида, когда в операции умножения на число участвуют не только вещественные числа. Забегая вперед, отметим, что все результаты этой главы будут справедливы и в общем случае.

что каждый раз, когда мы встречаемся с множествами, наделенными операциями, удовлетворяющими перечисленным аксиомам, мы вправе считать их линейными пространствами и переносить на них все результаты теории линейных пространств.

Элементы линейного пространства называются ***векторами.*** Использование этого термина не случайно, геометричность терминологии подчеркивает геометрический характер абстрактных алгебраических понятий, делает их "наглядными" и помогает уяснить, а зачастую и предвидеть не всегда очевидные факты.

Вектор θ называется ***нулевым*** вектором пространства, а вектор $(-a)$ – ***противоположным к вектору a.*** Нулевой вектор обозначают также символом $\mathbf{0}$.

Разностью векторов b и a линейного пространства V называется вектор $x\in V$ такой, что $a+x=b$. Обозначение: $b-a$.

Примеры. 1. Геометрические пространства V_1, V_2, V_3. Как следует из §12, множества V_1, V_2, V_3 всех векторов на прямой, на плоскости и в пространстве являются вещественными линейными пространствами относительно введенных в них операций сложения векторов и умножения вектора на действительное число. Эти пространства будем называть ***геометрическими.***

Для изображения геометрических пространств V_1, V_2, V_3 условимся все векторы откладывать от одной фиксированной точки O на прямой (на плоскости и в пространстве соответственно). При таком соглашении каждый свободный вектор будет однозначно определен своим концом. В этом смысле мы будем, говоря о свободном векторе, указывать только его конец. Точку O будем (пока условно) называть началом координат (рис. 1).

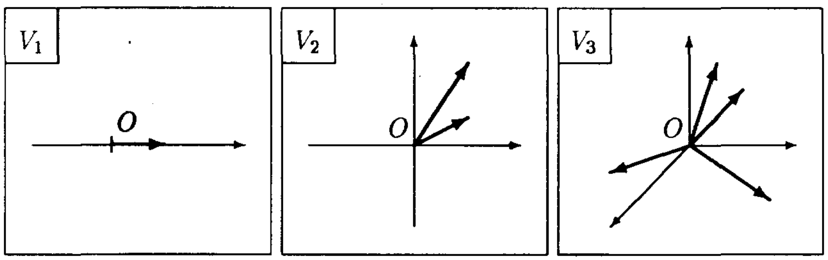


Рис. 1

2. Пространство вещественных матриц $\mathbb{R}^{m\times n}$. Как следует из теорем 2.1 и 2.2, множество $\mathbb{R}^{m\times n}$ всех вещественных матриц размера $m\times n$ является вещественным линейным пространством.

3. Арифметическое (координатное) пространство $\mathbb{R}^n$. Пусть $\mathbb{R}^n$ – множество всевозможных упорядоченных наборов n действительных чисел, называемых ***арифметическими векторами*** (или ***n-мерными векторами***). Если арифметические векторы записывать в виде $a=(a_1,\ldots,a_n)$, то $\mathbb{R}^n=\{a=(a_1,\ldots,a_n)\,|\,a_i\in\mathbb{R}, i=\overline{1,n}\}$.

Два арифметических вектора $a=(a_1,\ldots,a_n)$ и $b=(b_1,\ldots,b_n)$ называются ***равными***, если $a_i=b_i$, $i=\overline{1,n}$.

Операции над арифметическими векторами вводятся следующим образом: $a+b=(a_1+b_1,\ldots,a_n+b_n)$, $\alpha a=(\alpha a_1,\ldots,\alpha a_n)$, $\alpha\in\mathbb{R}$.

Нетрудно проверить, что $\mathbb{R}^n$ – вещественное линейное пространство относительно введенных операций.

Замечание 1. Отметим, что $\mathbb{R}^n$ – декартова n-я степень множества $\mathbb{R}$ всех вещественных чисел.

Замечание 2. Иногда из соображений удобства мы будем записывать арифметические векторы в виде столбцов.

Замечание 3. Арифметическое пространство $\mathbb{R}^n$ совпадает с пространством матриц $\mathbb{R}^{1\times n}$ или $\mathbb{R}^{n\times 1}$.

4. Пространства многочленов. ***Многочленом n-й степени*** от одной переменной t с вещественными коэффициентами называется выражение вида $f(t)=a_0+a_1t+a_2t^2+\ldots+a_nt^n$, где $a_i\in\mathbb{R}$, $i=\overline{0,n}$, причем $a_n\neq 0$. Число $0\in\mathbb{R}$ по определению считается многочленом с нулевыми коэффициентами и называется ***нулевым многочленом.*** Степень нулевого многочлена не определена.

Два многочлена $f(t)=\sum_{k=0}^n a_kt^k$ и $g(t)=\sum_{k=0}^m b_kt^k$ называются ***равными***, если $n=m$ и $a_k=b_k$, $k=\overline{0,n}$.

Суммой многочленов $f(t)=\sum_{k=0}^n a_kt^k$ и $g(t)=\sum_{k=0}^m b_kt^k$ называется многочлен $h(t)=\sum_{k\geq 0}(a_k+b_k)t^k$, где недостающие коэффициенты (a_k или b_k) заменяются нулями. Обозначение: $f(t)+g(t)$.

Произведением многочлена $f(t)=\sum_{k=0}^n a_kt^k$ ***на число*** $\alpha\in\mathbb{R}$ называется многочлен $\alpha f(t)=\sum_{k=0}^n \alpha a_kt^k$.

Нетрудно проверить, что множества M_n всех многочленов степени не выше n и множество M_∞ многочленов всех степеней, пополненные нулевым многочленом, образуют вещественные линейные пространства.

Простейшие свойства линейных пространств. Следующие свойства линейных пространств являются элементарными следствиями из аксиом.

1°. *В линейном пространстве существует единственный нулевой вектор*, так как если θ_1 и θ_2 – два нулевых вектора, то из аксиомы 3 следует, что $\theta_1=\theta_1+\theta_2=\theta_2$.

Замечание 4. Говоря о единственности нулевого вектора, мы не различаем равные векторы. В этом же смысле следует понимать и другие утверждения о единственности.

2°. *Для любого вектора линейного пространства существует единственный противоположный вектор*, так как если b и c – два противоположных вектора к вектору a, то, последовательно применяя аксиомы 3, 4, 2, получим, что $b=b+(a+c)=(b+a)+c=c$.

3°. *В линейном пространстве справедливы равенства*: $0a=\theta$, $\forall a\in V$ *и* $\alpha\theta=\theta$, $\forall\alpha\in V$.

Доказательство. Для доказательства первого равенства достаточно проверить, что $b+0a=b$, $\forall b\in V$. Это соотношение вытекает из следующей цепочки равенств, основанных на аксиомах 2–7:

$$b+0a=(b+\theta)+0a=b+((-a)+a)+0a=(b+(-a))+a+0a=$$

$$=(b+(-a))+1a+0a=(b+(-a))+(1+0)a=(b+(-a))+a=$$
$$=b+((-a)+a)=b+\theta=b.$$

Второе равенство доказывается с помощью первого и аксиомы 6: если a – произвольный вектор пространства, то $\alpha\theta=\alpha(0a)=(\alpha 0)a=0a=\theta$. ■

4°. *В линейном пространстве из равенства $\alpha a=\theta$ следует, что либо $\alpha=0$, либо $a=\theta$.*

В самом деле, как следует из свойства 3°, случай $\alpha=0$ возможен, если $\alpha a=\theta$. В случае когда $\alpha\neq 0$, на основании свойства 3° и аксиом 5, 6 получим

$$a=1a=\left(\frac{1}{\alpha}\alpha\right)a=\frac{1}{\alpha}(\alpha a)=\frac{1}{\alpha}\theta=\theta.$$

5°. *В линейном пространстве для любого вектора a противоположный вектор может быть получен как произведение*

$$-a=(-1)a.$$

Это утверждение вытекает из аксиом 3–5, 7 и свойства 3°, так как

$$a+(-1)a=1a+(-1)a=(1-1)a=0a=\theta.$$

6°. *Для любой пары векторов a и b линейного пространства существует, и притом единственная, разность $b-a$.*

Доказательство. Вектор $b+(-a)$ является разностью $b-a$ векторов a и b, так как на основании аксиом 1–4 и определения разности имеем

$$a+(b+(-a))=(a+(-a))+b=\theta+b=b.$$

При этом если c – любая другая разность $b-a$, то из аксиом 2–4 следует, что $c=c+\theta=c+(a+(-a))=(c+a)+(-a)=b+(-a)$. ■

§14. Линейная зависимость

Пусть $a_1,a_2,\ldots,a_k$ – векторы линейного пространства V и $\alpha_1,\alpha_2,\ldots,\alpha_k$ – действительные числа. Вектор $\alpha_1a_1+\alpha_2a_2+\ldots+\alpha_ka_k$ называется ***линейной комбинацией векторов*** $a_1,a_2,\ldots,a_k$ ***с коэффициентами*** $\alpha_1,\alpha_2,\ldots,\alpha_k$. Если вектор b является линейной комбинацией векторов $a_1,a_2,\ldots,a_k$, то говорят, что вектор b ***линейно выражается через векторы*** $a_1,a_2,\ldots,a_k$, при этом представление вектора b в виде $b=\alpha_1a_1+\ldots+\alpha_ka_k$ называют ***разложением вектора b по векторам*** $a_1,a_2,\ldots,a_k$. Очевидно, нулевой вектор линейно выражается через любой вектор, так как $\theta=0a$, $\forall a\in V$. Через нулевой вектор не может выражаться ни один ненулевой вектор, так как $\alpha\theta=\theta$, $\forall\alpha\in\mathbb{R}$.

Линейная комбинация называется ***тривиальной***, если все ее коэффициенты равны нулю, и ***нетривиальной***, если среди ее коэффициентов хотя бы один отличен от нуля. Очевидно, тривиальная комбинация любой системы векторов равна нулевому вектору.

Система векторов a_1, a_2, ..., a_k называется ***линейно зависимой***, если существует нетривиальная линейная комбинация этих векторов, равная нулевому вектору, т.е. если существуют числа $\alpha_1,\alpha_2,\ldots,\alpha_k$, одновременно не равные нулю, такие, что

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_k a_k = \theta, \tag{14.1}$$

и ***линейно независимой***, если нулевому вектору равна только тривиальная линейная комбинация этих векторов, т.е. если из равенства (14.1) следует, что $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$.

Теорема 14.1. *Система из одного вектора линейно зависима тогда и только тогда, когда этот вектор нулевой.*

Доказательство. Линейная зависимость системы из одного вектора a равносильна тому, что $\alpha a = \theta$ при некотором $\alpha \neq 0$, а это, в свою очередь, равносильно (§13, свойства 3, 4) тому, что $a = \theta$. ■

Теорема 14.2. *Система векторов $a_1, a_2, \ldots, a_k$, где $k > 1$, линейно зависима тогда и только тогда, когда хотя бы один из векторов этой системы линейно выражается через другие.*

Доказательство. Необходимость. Если система векторов $a_1, a_2, \ldots, a_k$ линейно зависима, то существуют числа α_1, α_2, $\ldots$, α_k, одновременно не равные нулю и такие, что

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_i a_i + \ldots + \alpha_k a_k = \theta. \tag{14.2}$$

Пусть $\alpha_i \neq 0$. Тогда в силу (14.2) $a_i = \sum_{s \neq i} (-\alpha_s/\alpha_i)\, a_s$. Так как $k > 1$, то для вектора a_i существует хотя бы один "другой" вектор системы.

Достаточность. Пусть $a_i = \sum_{s \neq i} \alpha_s a_s$. Тогда, перенеся правую часть этого равенства в левую, получим нетривиальную линейную комбинацию векторов системы, равную нулевому вектору:

$$-\alpha_1 a_1 - \ldots - \alpha_{i-1} a_{i-1} + 1 a_i - \alpha_{i+1} a_{i+1} - \ldots - \alpha_k a_k = \theta. \blacksquare$$

Эта теорема дает другое определение линейной зависимости системы векторов, в которой более одного вектора.

Теорема 14.3. *Если подсистема системы векторов линейно зависима, то и вся система линейно зависима.*

Доказательство. Не ограничивая общности рассуждений, можно считать подсистемой системы векторов $a_1, a_2, \ldots, a_s, \ldots, a_k$ ее первые s векторов. Из линейной зависимости $a_1, a_2, \ldots, a_s$ следует, что $\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_s a_s = \theta$ для некоторых чисел $\alpha_1, \alpha_2, \ldots, \alpha_s$, среди которых существует $\alpha_i \neq 0$. Тогда $\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_s a_s + 0a_{s+1} + \ldots + 0a_k = \theta$, причем $\alpha_i \neq 0$. Следовательно, система векторов $a_1, a_2, \ldots, a_s, \ldots, a_k$ линейно зависима. ■

Теорема 14.4. *Любая подсистема линейно независимой системы векторов линейно независима.*

В самом деле, если бы существовала линейно зависимая подсистема, то на основании теоремы 14.3 вся система была бы линейно зависимой. ■

Теорема 14.5. *Система векторов $a_1, a_2, \ldots, a_k$ линейно независима тогда и только тогда, когда любой вектор, являющийся линейной комбинацией этих векторов, имеет единственное разложение по этим векторам.*

Доказательство. Необходимость доказывается от противного. Пусть существует вектор b, который имеет два различных разложения по векторам $a_1, a_2, \ldots, a_k$: $b = \sum_{i=1}^{k} \alpha_i a_i$, $b = \sum_{i=1}^{k} \alpha'_i a_i$, $\exists \alpha_s \neq \alpha'_s$. Вычитая почленно одно равенство из другого, получим нетривиальную линейную комбинацию векторов $a_1, a_2, \ldots, a_k$, равную нулевому вектору. Отсюда следует линейная зависимость a_1, a_2, $\ldots$, a_k.

Достаточность также доказывается от противного. Пусть система $a_1, a_2, \ldots, a_k$ линейно зависима, тогда

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_k a_k = \theta \tag{14.3}$$

для некоторых чисел $\alpha_1, \alpha_2, \ldots, \alpha_k$, среди которых существует $\alpha_s \neq 0$. Но, с другой стороны,

$$0a_1 + 0a_2 + \ldots + 0a_k = \theta. \tag{14.4}$$

Мы получили два различных разложения (14.3) и (14.4) нулевого вектора θ по векторам $a_1, a_2, \ldots, a_k$. ■

Теорема 14.6. *Если система векторов $a_1, a_2, \ldots, a_k$ линейно независима, а система $a_1, a_2, \ldots, a_k, b$ линейно зависима, то вектор b линейно выражается через векторы $a_1, a_2, \ldots, a_k$.*

Доказательство. Из линейной зависимости системы векторов a_1, a_2, $\ldots$, a_k, b следует, что

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_k a_k + \alpha_0 b = \theta \tag{14.5}$$

для некоторых чисел $\alpha_1, \alpha_2, \ldots, \alpha_k, \alpha_0$, среди которых хотя бы одно отлично от нуля. Если $\alpha_0 = 0$, то ненулевой коэффициент α_s находится среди чисел $\alpha_1, \alpha_2, \ldots, \alpha_k$; при этом (14.5) переходит в равенство $\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_k a_k = \theta$, где $\alpha_s \neq 0$, $1 \leq s \leq k$, которое противоречит условию линейной независимости $a_1, a_2, \ldots, a_k$. Следовательно, $\alpha_0 \neq 0$; отсюда и из (14.5) получаем, что $b = \sum_{i=1}^{k} (-\alpha_i/\alpha_0)\, a_i$. ■

Примеры. 1. Арифметическое пространство $\mathbb{R}^n$. В арифметическом пространстве $\mathbb{R}^n$ единичные векторы

$$\begin{aligned} e_1 &= (1, 0, 0, \ldots, 0), \\ e_2 &= (0, 1, 0, \ldots, 0), \\ &\ldots \\ e_n &= (0, 0, 0, \ldots, 1) \end{aligned} \tag{14.6}$$

линейно независимы. Это следует из того, что линейная комбинация этих векторов с коэффициентами $\alpha_1, \alpha_2, \ldots, \alpha_n$ представляет собой арифметический вектор $(\alpha_1, \alpha_2, \ldots, \alpha_n)$, который равен нулевому вектору $\theta = (0, \ldots, 0)$ тогда и только тогда, когда $\alpha_i = 0$, $i = \overline{1, n}$.

2. Пространство многочленов. Многочлены $1, t, t^2, \ldots, t^n$ линейно независимы. Это следует из того, что линейная комбинация этих многочленов с коэффициентами $\alpha_0, \alpha_1, \ldots, \alpha_n$ представляет собой многочлен $\sum_{k=0}^{n} \alpha_k t^k$, который равен нулевому многочлену тогда и только тогда, когда $\alpha_k = 0$, $k = \overline{0, n}$.

§15. Геометрический смысл линейной зависимости

Будем рассматривать геометрические векторы на прямой, на плоскости и в пространстве. Выясним, что означает линейная зависимость геометрических векторов. Предварительно докажем несколько чисто геометрических утверждений.

Утверждение 1. *На прямой (на плоскости и в пространстве) существует ненулевой вектор (соответственно два неколлинеарных и три некомпланарных вектора).*

Доказательство. В случае прямой достаточно взять две несовпадающие точки O и A (рис. 1,*а*), тогда вектор $\mathbf{a} = \overrightarrow{OA} \neq \mathbf{0}$. На плоскости достаточно взять три точки O, A и B, не лежащие на одной прямой (рис. 1,*б*), тогда векторы $\mathbf{a} = \overrightarrow{OA}$ и $\mathbf{b} = \overrightarrow{OB}$ неколлинеарны. В пространстве достаточно взять четыре точки O, A, B, C, не лежащие в одной плоскости (рис. 1,*в*), тогда векторы $\mathbf{a} = \overrightarrow{OA}$, $\mathbf{b} = \overrightarrow{OB}$, $\mathbf{c} = \overrightarrow{OC}$ некомпланарны. ■

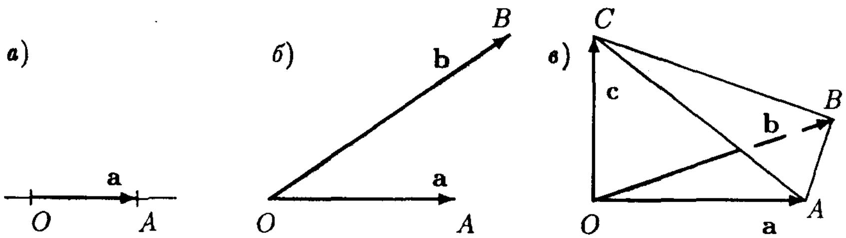


Рис. 1

Утверждение 2. *На прямой (на плоскости и в пространстве) всякий вектор линейно выражается через любой ненулевой вектор (соответственно любые два неколлинеарных и любые три некомпланарных вектора).*

Доказательство. 1. Пусть $\mathbf{a}$, $\mathbf{b}$ – векторы на прямой и $\mathbf{a} \neq \mathbf{0}$. Отложим их от одной точки O прямой. Пусть $\mathbf{a} = \overrightarrow{OA}$, $\mathbf{b} = \overrightarrow{OB}$ (рис. 2,*а*). Если $\mathbf{b} = \mathbf{0}$, то $\mathbf{b} = 0\,\mathbf{a}$. Если $\mathbf{b} \neq \mathbf{0}$, то, взяв

$$\alpha = \begin{cases} |\overrightarrow{OB}|/|\overrightarrow{OA}|, & \mathbf{a} \uparrow\uparrow \mathbf{b}, \\ -|\overrightarrow{OB}|/|\overrightarrow{OA}|, & \mathbf{a} \uparrow\downarrow \mathbf{b}, \end{cases}$$

согласно определению произведения вектора на число получим, что $\mathbf{b} = \alpha\,\mathbf{a}$.

2. Пусть $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ – векторы плоскости и $\mathbf{a}$, $\mathbf{b}$ неколлинеарны (значит, ни один из них не равен $\mathbf{0}$). Отложим эти векторы от одной точки O плоскости. Пусть $\mathbf{a} = \overrightarrow{OA}$, $\mathbf{b} = \overrightarrow{OB}$, $\mathbf{c} = \overrightarrow{OC}$ (рис. 2,*б*). Если $\mathbf{c} = \mathbf{0}$, то $\mathbf{c} = 0\,\mathbf{a} + 0\,\mathbf{b}$. Если $\mathbf{c} \neq \mathbf{0}$, то проведем из точки C прямые, параллельные прямым OB и OA, до пересечения с прямыми OA и OB соответственно. Пусть точки A_1, B_1 – точки пересечения этих прямых (существование точек пересечения следует из неколлинеарности $\overrightarrow{OA}$ и $\overrightarrow{OB}$). Тогда $\overrightarrow{OC} = \overrightarrow{OA_1} + \overrightarrow{OB_1}$. Отсюда и из первой части утверждения получим, что $\mathbf{c} = \alpha\,\mathbf{a} + \beta\,\mathbf{b}$.

3. Пусть $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ – векторы пространства и $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ некомпланарны (значит, попарно неколлинеарны и, тем более, ни один из них не равен $\mathbf{0}$). Отложим эти векторы от одной точки O. Пусть $\mathbf{a} = \overrightarrow{OA}$, $\mathbf{b} = \overrightarrow{OB}$, $\mathbf{c} = \overrightarrow{OC}$, $\mathbf{d} = \overrightarrow{OD}$ (рис. 2,*в*). Если $\mathbf{d} = \mathbf{0}$, то $\mathbf{d} = 0\,\mathbf{a} + 0\,\mathbf{b} + 0\,\mathbf{c}$. Если $\mathbf{d} \neq \mathbf{0}$, то проведем из точки D плоскости, параллельные плоскостям OBC, OAC и OAB (это плоскости, так как $\overrightarrow{OA}, \overrightarrow{OB}, \overrightarrow{OC}$ попарно неколлинеарны), до пересечения с прямыми OA, OB и OC соответственно. Пусть A_1, B_1, C_1 – точки пересечения (существование точек пересечения следует из некомпланарности $\overrightarrow{OA}$, $\overrightarrow{OB}$ и $\overrightarrow{OC}$). Тогда $\overrightarrow{OD} = \overrightarrow{OA_1} + \overrightarrow{OB_1} + \overrightarrow{OC_1}$. Отсюда и из первой части утверждения получим, что $\mathbf{c} = \alpha\,\mathbf{a} + \beta\,\mathbf{b} + \gamma\,\mathbf{c}$. ■

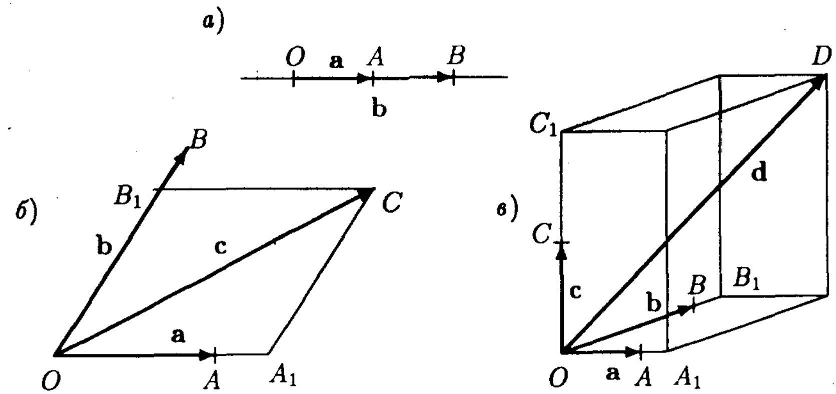


Рис. 2

Теорема 15.1. *Два вектора линейно зависимы тогда и только тогда, когда они коллинеарны.*

Доказательство. Необходимость. Пусть векторы $\mathbf{a}$, $\mathbf{b}$ линейно зависимы, тогда в силу теоремы 14.2 один из них линейно выражается через другой. Пусть $\mathbf{b} = \alpha\,\mathbf{a}$. Отсюда и из определения произведения вектора на число следует коллинеарность $\mathbf{a}$ и $\mathbf{b}$.

Достаточность. Пусть $\mathbf{a}$ и $\mathbf{b}$ коллинеарны, т.е. параллельны одной прямой. Будем считать, что $\mathbf{a} \neq 0$ (так как если $\mathbf{a} = 0$, то линейная зависимость $\mathbf{a}$, $\mathbf{b}$ следует из теорем 14.1 и 14.3). Отложим $\mathbf{a}$, $\mathbf{b}$ от одной точки. Тогда они окажутся на одной прямой, при этом, согласно утверждению 2, $\mathbf{b} = \alpha\,\mathbf{a}$. В силу теоремы 14.2 отсюда следует линейная зависимость $\mathbf{a}$, $\mathbf{b}$. ■

Следствие 1. *Любые два (значит, и более) вектора прямой линейно зависимы.*

Теорема 15.2. *Три вектора линейно зависимы тогда и только тогда, когда они компланарны.*

Доказательство. Необходимость. Пусть векторы $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ линейно зависимы, тогда один из них линейно выражается через другие. Пусть $\mathbf{c} = \alpha\,\mathbf{a} + \beta\,\mathbf{b}$. Если $\mathbf{a}$ и $\mathbf{b}$ коллинеарны, то $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ коллинеарны и, тем более, компланарны. Если $\mathbf{a}$ и $\mathbf{b}$ неколлинеарны, то отложим векторы $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ от одной точки (рис. 2,*б*). Тогда вектор $\mathbf{c}$, являясь диагональю параллелограмма, построенного на векторах $\alpha\,\mathbf{a}$ и $\beta\,\mathbf{b}$, окажется в той же плоскости, что и $\mathbf{a}$, $\mathbf{b}$. Значит, $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ компланарны.

Достаточность. Пусть $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ компланарны, т.е. параллельны одной плоскости. Будем считать, что $\mathbf{a}$, $\mathbf{b}$ неколлинеарны (так как если $\mathbf{a}$, $\mathbf{b}$ коллинеарны, то линейная зависимость $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ следует из линейной зависимости подсистемы). Отложим $\mathbf{a}$, $\mathbf{b}$ и $\mathbf{c}$ от одной точки. Тогда они окажутся в одной плоскости и на основании утверждения 2 будем иметь $\mathbf{c} = \alpha\,\mathbf{a} + \beta\,\mathbf{b}$. В силу теоремы 14.2 отсюда следует, что векторы $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ линейно зависимы. ■

Следствие 2. *Любые три (значит, и более) вектора плоскости линейно зависимы.*

Теорема 15.3. *Любые четыре вектора линейно зависимы.*

Доказательство. Будем считать, что в четверке векторов $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ векторы $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ некомпланарны (так как если $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ компланарны, то линейная зависимость $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ вытекает из линейной зависимости подсистемы). Тогда на основании утверждения 2 будем иметь $\mathbf{d} = \alpha\,\mathbf{a} + \beta\,\mathbf{b} + \gamma\,\mathbf{c}$. В силу теоремы 14.2 отсюда следует, что векторы $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ линейно зависимы. ■

Итак, понятие линейной зависимости является обобщением понятий коллинеарности и компланарности геометрических векторов.

Трудно не заметить логическое "однообразие" формулировок и доказательств изложенных выше теорем. Мы не старались уходить от повторений, чтобы подчеркнуть, как похожи по своей структуре три различных линейных пространства V_1, V_2 и V_3 – векторов на прямой, на плоскости и в пространстве. Резюмируем полученные результаты. Итак, для $n = 1, 2, 3$ в линейном пространстве V_n

1) существует линейно независимая система из n векторов, а любая система из большего числа векторов линейно зависима;

2) существует линейно независимая система из n векторов, а любой вектор пространства V_n линейно выражается через них;

3) максимальное число линейно независимых векторов равно n.

Мы выделили общие свойства, характерные для различных линейных пространств V_1, V_2 и V_3. Возникает естественный вопрос, не являются ли они общими для всех линейных пространств. Прежде чем ответить на этот вопрос, обратимся вновь к матрицам.

§16. Ранг матрицы

Понятие линейной зависимости векторов тесно связано с понятием матрицы. Если рассматривать строки и столбцы матрицы $A \in \mathbb{R}^{m\times n}$ как векторы арифметических пространств $\mathbb{R}^n$ и $\mathbb{R}^m$, то оказывается, что одна из важнейших характеристик матрицы определяется свойством линейной зависимости ее строк и столбцов. В свою очередь, эта характеристика будет удобным аппаратом дальнейшего исследования линейных пространств.

Ранг матрицы и линейная зависимость. *Рангом* ненулевой матрицы называется максимальный порядок ненулевых миноров этой матрицы. Ранг нулевой матрицы по определению считается равным нулю. Обозначение: $\operatorname{rg} A$, $\operatorname{rang} A$, r_A и др.

Из определения вытекают следующие факты:

1) ранг матрицы не превосходит ее размеров: если $A \in \mathbb{R}^{m\times n}$, то

$$\operatorname{rg} A \le \min(m, n); \tag{16.1}$$

2) равенство $\operatorname{rg} A = r > 0$ равносильно выполнению двух условий:

а) в матрице A существует ненулевой минор r-го порядка,

б) любой минор более высокого порядка равен нулю.

Пусть $\operatorname{rg} A = r > 0$. Любой ненулевой минор r-го порядка этой матрицы называется ***базисным минором***, а строки и столбцы, в которых расположен базисный минор, – ***базисными строками*** и ***столбцами***.

Разумеется, у матрицы может быть не один базисный минор, но все они имеют один и тот же порядок, равный рангу этой матрицы.

Теорема 16.1 (о базисном миноре). *Базисные строки (столбцы) линейно независимы. Любая строка (столбец) является линейной комбинацией базисных строк (столбцов).*

Доказательство. Докажем столбцовый вариант теоремы. Пусть $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ и $\operatorname{rg} A = r > 0$. Не нарушая общности рассуждений, будем считать, что базисный минор M_r находится в левом верхнем углу. Тогда базисными столбцами будут первые r столбцов: $a_1, a_2, \ldots, a_r$.

Линейная независимость базисных столбцов доказывается от противного. В самом деле, пусть это не так. Тогда на основании теоремы 14.2 один из базисных столбцов является линейной комбинацией других базисных столбцов. Но тогда и в миноре M_r один из столбцов будет линейной комбинацией других столбцов и $M_r = 0$. Это противоречит тому, что M_r – базисный минор.

Пусть теперь a_k – произвольный столбец матрицы A. Покажем, что он является линейной комбинацией столбцов $a_1, a_2, \ldots, a_r$. Будем считать, что $k > r$ (в случае $k \le r$ утверждение очевидно, так как $a_k = 0a_1 + \ldots + 0a_{k-1} + a_k + 0a_{k+1} + \ldots + 0a_r$). Найдем числа $\alpha_1, \alpha_2, \ldots, \alpha_r \in \mathbb{R}$ такие, что $a_k = \sum_{s=1}^{r} \alpha_s a_s$ или, в поэлементной записи,

$$a_{ik} = \sum_{s=1}^{r} \alpha_s a_{is}, \quad i = \overline{1, m}. \tag{16.2}$$

Для этого составим вспомогательную матрицу

$$\Delta_i = \begin{bmatrix} a_{11} & \dots & a_{1r} & a_{1k} \\ \dots & \dots & \dots & \dots \\ a_{r1} & \dots & a_{rr} & a_{rk} \\ a_{i1} & \dots & a_{ir} & a_{ik} \end{bmatrix},$$

полученную окаймлением минора M_r i-й строкой и k-м столбцом матрицы A. Очевидно, $|\Delta_i| = 0$, $i = \overline{1,m}$, так как в случае, когда $i \le r$, матрица Δ_i содержит две одинаковые строки, а в случае, когда $i > r$, $|\Delta_i|$ является минором $(r+1)$-го порядка матрицы A ранга r. Разложив $|\Delta_i|$ по последней строке, получим $0 = a_{i1}A_1 + a_{i2}A_2 + \ldots + a_{ir}A_r + a_{ik}M_r$, где $A_1, A_2, \ldots, A_r$, являясь алгебраическими дополнениями к элементам вычеркиваемой строки, не зависят от i. Отсюда, положив $\alpha_s = -A_s/M_r$, $s = \overline{1,r}$, приходим к (16.2). ■

Следствие 1 (*необходимое и достаточное условие равенства нулю определителя*). *Определитель матрицы равен нулю тогда и только тогда, когда какая-либо ее строка (столбец) является линейной комбинацией других ее строк (столбцов).*

Доказательство. Достаточность доказана в §4 (свойство 8). Необходимость вытекает из теоремы о базисном миноре, так как из равенства $|A| = 0$ следует, что $\operatorname{rg} A < n$ и, значит, в матрице A есть хотя бы одна небазисная строка (столбец), которая является линейной комбинацией других (т.е. базисных) строк (столбцов). ■

Пусть в линейном пространстве даны две системы векторов. Если каждый вектор одной системы линейно выражается через векторы другой, то говорят, что *первая система линейно выражается через вторую*. Очевидно, что "линейная выражаемость" обладает свойством транзитивности, т.е. если система векторов $a_1, \ldots, a_k$ линейно выражается через $b_1, \ldots, b_n$, а система векторов $b_1, \ldots, b_n$ – через $c_1, \ldots, c_m$, то $a_1, \ldots, a_k$ линейно выражается через $c_1, \ldots, c_m$.

Теорема 16.2. *Если в линейном пространстве бóльшая система векторов линейно выражается через меньшую, то большая система линейно зависима* [2].

Доказательство. Пусть $a_1, \ldots, a_m$ и $b_1, \ldots, b_n$ – две системы векторов в линейном пространстве, $m > n$ и $a_i = \sum_{j=1}^{n} c_{ij} b_j$, $i = \overline{1,m}$. Тогда в матрице

$$C = \begin{bmatrix} c_{11} & \dots & c_{1n} \\ \dots & \dots & \dots \\ c_{m1} & \dots & c_{mn} \end{bmatrix}$$

число строк больше числа столбцов. Из (16.1) и условия теоремы следует, что $\operatorname{rg} A \le n < m$. Значит, в матрице C существует хотя бы одна небазисная строка (например, i-я), которая линейно выражается через другие строки. Тогда из свойств операций в пространстве $\mathbb{R}^n$ следует, что вектор a_i линейно выражается через другие векторы

[2] Теорема представляет собой другую формулировку теоремы, известной как "основная теорема о линейной зависимости" [9].

системы $a_1, \ldots, a_m$. Отсюда и из теоремы 14.2 следует линейная зависимость системы векторов $a_1, \ldots, a_m$. ■

Теорема 16.3. *Ранг матрицы равен максимальному числу ее линейно независимых строк (столбцов).*

Доказательство. Пусть $\operatorname{rg} A = r$ и $r > 0$ (случай $r = 0$ очевиден). Тогда в матрице A существует r линейно независимых строк (столбцов) – это ее базисные строки (столбцы). При этом любая система из большего числа строк (столбцов) линейно зависима на основании теорем 16.1, 14.2. ■

Следствие 2. $\operatorname{rg} A = \operatorname{rg} A^T$.

Замечание 1. Теорема 16.3 дает два новых определения ранга матрицы.

Теорема 16.4. *Если все строки (столбцы) матрицы A линейно выражаются через строки (столбцы) матрицы B, то*

$$\operatorname{rg} A \le \operatorname{rg} B.$$

Доказательство. Пусть $\operatorname{rg} A = r$, $\operatorname{rg} B = s$ и пусть $a_1, \ldots, a_r$ и $b_1, \ldots, b_s$ – базисные столбцы матриц A и B. Докажем, что $r \le s$. Пусть $r > s$, тогда согласно условию теоремы система $a_1, \ldots, a_r$ линейно выражается через систему столбцов матрицы B, которая, в свою очередь, в силу теоремы 16.1 линейно выражается через систему базисных столбцов $b_1, \ldots, b_s$. Отсюда следует, что $a_1, \ldots, a_r$ линейно выражается через $b_1, \ldots, b_s$, причем $r > s$. Из теоремы 16.2 вытекает линейная зависимость $a_1, \ldots, a_r$. Это противоречит тому, что $a_1, \ldots, a_r$ – базисные столбцы. ■

Теорема 16.5. *Ранг произведения матриц не превосходит рангов сомножителей.*

Доказательство. Пусть $C = AB$. Из (2.3), (2.2) следует, что строки матрицы C линейно выражаются через строки B, а столбцы C – через столбцы A. Отсюда в силу теоремы 16.4 вытекает, что $\operatorname{rg} C \le \operatorname{rg} B$, $\operatorname{rg} C \le \operatorname{rg} A$. ■

Ранг матрицы и элементарные преобразования.

Теорема 16.6. *Ранг матрицы не изменяется при умножении ее на невырожденную матрицу.*

Доказательство. Пусть $C = AP$, $|P| \ne 0$. Тогда на основании теоремы 16.5 имеем $\operatorname{rg} C \le \operatorname{rg} A$. С другой стороны, матрица P обратима (теорема 5.2) и $A = CP^{-1}$. Снова воспользовавшись теоремой 16.5, получим, что $\operatorname{rg} A \le \operatorname{rg} C$. Оба этих неравенства говорят о том, что $\operatorname{rg} C = \operatorname{rg} A$. Аналогично доказывается, что $\operatorname{rg} QA = \operatorname{rg} A$, если $|Q| \ne 0$. ■

Теорема 16.7. *Элементарные преобразования матрицы не изменяют ее ранга.*

Доказательство. Это утверждение следует из теоремы 3.2 и невырожденности матриц (3.1) элементарных преобразований ($|P_{ij}| = -1$, $|D_i| = \alpha \ne 0$, $|L_{ij}| = 1$). ■

Теорема 16.8. *Ранг матрицы не изменится, если из системы ее строк (столбцов) вычеркнуть или приписать строку (соответственно столбец), которая является линейной комбинацией других строк (соответственно столбцов).*

Доказательство. Рассмотрим вариант теоремы, относящийся к вычеркиванию строки. Прежде чем вычеркивать строку, являющуюся линейной комбинацией других строк, вычтем из нее эту линейную комбинацию. Тогда вычеркиваемая строка станет нулевой, а ранг матрицы согласно теореме 16.7 не изменится. Затем вычеркнем нулевую строку. При этом ранг матрицы останется прежним, так как нулевая строка не влияет на порядок ее ненулевых миноров. Аналогично доказываются другие варианты теоремы. ■

Метод Гаусса вычисления ранга. Теоретическую основу этого метода для решения данной задачи (см. §4) составляют следующие факты:

– ранг верхней (нижней) трапециевидной матрицы равен количеству ненулевых строк (соответственно столбцов);

– элементарные преобразования матрицы не изменяют ее ранга (теорема 16.7);

– любая матрица элементарными преобразованиями строк и столбцов приводится к трапециевидной форме (теорема 3.1).

Метод Гаусса вычисления ранга матрицы состоит в приведении этой матрицы элементарными преобразованиями к верхней (нижней) трапециевидной форме и подсчете ее ненулевых строк (столбцов).

Отметим, что метод почти копирует метод Гаусса вычисления определителя, отличие лишь в том, что в данной задаче он реализуется несколько проще, так как никакое элементарное преобразование *не изменяет* ранга матрицы.

Замечание 2. Среди других методов вычисления ранга выделим *метод окаймления миноров*. Минор M_{k+1} $(k+1)$-го порядка называется окаймляющим минор M_k k-го порядка, если M_k получается из M_{k+1} вычеркиванием одной строки и одного столбца. Метод окаймления миноров основан на следующем утверждении (докажите!): *если в матрице A существует ненулевой минор M_r r-го порядка, а все окаймляющие его миноры равны нулю, то* $\operatorname{rg} A = r$. Метод окаймления миноров состоит в поиске такого минора M_r путем перебора. Безусловно, он уступает методу Гаусса из-за большого объема требуемых вычислений, однако в некоторых случаях бывает полезным. Во всяком случае идея окаймления матрицы лежит в основе многих методов вычислительной алгебры.

Эквивалентные матрицы. Две матрицы $A, B \in \mathbb{R}^{m\times n}$ называются *эквивалентными*, если существуют невырожденные матрицы P и Q такие, что $A = PBQ$. Обозначение: $A \sim B$.

Теорема 16.9. *Эквивалентность матриц является отношением эквивалентности на множестве матриц* $\mathbb{R}^{m\times n}$.

Доказательство. В самом деле, рефлексивность отношения следует из того, что $A = I_{m\times m} A I_{n\times n}$, $\forall A \in \mathbb{R}^{m\times n}$; симметричность – из того, что если $A = PBQ$, то $B = P^{-1}AQ^{-1}$; транзитивность – из того, что если $A = P_1BQ_1$, $B = P_2CQ_2$, то $A = P_1P_2CQ_2Q_1 = PCQ$, где $P = P_1P_2$, $Q = Q_2Q_1$. ■

Теорема 16.10. *Любая ненулевая матрица $A \in \mathbb{R}^{m\times n}$ ранга r эквивалентна матрице $I_r \in \mathbb{R}^{m\times n}$ вида*

$$I_r = \left[\begin{array}{cccc|ccc} 1 & 0 & \dots & 0 & & & \\ 0 & 1 & \dots & 0 & & O & \\ & & \ddots & & & & \\ 0 & 0 & \dots & 1 & & & \\ \hline & & & & 0 & & \\ & O & & & & \ddots & \\ & & & & & & 0 \end{array}\right]$$

(здесь все элементы, кроме первых r диагональных элементов, равных 1, *равны* 0*).*

Доказательство. Покажем, что матрица A элементарными преобразованиями строк и столбцов приводится к виду I_r. В самом деле, строчный вариант *основного процесса* (§3) приводит ее к верхней ступенчатой форме. Если к получившейся матрице применить столбцовый вариант основного процесса, то получим диагональную матрицу. Так как элементарные преобразования матрицы не изменяют ее ранга, то ровно r первых диагональных элементов этой матрицы отличны от нуля. Поделив каждую из первых r строк на диагональный элемент, получим матрицу I_r. На языке матриц элементарных преобразований это означает, что существуют матрицы элементарных преобразований $Q_1, \dots, Q_k$ и $P_1, \dots, P_s$ такие, что $I_r = Q_k \dots Q_1 A P_1 \dots P_s$ (теорема 3.2). Положив $Q = Q_k \dots Q_1$ и $P = P_1 \dots P_s$, получим, что $I_r = QAP$, где $|Q| \neq 0$, $|P| \neq 0$ в силу невырожденности матриц элементарных преобразований. ■

Теорема 16.11. *Две матрицы $A, B \in \mathbb{R}^{m\times n}$ эквивалентны тогда и только тогда, когда их ранги совпадают.*

Доказательство. Необходимость вытекает из определения и теоремы 16.6. Достаточность следует из теоремы 16.10 и транзитивности отношения эквивалентности. ■

§17. Базис и размерность

Базисом линейного пространства называется упорядоченная линейно независимая система векторов пространства, через которую линейно выражается любой вектор пространства. Согласно этому определению система векторов $e_1, \dots, e_n$ линейного пространства V образует базис V, если

- $e_1, \dots, e_n$ линейно независима,
- для любого вектора $x \in V$ существуют числа $x_1, \dots, x_n$ такие, что $x = x_1e_1 + \dots + x_ne_n$.

Как показано в §15, любой ненулевой вектор является базисом линейного пространства V_1 векторов на прямой, любая пара неколлинеарных векторов – базисом пространства V_2 векторов плоскости, любая тройка некомпланарных векторов – базисом пространства V_3.

Теорема 17.1. *Система векторов $e_1, \ldots, e_n$ линейного пространства является его базисом тогда и только тогда, когда она образует максимальную линейно независимую систему векторов этого пространства.*

Доказательство. Необходимость. Базис пространства является максимальной линейно независимой системой векторов в этом пространстве, так как любая бо́льшая система векторов линейно выражается через этот базис (т.е. через меньшую систему) и на основании теоремы 16.2 линейно зависима.

Достаточность. Пусть $e_1, \ldots, e_n$ – максимальная линейно независимая система векторов пространства V, тогда для любого вектора $x \in V$ система векторов $e_1, \ldots, e_n, x$ линейно зависима, так как содержит более чем n векторов. Из теоремы 14.6 следует, что вектор x является линейной комбинацией $e_1, \ldots, e_n$. Следовательно, векторы $e_1, \ldots, e_n$ образуют базис пространства V. ■

Теорема 17.1 дает другое определение базиса.

Итак, все базисы одного линейного пространства состоят из одинакового числа векторов, равного максимальному числу линейно независимых векторов этого пространства. Это означает, что число векторов базиса является характеристикой не столько базиса, сколько самого пространства. Число векторов базиса называется *размерностью* линейного пространства. Размерность нулевого пространства по определению считается равной нулю. Обозначение: $\dim V$. Из теоремы 17.1 следует, что размерность линейного пространства равна максимальному числу линейно независимых векторов этого пространства. Линейное пространство размерности n, где n – целое неотрицательное число, называется n-мерным пространством. Любое n-мерное пространство называется *конечномерным линейным пространством.*

Конечномерными пространствами не исчерпываются все линейные пространства: не во всяком пространстве можно указать максимальное число линейно независимых векторов. Линейное пространство называется *бесконечномерным*, если для любого $k \in \mathbb{N}$ в нем найдется линейно независимая система из k векторов. Из определения размерности и теоремы 17.1 вытекают следующие утверждения.

Утверждение 1. *В n-мерном пространстве любые n линейно независимых векторов образуют базис.*

Утверждение 2. *В n-мерном пространстве любая система из s векторов, где $s > n$, линейно зависима.*

Примеры. 1. Геометрические пространства V_1, V_2, V_3. Из результатов §15 следует, что $\dim V_n = n$, где $n = 1, 2, 3$.

2. Арифметическое пространство $\mathbb{R}^n$. В пространстве $\mathbb{R}^n$ единичные векторы (14.6) образуют базис, так как они линейно независимы и, как нетрудно проверить, любой вектор $a = (a_1, \ldots, a_n)$ пространства $\mathbb{R}^n$ линейно выражается через них: $a = a_1e_1 + \ldots + a_ne_n$. Итак, $\dim \mathbb{R}^n = n$.

3. Пространство матриц $\mathbb{R}^{m\times n}$. Рассмотрим матрицы E_{11}, $E_{12}, \ldots, E_{1n}, E_{21}, \ldots, E_{mn}$, у которых все элементы равны нулю, кроме одного: в матрице E_{ij} элемент в позиции (i, j) равен единице.

а) Эти матрицы линейно независимы, так как из того, что

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}E_{ij} = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ & \ldots & \\ a_{m1} & \ldots & a_{mn} \end{bmatrix} = O,$$

следует, что $a_{ij} = 0$, $\quad i = \overline{1, m}$, $j = \overline{1, n}$.

б) Любая матрица $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ является линейной комбинацией этих матриц (как видно из п."а"). Следовательно, эти матрицы образуют базис $\mathbb{R}^{m\times n}$ и $\dim \mathbb{R}^{m\times n} = mn$.

4. Пространство многочленов M_n. Многочлены $1, t, t^2, \ldots, t^n$ образуют базис M_n, так как они линейно независимы (§14), а любой многочлен $p(t) = a_0 + a_1t + \ldots + a_nt^n \in M_n$ является линейной комбинацией этих многочленов с коэффициентами $a_0, a_1, \ldots, a_n$. Итак, $\dim M_n = n + 1$.

5. Пространство M_∞ многочленов всех степеней (§13, пример 4) бесконечномерно, так как для любого $k \in \mathbb{N}$ можно указать k линейно независимых многочленов: $1, t, t^2, \ldots, t^{k-1}$.

Координаты вектора. Базис играет большую роль в изучении линейного пространства. С его помощью абстрактные векторы можно задавать в виде совокупности чисел, а операции над векторами сводить к операциям над числами.

Теорема 17.2. *Разложение вектора по базису единственно.*

Это утверждение вытекает из теоремы 14.5. ■

Коэффициенты разложения вектора по базису называются *координатами вектора в этом базисе.* Из определения базиса и теоремы 17.2 следует, что каждый вектор имеет координаты, при этом два вектора равны тогда и только тогда, когда равны их координаты.

Обозначение. Если $e_1, \ldots, e_n$ – базис пространства и

$$x = x_1e_1 + \ldots + x_ne_n, \tag{17.1}$$

то будем обозначать через x_e вектор-столбец из координат вектора x в этом базисе:

$$x_e = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^T. \tag{17.2}$$

Столбец x_e называют *координатным столбцом* вектора x в базисе $e_1, \ldots, e_n$. Положим

$$e = (e_1, e_2, \ldots, e_n). \tag{17.3}$$

Символ e будем понимать как обозначение базиса $e_1, \ldots, e_n$ и как матрицу-строку $(e_1, e_2, \ldots, e_n)$. В обозначениях (17.2) и (17.3) разложение (17.1) может быть записано как произведение строки e на столбец x_e:

$$x = ex_e. \tag{17.4}$$

Теорема 17.3. *При сложении векторов их координаты в одном базисе складываются, а при умножении вектора на число его координаты умножаются на это число.*

Доказательство. Пусть $x = \sum_{i=1}^{n} x_i e_i$, $y = \sum_{i=1}^{n} y_i e_i$. Тогда из аксиом линейного пространства следует, что $x+y = \sum_{i=1}^{n}(x_i+y_i)e_i$ и $\alpha x = \sum_{i=1}^{n} \alpha x_i e_i$. Следовательно, $(x+y)_e = x_e + y_e$ и $(\alpha x)_e = \alpha x_e$. ■

Переход к другому базису. Пусть $e = (e_1, e_2, \ldots, e_n)$ и $f = (f_1, f_2, \ldots, f_n)$ – два базиса n-мерного пространства V. Выясним, как меняются координаты вектора при переходе от базиса e к базису f.

Векторы второго базиса, как векторы пространства V, разлагаются по базису e; пусть

$$\begin{aligned} f_1 &= c_{11}e_1 + c_{21}e_2 + \ldots + c_{n1}e_n, \\ f_2 &= c_{12}e_1 + c_{22}e_2 + \ldots + c_{n2}e_n, \\ &\ldots \\ f_n &= c_{1n}e_1 + c_{2n}e_2 + \ldots + c_{nn}e_n. \end{aligned} \tag{17.5}$$

Коэффициенты c_{ij} этих разложений образуют матрицу

$$C = (c_{ij}) \in \mathbb{R}^{m \times n}, \tag{17.6}$$

которая называется ***матрицей перехода от базиса e к базису f***.

Обозначение: C или $C_{e \to f}$.

Замечание. Соотношения (17.5) в обозначениях (17.3) и (17.6) могут быть записаны компактно в виде

$$f = eC. \tag{17.7}$$

Теорема 17.4. *Матрица перехода к другому базису не вырождена.*

Доказательство. Пусть $|C| = 0$. Тогда (теорема 16.1, следствие) один из столбцов матрицы C является линейной комбинацией других ее столбцов. В силу свойства линейности координат отсюда следует, что один из векторов $f_1, \ldots, f_n$ линейно выражается через другие векторы этой системы. Это противоречит линейной независимости $f_1, \ldots, f_n$. ■

Теорема 17.5. *Если C – матрица перехода от базиса e к базису f, то C^{-1} – матрица перехода от базиса f к базису e.*

Доказательство. Пусть $f = eC$. Умножив обе части этого равенства на C^{-1}, получим, что $e = fC^{-1}$, откуда в силу (17.7) следует, что C^{-1} – матрица перехода от базиса f к базису e. ■

Теорема 17.6. *Координаты вектора x в базисах e и f связаны между собой соотношением*

$$x_e = Cx_f, \tag{17.8}$$

где C – матрица перехода от базиса e к базису f.

Доказательство. Из (17.4) и (17.7) имеем, что $x = ex_e$ и $x = fx_f = (eC)x_f = e(Cx_f)$. Отсюда в силу единственности разложения по базису получаем (17.8). ■

§18. Линейное подпространство и линейное многообразие

Линейное подпространство. Непустое подмножество L пространства V называется ***линейным подпространством пространства V***, если оно само является линейным пространством относительно законов композиции, действующих в V.

Теорема 18.1. *Непустое подмножество L пространства V является линейным подпространством этого пространства тогда и только тогда, когда имеют место импликации:*

$$\begin{aligned} a, b \in L \quad &\Rightarrow \quad a + b \in L; \\ a \in L, \alpha \in \mathbb{R} \quad &\Rightarrow \quad \alpha a \in L. \end{aligned} \tag{18.1}$$

Доказательство. Необходимость очевидна, так как если L – линейное пространство, то результаты выполнения операций над элементами из L находятся в L.

Достаточность. Соотношения (18.1) говорят о том, что операции сложения и умножения на число являются законами композиции в L. Остается проверить, что они подчиняются всем аксиомам линейного пространства. В действительности же необходимо проверить только аксиомы нуля и противоположного элемента, так как выполнение остальных аксиом очевидно. Возьмем произвольный элемент $a \in L$, тогда $0a = \theta$ и $(-1)a = -a$. Из (18.1) следует, что $\theta \in L$ и $-a \in L$. Итак, L – линейное подпространство пространства V. ■

Свойства (18.1) называют свойствами ***замкнутости множества L*** *относительно* указанных ***операций***.

Примеры. 1. Каждое линейное пространство обладает двумя подпространствами: нулевым подпространством (состоящим из одного нулевого вектора) и самим пространством. Эти подпространства называются ***тривиальными***.

2. Геометрическое пространство V_1 векторов на прямой имеет два тривиальных подпространства. Геометрическое пространство V_2 векторов на плоскости, кроме тривиальных подпространств, имеет бесконечно много нетривиальных: каждое из них является пространством всех векторов, концы которых лежат на прямой, проходящей через начало координат (или если отойти от соглашения об откладывании всех векторов от начала координат, то пространством всех векторов, параллельных некоторой прямой). В геометрическом пространстве V_3 векторов пространства каждая прямая и каждая плоскость, проходящие через начало координат, определяют линейное подпространство.

3. В пространстве многочленов M_n каждое пространство M_k, где $0 \leq k \leq n$, образует линейное подпространство.

Линейное многообразие. Примеры линейных подпространств в геометрических пространствах дают геометрический образ подпространства как прямой или плоскости, проходящей через начало координат. Возникает естественный вопрос, что же будет аналогом других прямых и плоскостей в произвольном линейном пространстве. Ответим на этот вопрос.

Пусть V – линейное пространство, L – некоторое его подпространство, x_0 – некоторый вектор пространства V. Множество H всевозможных векторов вида $x_0 + x$, где $x \in L$, называется ***линейным многообразием*** (или ***линейным аффинным многообразием***, или ***линейным точечным многообразием***) ***пространства*** V, ***полученным сдвигом подпространства*** L ***на вектор*** x_0. Вектор x_0 называется ***вектором сдвига***, а подпространство L – ***направляющим подпространством*** линейного многообразия H.

Обозначение: $H = x_0 + L$. Итак, $x_0 + L = \{x_0 + x | x \in L\}$.

Очевидно, линейное подпространство L является частным случаем линейного многообразия, когда вектор сдвига $x_0 \in L$.

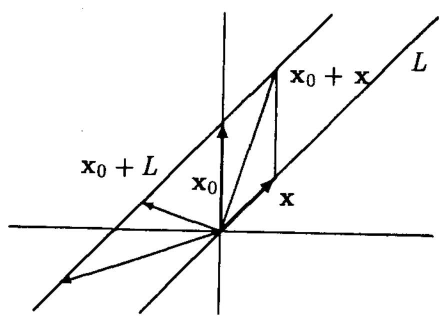


Рис. 1

Пример. Пусть на плоскости V_2 задано некоторое подпространство L, т.е. совокупность всех векторов $\mathbf{x}$, концы которых лежат на прямой, проходящей через начало координат. И пусть $\mathbf{x}_0$ – некоторый фиксированный вектор плоскости (рис. 1). Тогда нетрудно показать, что линейное многообразие $\mathbf{x}_0 + L$ есть множество векторов, концы которых лежат на прямой, полученной из прямой L сдвигом ее на вектор $\mathbf{x}_0$.

Из определения вытекают следующие факты.

1°. *Вектор сдвига* x_0 ***принадлежит линейному многообразию***, так как $x_0 = x_0 + \theta$, $\theta \in L$.

2°. ***Разность двух векторов линейного многообразия принадлежит направляющему подпространству***, так как если $z_1 = x_0 + x$, $x \in L$, $z_2 = x_0 + y$, $y \in L$, то $z_1 - z_2 = x - y \in L$.

Теорема 18.2. *Два линейных многообразия* $H_1 = x_1 + L_1$ *и* $H_2 = x_2 + L_2$ *совпадают тогда и только тогда, когда* $L_1 = L_2 = L$ *и* $x_1 - x_2 \in L$.

Доказательство. Необходимость. Пусть $H_1 = H_2$, тогда $x_1 \in H_2$ и, в силу свойства 2°, $x_1 - x_2 \in L_2$. Покажем, что $L_1 = L_2$. Для произвольного вектора $x \in L_1$ имеем $x_1 + x \in H_1$. Так как $H_1 = H_2$, то существует вектор $y \in L_2$ такой, что $x_1 + x = x_2 + y$. Таким

образом, $x = (x_2 - x_1) + y$, где $x_2 - x_1 \in L_2$, $y \in L_2$. Отсюда следует, что $x \in L_2$ и $L_1 \subset L_2$. Аналогично доказывается, что $L_2 \subset L_1$. Оба вложения говорят о том, что $L_1 = L_2$.

Достаточность. Пусть $L_1 = L_2 = L$ и $x_1 - x_2 \in L$. Тогда для произвольного вектора $z \in H_1$ имеем $z = x_1 + x$, где $x \in L$, или $z = x_2 + (x_1 - x_2) + x$, $x \in L$. Так как $x_1 - x_2 \in L$ и $x \in L$, то $(x_1 - x_2) + x \in L$, следовательно, $z \in H_2$ и $H_1 \subset H_2$. Аналогично получаем, что $H_2 \subset H_1$. Это означает, что $H_1 = H_2$. ■

Следствие 1. ***Вектором сдвига может быть любой вектор линейного многообразия.***

Действительно, если x_1 – произвольный вектор линейного многообразия $H = x_0 + L$, то $x_1 - x_0 \in L$ и $H = x_1 + L$.

Следствие 2. ***Линейное многообразие может быть получено сдвигом единственного направляющего подпространства.*** Это утверждение непосредственно вытекает из теоремы.

Следствие 2 позволяет переносить некоторые характеристики направляющего подпространства на само линейное многообразие. ***Размерностью линейного многообразия*** называется размерность его направляющего подпространства. Обозначение: $\dim H$. Линейное многообразие размерности единица называется ***прямой в линейном пространстве***, размерности $(n - 1)$, где $n = \dim V$, – ***гиперплоскостью***, а размерности k, $1 < k < n - 1$, – ***k-мерной плоскостью***.

Пример. Из определения следует, что прямая на плоскости V_2 и плоскость в пространстве V_3 – гиперплоскости в пространствах V_2 и V_3. Естественно ожидать, что, представляя собой один и тот же объект, они обладают общими свойствами. В этом мы убедимся в последующих главах.

Глава V. Векторная алгебра

Рассмотрим линейные пространства V_1, V_2, V_3 векторов на прямой, на плоскости и в пространстве. В §15, 17 было доказано, что:

- $\dim V_1 = 1$, а любой ненулевой вектор $\mathbf{e}_1$ является базисом V_1,
- $\dim V_2 = 2$, а любая пара неколлинеарных векторов $\mathbf{e}_1$, $\mathbf{e}_2$ является базисом V_2,
- $\dim V_3 = 3$, а любая тройка некомпланарных векторов $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ является базисом V_3.

При доказательстве теорем §15 практически показано, как находятся координаты векторов в этих пространствах. Вернемся к этому вопросу.

§19. Координаты вектора

1. Пусть $\mathbf{e}_1$ – базис V_1 и $\mathbf{a}$ – произвольный вектор из V_1. Отложив эти векторы от одной точки O прямой V_1 (§15, рис. 2,*а*), так что $\mathbf{e}_1 = \overrightarrow{OE}$, $\mathbf{a} = \overrightarrow{OA_1}$, получим, что $\mathbf{a} = x\,\mathbf{e}_1$, где

$$x = \begin{cases} |\overrightarrow{OA_1}|/|\overrightarrow{OE_1}|, & \mathbf{a} \uparrow\uparrow \mathbf{e}_1, \\ -|\overrightarrow{OA_1}|/|\overrightarrow{OE_2}|, & \mathbf{a} \uparrow\downarrow \mathbf{e}_1. \end{cases} \tag{19.1}$$

Введем на прямой V_1 направление: пусть положительное направление на прямой совпадает с направлением базисного вектора $\mathbf{e}_1$. Тогда согласно (19.1) и (10.1) получим

$$x = \frac{(\overrightarrow{OA_1})}{|\mathbf{e}_1|} = \frac{(\mathbf{a})}{|\mathbf{e}_1|}. \tag{19.2}$$

Ось, положительное направление которой совпадает с направлением вектора $\mathbf{e}$, будем называть *осью, определенной вектором* $\mathbf{e}$.

2. Пусть $\mathbf{e}_1$, $\mathbf{e}_2$ – базис V_2 и $\mathbf{a}$ – произвольный вектор из V_2. Отложив эти векторы от одной точки O плоскости V_2 (§15, рис. 2,*б*), так что $\mathbf{e}_1 = \overrightarrow{OE_1}$, $\mathbf{e}_2 = \overrightarrow{OE_2}$, $\mathbf{a} = \overrightarrow{OA}$, и введя направления на прямых OE_1 и OE_2, совпадающие с направлениями базисных векторов $\mathbf{e}_1$ и $\mathbf{e}_2$, получим в соответствии с (19.2), что $\mathbf{a} = x\,\mathbf{e}_1 + y\,\mathbf{e}_2$,

$$x = \frac{(\overrightarrow{OA_1})}{|\mathbf{e}_1|}, \quad y = \frac{(\overrightarrow{OA_2})}{|\mathbf{e}_2|}, \tag{19.3}$$

где A_1, A_2 – проекции точки A на прямые OE_1 и OE_2 параллельно соответственно прямым OE_2 и OE_1.

3. Пусть $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ – базис V_3 и $\mathbf{a}$ – произвольный вектор из V_3. Поступая аналогично (§15, рис. 2,*в*), получим $\mathbf{a} = x\,\mathbf{e}_1 + y\,\mathbf{e}_2 + z\,\mathbf{e}_3$,

$$x = \frac{(\overrightarrow{OA_1})}{|\mathbf{e}_1|}, \quad y = \frac{(\overrightarrow{OA_2})}{|\mathbf{e}_2|}, \quad z = \frac{(\overrightarrow{OA_3})}{|\mathbf{e}_3|}, \tag{19.4}$$

где A_1, A_2, A_3 – проекции точки A на прямые OE_1, OE_2 и OE_3, параллельные соответственно плоскостям OE_2E_3, OE_1E_3 и OE_1E_2.

§20. Координаты точки

Аффинная система координат. Пусть в пространстве V_3 (на плоскости V_2 или на прямой V_1) зафиксирована некоторая точка O, называемая *полюсом*. Для любой точки A вектор $\mathbf{r}_A = \overrightarrow{OA}$ называется *радиус-вектором точки A относительно полюса O*. Задание точки ее радиус-вектором определяет, очевидно, биективное отображение. Тот факт, что точка A имеет радиус-вектор $\mathbf{r}$, обозначают символом $A(\mathbf{r})$.

Если в пространстве V_3 зафиксированы точка O и базис $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, то говорят, что в пространстве задана *аффинная система координат* (или *общая декартова система координат*) $\{O; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Точка O называется *началом координат*; оси, проходящие через начало координат и определенные векторами $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, называются *осями координат* и обозначаются Ox (ось абсцисс), Oy (ось ординат), Oz (ось аппликат) соответственно. Плоскость, определяемая осями координат Ox и Oy (Ox и Oz, Oy и Oz), называется *координатной плоскостью* Oxy (Oxz, Oyz соответственно). В этой терминологии аффинная система координат обозначается также символом $Oxyz$.

Координатами точки A в аффинной системе координат $\{O; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ называются координаты радиус-вектора $\mathbf{r}_A$ этой точки в базисе $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$. Тот факт, что точка A имеет координаты x, y, z, обозначают символом $A(x, y, z)$. Итак,

$$\mathbf{r}_A = x\,\mathbf{e}_1 + y\,\mathbf{e}_2 + z\,\mathbf{e}_3 \iff A(x, y, z). \tag{20.1}$$

Замечание 1. Из определения следует, что любая точка A пространства в заданной системе координат имеет координаты, причем точки $A_1(x_1, y_1, z_1)$ и $A_2(x_2, y_2, z_2)$ совпадают тогда и только тогда, когда $x_1 = x_2$, $y_1 = y_2$, $z_1 = z_2$.

Замечание 2. Координаты точки $A(x, y, z)$ определяются соотношениями (19.4). При этом, как легко видеть, проекции A_1, A_2, A_3 точки A имеют координаты $A_1(x, 0, 0)$, $A_2(0, y, 0)$, $A_3(0, 0, z)$.

Аналогично определяются аффинные системы координат $\{O; \mathbf{e}_1, \mathbf{e}_2\}$ на плоскости V_2 и $\{O; \mathbf{e}_1\}$ на прямой V_1, а также координаты точки $A(x, y)$ и $A(x)$ соответственно. При этом имеют место очевидные аналоги соотношения (20.1) и обоих замечаний. В дальнейшем все факты будем излагать только в терминах V_3.

Теорема 20.1. *Если $A(x_1, y_1, z_1)$ и $B(x_2, y_2, z_2)$ – точки пространства, заданные своими координатами в системе координат $\{O; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, то вектор $\mathbf{a} = \overrightarrow{AB}$ в базисе $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ имеет координаты $\mathbf{a} = \{x_2 - x_1, y_2 - y_1, z_2 - z_1\}$.*

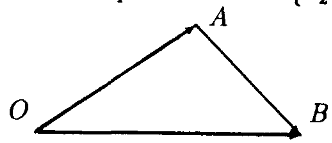


Рис. 1

Доказательство. Действительно, $\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA}$ (рис. 1) и, следовательно, $\mathbf{a} = \mathbf{r}_B - \mathbf{r}_A$. Так как $\mathbf{r}_A = \{x_1, y_1, z_1\}$, $\mathbf{r}_B = \{x_2, y_2, z_2\}$, то в силу свойства линейности координат отсюда следует утверждение теоремы. ■

Деление отрезка в данном отношении. Говорят, что точка $M \neq B$ делит отрезок $[AB]$ в отношении λ, если $\overrightarrow{AM} = \lambda \overrightarrow{MB}$ (рис. 2).

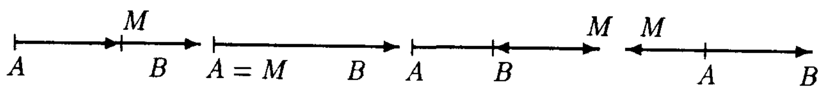


Рис. 2

Обозначение: $(ABM) = \lambda$. Из определения следует, что точка M расположена на прямой AB, при этом (рис. 2):

1) если M – внутренняя точка отрезка $[AB]$, то $\lambda > 0$;
2) если $M = A$, то $\lambda = 0$;
3) если M расположена вне отрезка $[AB]$, то $\lambda < 0$.

Заметим, что других вариантов расположения точки M не может быть и что ни в одном из возможных вариантов λ не равно -1.

Теорема 20.2. *Пусть $A(\mathbf{r}_1)$, $B(\mathbf{r}_2)$, $M(\mathbf{r}_3)$ – точки пространства и $(ABM) = \lambda$. Тогда*

$$\mathbf{r}_3 = \frac{\mathbf{r}_1 + \lambda \mathbf{r}_2}{1 + \lambda}. \qquad (20.2)$$

Доказательство. Условие $(ABM) = \lambda$ означает, что $\overrightarrow{AM} = \lambda \overrightarrow{MB}$ или $\mathbf{r}_3 - \mathbf{r}_1 = \lambda(\mathbf{r}_2 - \mathbf{r}_3)$. Отсюда следует (20.2). ■

Следствие. Соотношение (20.2) *в координатной форме имеет следующий вид: для $A(x_1, y_1, z_1)$, $B(x_2, y_2, z_2)$, $M(x_3, y_3, z_3)$*

$$x_3 = \frac{x_1 + \lambda x_2}{1 + \lambda}, \quad y_3 = \frac{y_1 + \lambda y_2}{1 + \lambda}, \quad z_3 = \frac{z_1 + \lambda z_2}{1 + \lambda}. \qquad (20.3)$$

Прямоугольные координаты. Базис $\mathbf{e}_1, \dots, \mathbf{e}_n$, где $n = 1, 2, 3$, называется *ортонормированным*, если векторы базиса

1) имеют единичную длину

и, в случае $n > 1$,

2) попарно перпендикулярны.

Аффинная система координат $\{O; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, соответствующая ортонормированному базису $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, называется *прямоугольной декартовой системой координат*.

§21. Проекции вектора и координаты

Проекции вектора на плоскости. Пусть на плоскости P даны две непараллельные прямые l и L. *Проекцией направленного отрезка $\overrightarrow{AB}$ на прямую l параллельно прямой L* называется (рис. 1) направленный отрезок $\overrightarrow{A_1B_1}$, где A_1, B_1 – проекции точек A и B на прямую l параллельно прямой L. Обозначение: $pr_l^L \overrightarrow{AB}$.

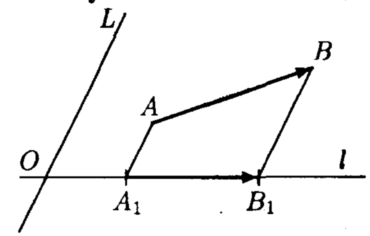


Рис. 1

Теорема 21.1. *Проекции равных направленных отрезков равны.*

Доказательство. Пусть $\overrightarrow{AB} = \overrightarrow{CD}$, покажем, что $pr_l^L \overrightarrow{AB} = pr_l^L \overrightarrow{CD}$. Введем на плоскости систему координат $\{O; \mathbf{e}_1, \mathbf{e}_2\}$, где O – точка пересечения прямых l и L, $\mathbf{e}_1$ – базис на прямой l, $\mathbf{e}_2$ – базис на прямой L. Пусть точки A, B, C, D имеют координаты $A(x_A, y_A)$, $B(x_B, y_B)$, $C(x_C, y_C)$, $D(x_D, y_D)$. Так как $\mathbf{a} = \overrightarrow{AB} = \overrightarrow{CD}$, то согласно теореме 20.1

$$x_B - x_A = x_D - x_C. \qquad (21.1)$$

Так как точки A_1, B_1, C_1, D_1 имеют координаты $A_1(x_A, 0)$, $B_1(x_B, 0)$, $C_1(x_C, 0)$, $D_1(x_D, 0)$ (§20, замечание 2), то $\mathbf{b} = \overrightarrow{A_1B_1} = \{x_B - x_A, 0\}$, $\mathbf{c} = \overrightarrow{C_1D_1} = \{x_D - x_C, 0\}$. Согласно (21.1) отсюда следует, что $\mathbf{b} = \mathbf{c}$. Это означает, что $\overrightarrow{A_1B_1}$ и $\overrightarrow{C_1D_1}$ порождают один и тот же вектор и поэтому равны. ■

Проекцией вектора $\mathbf{a} = \overrightarrow{AB}$ на прямую l параллельно прямой L называется вектор, порожденный $pr_l^L \overrightarrow{AB}$. Обозначение: $pr_l^L \mathbf{a}$. Корректность определения вытекает из теоремы 21.1.

Теорема 21.2. *Проекция вектора на прямую l параллельно прямой L обладает свойством линейности:*

1) $pr_l^L(\mathbf{a}+\mathbf{b}) = pr_l^L\mathbf{a} + pr_l^L\mathbf{b}, \quad \forall \mathbf{a}, \mathbf{b}$,

2) $pr_l^L(\alpha\mathbf{a}) = \alpha \cdot pr_l^L\mathbf{a}, \quad \forall \mathbf{a}, \ \forall \alpha \in \mathbb{R}$.

Доказательство. Пусть в базисе $\mathbf{e}_1$, $\mathbf{e}_2$, рассмотренном выше, векторы $\mathbf{a}, \mathbf{b}$ имеют координаты $\mathbf{a} = \{a_1, a_2\}$, $\mathbf{b} = \{b_1, b_2\}$. Тогда $pr_l^L\mathbf{a} = \{a_1, 0\}$, $pr_l^L\mathbf{b} = \{b_1, 0\}$ и в силу линейности координат $pr_l^L\mathbf{a} + pr_l^L\mathbf{b} = \{a_1 + b_1, 0\}$. С другой стороны, $\mathbf{a}+\mathbf{b} = \{a_1+b_1, a_2+b_2\}$ и $pr_l^L(\mathbf{a}+\mathbf{b}) = \{a_1+b_1, 0\}$. Следовательно, $pr_l^L(\mathbf{a}+\mathbf{b}) = pr_l^L\mathbf{a} + pr_l^L\mathbf{b}$. Аналогично доказывается второе условие линейности. ■

Замечание 1. Формулы (19.3) для координат вектора $\mathbf{a}$ в базисе $\mathbf{e}_1$, $\mathbf{e}_2$ плоскости V_2 могут быть записаны в терминах проекций вектора на ось в виде

$$x = \frac{(pr_x\,\mathbf{a})}{|\mathbf{e}_1|}, \qquad y = \frac{(pr_y\,\mathbf{a})}{|\mathbf{e}_2|}, \tag{21.2}$$

где $pr_x\,\mathbf{a}$ и $pr_y\,\mathbf{a}$ – проекции вектора $\mathbf{a}$ на оси, определенные векторами $\mathbf{e}_1$ и $\mathbf{e}_2$ соответственно (т.е. оси координат Ox и Oy), параллельно другой оси (т.е. оси Oy и Ox соответственно).

Проекции вектора в пространстве. Пусть в пространстве заданы плоскость π и непараллельная ей прямая l. *Проекцией направленного отрезка $\overrightarrow{AB}$ на прямую l (на плоскость π) параллельно плоскости π* (соответственно *прямой l*) называется (рис. 2) направленный отрезок $\overrightarrow{A_1B_1}$ ($\overrightarrow{A_2B_2}$), где A_1 и B_1 (A_2 и B_2) – проекции точек A и B на прямую l (плоскость π) параллельно плоскости π (прямой l). Обозначение: $pr_l^\pi\overrightarrow{AB}$, $pr_\pi^l\overrightarrow{AB}$.

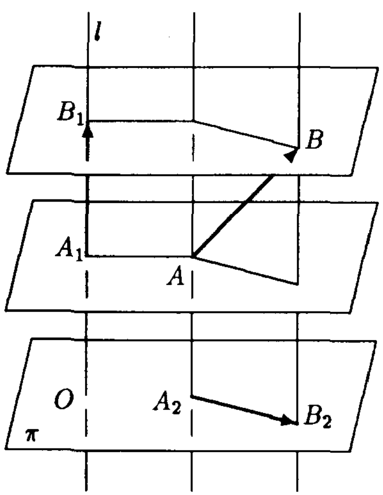


Рис. 2

Для обеих проекций справедливо утверждение теоремы 21.1: ***проекции равных направленных отрезков равны.*** Доказательство утверждения для проекций в пространстве отличается от доказательства теоремы 21.1 только тем, что система координат $\{O;\, \mathbf{e}_1,\, \mathbf{e}_2,\, \mathbf{e}_3\}$ состоит из точки O пересечения прямой l с плоскостью π, базиса $\mathbf{e}_1$, $\mathbf{e}_2$ плоскости π и базиса $\mathbf{e}_3$ прямой l.

Проекцией вектора $\mathbf{a} = \overrightarrow{AB}$ на прямую l (плоскость π) параллельно плоскости π (прямой l) называется вектор, порожденный $pr_l^\pi\overrightarrow{AB}$ ($pr_\pi^l\overrightarrow{AB}$). Обозначение: $pr_l^\pi\,\mathbf{a}$, $pr_\pi^l\,\mathbf{a}$. *Обе проекции вектора обладают свойством линейности.* Доказательство этого факта повторяет доказательство теоремы 21.2 с той лишь разницей, что рассматривается система координат $\{O;\, \mathbf{e}_1,\, \mathbf{e}_2,\, \mathbf{e}_3\}$, упомянутая выше.

Замечание 2. Формулы (19.4) для координат вектора $\mathbf{a} \in V_3$ в базисе $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ могут быть записаны в терминах проекций вектора на ось в виде

$$x = \frac{(pr_x\,\mathbf{a})}{|\mathbf{e}_1|}, \qquad y = \frac{(pr_y\,\mathbf{a})}{|\mathbf{e}_2|}, \qquad z = \frac{(pr_z\,\mathbf{a})}{|\mathbf{e}_3|}, \tag{21.3}$$

где $pr_x\,\mathbf{a}$, $pr_y\,\mathbf{a}$, $pr_z\,\mathbf{a}$ – проекции вектора $\mathbf{a}$ на оси, определенные базисными векторами $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ (т.е. оси координат Ox, Oy, Oz), параллельно координатным плоскостям Oyz, Oxz, Oxy соответственно.

Теорема 21.3. *На плоскости (в пространстве) величина проекции вектора на ось параллельно прямой (соответственно плоскости) обладает свойством линейности.*

Утверждение теоремы следует из того, что величины рассматриваемых проекций пропорциональны координатам (согласно (21.2) и (21.3)), которые обладают свойством линейности (теорема 17.3). ■

Ортогональные проекции. Мы определили три различные проекции вектора. Во всех трех случаях, если $l \perp L$ или $l \perp \pi$, проекции вектора называются *ортогональными проекциями.*

§22. Скалярное произведение

Определение и основные факты. Пусть $\mathbf{a}$ и $\mathbf{b}$ – ненулевые векторы. Отложим их от одной точки O. Пусть $\mathbf{a} = \overrightarrow{OA}$, $\mathbf{b} = \overrightarrow{OB}$. *Углом между векторами* $\mathbf{a}$ и $\mathbf{b}$ называется наименьший угол между лучами $[OA)$ и $[OB)$. Обозначение: $(\widehat{\mathbf{a}, \mathbf{b}})$. Корректность определения очевидна. Из определения следует, что $0 \le (\widehat{\mathbf{a}, \mathbf{b}}) \le \pi$.

Скалярным произведением ненулевых *векторов* $\mathbf{a}$ и $\mathbf{b}$ называется число, равное произведению длин этих векторов на косинус угла между ними. Если один из векторов $\mathbf{a}$ или $\mathbf{b}$ нулевой, то скалярное произведение этих векторов по определению считается равным нулю. Обозначение: $(\mathbf{a}, \mathbf{b})$. Итак, для ненулевых векторов $\mathbf{a}$ и $\mathbf{b}$

$$(\mathbf{a}, \mathbf{b}) = |\mathbf{a}| \cdot |\mathbf{b}| \cos(\widehat{\mathbf{a}, \mathbf{b}}). \tag{22.1}$$

Обозначим через $pr_{\mathbf{a}}\,\mathbf{b}$ ортогональную проекцию вектора $\mathbf{b}$ на ось, определенную вектором $\mathbf{a} \neq \mathbf{0}$ (рис. 1).

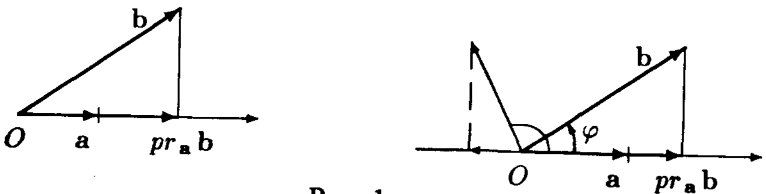


Рис. 1

Теорема 22.1. *Если* $\mathbf{a} \neq 0$, *то для любого вектора* $\mathbf{b}$

$$(\mathbf{a},\,\mathbf{b}) = |\mathbf{a}|\,(pr_{\mathbf{a}}\,\mathbf{b}) = (\mathbf{a}, pr_{\mathbf{a}}\,\mathbf{b}). \tag{22.2}$$

Доказательство. Утверждение теоремы очевидно для $\mathbf{b} = \mathbf{0}$ и $(\widehat{\mathbf{a},\,\mathbf{b}}) = \pi/2$. Пусть $\mathbf{b} \neq 0$ и $\varphi = (\widehat{\mathbf{a},\,\mathbf{b}}) \neq \pi/2$. Тогда (рис. 1)

$$|pr_{\mathbf{a}}\,\mathbf{b}| = |\,\mathbf{b}|\cdot|\cos\varphi| = \begin{cases} |\,\mathbf{b}|\cos\varphi, & pr_{\mathbf{a}}\,\mathbf{b} \uparrow\uparrow \mathbf{a}, \\ -|\,\mathbf{b}|\cos\varphi, & pr_{\mathbf{a}}\,\mathbf{b} \uparrow\downarrow \mathbf{a}. \end{cases}$$

Отсюда следует, что $(pr_{\mathbf{a}}\,\mathbf{b}) = |\,\mathbf{b}|\cos\varphi$ и, тем самым, $(\mathbf{a},\,\mathbf{b}) = |\,\mathbf{a}|\,(pr_{\mathbf{a}}\,\mathbf{b})$. Вторая часть соотношения (22.2) вытекает из первой, так как $pr_{\mathbf{a}}\,pr_{\mathbf{a}}\,\mathbf{b} = pr_{\mathbf{a}}\,\mathbf{b}$. ■

Теорема 22.2. *Для любых векторов* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ *и числа* $\alpha \in \mathbb{R}$

1) $(\mathbf{a},\,\mathbf{b}) = (\mathbf{b},\,\mathbf{a})$;
2) $(\mathbf{a}+\mathbf{b},\,\mathbf{c}) = (\mathbf{a},\,\mathbf{c}) + (\mathbf{b},\,\mathbf{c})$;
3) $(\alpha\,\mathbf{a},\,\mathbf{b}) = \alpha(\mathbf{a},\,\mathbf{b})$;
4) $(\mathbf{a},\,\mathbf{a}) \geq 0$, *причем*
$(\mathbf{a},\,\mathbf{a}) = 0$ *тогда и только тогда, когда* $\mathbf{a} = 0$.

Доказательство. Свойства 1 и 4 очевидны. Докажем свойство 2. Имеем $(\mathbf{a}+\mathbf{b},\,\mathbf{c}) = (\mathbf{c},\,\mathbf{a}+\mathbf{b}) = |\,\mathbf{c}|\,(pr_{\mathbf{c}}(\mathbf{a}+\mathbf{b})) = \{$согласно теореме 21.3$\} = |\,\mathbf{c}|\,(pr_{\mathbf{c}}\,\mathbf{a}) + |\,\mathbf{c}|\,(pr_{\mathbf{c}}\,\mathbf{b}) = (\mathbf{a},\,\mathbf{c}) + (\mathbf{b},\,\mathbf{c})$. Аналогично проверяется свойство 3. ■

Следствие 1. *Из свойств 1–3 следует, что скалярное произведение линейно и по второму множителю:*

$$(\mathbf{a},\,\mathbf{b}+\mathbf{c}) = (\mathbf{a},\,\mathbf{b}) + (\mathbf{a},\,\mathbf{c}), \quad (\mathbf{a}, \alpha\,\mathbf{b}) = \alpha(\mathbf{a},\,\mathbf{b}).$$

Скалярное произведение в координатах. Свойство **линейности** скалярного произведения позволяет перемножать линейные **комбинации** векторов (т.е. находить их скалярное произведение):

$$\left(\sum_{i=1}^{k}\alpha_i\,\mathbf{a}_i, \sum_{j=1}^{m}\beta_j\,\mathbf{b}_j\right) = \sum_{i=1}^{k}\sum_{j=1}^{m}\alpha_i\beta_j(\mathbf{a}_i,\,\mathbf{b}_j). \tag{22.3}$$

Это, в частности, означает, что скалярное произведение векторов может быть вычислено по их координатам, если известна "таблица умножения" базисных векторов. Возможность подобного вычисления

скалярного произведения векторов позволяет определить по координатам длину вектора и угол между векторами, так как $|\,\mathbf{a}| = \sqrt{(\mathbf{a},\,\mathbf{a})}$, $\cos(\widehat{\mathbf{a},\,\mathbf{b}}) = (\mathbf{a},\,\mathbf{b})/(|\,\mathbf{a}|\cdot|\,\mathbf{b}|)$.

Задача вычисления скалярного произведения векторов по их координатам существенно упрощается, если рассматривается ортонормированный базис. Векторы $\mathbf{a}$ и $\mathbf{b}$ называются ***ортогональными***, если $(\mathbf{a},\,\mathbf{b}) = 0$. Из определения следует, что векторы $\mathbf{a}$ и $\mathbf{b}$ ортогональны тогда и только тогда, когда либо один из них нулевой, либо они перпендикулярны. В терминах ортогональности векторов ортонормированность базиса $\mathbf{e}_1, \ldots, \mathbf{e}_n$, где $n = 1, 2, 3$, означает, что

$$(\mathbf{e}_i,\,\mathbf{e}_j) = \begin{cases} 0, & i \neq j, \\ 1, & i = j. \end{cases} \tag{22.4}$$

Теорема 22.3. *Скалярное произведение векторов* $\mathbf{a} = \alpha_1\,\mathbf{e}_1 + \alpha_2\,\mathbf{e}_2 + \alpha_3\,\mathbf{e}_3$ *и* $\mathbf{b} = \beta_1\,\mathbf{e}_1 + \beta_2\,\mathbf{e}_2 + \beta_3\,\mathbf{e}_3$ ***равно сумме попарных произведений координат***

$$(\mathbf{a},\,\mathbf{b}) = \sum_{i=1}^{3}\alpha_i\beta_i \tag{22.5}$$

тогда и только тогда, когда $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ – ***ортонормированный базис.***

Доказательство. Необходимость вытекает из (22.3) и (22.4).

Достаточность. Соотношения (22.4) вытекают из (22.5), если учесть, что векторы $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ в базисе $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ имеют координаты $(1,0,0)$, $(0,1,0)$, $(0,0,1)$ соответственно. ■

Следствие 2. *Если векторы* $\mathbf{a} = \{\alpha_1, \alpha_2, \alpha_3\}$, $\mathbf{b} = \{\beta_1, \beta_2, \beta_3\}$ *заданы координатами в ортонормированном базисе, то*

$$|\,\mathbf{a}| = \sqrt{\alpha_1^2 + \alpha_2^2 + \alpha_3^2};\ \cos(\widehat{\mathbf{a},\,\mathbf{b}}) = \frac{\alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3}{|\,\mathbf{a}|\,|\,\mathbf{b}|}. \tag{22.6}$$

Следствие 3. *В* ***прямоугольной*** *декартовой системе координат расстояние* $\rho(A, B)$ *между точками* $A(x_1, y_1, z_1)$ *и* $B(x_2, y_2, z_2)$ *равно*

$$\rho(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

Это равенство следует из того, что $\rho(A, B) = |\overrightarrow{AB}|$ и $\overrightarrow{AB} = \{x_2 - x_1, y_2 - y_1, z_2 - z_1\}$.

§23. Векторное и смешанное произведения

Ориентация в вещественном линейном пространстве. В §19 при рассмотрении оси, определенной ненулевым вектором $\mathbf{e}$, мы фактически ввели ориентацию на прямой, назвав положительным то направление, которое совпадает с направлением вектора **e**. Прежде чем вводить определение ориентации, общее для прямой, плоскости и пространства, снова обратимся к прямой. Каждый ненулевой вектор **e** прямой образует базис, и переход от одного базиса к другому осуществляется умножением этого вектора на ненулевое число. Так как это число либо положительно, либо отрицательно, то все базисы на прямой разбиваются ровно на два класса: базисные векторы одного класса имеют одинаковое направление, а любые два базисных вектора разных классов имеют противоположное направление. Тот факт, что базисные векторы относятся к одному классу, на алгебраическом языке означает, что они отличаются *положительным множителем*. Именно этим мы и воспользуемся для определения ориентации.

Два базиса $e = (e_1, \dots, e_n)$ и $e' = (e'_1, \dots, e'_n)$ линейного пространства V называются ***одинаково ориентированными***, если матрица перехода $C_{e \to e'}$ имеет положительный определитель, и ***противоположно ориентированными*** – в противном случае.

Замечание 1. Из определения и из свойств определителя следует, что два базиса, получающиеся друг из друга

– перестановкой двух их векторов или

– умножением какого-либо вектора на отрицательное число,

противоположно ориентированы.

Теорема 23.1. *Отношение одинаковой ориентированности является отношением эквивалентности на множестве всех базисов пространства V.*

Доказательство. Действительно, рефлексивность отношения следует из того, что переход от базиса к самому себе осуществляется с помощью единичной матрицы, симметричность – из теоремы 17.5 и очевидного факта: $|C^{-1}| \cdot |C| > 0$, транзитивность – из того, что если $e' = eC$, $e'' = e'D$, то $e'' = e(CD)$, при этом $|CD| = |C| \cdot |D| > 0$. ■

Так как определитель матрицы перехода от одного базиса к другому либо положителен, либо отрицателен, то множество всех базисов пространства разбивается отношением одинаковой ориентированности ровно на два непересекающихся класса (класса эквивалентности) так, что всякий базис принадлежит одному и только одному классу, два базиса одного класса одинаково ориентированы, а любые два базиса из разных классов противоположно ориентированы.

Один из классов называют классом *правых* (или *положительно ориентированных*) *базисов*, а другой – *левых* (*отрицательно ориентированных*). Каждый из этих двух классов называется ***ориентацией пространства***. Вещественное линейное пространство с выбранной на нем ориентацией называется ***ориентированным пространством***. Так как класс эквивалентности порождается любым своим представителем, то для того, чтобы ориентировать линейное пространство, достаточно задать один какой-нибудь базис пространства и объявить положительно ориентированными все одноименные с ним базисы.

Класс правых базисов на плоскости V_2 и в пространстве V_3 обычно выбирают следующим образом:

– упорядоченную пару неколлинеарных векторов $\mathbf{e}_1$, $\mathbf{e}_2$ плоскости называют *правой* (*положительно ориентированной*), если кратчайший поворот (рис. 1) от $\mathbf{e}_1$ к $\mathbf{e}_2$ выполняется против часовой стрелки, и *левой* (*отрицательно ориентированной*) – в противном случае (начала векторов считаются совмещенными);

– упорядоченную тройку некомпланарных векторов $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ пространства называют *правой* (*положительно ориентированной*), если из конца вектора $\mathbf{e}_3$ (рис. 2) кратчайший поворот от $\mathbf{e}_1$ к $\mathbf{e}_2$ виден против часовой стрелки, и *левой* (*отрицательно ориентированной*) – в противном случае (начала векторов тройки считаются совмещенными).

Рис. 1

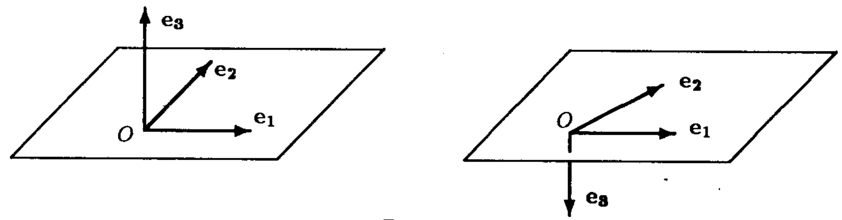


Рис. 2

Практически задание ориентации в геометрических пространствах означает задание направления движения на прямой (слева направо или наоборот), направления вращения на плоскости (против часовой стрелки или наоборот) или винта в пространстве (правого или левого).

Определения, основные факты. Пусть в пространстве V_3 выбрана ориентация. Базисы, задающие эту ориентацию, назовем правыми (положительными).

Векторным произведением ненулевых векторов **a** и **b** называется вектор **c** такой, что:

1) $|\mathbf{c}| = |\mathbf{a}| \cdot |\mathbf{b}| \sin(\widehat{\mathbf{a}, \mathbf{b}})$,

2) **c** ортогонален каждому из векторов **a** и **b**

и если $\mathbf{c} \neq \mathbf{0}$, то

3) **c** направлен так, что упорядоченная тройка **a**, **b**, **c** – правая.

Если один из векторов **a** или **b** нулевой, то векторное произведение считается равным **0**. Обозначение: $[\mathbf{a}, \mathbf{b}]$.

Теорема 23.2 (критерий коллинеарности). *Векторы* **a** *и* **b** *коллинеарны тогда и только тогда, когда* $[\mathbf{a}, \mathbf{b}] = \mathbf{0}$.

Доказательство. Действительно, векторы **a** и **b** коллинеарны тогда и только тогда, когда либо $\mathbf{a} = \mathbf{0}$, либо $\mathbf{b} = \mathbf{0}$, либо $\sin(\widehat{\mathbf{a}, \mathbf{b}}) = 0$; это равносильно тому, что $|[\mathbf{a}, \mathbf{b}]| = 0$, т.е. $[\mathbf{a}, \mathbf{b}] = \mathbf{0}$. ■

Замечание 2. Из теоремы 23.2 следует, что определение векторного произведения $[\mathbf{a}, \mathbf{b}]$ для коллинеарных векторов **a** и **b** заканчивается требованием 1. Если же **a** и **b** не коллинеарны, то условия 1 и 2 означают, что:

а) $|[\mathbf{a}, \mathbf{b}]| = S_{\mathbf{ab}}$, где $S_{\mathbf{ab}}$ – площадь параллелограмма, построенного на векторах **a** и **b**;

б) вектор $[\mathbf{a}, \mathbf{b}]$ перпендикулярен плоскости $\pi(\mathbf{a}, \mathbf{b})$, определяемой векторами **a** и **b**.

Теорема 23.3. *Векторное произведение антикоммутативно, т.е.* $[\mathbf{a}, \mathbf{b}] = -[\mathbf{b}, \mathbf{a}]$, $\forall \mathbf{a}, \mathbf{b}$.

Доказательство. Утверждение теоремы очевидно, если **a** и **b** коллинеарны. Пусть **a** и **b** не коллинеарны, тогда $[\mathbf{a}, \mathbf{b}] \neq \mathbf{0}$, $[\mathbf{b}, \mathbf{a}] \neq \mathbf{0}$, при этом $|[\mathbf{a}, \mathbf{b}]| = |[\mathbf{b}, \mathbf{a}]| = S_{\mathbf{ab}}$ и $[\mathbf{a}, \mathbf{b}]$, $[\mathbf{b}, \mathbf{a}]$ перпендикулярны плоскости $\pi(\mathbf{a}, \mathbf{b})$. Значит, либо $[\mathbf{a}, \mathbf{b}] = [\mathbf{b}, \mathbf{a}]$, либо $[\mathbf{a}, \mathbf{b}] = -[\mathbf{b}, \mathbf{a}]$. Но вектор $[\mathbf{b}, \mathbf{a}] \neq [\mathbf{a}, \mathbf{b}]$, так как тройка векторов **b**, **a**, $[\mathbf{b}, \mathbf{a}]$ – правая (по определению векторного произведения) и, следовательно, тройка **a**, **b**, $[\mathbf{b}, \mathbf{a}]$ – левая (см. замечание 1). ■

Смешанным произведением векторов **a**, **b** и **c** называется число, равное скалярному произведению векторного произведения **a** и **b** на вектор **c**. Обозначение: $(\mathbf{a}, \mathbf{b}, \mathbf{c})$. Итак, $(\mathbf{a}, \mathbf{b}, \mathbf{c}) = ([\mathbf{a}, \mathbf{b}], \mathbf{c})$.

Теорема 23.4 (критерий компланарности). *Векторы* **a**, **b**, **c** *компланарны тогда и только тогда, когда* $(\mathbf{a}, \mathbf{b}, \mathbf{c}) = 0$.

Доказательство. Необходимость. Пусть **a**, **b**, **c** компланарны. Будем считать, что **a** и **b** не коллинеарны и $\mathbf{c} \neq \mathbf{0}$ (в каждом из этих случаев, очевидно, $(\mathbf{a}, \mathbf{b}, \mathbf{c}) = 0$). Тогда **a**, **b**, **c** параллельны плоскости $\pi(\mathbf{a}, \mathbf{b})$, причем $[\mathbf{a}, \mathbf{b}] \perp \pi(\mathbf{a}, \mathbf{b})$. Следовательно, $([\mathbf{a}, \mathbf{b}], \mathbf{c}) = 0$.

Достаточность. Пусть $(\mathbf{a}, \mathbf{b}, \mathbf{c}) = 0$. Тогда либо $|[\mathbf{a}, \mathbf{b}]| = 0$, либо $|\mathbf{c}| = 0$, либо $\cos\varphi = 0$, где φ – угол между векторами $[\mathbf{a}, \mathbf{b}]$ и **c**. Это означает, что либо **a** и **b** коллинеарны, либо $\mathbf{c} = \mathbf{0}$, либо **c** параллелен плоскости $\pi(\mathbf{a}, \mathbf{b})$. Во всех этих случаях **a**, **b** и **c** компланарны. ■

Теорема 23.5. *Смешанное произведение некомпланарных векторов* **a**, **b** *и* **c** *равно по абсолютной величине объему V параллелепипеда, построенного на приведенных к общему началу векторах* **a**, **b**, **c**. *Причем*

$$(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \begin{cases} V, & \text{если } \mathbf{a}, \mathbf{b}, \mathbf{c} \text{ – правая тройка;} \\ -V, & \text{если } \mathbf{a}, \mathbf{b}, \mathbf{c} \text{ – левая тройка.} \end{cases}$$

Доказательство. Из некомпланарности векторов **a**, **b** и **c** следует, что **a** и **b** не коллинеарны и $\mathbf{c} \neq \mathbf{0}$.

Отложив векторы **a**, **b**, **c** от одной точки O (рис. 3), получим параллелепипед, ребрами которого являются эти векторы. Обозначим через h высоту параллелепипеда, опущенную из конца вектора **c**. Тогда $V = S_{\mathbf{ab}}h$. Отсюда и из замечания 2 следует, что

$$V = |[\mathbf{a}, \mathbf{b}]| \cdot |\mathbf{c}| \cdot |\cos\varphi| = |(\mathbf{a}, \mathbf{b}, \mathbf{c})|. \tag{23.1}$$

Знак $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ определяется только знаком $\cos\varphi$, но $\cos\varphi > 0$ тогда и только тогда, когда векторы $[\mathbf{a}, \mathbf{b}]$ и **c** направлены в одну сторону от плоскости $\pi(\mathbf{a}, \mathbf{b})$, т.е. тогда и только тогда, когда тройка **a**, **b**, **c** – правая. В силу (23.1) отсюда следует утверждение теоремы. ■

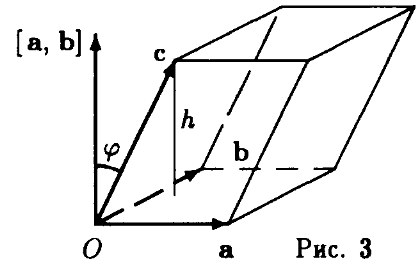


Рис. 3

Теорема 23.6. *Для любых векторов* **a**, **b**, **c**

$$([\mathbf{a}, \mathbf{b}], \mathbf{c}) = (\mathbf{a}, [\mathbf{b}, \mathbf{c}]). \tag{23.2}$$

Доказательство. Утверждение очевидно для компланарных векторов **a**, **b**, **c** (в силу теоремы 23.4). Пусть **a**, **b**, **c** – не компланарны. Тогда тройки **a**, **b**, **c** и **b**, **c**, **a** одинаково ориентированы (см. замечание 1). Так как $(\mathbf{a}, [\mathbf{b}, \mathbf{c}]) = (\mathbf{b}, \mathbf{c}, \mathbf{a})$, то из теоремы 23.5 следует (23.2). ■

Следствие 1. Для любых векторов **a**, **b**, **c** *имеют место равенства*

$$\begin{gathered}(\mathbf{a}, \mathbf{b}, \mathbf{c}) = (\mathbf{b}, \mathbf{c}, \mathbf{a}) = (\mathbf{c}, \mathbf{a}, \mathbf{b}) = \\ = -(\mathbf{b}, \mathbf{a}, \mathbf{c}) = -(\mathbf{a}, \mathbf{c}, \mathbf{b}) = -(\mathbf{c}, \mathbf{b}, \mathbf{a}).\end{gathered} \tag{23.3}$$

Следствие 2. Смешанное произведение линейно по каждому из сомножителей. Это утверждение вытекает из (23.3) и линейности скалярного произведения. ■

Теорема 23.7. *Векторное произведение линейно по каждому из сомножителей.*

Доказательство. В силу теоремы 23.3 достаточно показать, что для любых векторов **a**, **b**, **c** и любого числа $\alpha \in \mathbb{R}$ имеют место равенства $[\mathbf{a} + \mathbf{b}, \mathbf{c}] = [\mathbf{a}, \mathbf{c}] + [\mathbf{b}, \mathbf{c}]$ и $[\alpha\mathbf{a}, \mathbf{b}] = \alpha[\mathbf{a}, \mathbf{b}]$. Пусть $\mathbf{d} = [\mathbf{a} + \mathbf{b}, \mathbf{c}] - [\mathbf{a}, \mathbf{c}] - [\mathbf{b}, \mathbf{c}]$. Тогда $(\mathbf{d}, \mathbf{d}) = (\mathbf{a} + \mathbf{b}, \mathbf{c}, \mathbf{d}) - (\mathbf{a}, \mathbf{c}, \mathbf{d}) - (\mathbf{b}, \mathbf{c}, \mathbf{d})$. Из линейности смешанного произведения следует, что $(\mathbf{d}, \mathbf{d}) = (\mathbf{a}, \mathbf{c}, \mathbf{d}) + (\mathbf{b}, \mathbf{c}, \mathbf{d}) - (\mathbf{a}, \mathbf{c}, \mathbf{d}) - (\mathbf{b}, \mathbf{c}, \mathbf{d}) = 0$. Это доказывает первое из требуемых равенств. Второе равенство доказывается аналогично. ■

Векторное и смешанное произведения в прямоугольных координатах. Пусть $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ – ортонормированный базис пространства и пусть $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ – правая тройка.

1. Найдем координаты векторного произведения $[\mathbf{a}, \mathbf{b}]$, если векторы $\mathbf{a} = \{a_1, a_2, a_3\}$ и $\mathbf{b} = \{b_1, b_2, b_3\}$ заданы своими координатами в базисе $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$. Согласно теоремам 23.2 и 23.7 имеем $[\mathbf{a}, \mathbf{b}] = [a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3, b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + b_3\mathbf{e}_3] = a_1b_2[\mathbf{e}_1, \mathbf{e}_2] +$

$a_1b_3[\mathbf{e}_1, \mathbf{e}_3] + a_2b_1[\mathbf{e}_2, \mathbf{e}_1] + a_2b_3[\mathbf{e}_2, \mathbf{e}_3] + a_3b_1[\mathbf{e}_3, \mathbf{e}_1] + a_3b_2[\mathbf{e}_3, \mathbf{e}_2]$.
Отсюда в силу теоремы 23.3 следует, что

$$[\mathbf{a}, \mathbf{b}] = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} [\mathbf{e}_2, \mathbf{e}_3] - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} [\mathbf{e}_3, \mathbf{e}_1] + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} [\mathbf{e}_1, \mathbf{e}_2].$$

Пусть $[\mathbf{a}, \mathbf{b}] = x\,\mathbf{e}_1 + y\,\mathbf{e}_2 + z\,\mathbf{e}_3$. Тогда, применяя (22.4), теоремы 23.3–23.5 и замечание 1 (§23), получаем, что $x = ([\mathbf{a}, \mathbf{b}], \mathbf{e}_1) =$

$$= \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} (\mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_1) = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}.$$

Аналогично $y = -\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}$, $z = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$. Итак,

$$[\mathbf{a}, \mathbf{b}] = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \mathbf{e}_1 - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \mathbf{e}_2 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \mathbf{e}_3, \qquad (23.4)$$

или, в условной записи в виде мнемонического определителя,

$$[\mathbf{a}, \mathbf{b}] = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \qquad (23.5)$$

(имеется в виду разложение этого определителя по первой строке).

2. Из (23.4) и (22.5) непосредственно находится и смешанное произведение векторов $\mathbf{a} = \{a_1, a_2, a_3\}$, $\mathbf{b} = \{b_1, b_2, b_3\}$, $\mathbf{c} = \{c_1, c_2, c_3\}$, заданных своими координатами в ортонормированном базисе:

$$(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}. \qquad (23.6)$$

Замечание 3. Если исходный базис $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ отрицательно ориентирован, то $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = -1$ (теорема 23.5) и, следовательно, в соотношениях (23.4) – (23.6) следует поменять знаки на противоположные:

$$[\mathbf{a}, \mathbf{b}] = -\left(\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \mathbf{e}_1 - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \mathbf{e}_2 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \mathbf{e}_3 \right),$$

$$[\mathbf{a}, \mathbf{b}] = -\begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}, \quad (\mathbf{a}, \mathbf{b}, \mathbf{c}) = -\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

§24. Преобразование координат

Преобразование аффинной системы координат. Пусть в пространстве даны две аффинные (общие декартовы) системы координат: $\{O; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ и $\{O'; \mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'\}$. Первую из них будем называть "старой" системой координат, а вторую – "новой". Пусть

известно положение новой системы координат относительно старой: $O'(\alpha, \beta, \gamma)$ и $C = (c_{ij})$ – матрица перехода от базиса e к базису e', т.е.

$$\begin{aligned} \mathbf{e}_1' &= c_{11}\,\mathbf{e}_1 + c_{21}\,\mathbf{e}_2 + c_{31}\,\mathbf{e}_3, \\ \mathbf{e}_2' &= c_{12}\,\mathbf{e}_1 + c_{22}\,\mathbf{e}_2 + c_{32}\,\mathbf{e}_3, \\ \mathbf{e}_3' &= c_{13}\,\mathbf{e}_1 + c_{23}\,\mathbf{e}_2 + c_{33}\,\mathbf{e}_3. \end{aligned} \qquad (24.1)$$

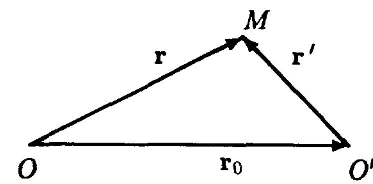


Рис. 1

Пусть M – произвольная точка пространства. Исследуем, как изменяются ее координаты при переходе к новой системе координат. Обозначим через (x, y, z) и (x', y', z') координаты точки M в старой и новой системах координат, а через $\mathbf{r}$, $\mathbf{r}'$ – радиус-векторы точки M относительно полюсов O и O' соответственно; через $\mathbf{r}_0$ – радиус-вектор точки O' относительно полюса O. Тогда по определению координат точки

$$\mathbf{r}_e = \begin{bmatrix} x \\ y \\ z \end{bmatrix}; \quad \mathbf{r}'_{e'} = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}; \quad (\mathbf{r}_0)_e = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}.$$

Равенство $\overrightarrow{OM} = \overrightarrow{OO'} + \overrightarrow{O'M}$ означает (рис. 1), что $\mathbf{r} = \mathbf{r}_0 + \mathbf{r}'$. Перейдя в этом векторном равенстве к равенству координат в базисе e, получаем с учетом (17.8), что

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} + C \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad \text{или} \quad \begin{aligned} x &= \alpha + c_{11}x' + c_{12}y' + c_{13}z', \\ y &= \beta + c_{21}x' + c_{22}y' + c_{23}z', \\ z &= \gamma + c_{31}x' + c_{32}y' + c_{33}z'. \end{aligned} \qquad (24.2)$$

Соотношения (24.2) называются *формулами преобразования координат*. Эти формулы выражают старые координаты точки через новые.

Ортогональная матрица. Вещественная матрица C называется *ортогональной*, если

$$CC^T = C^TC = I. \qquad (24.3)$$

Из определения следует, что:
1) C – квадратная матрица;
2) $C^{-1} = C^T$;
3) C^{-1} – ортогональная матрица;
4) $|C| = \pm 1$;
5) матрица $C = (c_{ij}) \in \mathbb{R}^{n\times n}$ ортогональна тогда и только тогда, когда

$$\sum_{k=1}^{n} c_{ik}c_{jk} = \begin{cases} 1, & i = j, \\ 0, & i \neq j \end{cases} \quad \text{и} \quad \sum_{k=1}^{n} c_{ki}c_{kj} = \begin{cases} 1, & i = j, \\ 0, & i \neq j; \end{cases} \qquad (24.4)$$

6) произведение ортогональных матриц – ортогональная матрица.

Теорема 24.1. *Матрица $C = (c_{ij})$ является матрицей перехода от одного ортонормированного базиса к другому тогда и только тогда, когда она ортогональна.*

Доказательство. Пусть $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ и $\mathbf{e}_1{}'$, $\mathbf{e}_2{}'$, $\mathbf{e}_3{}'$ – базисы пространства. Из (24.1) следует, что столбцы матрицы перехода C являются координатами векторов $\mathbf{e}_1{}'$, $\mathbf{e}_2{}'$, $\mathbf{e}_3{}'$ в базисе $e = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$. Согласно теореме 22.3 отсюда получаем, что $(\mathbf{e}_i', \mathbf{e}_j') = \sum_{k=1}^{3} c_{ki}c_{kj}$ тогда и только тогда, когда $e = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ – ортонормированный базис. Следовательно, условие ортонормированности базиса e' равносильно условию (24.4) ортогональности матрицы C. Аналогично рассматривается случай плоскости или прямой. ■

Преобразование прямоугольной декартовой системы координат на плоскости. Пусть $\{O; \mathbf{e}_1, \mathbf{e}_2\}$ и $\{O'; \mathbf{e}_1{}', \mathbf{e}_2{}'\}$ – прямоугольные декартовы системы координат на плоскости, т.е. $e = (\mathbf{e}_1, \mathbf{e}_2)$ и $e' = (\mathbf{e}_1{}', \mathbf{e}_2{}')$ – ортонормированные базисы и матрица перехода $C = (c_{ij}) \in \mathbb{R}^{2\times 2}$ ортогональна. Согласно (24.4) имеем

$$c_{11}^2 + c_{21}^2 = 1, \quad c_{11}c_{12} + c_{21}c_{22} = 0, \quad c_{12}^2 + c_{22}^2 = 1.$$

Из первого равенства следует существование такого φ, что $c_{11} = \cos\varphi$, $c_{21} = \sin\varphi$. Из второго равенства следует, что $c_{12} = k\sin\varphi$, $c_{22} = -k\cos\varphi$. Отсюда и из третьего равенства получим, что $k = \pm 1$.

Итак, ортогональная матрица второго порядка определяется лишь единственным параметром φ, при этом либо

$$C = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}, \quad \text{либо} \quad C = \begin{bmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{bmatrix}. \qquad (24.5)$$

1. Пусть $C = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$, тогда $|C| = 1$ и

$$\mathbf{e}_1{}' = \cos\varphi\cdot\mathbf{e}_1 + \sin\varphi\cdot\mathbf{e}_2, \quad \mathbf{e}_2{}' = -\sin\varphi\cdot\mathbf{e}_1 + \cos\varphi\cdot\mathbf{e}_2.$$

Умножая скалярно обе части этих равенств на $\mathbf{e}_1$ и $\mathbf{e}_2$, получим (рис. 2,*а*), что $(\widehat{\mathbf{e}_1{}', \mathbf{e}_1}) = \varphi$, $(\widehat{\mathbf{e}_1{}', \mathbf{e}_2}) = \frac{\pi}{2} - \varphi$, $(\widehat{\mathbf{e}_2{}', \mathbf{e}_1}) = \frac{\pi}{2} + \varphi$, $(\widehat{\mathbf{e}_2{}', \mathbf{e}_2}) = \varphi$ (все углы отсчитываются в направлении кратчайшего поворота от $\mathbf{e}_1$ к $\mathbf{e}_2$).

Это означает, что базисы e и e' одинаково ориентированы и что система координат $\{O'; \mathbf{e}_1{}', \mathbf{e}_2{}'\}$ может быть совмещена с $\{O; \mathbf{e}_1, \mathbf{e}_2\}$ путем переноса начала и поворота на угол φ вокруг начала. При этом формулы преобразования координат имеют вид

$$\begin{cases} x = \alpha + x'\cos\varphi - y'\sin\varphi, \\ y = \beta + x'\sin\varphi + y'\cos\varphi. \end{cases} \qquad (24.6)$$

2. Пусть $C = \begin{bmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{bmatrix}$, тогда $|C| = -1$ и

$$\mathbf{e}_1{}' = \cos\varphi\cdot\mathbf{e}_1 + \sin\varphi\cdot\mathbf{e}_2, \quad \mathbf{e}_2{}' = \sin\varphi\cdot\mathbf{e}_1 - \cos\varphi\cdot\mathbf{e}_2.$$

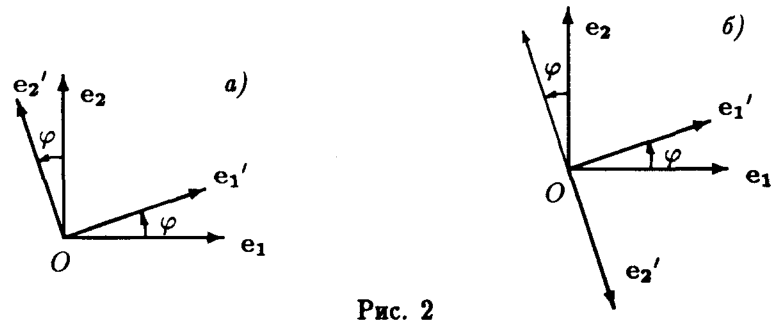


Рис. 2

Поступая аналогично п. 1, получим (рис. 2,*б*), что

$(\widehat{\mathbf{e}_1{}', \mathbf{e}_1}) = \varphi$, $\quad (\widehat{\mathbf{e}_1{}', \mathbf{e}_2}) = (\widehat{\mathbf{e}_2{}', \mathbf{e}_1}) = \frac{\pi}{2} - \varphi$, $\quad (\widehat{\mathbf{e}_2{}', \mathbf{e}_2}) = \pi - \varphi$.

В этом случае базисы e и e' противоположно ориентированы и система координат $\{O'; \mathbf{e}_1{}', \mathbf{e}_2{}'\}$ не может быть совмещена с $\{O; \mathbf{e}_1, \mathbf{e}_2\}$ путем переноса начала и поворота; после поворота на угол φ нужно выполнить отражение относительно оси $\mathbf{e}_1$ (т.е. изменить направление $\mathbf{e}_2$). При этом формулы преобразования координат имеют вид

$$\begin{cases} x = \alpha + x'\cos\varphi + y'\sin\varphi, \\ y = \beta + x'\sin\varphi - y'\cos\varphi. \end{cases}$$

§25. Полярные координаты

Декартовы системы координат – не единственный способ определять положение точки с помощью чисел и тем самым применять числовые расчеты для решения геометрических задач. Для этой цели используются и другие типы координатных систем. Опишем некоторые из них.

Полярные координаты на плоскости. Полярная система координат на плоскости состоит из точки O плоскости и исходящего из нее луча l. Точка O называется ***полюсом***, луч l — ***полярной осью***. Обозначение: $\{O; l\}$. Задание полярной системы координат позволяет однозначно определить положение любой точки плоскости с помощью двух чисел (рис. 1):

1) расстояния r от точки M до полюса O
и, если $M \neq O$,

2) угла φ между лучом $[OM)$ и полярной осью l (угол отсчитывается от оси l против часовой стрелки).

Число r называется ***полярным радиусом*** точки M, а угол φ — ***полярным углом***. Упорядоченная пара чисел (r, φ) называется ***полярными координатами*** точки M. Обозначение: $M(r, \varphi)$.

Из определения следует, что:

1) $r \geq 0$ для любой точки M плоскости и $r = 0 \iff M = O$;

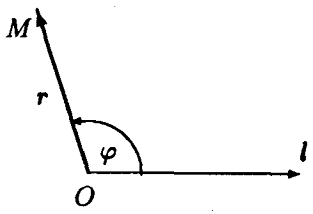


Рис. 1

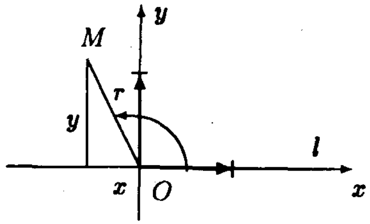


Рис. 2

2) $0 \le \varphi < 2\pi$ для любой точки плоскости, кроме полюса, и для полюса угол φ не определен;

3) любая точка плоскости, отличная от полюса, имеет полярные координаты, при этом две точки плоскости совпадают тогда и только тогда, когда совпадают их полярные координаты.

С каждой полярной системой координат связана некоторая прямоугольная декартова система координат. В этой системе (рис. 2) начало совпадает с полюсом, положительная полуось абсцисс – с полярной осью, положительная полуось ординат получается вращением полярной оси на угол $\pi/2$ против часовой стрелки. Полученную таким образом систему координат Oxy называют ***системой координат, определенной полярной системой координат*** $\{O; l\}$. Очевидно, что такое соответствие между полярными и прямоугольными системами координат биективно.

Как легко видеть из рис. 2, прямоугольные координаты (x, y) точки M и ее полярные координаты (r, φ) связаны соотношениями

$$\left\{\begin{array}{lcl} x & = & r\cos\varphi, \\ y & = & r\sin\varphi \end{array}\right. \quad \text{и} \quad \left\{\begin{array}{lcl} r & = & \sqrt{x^2+y^2}, \\ \cos\varphi & = & \dfrac{x}{r} = \dfrac{x}{\sqrt{x^2+y^2}}, \\ \sin\varphi & = & \dfrac{y}{r} = \dfrac{y}{\sqrt{x^2+y^2}}. \end{array}\right. \qquad (25.1)$$

Полярные координаты в пространстве. Обобщением полярной системы координат в пространстве являются цилиндрические и сферические системы координат.

Пусть в пространстве задана плоскость π и перпендикулярная к ней ось Oz. Пусть O – точка пересечения оси Oz и плоскости π.

И та и другая система координат состоят из полярной системы координат $\{O; l\}$ плоскости π и оси Oz. Обозначение: $\{O; l; z\}$. Точка O называется ***полюсом***, ось l – ***полярной осью***, ось Oz – ***зенитной осью***.

Цилиндрическими координатами точки M, не лежащей на зенитной оси, называются упорядоченная тройка чисел (r, φ, z), где (r, φ) – полярные координаты ортогональной проекции P точки M на плоскость π, z – координата на зенитной оси ортогональной проекции Q точки M на зенитную ось (рис. 3). Для точек зенитной оси $r = 0$ и φ не определен.

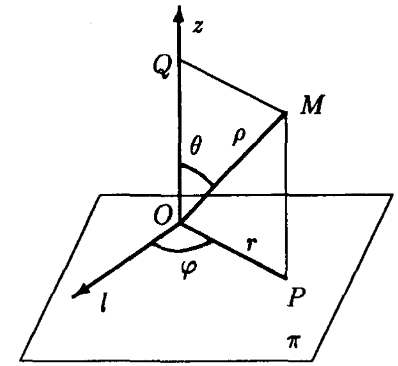


Рис. 3

Сферическими координатами точки M, не лежащей на зенитной оси, называется (рис. 3) упорядоченная тройка чисел (ρ, φ, θ), где ρ – расстояние от точки M до полюса O, θ – угол между радиус-вектором $\overrightarrow{OM}$ и положительной полуосью Oz (отсчитываемый от $\overrightarrow{OM}$ против часовой стрелки, если смотреть из полярной оси), φ – полярный угол точки P. При этом φ называется долготой точки M, θ – широтой, ρ – радиусом. Долгота не определена для всех точек зенитной оси, широта не определена для полюса.

Глава VI. Системы линейных алгебраических уравнений

§26. Постановка задачи

Терминология. *Системой m линейных алгебраических уравнений с n неизвестными* называется совокупность соотношений

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m, \end{cases} \tag{26.1}$$

где a_{ij}, b_i $(i = \overline{1,m},\ j = \overline{1,n})$ – заданные вещественные числа, а $x_1, \ldots, x_n$ – неизвестные величины. Числа a_{ij} называются *коэффициентами системы*, а b_i – *свободными членами*.

Упорядоченная совокупность чисел $c_1, \ldots, c_n \in \mathbb{R}$ называется *решением системы*, если при подстановке этих чисел в систему вместо неизвестных $x_1, \ldots, x_n$ соответственно каждое уравнение обращается в тождество.

Система уравнений называется *совместной*, если она имеет хотя бы одно решение, и *несовместной*, если не имеет ни одного решения. Система уравнений называется *определенной*, если она имеет единственное решение, и *неопределенной*, если имеет более одного решения.

Исследовать и решить систему – это значит:

- установить, совместна она или несовместна;
- если она совместна, установить, является она определенной или неопределенной, при этом:

– в случае определенной системы найти единственное ее решение;

– в случае неопределенной системы описать множество всех ее решений.

Компактная запись системы. Коэффициенты системы образуют матрицу $A = (a_{ij}) \in \mathbb{R}^{m\times n}$, называемую *основной матрицей системы*, свободные члены образуют столбец $b = (b_1, \ldots, b_m)^T \in \mathbb{R}^m$, называемый *столбцом свободных членов*, а неизвестные – столбец $x = (x_1, \ldots, x_n)^T$, называемый *столбцом неизвестных*. В этих обозначениях система (26.1) может быть записана в виде

$$Ax = b \tag{26.2}$$

или

$$x_1 a_1 + \ldots + x_n a_n = b, \tag{26.3}$$

где a_i $(i = \overline{1,n})$ – столбцы матрицы A.

В записи (26.3) система уравнений приобретает новый смысл: совместность системы равносильна тому, что столбец свободных членов является линейной комбинацией столбцов матрицы A, причем коэффициентами линейной комбинации служат компоненты $x_1, \ldots, x_n$ решения.

Эквивалентность систем. Две системы линейных алгебраических уравнений с одинаковым числом неизвестных называются *эквивалентными*, если множества всех решений этих систем совпадают.

Теорема 26.1. *Умножение обеих частей системы $Ax = b$ слева на невырожденную матрицу приводит ее к эквивалентной системе.*

Доказательство. Пусть $Q \in \mathbb{R}^{m\times m}$ и $|Q| \neq 0$. Рассматриваемые системы имеют вид $Ax = b$ и $QAx = Qb$. Если $c \in \mathbb{R}^n$ – решение первой системы, то $Ac \equiv b$ и, следовательно, $QAc \equiv Qb$, откуда следует, что c – решение второй системы. С другой стороны, если c – решение второй системы, то $QAc \equiv Qb$. Умножив обе части этого тождества слева на Q^{-1}, получаем тождество $Ac \equiv b$, откуда следует, что c – решение первой системы. ■

§27. Системы с квадратной невырожденной матрицей

Прежде чем рассматривать системы общего вида, исследуем простейший класс систем (26.2), когда число уравнений совпадает с числом неизвестных и $|A| \neq 0$.

Теорема 27.1. *Система линейных алгебраических уравнений с квадратной невырожденной матрицей совместна и имеет единственное решение.*

Доказательство. В силу невырожденности матрицы A для нее существует обратная матрица A^{-1}. Непосредственной проверкой легко установить, что вектор

$$x = A^{-1}b \tag{27.1}$$

является решением системы (26.2). Это решение единственно, так как если y – другое решение системы (26.2), то $Ax \equiv Ay$. Умножив обе части этого тождества слева на A^{-1}, получим, что $x = y$. ■

Правило Крамера. Решение (27.1) может быть записано покомпонентно, если воспользоваться явным выражением (5.3) для обратной матрицы. Действительно, $x = \frac{1}{|A|}\widehat{A}b$ или, в соответствии с (5.1),

$$x_i = \frac{A_{1i}b_1 + A_{2i}b_2 + \ldots + a_{ni}b_n}{|A|}, \quad i = \overline{1,n}.$$

Эти соотношения в свете свойств определителя означают, что

$$x_i = |A_i|/|A|, \quad i = \overline{1,n}, \tag{27.2}$$

где A_i получается из матрицы A заменой ее i-го столбца столбцом свободных членов.

Формулы (27.2) называются ***правилом Крамера***.

Замечание. Правило Крамера дает решение системы в явном виде и в некотором смысле носит алгоритмический характер. Однако правило Крамера полезно лишь в теоретических исследованиях и противопоказано для практического использования в приложениях. В самом деле, для решения систем n-го порядка по правилу Крамера требуется вычислить $(n+1)$ определителей n-го порядка, тогда как большинство современных методов решения систем по объему вычислений равносильны вычислению одного определителя. В § 29 мы приведем описание одного из таких методов – метода Гаусса.

Продолжим теоретическое исследование систем, так как вопрос о решении систем с прямоугольной матрицей или квадратной, но вырожденной матрицей остается открытым.

§28. Системы общего вида

Совместность системы. Пусть теперь

$$Ax = b \tag{28.1}$$

– система общего вида и $A = (a_{ij}) \in \mathbb{R}^{m\times n}$. Исследование системы следует начать с вопроса о ее совместности. Для этой цели составим матрицу B, приписав к матрице A столбец свободных членов:

$$B = [A|b].$$

Матрица B называется ***расширенной матрицей*** системы (28.1).

Теорема 28.1 (теорема Кронекера–Капелли). *Система линейных алгебраических уравнений совместна тогда и только тогда, когда ранг основной матрицы равен рангу расширенной матрицы.*

Доказательство. Необходимость. Пусть система (28.1) совместна. Тогда из (26.3) следует, что существуют числа $x_1, \ldots, x_n \in \mathbb{R}$ такие, что $b = x_1a_1 + \ldots + x_na_n$. Следовательно, столбец b является линейной комбинацией столбцов $a_1, \ldots, a_n$ матрицы A. Из теоремы 16.8 следует, что $\operatorname{rg} A = \operatorname{rg} B$.

Достаточность. Пусть $\operatorname{rg} A = \operatorname{rg} B = r$. Возьмем в матрице A какой-нибудь базисный минор. Так как $\operatorname{rg} B = r$, то он же будет базисным минором и матрицы B. Тогда согласно теореме 16.1 о базисном миноре последний столбец матрицы B будет линейной комбинацией базисных столбцов, т.е. столбцов матрицы A. Следовательно, столбец свободных членов системы является линейной комбинацией столбцов матрицы A. Это означает совместность системы. ∎

Схема исследования совместной системы. Теорема Кронекера–Капелли устанавливает совместность системы. Перейдем к исследованию совместной системы. Итак, пусть система уравнений

$$\begin{cases} a_{11}x_1 + \ldots + a_{1r}x_r + a_{1,r+1}x_{r+1} + \ldots + a_{1n}x_n = b_1, \\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots \\ a_{r1}x_1 + \ldots + a_{rr}x_r + a_{r,r+1}x_{r+1} + \ldots + a_{rn}x_n = b_r, \\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots \\ a_{m1}x_1 + \ldots + a_{mr}x_r + a_{m,r+1}x_{r+1} + \ldots + a_{mn}x_n = b_m \end{cases} \tag{$*$}$$

совместна и $\operatorname{rg} A = \operatorname{rg} B = r$. Не нарушая общности рассуждений, будем считать, что базисный минор матрицы A находится в левом верхнем углу, так что

$$\begin{vmatrix} a_{11} & \ldots & a_{1r} \\ \ldots & \ldots & \ldots \\ a_{r1} & \ldots & a_{rr} \end{vmatrix} \neq 0. \tag{28.2}$$

Рассмотрим укороченную систему из первых r уравнений системы $(*)$, т.е. из уравнений, коэффициенты которых входят в базисный минор:

$$\begin{cases} a_{11}x_1 + \ldots + a_{1r}x_r + a_{1,r+1}x_{r+1} + \ldots + a_{1n}x_n = b_1, \\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots \\ a_{r1}x_1 + \ldots + a_{rr}x_r + a_{r,r+1}x_{r+1} + \ldots + a_{rn}x_n = b_r. \end{cases} \tag{$**$}$$

Теорема 28.2. *Укороченная система эквивалентна исходной системе:* $(*) \sim (**)$.

Доказательство. Обе системы содержат одинаковое число неизвестных. Очевидно, что любое решение системы $(*)$ является решением системы $(**)$. Покажем, что верно и обратное. Действительно, в расширенной матрице B системы $(*)$ первые r строк являются базисными. Следовательно, все остальные строки согласно теореме о базисном миноре будут линейными комбинациями этих строк. Это означает, что каждое уравнение системы $(*)$, начиная с $(r+1)$-го, будет линейной комбинацией (т.е. следствием) первых r уравнений этой системы. Отсюда вытекает, что каждое решение первых r уравнений системы $(*)$ обращает в тождества все последующие уравнения этой системы. ∎

Итак, задача исследования системы $(*)$ упрощена, теперь достаточно изучить укороченную систему $(**)$. Перейдем к этой задаче.

Если $r = n$, то система $(**)$ имеет единственное решение как система с квадратной невырожденной матрицей (§27).

Пусть $r < n$. Неизвестные $x_1, \ldots, x_r$, коэффициенты при которых входят в базисный минор, назовем ***главными***, а остальные неизвестные $x_{r+1}, \ldots, x_n$ – ***свободными***.

Запишем систему $(**)$ в виде

$$\begin{cases} a_{11}x_1 + \ldots + a_{1r}x_r = b_1 - a_{1,r+1}x_{r+1} - \ldots - a_{1n}x_n, \\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots \\ a_{r1}x_1 + \ldots + a_{rr}x_r = b_r - a_{r,r+1}x_{r,r+1} - \ldots - a_{rn}x_n. \end{cases} \tag{28.3}$$

Придав свободным неизвестным $x_{r+1}, \ldots, x_n$ произвольные значения $c_{r+1}, \ldots, c_n$, получим систему уравнений относительно неизвестных $x_1, \ldots, x_r$:

$$\left\{\begin{array}{l} a_{11}x_1 + \ldots + a_{1r}x_r = b_1 - a_{1,r+1}c_{r+1} - \ldots - a_{1n}c_n, \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ a_{r1}x_1 + \ldots + a_{rr}x_r = b_r - a_{r,r+1}c_{r,r+1} - \ldots - a_{rn}c_n \end{array}\right. \qquad (***)$$

: квадратной невырожденной (согласно (28.2)) матрицей. Эта систе-иа имеет (§27) единственное решение $c_1, \ldots, c_r$. Очевидно, что сово-купность $(c_1, \ldots, c_r, c_{r+1}, \ldots, c_n)$ является решением системы (**).

Теорема 28.3. *Придавая свободным неизвестным произ-зольные значения и вычисляя значения главных неизвестных из си-стемы (***), можно получить все решения системы (**).*

Доказательство. Пусть $(c_1, \ldots, c_r, c_{r+1}, \ldots, c_n)$ – произволь-юе решение системы (**). Покажем, что оно может быть получено указанным путем. Возьмем числа $c_{r+1}, \ldots, c_n$ в качестве значений для свободных неизвестных $x_{r+1}, \ldots, x_n$ и будем вычислять значения главных неизвестных, решая систему (***). Так как $(c_1, \ldots, c_r, c_{r+1}, \ldots, c_n)$ – решение системы (**), то $(c_1, \ldots, c_r)$ – решение системы (***). Но система (***) имеет единственное решение, следовательно, в качестве значений главных неизвестных мы можем получить только числа $(c_1, \ldots, c_r)$. ■

Итак, мы нашли правило, которое позволяет получить любое ре-шение системы (**), а следовательно, и произвольной системы ли-нейных алгебраических уравнений.

Теорема 28.4. *Система алгебраических уравнений с n не-известными имеет единственное решение тогда и только тогда, когда* $\operatorname{rg} A = \operatorname{rg} B = n$.

Доказательство теоремы фактически содержится в описан-ном выше правиле для получения решения системы. Действительно, если $\operatorname{rg} A = \operatorname{rg} B = n$, то, как указано выше, система имеет единствен-ное решение. Если же $\operatorname{rg} A = \operatorname{rg} B < n$, то среди неизвестных будет хотя бы одно свободное неизвестное. Придавая ему произвольные зна-чения, получим бесконечно много решений системы. ■

Общее решение системы. Как указывалось в §26, решить си-стему – значит описать множество всех ее решений. В случае опреде-ленной системы для этого достаточно найти то единственное решение, которым она обладает. В случае же неопределенной системы необхо-димо найти способ, позволяющий описать бесконечное множество ее решений. Один из таких способов состоит в следующем.

Решим систему (28.3) относительно главных неизвестных:

$$\begin{array}{l} x_1 = f_1(x_{r+1}, \ldots, x_n), \\ \ldots \quad \ldots \quad \ldots \quad \ldots \\ x_r = f_r(x_{r+1}, \ldots, x_n), \end{array} \qquad (28.4)$$

где $f_1, \ldots, f_r$ – некоторые однозначно (в силу теоремы 27.1) опреде-ляемые из (28.3) функции.

Соотношения (28.4) при произвольных $x_{r+1}, \ldots, x_n$ описывают множество всех решений системы и называются ***общим решением си-стемы.*** В отличие от общего, конкретное решение $x = (c_1, \ldots, c_n)^T$, где c_i, $i = \overline{1, n}$, – известные числа, называется ***частным решением.***

Однородные системы. Система линейных алгебраических ура-внений с нулевой правой частью называется *однородной.*

Все результаты общей теории справедливы и для этого частного случая, однако здесь имеет место некоторая специфика, на которой мы остановимся подробнее. Из теоремы Кронекера–Капелли следует, что однородная система

$$Ax = 0 \qquad (28.5)$$

всегда совместна, так как ранг расширенной матрицы $[A|0]$, очевид-но, равен рангу матрицы A. Впрочем, это видно и непосредственно: однородная система заведомо имеет решение $(0, \ldots, 0)^T$, называемое *тривиальным.* Согласно общей теории для однородных систем имеют место следующие теоремы.

Теорема 28.5. *Однородная система* (28.5) *с n неизвест-ными имеет нетривиальное решение тогда и только тогда, когда* $\operatorname{rg} A < n$.

Утверждение теоремы следует из теоремы 28.4, так как наличие нетривиального решения для однородной системы равносильно ее не-определенности. ■

Теорема 28.6. *Однородная система* (28.5) *с квадратной матрицей имеет нетривиальное решение тогда и только тогда, ко-гда* $|A| = 0$.

§29. Метод Гаусса исследования и решения систем

В соответствии с общими принципами метода Гаусса решения матричных задач (§4) укажем тип простейших систем линейных урав-нений, тип эквивалентных преобразований системы, а также пока-жем, что произвольная система линейных алгебраических уравнений указанными преобразованиями приводится к указанному типу.

Системы с трапециевидной матрицей. Пусть система

$$Ax = b \qquad (29.1)$$

– система с верхней трапециевидной матрицей A. Расширенная мат-рица этой системы имеет вид

$$\left[\begin{array}{ccccccc|c} a_{11} & a_{12} & \ldots & a_{1r} & a_{1,r+1} & \ldots & a_{1n} & b_1 \\ 0 & a_{22} & \ldots & a_{2r} & a_{2,r+1} & \ldots & a_{2n} & b_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & a_{rr} & a_{r,r+1} & \ldots & a_{rn} & b_r \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 & b_{r+1} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 & b_m \end{array}\right], \qquad (29.2)$$

где $a_{ii} \neq 0$, $i = \overline{1, r}$.

Мы не случайно выбрали системы с верхней трапециевидной матрицей. Для таких систем чрезвычайно просто устанавливается совместность и достаточно просто находится решение.

Теорема 29.1. *Система* (29.1) *с верхней трапециевидной матрицей совместна тогда и только тогда, когда* $b_k = 0$ *при* $k > r$.

Доказательство. Действительно, ранг трапециевидной матрицы A равен числу ненулевых строк, так как минор r-го порядка, расположенный в левом верхнем углу, отличен от нуля, а все миноры более высокого порядка содержат нулевые строки и поэтому равны нулю. Ранг расширенной матрицы (29.2) равен r тогда и только тогда, когда $b_k = 0$, $k = \overline{r+1, n}$, так как наличие хотя бы одного ненулевого элемента b_i, где $r < i \leq m$, означает наличие ненулевого минора $(r+1)$-го порядка:

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1r} & b_1 \\ 0 & a_{22} & \dots & a_{2r} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{rr} & b_r \\ 0 & 0 & \dots & 0 & b_i \end{vmatrix}.$$

Отсюда на основании теоремы Кронекера–Капелли следует утверждение теоремы. ■

Итак, совместность системы с верхней трапециевидной матрицей устанавливается чисто "визуально".

Пусть теперь система (29.1) совместна. Реализация всех пунктов общей теории исследования и решения совместной системы (§28) для системы (29.1) также проста.

1°. В качестве базисного минора матрицы A всегда можно взять минор, расположенный в левом верхнем углу.

2°. Укороченная система состоит из первых r уравнений.

3°. Если $r = n$, то система (29.1) станет системой с треугольной матрицей

$$\begin{cases} a_{11}x_1 + a_{12} + \ldots + a_{1r}x_r = b_1, \\ \qquad a_{22} + \ldots + a_{2r}x_r = b_2, \\ \qquad\qquad \ldots \quad \ldots \quad \ldots \\ \qquad\qquad\qquad a_{rr}x_r = b_r, \end{cases}$$

которая имеет единственное решение (§27); найти его не представляет труда: решая последовательно уравнения системы снизу вверх, мы каждый раз будем иметь дело с уравнением, содержащим только одно неизвестное.

4°. Если $r < n$, то неизвестные $x_{r+1}, \ldots, x_n$ будут свободными и система относительно главных неизвестных будет иметь вид

$$\begin{cases} a_{11}x_1 + \ldots + a_{1r}x_r &= b_1 - a_{1,r+1}x_{r+1} - \ldots - a_{1n}x_n, \\ \ldots \quad \ldots \quad \ldots & \ldots \quad \ldots \quad \ldots \quad \ldots \\ \qquad\quad a_{rr}x_r &= b_r - a_{r,r+1}x_{r+1} - \ldots - a_{rn}x_n. \end{cases} \tag{29.3}$$

Общее и частное решения исходной системы находятся из системы (29.3) с треугольной матрицей.

Элементарные преобразования системы уравнений. *Элементарными преобразованиями системы уравнений* называются преобразования следующих типов:

1) перестановка местами двух уравнений системы;

2) умножение какого-либо уравнения системы на число $\alpha \neq 0$;

3) прибавление к одному уравнению системы другого ее уравнения, умноженного на любое число β.

Теорема 29.2. *Элементарные преобразования системы линейных алгебраических уравнений приводят ее к эквивалентной системе.*

Доказательство. Элементарные преобразования системы уравнений означают элементарные преобразования строк расширенной матрицы B, которые, как известно (§3), равносильны умножению матрицы B слева на матрицы элементарных преобразований. Это, в свою очередь, согласно трактовке (2.3) операции умножения матриц равносильно умножению слева обеих частей системы уравнений на невырожденные матрицы элементарных преобразований, что приводит к эквивалентной системе (теорема 26.1). ■

Замечание 1. Рассматриваемые в теореме преобразования относятся *только к строкам* расширенной матрицы. Очевидно, что элементарные преобразования столбцов расширенной матрицы, вообще говоря, лишены смысла. Однако можно рассматривать элементарные преобразования столбцов основной матрицы A, в частности перестановки ее столбцов, которые означают перенумерацию неизвестных системы.

Приведение системы общего вида к системе с верхней трапециевидной матрицей. Как следует из теоремы 3.1 (и замечания 1 к ней), матрица A системы уравнений (28.1) общего вида элементарными преобразованиями строк и перестановками столбцов приводится к верхней трапециевидной форме (1.3). Если используемые при этом элементарные преобразования строк матрицы A применить к строкам всей расширенной матрицы B, то на основании теоремы 29.2 (и замечания к ней) мы придем к системе с верхней трапециевидной матрицей, решения которой отличаются от решений исходной системы только нумерацией неизвестных. *Метод Гаусса* исследования и решения системы уравнений состоит в приведении ее к системе с верхней трапециевидной матрицей с последующим исследованием и решением получившейся системы. При этом, если в процессе преобразования использовались перестановки столбцов основной матрицы A, то в полученных решениях необходимо восстановить исходную нумерацию неизвестных.

Процесс приведения системы к системе с трапециевидной матрицей называется *прямым ходом* метода Гаусса, а процесс решения системы с треугольной матрицей – *обратным ходом*.

По объему вычислений обратный ход оказывается несущественным в методе Гаусса, так как он требует выполнения $O(n^2)$ операций умножения, тогда как для прямого хода требуется выполнить $\frac{n^3}{3} + O(n^2)$ умножений (§4). Таким образом, метод Гаусса решения системы линейных алгебраических уравнений по объему вычислений равносилен вычислению одного определителя (см. §27, замечание).

Замечание 2. В основе метода Гаусса вычисления определителя, вычисления ранга матрицы и решения системы линейных алгебраических уравнений лежит один и тот же алгоритм – основной процесс приведения матрицы к ступенчатой форме. Однако, как отмечалось в §4, 16, 29, реализация этого алгоритма для каждой из этих задач имеет свою специфику. Наиболее "свободно" он реализуется для задачи вычисления ранга матрицы, так как любые элементарные преобразования и строк, и столбцов матрицы не изменяют ее ранга. В задаче вычисления определителя по-прежнему допускаются преобразования как строк, так и столбцов, но уже необходимо следить за изменением определителя, если в преобразованиях участвовали перестановки строк (или столбцов) или умножение строки (столбца) на число $\alpha \neq 0$. Аккуратнее следует относиться к задаче решения систем уравнений, где допускаются преобразования *только строк расширенной матрицы* и *перестановки столбцов только основной матрицы*.

§30. Геометрические свойства решений системы

Рассмотрим свойства решений системы линейных уравнений с точки зрения теории линейного пространства.

Пусть дана система линейных алгебраических уравнений

$$Ax = b \tag{30.1}$$

с матрицей $A \in \mathbb{R}^{m\times n}$. Каждое решение этой системы можно рассматривать как вектор арифметического пространства $\mathbb{R}^n$. Совокупность всех решений системы (30.1) образует некоторое подмножество пространства $\mathbb{R}^n$. Выясним, какими свойствами оно обладает.

Линейное подпространство решений однородной системы.

Теорема 30.1. *Множество всех решений однородной системы $Ax = 0$ с n неизвестными является линейным подпространством арифметического пространства $\mathbb{R}^n$.*

Доказательство. Обозначим $L = \{x \in \mathbb{R}^n | Ax = 0\}$. Легко проверить, что если $x \in L$, $y \in L$, то $x + y \in L$, и, кроме того, если $\alpha \in \mathbb{R}$, то $\alpha x \in L$. В силу теоремы 18.1 отсюда следует, что L – линейное подпространство пространства $\mathbb{R}^n$. ■

Эта теорема позволяет "нарисовать" множество решений однородной системы, так как геометрическим образом линейного подпространства являются (§18) прямая и плоскость, проходящие через начало координат.

Произвольный базис подпространства решений однородной системы линейных уравнений называется *фундаментальной системой решений*. Понятно, что фундаментальная система решений существует лишь в том случае, когда однородная система имеет нетривиальное решение. При этом система уравнений может обладать многими фундаментальными системами решений. Однако все эти системы состоят из одинакового числа векторов, равного максимальному числу линейно независимых решений однородной системы. Определим это число.

Теорема 30.2. *Размерность пространства решений однородной системы $Ax = 0$ с n неизвестными равна $n - r$, где $r = \operatorname{rg} A$.*

Доказательство. Построим фундаментальную систему решений. Для этого воспользуемся аппаратом свободных неизвестных. Пусть $x_1, \ldots, x_r$ – главные неизвестные. Придадим свободным неизвестным $x_{r+1}, \ldots, x_n$ следующие $n - r$ наборов решений: $(1, 0, \ldots, 0)$, $(0, 1, \ldots, 0)$, …, $(0, 0, \ldots, 1)$. Для каждого из этих наборов найдем соответствующие значения r главных неизвестных. Тем самым найдем $n - r$ решений системы:

$$\begin{array}{lcl} e_1 & = & (c_{11}, \; c_{12}, \; \ldots, \; c_{1r}, \; 1, 0, \ldots, 0)^T, \\ e_2 & = & (c_{21}, \; c_{22}, \; \ldots, \; c_{2r}, \; 0, 1, \ldots, 0)^T, \\ \ldots & & \ldots \\ e_{n-r} & = & (c_{n-r,1}, c_{n-r,2}, \ldots, c_{n-r,r}, \; 0, 0, \ldots, 1)^T. \end{array} \tag{30.2}$$

Решения $e_1, \ldots, e_{n-r}$ обладают следующими свойствами.

1. Решения $e_1, \ldots, e_{n-r}$ линейно независимы, так как матрица

$$C = \left[\begin{array}{cccc|cccc} c_{11} & c_{12} & \ldots & c_{1r} & 1 & 0 & \ldots & 0 \\ c_{21} & c_{22} & \ldots & c_{2r} & 0 & 1 & \ldots & 0 \\ & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ c_{n-r,1} & c_{n-r,2} & \ldots & c_{n-r,r} & 0 & 0 & \ldots & 1 \end{array}\right]$$

имеет ранг, равный $n - r$ (ибо содержит минор $(n - r)$-го порядка, отличный от нуля), а ранг матрицы согласно теореме 16.3 равен максимальному числу линейно независимых строк.

2. Любое решение $x = (x_1, \ldots, x_r, x_{r+1}, \ldots, x_n)^T$ является линейной комбинацией $e_1, \ldots, e_{n-r}$, и, более того,

$$x = x_{r+1}e_1 + \cdots + x_n e_{n-r}. \tag{30.3}$$

Действительно, обозначим $y = x - x_{r+1}e_1 - \ldots - x_n e_{n-r}$. Пусть $y = (y_1, \ldots, y_n)^T$. Очевидно, что y – решение системы. Найдем его, пользуясь общей теорией решения систем: придадим свободным неизвестным значения $y_{r+1}, \ldots, y_n$. Пользуясь соотношением (30.2), легко показать, что $y_{r+1} = \ldots = y_n = 0$. Следовательно, укороченная система будет однородной системой, имеющей единственное и, очевидно, тривиальное решение $y_1 = \ldots = y_r = 0$. Таким образом, $y = 0$. Отсюда вытекает (30.3). Из свойств 1 и 2 следует, что $e_1, \ldots, e_{n-r}$ – фундаментальная система решений. ■

Построенная фундаментальная система решений называется *нормальной фундаментальной системой решений*. Общий принцип построения фундаментальной системы решений следует из теоремы 30.2: так как размерность пространства решений однородной системы равна $n - r$, то для построения фундаментальной системы решений достаточно найти любые $n - r$ линейно независимых решений (§ 17, утверждение 1). Для этого достаточно свободным неизвестным придать $n - r$ линейно независимых наборов значений, т.е. наборов вида

$(c_{1,r+1},\ldots,c_{1n}),\ \ldots,\ (c_{n-r,r+1},\ldots,c_{n-r,n})$, для которых

$$\begin{vmatrix} c_{1,r+1} & \ldots & c_{1n} \\ \ldots & \ldots & \ldots \\ c_{n-r,r+1} & \ldots & c_{n-r,n} \end{vmatrix} \neq 0. \tag{30.4}$$

Если для каждого из этих наборов найти соответствующие значения главных неизвестных, то получим $n-r$ решений системы:

$$\begin{array}{lllllll} e_1 = & (c_{11}, & \ldots, & c_{1r}, & c_{1,r+1}, & \ldots, & c_{1n})^T, \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ e_{n-r} = & (c_{n-r,1}, & \ldots, & c_{n-r,r}, & c_{n-r,r+1}, & \ldots, & c_{n-r,n})^T, \end{array}$$

линейно независимых вследствие (30.4).

Общее решение однородной системы. Фундаментальная система решений $e_1,\ldots,e_{n-r}$ однородной системы линейных уравнений позволяет записать любое решение системы в общем виде:

$$x = \alpha_1 e_1 + \ldots + \alpha_{n-r}e_{n-r}\,, \quad \forall \alpha_1,\ldots,\alpha_{n-r} \in \mathbb{R}. \tag{30.5}$$

Представление (30.5) решения называется ***общим решением однородной системы*** уравнений ***через фундаментальную систему решений*** (в отличие от общего решения (28.4) через свободные неизвестные).

Линейное многообразие решений неоднородной системы. Пусть

$$Ax = b \tag{30.6}$$

– неоднородная система линейных алгебраических уравнений. Однородная система

$$Ax = 0, \tag{30.7}$$

полученная из системы (30.6) заменой свободных членов нулями, называется ***приведенной однородной*** системой для системы (30.6).

Между решениями обеих систем существует тесная связь. Легко проверить следующие факты.

1°. *Сумма решений неоднородной и приведенной однородной систем является решением неоднородной системы.*

2°. *Разность двух решений неоднородной системы является решением приведенной однородной системы.*

Теорема 30.3. *Множество всех решений неоднородной системы является линейным многообразием, полученным сдвигом подпространства решений приведенной однородной системы на частное решение неоднородной системы.*

Доказательство. Пусть H – множество всех решений неоднородной системы (30.6), c – частное его решение и L – множество всех решений приведенной системы (30.7). Покажем, что $H = c + L$. Действительно, если $z \in H$, то $z = c+(z-c) = c+x$, где $x = z-c$. Так как $x \in L$ (как разность двух решений системы (30.6)), то $z \in c + L$ и, следовательно, $H \subset c + L$. С другой стороны, если $z \in c + L$, то $z = c + x$, $x \in L$. Значит, $z \in H$ как сумма решений неоднородной и приведенной систем. Следовательно, $c + L \subset H$ и $H = c + L$. ■

Общее решение неоднородной системы. Итак, найдя одно (частное) решение неоднородной системы и прибавляя его к каждому решению приведенной системы, можно получить все решения неоднородной системы. В силу (30.5) это позволяет записать решение неоднородной системы в общем виде следующим образом:

$$z = c + \alpha_1 e_1 + \ldots + \alpha_{n-r}e_{n-r}, \quad \forall \alpha_1,\ldots,\alpha_{n-r} \in \mathbb{R}, \tag{30.8}$$

где c – частное решение (30.6), а $e_1,\ldots,e_{n-r}$ – фундаментальная система решений (30.7). Представление (30.8) решения называется ***общим решением неоднородной*** системы уравнений ***через фундаментальную систему решений***. Итак, общее решение неоднородной системы есть сумма частного решения неоднородной системы и общего решения приведенной однородной системы.

Замечание. В соответствии с теоремой 30.3 геометрическим образом множества всех решений неоднородной системы уравнений служат прямая и плоскость, не проходящие через начало координат (§18).

Глава VII. Алгебраические линии и поверхности первого порядка

§31. Понятие об уравнениях линии и поверхности

Пусть Oxy и $Oxyz$ – аффинные системы координат на плоскости и в пространстве. Уравнение

$$F(x,y)=0, \tag{31.1}$$

соответственно

$$F(x,y,z)=0, \tag{31.2}$$

называется ***уравнением линии $\mathcal{L}$ на плоскости*** (***поверхности π в пространстве***) в заданной системе координат, если этому уравнению удовлетворяют координаты всех точек линии $\mathcal{L}$ (поверхности π), и только они. Очевидно, что два уравнения в заданной системе координат определяют одну и ту же линию (поверхность) тогда и только тогда, когда они эквивалентны.

Алгебраическим одночленом относительно переменных x, y (соответственно x, y, z) с вещественным коэффициентом λ называется выражение

$$\lambda x^p y^q \qquad (\lambda x^p y^q z^r), \tag{31.3}$$

где p, q, r – целые неотрицательные числа. Если $\lambda \neq 0$, то число $p+q$ $(p+q+r)$ называется *степенью одночлена*. *Алгебраическим многочленом* относительно переменных x, y (x, y, z) с вещественными коэффициентами называется конечная сумма алгебраических одночленов (31.3). Наибольшая степень одночленов, входящих в многочлен, называется *степенью многочлена*.

Линия на плоскости (поверхность в пространстве) называется *алгебраической*, если в некоторой аффинной системе координат она определяется уравнением (31.1) (соответственно (31.2)), где $F(x,y)$ (соответственно $F(x,y,z)$) – алгебраический многочлен от переменных x, y (x, y, z) с вещественными коэффициентами. Степень многочлена $F(x,y)$ (соответственно $F(x,y,z)$) называется *порядком линии* (*поверхности*).

Корректность этого определения вытекает из следующей теоремы.

Теорема 31.1 (**об инвариантности порядка**). *При переходе от одной аффинной системы координат к другой алгебраическая линия (поверхность) остается алгебраической и порядок ее не изменяется.*

Доказательство проведем для линии. Пусть на плоскости в аффинной системе координат Oxy линия $\mathcal{L}$ определяется уравнением (31.1), где $F(x,y)$ – алгебраический многочлен степени n. При переходе к новой системе координат $O'x'y'$ уравнение (31.1) преобразуется в уравнение $F'(x',y')=0$. Покажем, что $F'(x',y')$ – тоже алгебраический многочлен и его степень $n'=n$. Для этого в каждый одночлен $\lambda x^p y^q$ многочлена $F(x,y)$ вместо x и y подставим их выражения через x', y' согласно формулам преобразования координат (§24). Тогда $\lambda x^p y^q = \lambda(c_{11}x' + c_{12}y' + \alpha)^p(c_{21}x' + c_{22}y' + \beta)^q$. Следовательно, одночлен $\lambda x^p y^q$ преобразуется в алгебраический многочлен от переменных x', y', степень которого не превосходит $p+q$. При этом многочлен $F(x,y)$ преобразуется в алгебраический многочлен $F'(x',y')$, степень которого $n' \leq n$. Если в этих рассуждениях поменять ролями системы координат, то получим, что $n \leq n'$, т.е. $n'=n$. ■

§32. Уравнения прямой на плоскости, плоскости в пространстве

Этот и несколько следующих параграфов посвящены прямой на плоскости и плоскости в пространстве. Мы будем изучать их одновременно, так как у них много общего. И это понятно, ведь и прямая на плоскости, и плоскость в пространстве представляют собой один и тот же объект в линейном пространстве, называемый гиперплоскостью (§18). В тех случаях, когда сходные теоремы имеют, по существу, одинаковые доказательства, мы будем доказывать только одну из них.

В этом параграфе рассматриваются различные типы уравнений прямой на плоскости (плоскости в пространстве), каждый тип уравнения определяется тем геометрическим заданием, которое однозначно определяет прямую (плоскость).

Канонические уравнения. Ненулевой вектор, коллинеарный прямой, называется ее ***направляющим вектором***. Из аксиом геометрии следует, что через любую точку проходит единственная прямая с заданным направляющим вектором.

Пусть прямая l с направляющим вектором $\mathbf{a}$ проходит через точку M_0. Очевидно, точка M лежит на прямой l (рис. 1) тогда и только тогда, когда векторы $\overrightarrow{M_0M}$ и $\mathbf{a}$ коллинеарны, т.е. линейно зависимы (§15). С учетом условия $\mathbf{a} \neq \mathbf{0}$ это равносильно тому, что вектор $\overrightarrow{M_0M}$ линейно

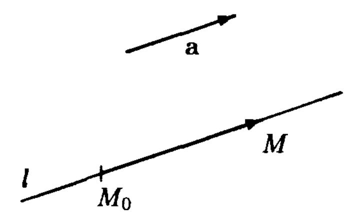


Рис. 1

выражается через вектор $\mathbf{a}$:

$$\overrightarrow{M_0M} = t\,\mathbf{a}, \quad t \in \mathbb{R}. \tag{32.1}$$

Таким образом, условию (32.1) удовлетворяют все точки M прямой l, и только они.

Теорема 32.1. *На плоскости в аффинной системе координат Oxy уравнение прямой l, проходящей через точку $M_0(x_0, y_0)$, с направляющим вектором $\mathbf{a} = \{m, n\}$ имеет вид*

$$\begin{vmatrix} x - x_0 & y - y_0 \\ m & n \end{vmatrix} = 0 \tag{32.2}$$

или

$$\frac{x - x_0}{m} = \frac{y - y_0}{n}. \tag{32.3}$$

Доказательство. Пусть точка M имеет координаты (x, y), тогда $\overrightarrow{M_0M} = \{x - x_0, y - y_0\}$. Условие (32.1) в силу линейности координат означает, что в определителе (32.2) первая строка линейно выражается через вторую, а это равносильно равенству (32.2). Итак, уравнению (32.2) удовлетворяют координаты (x, y) всех точек прямой l, и только они. Равенство нулю определителя второго порядка равносильно пропорциональности его строк, т.е. условию (32.3). ■

Замечание 1. Уравнение (32.3) означает лишь пропорциональность и в случае, когда $m = 0$ или $n = 0$, равносильно уравнению $x - x_0 = 0$ или $y - y_0 = 0$ соответственно.

Уравнения (32.2), (32.3) называются ***каноническими уравнениями прямой на плоскости.***

Следствие 1. *Уравнение прямой, проходящей через две различные точки $M_0(x_0, y_0)$ и $M_1(x_1, y_1)$, имеет вид*

$$\begin{vmatrix} x - x_0 & y - y_0 \\ x_1 - x_0 & y_1 - y_0 \end{vmatrix} = 0.$$

Это следует из того, что вектор $\overrightarrow{M_0M_1}$ является направляющим вектором прямой.

Без принципиальных изменений может быть получено аналогичное уравнение плоскости в пространстве. Два неколлинеарных вектора, параллельных плоскости, называются ее ***направляющими векторами.*** Из аксиом геометрии следует, что через любую точку проходит единственная плоскость с заданными направляющими векторами.

Пусть плоскость π с направляющими векторами $\mathbf{p}_1$ и $\mathbf{p}_2$ проходит через точку M_0. Очевидно, точка M лежит в плоскости π тогда и только тогда, когда (рис. 2) векторы $\overrightarrow{M_0M}$, $\mathbf{p}_1$, $\mathbf{p}_2$ компланарны, т.е. линейно зависимы (§15).

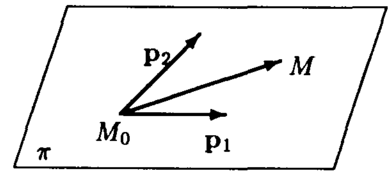


Рис. 2

С учетом условия неколлинеарности векторов $\mathbf{p}_1$ и $\mathbf{p}_2$ это равносильно тому, что вектор $\overrightarrow{M_0M}$ линейно выражается через $\mathbf{p}_1$ и $\mathbf{p}_2$:

$$\overrightarrow{M_0M} = u\,\mathbf{p}_1 + v\,\mathbf{p}_2, \quad u, v \in \mathbb{R}. \tag{32.4}$$

Теорема 32.2. *В пространстве в аффинной системе координат $Oxyz$ уравнение плоскости π, проходящей через точку $M_0(x_0, y_0, z_0)$, с направляющими векторами $\mathbf{p}_1 = \{m_1, n_1, k_1\}$ и $\mathbf{p}_2 = \{m_2, n_2, k_2\}$ имеет вид*

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ m_1 & n_1 & k_1 \\ m_2 & n_2 & k_2 \end{vmatrix} = 0. \tag{32.5}$$

Доказательство теоремы повторяет доказательство теоремы 32.1.

Уравнение (32.5) называется ***каноническим уравнением плоскости.***

Следствие 2. *Уравнение плоскости, проходящей через три точки $M_0(x_0, y_0, z_0)$, $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$, не лежащие на одной прямой, имеет вид*

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \end{vmatrix} = 0.$$

Параметрические уравнения. Этот тип уравнений представляет собой другую форму записи условий (32.1) и (32.4). Пусть $\mathbf{r} = \overrightarrow{OM}$, $\mathbf{r}_0 = \overrightarrow{OM_0}$ - радиус-векторы точек M и M_0 относительно полюса O. Тогда $\overrightarrow{M_0M} = \mathbf{r} - \mathbf{r}_0$ и условие (32.1) может быть записано в виде

$$\mathbf{r} = \mathbf{r}_0 + t\,\mathbf{a}, \quad t \in \mathbb{R}, \tag{32.6}$$

или, в координатной форме, в системе координат Oxy

$$\begin{cases} x = x_0 + tm, \\ y = y_0 + tn, \ t \in \mathbb{R}. \end{cases} \tag{32.7}$$

Уравнения (32.6), (32.7) называются ***параметрическими уравнениями прямой*** на плоскости в векторной и координатной формах.

Аналогично условие (32.4) может быть записано в виде

$$\mathbf{r} = \mathbf{r}_0 + u\,\mathbf{p}_1 + v\,\mathbf{p}_2, \quad u, v \in \mathbb{R}, \tag{32.8}$$

или, в координатной форме, в системе координат $Oxyz$

$$\left\{ \begin{array}{lcl} x &=& x_0 + um_1 + vm_2, \\ y &=& y_0 + un_1 + vn_2, \\ z &=& z_0 + uk_1 + vk_2, \quad u, v \in \mathbb{R}. \end{array} \right. \tag{32.9}$$

Уравнения (32.8), (32.9) называются ***параметрическими уравнениями плоскости*** в векторной и координатной формах.

Общие уравнения.

Теорема 32.3. *Линия на плоскости является прямой тогда и только тогда, когда она является алгебраической линией первого порядка.*

Доказательство. Необходимость. Пусть l – прямая на плоскости, проходящая через точку M_0 и параллельная ненулевому вектору $\mathbf{a}$. Пусть Oxy – произвольная аффинная система координат и $\mathbf{a} = \{m, n\}$, $M_0(x_0, y_0)$. Тогда прямая l описывается каноническим уравнением (32.2) или, что то же самое, уравнением $n(x - x_0) - m(y - y_0) = 0$, которое, если положить $A = n$, $B = -m$, $C = -nx_0 + my_0$, может быть записано в виде

$$Ax + By + C = 0. \tag{32.10}$$

Так как вектор $\mathbf{a} = \{m, n\} \neq \mathbf{0}$, то по крайней мере один из коэффициентов A или B отличен от нуля. Поэтому левая часть уравнения (32.10) представляет собой алгебраический многочлен первой степени. Следовательно, любая прямая на плоскости является алгебраической линией первого порядка.

Достаточность. Пусть в аффинной системе координат Oxy линия l определяется уравнением (32.10). Это уравнение имеет частное решение $x_0 = -\dfrac{AC}{A^2 + B^2}$, $y_0 = -\dfrac{BC}{A^2 + B^2}$, ибо $Ax_0 + By_0 + C \equiv 0$.

Вычитая последнее равенство из (32.10), получим, что $A(x - x_0) + B(y - y_0) = 0$ или, что то же самое,

$$\begin{vmatrix} x - x_0 & y - y_0 \\ -B & A \end{vmatrix} = 0.$$

В силу теоремы 32.1 это уравнение определяет прямую, проходящую через точку $M_0(x_0, y_0)$, с направляющим вектором $\mathbf{a} = \{-B, A\}$. ■

Уравнение (32.10) называется ***общим уравнением прямой на плоскости***. Вектор $\mathbf{n} = \{A, B\}$ называется ***вектором нормали к прямой относительно уравнения*** (32.10).

Теорема 32.4. *Поверхность в пространстве является плоскостью тогда и только тогда, когда она является алгебраической поверхностью первого порядка.*

Доказательство теоремы повторяет доказательство теоремы 32.3 с той лишь разницей, что уравнение плоскости имеет вид

$$Ax + By + Cz + D = 0, \quad \text{где} \quad A^2 + B^2 + C^2 \neq 0, \tag{32.11}$$

$$x_0 = -\frac{AD}{A^2 + B^2 + C^2}, \quad y_0 = -\frac{BD}{A^2 + B^2 + C^2}, \quad z_0 = -\frac{CD}{A^2 + B^2 + C^2},$$

а уравнение (32.11) может быть записано в виде

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ -B & A & 0 \\ -C & 0 & A \end{vmatrix} = 0, \quad \text{если} \quad A \neq 0$$

(случаи $B \neq 0$ и $C \neq 0$ рассматриваются аналогично). ■

Уравнение (32.11) называется ***общим уравнением плоскости*** в пространстве. Вектор $\mathbf{n} = \{A, B, C\}$ называется ***вектором нормали к плоскости относительно уравнения*** (32.11).

Общее уравнение прямой (плоскости) называется ***полным***, если все коэффициенты A, B, C (соответственно A, B, C, D) отличны от нуля.

Теорема 32.5. *В аффинной системе координат Oxy на плоскости ($Oxyz$ в пространстве) вектор $\mathbf{a} = \{m, n\}$ (соответственно $\mathbf{a} = \{m, n, k\}$) параллелен прямой (плоскости), заданной общим уравнением* (32.10) *(соответственно* (32.11)*), тогда и только тогда, когда*

$$Am + Bn = 0, \tag{32.12}$$

(соответственно

$$Am + Bn + Ck = 0). \tag{32.13}$$

Доказательство (для прямой). Как следует из доказательства теоремы 32.3, вектор $\mathbf{b} = \{-B, A\}$ является направляющим вектором прямой. Это означает, что вектор $\mathbf{a}$ параллелен этой прямой тогда и только тогда, когда $\mathbf{a}$ коллинеарен $\mathbf{b}$, т.е. когда $\begin{vmatrix} m & n \\ -B & A \end{vmatrix} = 0$ или, что то же самое, $Am + Bn = 0$. ■

Замечание 2. Левые части условий (32.12), (32.13) можно рассматривать как скалярные произведения вектора нормали $\mathbf{n}$ и вектора $\mathbf{a}$ в ортонормированном базисе. Таким образом, в ***прямоугольной декартовой системе координат вектор нормали*** $\mathbf{n} = \{A, B\}$ ***к прямой*** (32.10) ***(соответственно*** $\mathbf{n} = \{A, B, C\}$ ***к плоскости*** (32.11)***) перпендикулярен этой прямой (плоскости)***.

Уравнения в отрезках. Полные уравнения (32.10) и (32.11) прямой на плоскости и плоскости в пространстве могут быть записаны в следующем виде:

$$\frac{x}{-C/A}+\frac{y}{-C/B}=1 \quad \text{и} \quad \frac{x}{-D/A}+\frac{y}{-D/B}+\frac{z}{-D/C}=1.$$

Полагая $a=-C/A$, $b=-C/B$ (для прямой) и $a=-D/A$, $b=-D/B$, $c=-D/C$ (для плоскости), получим эквивалентные уравнения

$$\frac{x}{a}+\frac{y}{b}=1 \quad \text{и} \quad \frac{x}{a}+\frac{y}{b}+\frac{z}{c}=1,$$

называемые ***уравнениями*** прямой и соответственно плоскости ***в отрезках***. Числа a, b, c в этих уравнениях имеют простой геометрический смысл (рис. 3): они равны величинам отрезков, которые отсекает прямая (плоскость) на осях координат.

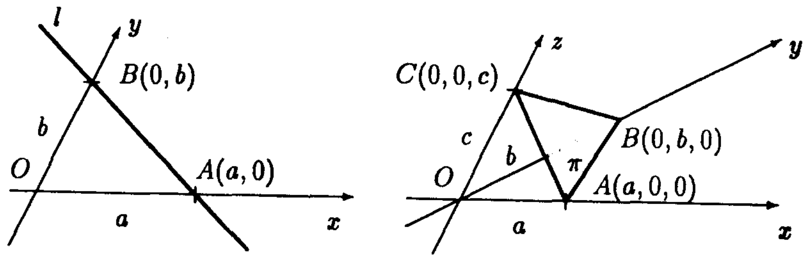


Рис. 3

Векторные уравнения. 1. Параметрическое уравнение (32.6) представляет собой векторное уравнение прямой на плоскости через направляющий вектор.

Аналогично параметрическое уравнение (32.8) представляет собой векторное уравнение плоскости в пространстве через направляющие векторы. Оно порождает другие формы векторных уравнений плоскости. В самом деле, это уравнение означает компланарность векторов $\mathbf{r}-\mathbf{r}_0$, $\mathbf{p}_1$ и $\mathbf{p}_2$, что согласно критерию компланарности (теорема 23.4) равносильно равенству

$$(\mathbf{r}-\mathbf{r}_0, \mathbf{p}_1, \mathbf{p}_2)=0 \tag{32.14}$$

или, в силу линейности смешанного произведения,

$$(\mathbf{r}, \mathbf{p}_1, \mathbf{p}_2)=D, \tag{32.15}$$

где D – константа, равная $(\mathbf{r}_0, \mathbf{p}_1, \mathbf{p}_2)$.

Уравнения (32.14), (32.15) представляют собой векторные уравнения плоскости через смешанное произведение.

2. Из аксиом геометрии следует, что на плоскости (в пространстве) через заданную точку проходит единственная прямая (соответственно плоскость), перпендикулярная заданному вектору.

Теорема 32.6. ***Уравнение прямой на плоскости (плоскости в пространстве), проходящей через точку $M_0(\mathbf{r}_0)$ перпендикулярно вектору $\mathbf{n}$, имеет вид***

$$(\mathbf{r}-\mathbf{r}_0, \mathbf{n})=0 \tag{32.16}$$

или, что то же самое,

$$(\mathbf{r}, \mathbf{n})=D, \tag{32.17}$$

где D – константа, равная $(\mathbf{r}_0, \mathbf{n})$.

Доказательство. Утверждение теоремы вытекает из того, что точка $M(\mathbf{r})$ лежит на прямой (на плоскости) тогда и только тогда, когда векторы $\overrightarrow{M_0M}$ и $\mathbf{n}$ ортогональны. ■

Уравнения (32.16), (32.17) представляют собой векторные уравнения прямой на плоскости и плоскости в пространстве через нормали.

§33. Взаимное расположение прямых на плоскости (плоскостей в пространстве)

Взаимное расположение двух прямых (плоскостей). Пусть на плоскости (в пространстве) две прямые (две плоскости) заданы общими уравнениями в аффинной системе координат Oxy $(Oxyz)$:

$$l_i: A_ix+B_iy+C_i=0, \quad A_i^2+B_i^2\neq 0, \quad i=1,2, \tag{33.1}$$

и соответственно

$$\pi_i: A_ix+B_iy+C_iz+D_i=0, \quad A_i^2+B_i^2+C_i^2\neq 0, \quad i=1,2. \tag{33.2}$$

Составим матрицы из коэффициентов этих уравнений:

$$A=\begin{bmatrix} A_1 & B_1 \\ A_2 & B_2 \end{bmatrix}, \quad B=\begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{bmatrix} \tag{33.3}$$

для прямых и

$$A=\begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{bmatrix}, \quad B=\begin{bmatrix} A_1 & B_1 & C_1 & D_1 \\ A_2 & B_2 & C_2 & D_2 \end{bmatrix} \tag{33.4}$$

для плоскостей. Очевидно, что в обоих случаях $1\leq \operatorname{rg} A\leq 2$, $1\leq \operatorname{rg} B\leq 2$, $\operatorname{rg} A\leq \operatorname{rg} B$. Следовательно, для рангов матриц A и B возможны только следующие три набора значений:

rg A	rg B
1	2
1	1
2	2

(33.5)

Теорема 33.1. *Прямые l_1 и l_2 на плоскости **совпадают** тогда и только тогда, когда*

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2}, \tag{33.6}$$

параллельны и не совпадают** тогда и только **тогда, когда

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} \neq \frac{C_1}{C_2}, \tag{33.7}$$

***пересекаются** тогда и только тогда, когда*

$$\frac{A_1}{A_2} \neq \frac{B_1}{B_2}. \tag{33.8}$$

Доказательство. Рассмотрим систему линейных алгебраических уравнений

$$\begin{cases} A_1x + B_1y + C_1 = 0, \\ A_2x + B_2y + C_2 = 0 \end{cases} \tag{33.9}$$

относительно неизвестных x, y. Для этой системы матрицы A и B из (33.3) являются основной и расширенной матрицами соответственно.

Согласно (33.5) эта система совместна только в двух случаях: когда $\mathrm{rg}\,A = \mathrm{rg}\,B = 1$ и $\mathrm{rg}\,A = \mathrm{rg}\,B = 2$. В случае когда $\mathrm{rg}\,A = 1$, а $\mathrm{rg}\,B = 2$, система несовместна. В первом случае решения системы образуют одномерное линейное многообразие (§30), а во втором – нульмерное, т.е. система имеет единственное решение. Перейдем к прямым l_1 и l_2. Совпадение прямых l_1 и l_2 означает, что решения системы (33.9) образуют одномерное линейное многообразие, т.е. что $\mathrm{rg}\,A = \mathrm{rg}\,B = 1$ или, что то же самое, $\mathrm{rg}\,B = 1$. Это условие равносильно (33.6) в силу теоремы о базисном миноре. Параллельность прямых l_1 и l_2 (и их несовпадение) означает, что система (33.9) несовместна, т.е. что $\mathrm{rg}\,A = 1$, $\mathrm{rg}\,B = 2$. Это условие равносильно (33.7). Пересекаемость прямых l_1 и l_2 означает, что система (33.9) имеет единственное решение, т.е. что $\mathrm{rg}\,A = \mathrm{rg}\,B = 2$ или, что то же самое, $\mathrm{rg}\,A = 2$. Это условие равносильно (33.8). ■

Теорема 33.2. *Плоскости π_1 и π_2 **совпадают** тогда и только тогда, когда*

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2} = \frac{D_1}{D_2},$$

параллельны и не совпадают** тогда и только **тогда, когда

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2} \neq \frac{D_1}{D_2},$$

***пересекаются** тогда и только тогда, когда коэффициенты A_1, B_1, C_1 и A_2, B_2, C_2 **не пропорциональны**.*

Доказательство теоремы представляет собой естественный аналог доказательства теоремы 33.1 и не содержит принципиальных отличий. Отметим лишь, что в случае, когда $\mathrm{rg}\,A = \mathrm{rg}\,B = 1$, решения системы образуют двумерное линейное многообразие (т.е. плоскости π_1 и π_2 совпадают), а в случае, когда $\mathrm{rg}\,A = \mathrm{rg}\,B = 2$, решения системы образуют одномерное линейное многообразие (т.е. плоскости пересекаются по прямой). ■

Замечание. Теоремам 33.1 и 33.2 можно дать единую формулировку в терминах матриц (33.3) и (33.4): *прямые l_1 и l_2 (плоскости π_1 и π_2)*

совпадают $\iff \mathrm{rg}\,B = 1$;

параллельны и не совпадают $\iff \mathrm{rg}\,A = 1, \mathrm{rg}\,B = 2$;

пересекаются $\iff \mathrm{rg}\,A = 2$.

Пучок прямых (плоскостей). Множество всех прямых плоскости, проходящих через данную точку M_0, называется ***пучком прямых*** с центром в точке M_0. Обозначение: $\pi(M_0)$. Пусть l_1 и l_2 – две несовпадающие прямые пучка $\pi(M_0)$, заданные уравнениями (33.1) в некоторой аффинной системе координат Oxy. Положим $F_i(x, y) = A_ix + B_iy + C_i$, где $i = 1, 2$.

Теорема 33.3. *Прямая принадлежит пучку $\pi(M_0)$ тогда и только тогда, когда она определяется уравнением*

$$\alpha F_1(x, y) + \beta F_2(x, y) = 0 \tag{33.10}$$

при некоторых $\alpha, \beta \in \mathbb{R}$, одновременно не равных нулю.

Доказательство. Достаточность очевидна, так как уравнение (33.10) является в силу (33.8) уравнением первой степени и определяет прямую. Она проходит через точку $M_0(x_0, y_0)$, поскольку $F_1(x_0, y_0) = 0$ и $F_2(x_0, y_0) = 0$.

Необходимость. Пусть прямая $l \in \pi(M_0)$. Возьмем на этой прямой точку $M_1(x_1, y_1)$, отличную от точки $M_0(x_0, y_0)$. Положим $\alpha = A_2x_1 + B_2y_1 + C_2$, $\beta = -(A_1x_1 + B_1y_1 + C_1)$. Поскольку точка M_1 не может одновременно лежать на l_1 и l_2 (ибо прямые l_1 и l_2 не совпадают), то по крайней мере одно из чисел α или β отлично от нуля. Тогда уравнение $\alpha F_1(x, y) + \beta F_2(x, y) = 0$ согласно условию (33.8) является уравнением первой степени и определяет прямую. Очевидно, она проходит через точки M_0 (так как прямые l_1, l_2 проходят через M_0) и M_1 (согласно выбору α, β). Следовательно, эта прямая совпадает с l. ■

Итак, любая прямая пучка $\pi(M_0)$ определяется двумя пересекающимися прямыми l_1 и l_2 этого пучка. Каждая пара чисел α, β, где $\alpha^2 + \beta^2 \neq 0$, определяет единственную прямую пучка. Уравнение (33.10) называется ***уравнением пучка прямых***, проходящих через точку пересечения прямых (33.1).

Множество всех плоскостей пространства, проходящих через прямую l, называется ***пучком плоскостей*** с осью l. Обозначение: $\pi(l)$. Пусть π_1 и π_2 – две пересекающиеся плоскости пучка $\pi(l)$, заданные уравнениями (33.2) в некоторой аффинной системе координат $Oxyz$. Положим $F_i(x,y,z) = A_i x + B_i y + C_i z + D_i$, где $i = 1, 2$.

Теорема 33.4. *Плоскость принадлежит пучку $\pi(l)$ тогда и только тогда, когда она определяется уравнением*

$$\alpha F_1(x,y,z) + \beta F_2(x,y,z) = 0 \tag{33.11}$$

при некоторых $\alpha, \beta \in \mathbb{R}$, одновременно не равных нулю.

Доказательство теоремы, по существу, повторяет доказательство теоремы 33.3. Отметим лишь, что в данном случае точка M_1 – любая точка плоскости, не лежащая на оси l. ■

Уравнение (33.11) называется ***уравнением пучка плоскостей***, проходящих через прямую пересечения плоскостей (33.2).

§34. Полуплоскости и полупространства

Из аксиом геометрии следует, что каждая прямая l на плоскости разбивает эту плоскость на две полуплоскости, при этом точки M_1 и M_2, не лежащие на прямой l, принадлежат одной полуплоскости, если отрезок $[M_1M_2]$ не имеет общих точек с прямой l, в противном случае они принадлежат разным полуплоскостям. Выясним, как этот геометрический факт описывается в аналитической форме.

Пусть прямая l в аффинной системе координат Oxy определяется уравнением

$$Ax + By + C = 0. \tag{34.1}$$

Теорема 34.1. *Точки $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$ принадлежат разным полуплоскостям относительно прямой l тогда и только тогда, когда*

$$(Ax_1 + By_1 + C)(Ax_2 + By_2 + C) < 0. \tag{34.2}$$

Доказательство. Предварительно заметим, что точка $M_0(x_0, y_0)$ является внутренней точкой отрезка $[M_1M_2]$ тогда и только тогда, когда $\overrightarrow{M_1M_0} = t\overrightarrow{M_0M_2}$, где $0 < t < 1$, т.е.

$$x_0 = x_1 + tx_2,\ y_0 = y_1 + ty_2, \quad 0 < t < 1.$$

Точки $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$ принадлежат разным полуплоскостям тогда и только тогда, когда существует точка $M_0(x_0, y_0)$, общая для прямой l и отрезка $[M_1M_2]$, причем точка M_0 является внутренней точкой отрезка $[M_1M_2]$, т.е.

$$\begin{cases} Ax_0 + By_0 + C = 0, \\ x_0 = x_1 + tx_2,\ y_0 = y_1 + ty_2, \quad 0 < t < 1. \end{cases}$$

С учетом очевидного тождества $C = tC + (1-t)C$ получим, что точки M_1, M_2 принадлежат разным полуплоскостям тогда и только тогда, когда существует число t такое, что

$$t(Ax_1 + By_1 + C) + (1-t)(Ax_2 + By_2 + C) = 0, \quad 0 < t < 1,$$

или, в обозначениях $Ax_1 + By_1 + C = F_1$, $Ax_2 + By_2 + C = F_2$,

$$(1-t)F_1 + tF_2 = 0, \qquad 0 < t < 1.$$

Это равносильно тому, что $F_1F_2 < 0$. ■

Итак, для координат (x, y) всех точек одной полуплоскости выполняется неравенство $Ax + By + C > 0$, а другой – неравенство $Ax + By + C < 0$. Полуплоскость, для точек $M(x, y)$ которой $Ax + By + C > 0$, называется ***положительной полуплоскостью*** относительно уравнения (34.1) прямой l и обозначается символом π_+, а полуплоскость, для точек которой $Ax + By + C < 0$, – ***отрицательной полуплоскостью*** и обозначается π_-.

Теорема 34.2. *Вектор нормали $\mathbf{n} = \{A, B\}$ к прямой l: $Ax + By + C = 0$, отложенный от любой точки прямой, направлен в сторону положительной полуплоскости.*

Доказательство. Пусть $M_0(x_0, y_0)$ – произвольная точка прямой l. Отложим вектор $\mathbf{n}$ от точки M_0, пусть конец вектора $\mathbf{n}$ совпадает с точкой $M_1(x_1, y_1)$. Тогда $\mathbf{n} = \{x_1 - x_0, y_1 - y_0\} = \{A, B\}$ и, следовательно, $x_1 = A + x_0$, $y_1 = B + y_0$. Подставив координаты x_1, y_1 точки M_1 в левую часть уравнения прямой l, получаем, что $Ax_1 + By_1 + C = A^2 + B^2 + Ax_0 + By_0 + C = A^2 + B^2 > 0$, так как $Ax_0 + By_0 + C = 0$ и $\mathbf{n} \neq \mathbf{0}$. ■

Аналогичные факты имеют место и в пространстве. Плоскость

$$\pi: \quad Ax + By + Cz + D = 0 \tag{34.3}$$

разбивает пространство на два полупространства.

Теорема 34.3. *Точки $M_1(x_1, y_1, z_1)$ и $M_2(x_2, y_2, z_2)$ принадлежат разным полупространствам относительно плоскости π тогда и только тогда, когда*

$$(Ax_1 + By_1 + Cz_1 + D)(Ax_2 + By_2 + Cz_2 + D) < 0.$$

Доказательство теоремы принципиально не отличается от доказательства теоремы 34.1. ■

Полупространство, для точек $M(x, y, z)$ которого $Ax + By + Cz + D > 0$, называется ***положительным полупространством*** относительно уравнения (34.3) плоскости π, а полупространство, для точек которого $Ax + By + Cz + D < 0$, – ***отрицательным полупространством***.

Теорема 34.4. *Вектор нормали $\mathbf{n} = \{A, B, C\}$ к плоскости π: $Ax + By + Cz + D = 0$, отложенный от любой точки плоскости, направлен в сторону положительного полупространства.*

Доказательство теоремы аналогично доказательству теоремы 34.2. ■

§35. Прямая на плоскости (плоскость в пространстве) в прямоугольной декартовой системе координат

Рассматривается прямоугольная декартова система координат Oxy на плоскости ($Oxyz$ в пространстве), определяемая ортонормированным базисом $\mathbf{e}_1, \mathbf{e}_2$ ($\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ соответственно). Известно (§32, замечание 2), что вектор нормали $\mathbf{n} = \{A, B\}$ (соответственно $\mathbf{n} = \{A, B, C\}$) перпендикулярен прямой l, заданной уравнением (32.10) (соответственно плоскости π, заданной уравнением (32.11)).

Расстояние от точки до прямой (до плоскости).

Теорема 35.1. *В прямоугольной декартовой системе координат Oxy расстояние $\rho(M_0, l)$ от точки $M_0(x_0, y_0)$ до прямой* (32.10) *определяется формулой*

$$\rho(M_0, l) = \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}. \tag{35.1}$$

Доказательство. Пусть $M_0 \notin l$ (для точек $M_0 \in l$ равенство (35.1) очевидно). Опустим из точки M_0 перпендикуляр на прямую l, пусть $M_1(x_1, y_1)$ – основание перпендикуляра. Тогда $\overrightarrow{M_1M_0} = \alpha\,\mathbf{n}$ и

$$\rho(M_0, l) = |\overrightarrow{M_1M_0}| = |\alpha| \cdot |\mathbf{n}| = |\alpha|\sqrt{A^2 + B^2}. \tag{35.2}$$

Найдем α. Так как $x_0 = x_1 + \alpha A$, $y_0 = y_1 + \alpha B$, то $Ax_0 + By_0 + C = Ax_1 + By_1 + C + \alpha(A^2 + B^2) = \alpha(A^2 + B^2)$. Отсюда $\alpha = (Ax_0 + By_0 + C)/(A^2 + B^2)$. Подставив это значение α в (35.2), получим (35.1). ■

Теорема 35.2. *В прямоугольной декартовой системе координат $Oxyz$ расстояние от точки $M_0(x_0, y_0, z_0)$ до плоскости* (32.11) *определяется формулой*

$$\rho(M_0, \pi) = \frac{|Ax_0 + By_0 + Cz_0 + D|}{\sqrt{A^2 + B^2 + C^2}}.$$

Угол между прямыми (между плоскостями). Пусть прямые l_1 и l_2 (плоскости π_1 и π_2) заданы уравнениями (33.1) (соответственно (33.2)).

Вообще говоря, две пересекающиеся прямые l_1 и l_2 (плоскости π_1 и π_2) образуют два угла, в сумме равные π. Достаточно определить один их них. Так как векторы нормали $\mathbf{n}_1$ и $\mathbf{n}_2$ перпендикулярны прямым (плоскостям), то угол $\varphi = (\widehat{\mathbf{n}_1, \mathbf{n}_2})$ совпадает с одним из углов между прямыми l_1 и l_2 (между плоскостями π_1 и π_2). Итак, согласно (22.6) *угол φ между прямыми* (33.1), *совпадающий с углом между их нормалями, определяется формулой*

$$\cos\varphi = \frac{A_1A_2 + B_1B_2}{\sqrt{A_1^2 + B_1^2}\sqrt{A_2^2 + B_2^2}},$$

угол φ между плоскостями (33.2), *совпадающий с углом между их нормалями, определяется формулой*

$$\cos\varphi = \frac{A_1A_2 + B_1B_2 + C_1C_2}{\sqrt{A_1^2 + B_1^2 + C_1^2}\sqrt{A_2^2 + B_2^2 + C_2^2}}.$$

В частности, *прямые l_1 и l_2 (плоскости π_1 и π_2) перпендикулярны тогда и только тогда, когда*

$$A_1A_2 + B_1B_2 = 0 \qquad (A_1A_2 + B_1B_2 + C_1C_2 = 0).$$

§36. Прямая в пространстве

Уравнения прямой. 1. Векторное уравнение (32.6) прямой на плоскости остается справедливым и для прямой в пространстве, так как и в пространстве прямая однозначно определяется точкой и направляющим вектором. Итак (§32), если в пространстве зафиксирован полюс O, то уравнение прямой, проходящей через точку $M_0(\mathbf{r}_0)$, с направляющим вектором $\mathbf{a}$ имеет вид

$$\mathbf{r} = \mathbf{r}_0 + t\,\mathbf{a}, \quad t \in \mathbb{R}. \tag{36.1}$$

Пусть в аффинной системе координат точка M_0 имеет координаты (x_0, y_0, z_0), а вектор $\mathbf{a} = \{m, n, k\}$. Тогда уравнение (36.1) может быть записано в координатной форме:

$$\begin{cases} x = x_0 + tm, \\ y = y_0 + tn, \\ z = z_0 + tk, \quad t \in \mathbb{R}. \end{cases} \tag{36.2}$$

Уравнения (36.1), (36.2) называются ***параметрическими уравнениями прямой*** в пространстве в векторной и координатной формах соответственно.

2. Уравнение (36.1), означающее коллинеарность векторов $\mathbf{r} - \mathbf{r}_0$ и $\mathbf{a}$, может быть записано и в терминах пропорциональности координат векторов $\mathbf{r} - \mathbf{r}_0 = \{x - x_0, y - y_0, z - z_0\}$ и $\mathbf{a} = \{m, n, k\}$:

$$\frac{x - x_0}{m} = \frac{y - y_0}{n} = \frac{z - z_0}{k}. \tag{36.3}$$

Уравнения (36.3) называются ***каноническими уравнениями прямой*** в пространстве.

Так же как и для канонического уравнения (32.3) прямой на плоскости, уравнения (36.3) говорят лишь о пропорциональности координат векторов $\mathbf{r} - \mathbf{r}_0$ и $\mathbf{a}$. Если, например, $m = 0$, то уравнения (36.3) переходят в уравнения

$$x - x_0 = 0, \quad \frac{y - y_0}{n} = \frac{z - z_0}{k}.$$

Если же $m = 0$, $n = 0$, то уравнения (36.3) переходят в уравнения

$$x - x_0 = 0, \quad y - y_0 = 0,$$

т.е. прямая является линией пересечения плоскостей $x - x_0 = 0$ и $y - y_0 = 0$.

3. Для прямой, проходящей через две различные точки $M_0(x_0, y_0, z_0)$ и $M_1(x_1, y_1, z_1)$, легко получить уравнения (36.1)–(36.3), так как в качестве направляющего вектора может быть взят вектор $\mathbf{a} = \overrightarrow{M_0M_1} = \{x_1 - x_0, y_1 - y_0, z_1 - z_0\}$.

4. Векторное уравнение (36.1) порождает и другие формы векторных уравнений прямой в пространстве. В самом деле, это уравнение означает коллинеарность векторов $\mathbf{r} - \mathbf{r}_0$ и $\mathbf{a}$, что согласно критерию коллинеарности (теорема 23.2) равносильно равенству

$$[\mathbf{r} - \mathbf{r}_0, \mathbf{a}] = \mathbf{0} \tag{36.4}$$

или, в силу линейности векторного произведения,

$$[\mathbf{r}, \mathbf{a}] = \mathbf{M}, \quad \text{где } \mathbf{M} = [\mathbf{r}_0, \mathbf{a}]. \tag{36.5}$$

Заметим, что уравнения (36.4) и (36.5) лишены смысла для прямой на плоскости; по своей структуре они близки к уравнениям (32.14) и (32.15) плоскости в пространстве.

5. Каждая прямая может быть представлена как пересечение двух плоскостей. Практически уравнения (36.3) задают прямую именно таким образом, так как они эквивалентны системе из двух линейных уравнений, каждое из которых определяет плоскость. Дадим общую формулировку этого факта. В аффинной системе координат $Oxyz$ прямая l, являющаяся линией пересечения плоскостей

$$\pi_i : A_i x + B_i y + C_i z + D_i = 0, \qquad i = 1, 2,$$

определяется системой уравнений

$$\begin{cases} A_1 x + B_1 y + C_1 z + D_1 = 0, \\ A_2 x + B_2 y + C_2 z + D_2 = 0, \end{cases} \tag{36.6}$$

где

$$\operatorname{rg} \begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{bmatrix} = 2. \tag{36.7}$$

Условие (36.7) означает, что плоскости π_1 и π_2 пересекаются (§33, замечание). Систему (36.6) называют ***общими уравнениями*** прямой в пространстве.

Важно уметь переходить от одного типа уравнения прямой к другому. Переход от каждого из уравнений (36.1)–(36.5) к любому другому очевиден. Для перехода от (36.6) к (36.1)–(36.5) необходимо найти точку M_0 и направляющий вектор $\mathbf{a}$. Координаты точки M_0 можно найти как частное решение системы (36.6). Найдем вектор $\mathbf{a}$.

Теорема 36.1. *Если в аффинной системе координат $Oxyz$ прямая l задана общими уравнениями* (36.6), *то вектор*

$$\mathbf{a} = \left\{ \begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix}, -\begin{vmatrix} A_1 & C_1 \\ A_2 & C_2 \end{vmatrix}, \begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} \right\} \tag{36.8}$$

является направляющим вектором этой прямой.

Доказательство. Отметим прежде всего, что $\mathbf{a} \neq \mathbf{0}$ в силу условия (36.7). Далее, вектор $\mathbf{a}$ параллелен плоскостям π_1 и π_2, так как разложения определителей

$$\begin{vmatrix} A_1 & B_1 & C_1 \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{vmatrix} = 0 \quad \text{и} \quad \begin{vmatrix} A_2 & B_2 & C_2 \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{vmatrix} = 0$$

по первой строке совпадают с условиями (32.13) параллельности вектора $\mathbf{a}$ плоскостям π_1 и π_2. Итак, ненулевой вектор $\mathbf{a}$ параллелен каждой из плоскостей π_1 и π_2. Следовательно, он является направляющим вектором линии их пересечения. ∎

Замечание 1. Для запоминания координат вектора $\mathbf{a}$ может быть использован мнемонический определитель

$$\mathbf{a} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{vmatrix},$$

где $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ – базис, соответствующий системе координат $Oxyz$. Разложение этого определителя по первой строке совпадает с разложением вектора $\mathbf{a}$ по базису $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$.

Замечание 2. Теорема 36.1 относится к аффинной системе координат. Очевидно, что в прямоугольной декартовой системе координат, соответствующей ортонормированному базису $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, вектор (36.8) может быть получен как векторное произведение $[\mathbf{n}_1, \mathbf{n}_2]$ нормалей $\mathbf{n}_1$ и $\mathbf{n}_2$ к плоскостям π_1 и π_2.

Взаимное расположение двух прямых в пространстве. Пусть каждая из прямых l_i, $i = 1, 2$, задана точкой $M_i(x_i, y_i, z_i)$ и направляющим вектором $\mathbf{a}_i = \{m_i, n_i, k_i\}$ в некоторой аффинной системе координат $Oxyz$. Положим $\mathbf{b} = \overrightarrow{M_1M_2}$.

Теорема 36.2. *Прямые l_1 и l_2 лежат в одной плоскости тогда и только тогда, когда*

$$\begin{vmatrix} x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ m_1 & n_1 & k_1 \\ m_2 & n_2 & k_2 \end{vmatrix} = 0. \tag{36.9}$$

Доказательство. Действительно, прямые l_1 и l_2 лежат в одной плоскости тогда и только тогда, когда векторы $\mathbf{b}$, $\mathbf{a}_1$, $\mathbf{a}_2$ компланарны, т.е. линейно зависимы (теорема 15.2). Согласно теореме 14.2 это равносильно тому, что один из них линейно выражается через другие. В силу линейности координат то же относится и к их координатам, что означает (§16, следствие 1) равенство нулю определителя, строками которого являются эти координаты. ■

Замечание 3. В прямоугольной декартовой системе координат условие (36.9) может быть получено из критерия компланарности векторов $\mathbf{b}$, $\mathbf{a}_1$, $\mathbf{a}_2$ (теорема 23.4 и соотношение (23.6)).

Теорема 36.3. *Прямые l_1 и l_2 совпадают тогда и только тогда, когда*

$$\operatorname{rg}\begin{bmatrix} x_2-x_1 & y_2-y_1 & z_2-z_1 \\ m_1 & n_1 & k_1 \\ m_2 & n_2 & k_2 \end{bmatrix} = 1, \qquad (36.10)$$

параллельны и не совпадают тогда и только тогда, когда

$$\operatorname{rg}\begin{bmatrix} m_1 & n_1 & k_1 \\ m_2 & n_2 & k_2 \end{bmatrix} = 1, \quad \operatorname{rg}\begin{bmatrix} x_2-x_1 & y_2-y_1 & z_2-z_1 \\ m_1 & n_1 & k_1 \\ m_2 & n_2 & k_2 \end{bmatrix} = 2, \qquad (36.11)$$

пересекаются тогда и только тогда, когда

$$\operatorname{rg}\begin{bmatrix} m_1 & n_1 & k_1 \\ m_2 & n_2 & k_2 \end{bmatrix} = \operatorname{rg}\begin{bmatrix} x_2-x_1 & y_2-y_1 & z_2-z_1 \\ m_1 & n_1 & k_1 \\ m_2 & n_2 & k_2 \end{bmatrix} = 2. \qquad (36.12)$$

Доказательство. Условие (36.10) равносильно тому, что векторы $\mathbf{b}$, $\mathbf{a}_1$, $\mathbf{a}_2$ коллинеарны; условие (36.11) – тому, что векторы $\mathbf{a}_1$, $\mathbf{a}_2$ коллинеарны, а векторы $\mathbf{b}$, $\mathbf{a}_1$, $\mathbf{a}_2$ компланарны, но не коллинеарны; условие (36.12) – тому, что векторы $\mathbf{a}_1$, $\mathbf{a}_2$ коллинеарны, а векторы $\mathbf{b}$, $\mathbf{a}_1$, $\mathbf{a}_2$ компланарны. Отсюда следует утверждение теоремы. ■

Взаимное расположение прямой и плоскости. Пусть в пространстве в некоторой аффинной системе координат $Oxyz$ заданы плоскость

$$\pi:\ Ax + By + Cz + D = 0$$

и прямая

$$l:\ \mathbf{r} = \mathbf{r}_0 + t\,\mathbf{a},$$

где $\mathbf{r}_0 = \{x_0, y_0, z_0\}$, $\mathbf{a} = \{m, n, k\}$.

Теорема 36.4. *Прямая l лежит в плоскости π тогда и только тогда, когда*

$$\begin{cases} Am + Bn + Ck = 0, \\ Ax_0 + By_0 + Cz_0 + D = 0; \end{cases}$$

прямая l параллельна плоскости π, но не лежит в ней тогда и только тогда, когда

$$\begin{cases} Am + Bn + Ck = 0, \\ Ax_0 + By_0 + Cz_0 + D \neq 0; \end{cases}$$

прямая l пересекает плоскость π тогда и только тогда, когда

$$Am + Bn + Ck \neq 0.$$

Утверждение теоремы следует из критерия (32.13) параллельности вектора $\mathbf{a}$ и плоскости π.

Прямая в пространстве в прямоугольной декартовой системе координат. Пусть $Oxyz$ – прямоугольная декартова система координат пространства. 1. Углом между двумя прямыми в пространстве называется любой из углов между параллельными им прямыми, проходящими через какую-либо точку пространства. Таким образом, две прямые в пространстве образуют между собой два различных (если они не перпендикулярны) угла, в сумме равные π. Очевидно, что угол между направляющими векторами прямых равен одному из этих углов. Следовательно, угол φ между прямыми l_i: $\mathbf{r} = \mathbf{r}_i + t\,\mathbf{a}_i$, $i = 1, 2$, совпадающий с углом между их направляющими векторами $\mathbf{a}_i = \{m_i, n_i, k_i\}$, вычисляется согласно (22.6) по формуле

$$\cos\varphi = \frac{m_1m_2 + n_1n_2 + k_1k_2}{\sqrt{m_1^2+n_1^2+k_1^2}\sqrt{m_2^2+n_2^2+k_2^2}}.$$

2. Углом между прямой и плоскостью (если они не перпендикулярны) называется меньший из углов между этой прямой и ее ортогональной проекцией на плоскость. Если же прямая и плоскости перпендикулярны, то угол между ними считается равным $\pi/2$. Угол φ между прямой l: $\mathbf{r} = \mathbf{r}_0 + t\,\mathbf{a}$ и плоскостью π: $Ax + By + Cz + D = 0$ находится как дополнительный к углу между направляющим вектором прямой $\mathbf{a} = \{m, n, k\}$ и вектором нормали к плоскости $\mathbf{n} = \{A, B, C\}$ и вычисляется согласно (22.6) по формуле

$$\sin\varphi = \frac{|Am + Bn + Ck|}{\sqrt{m^2+n^2+k^2}\sqrt{A^2+B^2+C^2}}, \qquad 0 \le \varphi \le \pi/2.$$

3. Расстояние $\rho(M_1, l)$ от точки $M_1(\mathbf{r}_1)$ до прямой $l:\ \mathbf{r} = \mathbf{r}_0 + t\,\mathbf{a}$ находится как высота h (рис. 1) параллелограмма, построенного на векторах $\mathbf{a}$ и $\overrightarrow{M_1M_2}$, площадь и основание которого известны (§23):

$$\rho(M_1, l) = \frac{|[\mathbf{a}, \mathbf{r}_1 - \mathbf{r}_0]|}{|\mathbf{a}|}.$$

Соотношения (23.5), (22.6) позволяют вычислить $\rho(M_1, l)$ по прямоугольным координатам.

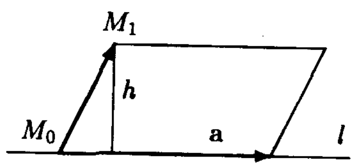


Рис. 1

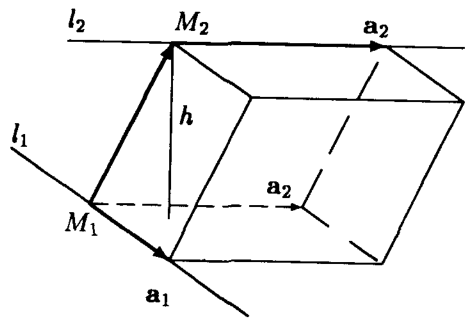


Рис. 2

4. ***Расстоянием между скрещивающимися прямыми*** l_i: $\mathbf{r} = \mathbf{r}_i + t\,\mathbf{a}_i$, $i = 1, 2$, называется расстояние между параллельными плоскостями, в которых лежат прямые l_1 и l_2. Это расстояние $\rho(l_1, l_2)$ находится как высота параллелепипеда (рис. 2), построенного на векторах $\overrightarrow{M_1M_2}$, $\mathbf{a}_1$, $\mathbf{a}_2$, объем и площадь основания которого известны (§23):

$$\rho(l_1, l_2) = \frac{|(\mathbf{r}_2 - \mathbf{r}_1, \mathbf{a}_1, \mathbf{a}_2)|}{|[\mathbf{a}_1, \mathbf{a}_2]|}.$$

Соотношения (23.6), (22.6) позволяют вычислить $\rho(l_1, l_2)$ по прямоугольным координатам.

Глава VIII. Элементы общей алгебры

В этой главе изучаются фундаментальные для всей алгебры понятия группы, кольца и поля. Они дают возможность с общих позиций взглянуть на рассмотренные в предыдущих главах понятия и факты и ввести, кроме того, новые.

§37. Группа

Множество G с заданной на нем алгебраической операцией $*$ называется *группой*, если:

1) операция ассоциативна: $(a * b) * c = a * (b * c)$, $\forall a, b, c \in G$;

2) операция обладает нейтральным элементом $e \in G$: $a * e = e * a = a$, $\forall a \in G$;

3) для любого элемента $a \in G$ существует симметричный элемент $a' \in G : a * a' = a' * a = e$.

Обозначение: G или $\langle G, * \rangle$. Условия 1–3 называются ***аксиомами группы***. Группа с коммутативной операцией называется ***коммутативной*** или ***абелевой***.

О терминологии. Обычно групповую операцию $*$ обозначают специальными символами, чаще всего символами $\cdot$ или $+$, называя $a \cdot b$ (или просто ab) произведением, а $a + b$ – суммой элементов $a, b \in G$. В первом случае нейтральный элемент называют *единицей* (обозначение: 1), симметричный элемент – *обратным* (обозначение: a^{-1}), а саму группу G – *мультипликативной*. Во втором случае нейтральный элемент называют *нулем* (обозначение: 0), симметричный – *противоположным* (обозначение: $-a$), а группу G – *аддитивной*. Если относительно групповой операции нет никаких оговорок, то обычно используют терминологию мультипликативной группы.

Как следует из определения, группа – это абстракция множества, наделенного одной алгебраической операцией. Большинство изученных нами множеств являются конкретными проявлениями этого понятия, т.е. представляют собой группы. Действительно:

а) $\langle \mathbb{Z}, + \rangle$; $\langle \mathbb{Q}, + \rangle$; $\langle \mathbb{R}, + \rangle$; $\langle V_n, + \rangle$, где $n = 1, 2, 3$; $\langle \mathbb{R}^{m \times n}, + \rangle$; любое линейное пространство V – это аддитивные абелевы группы;

б) $\langle \mathbb{Q} \backslash 0, \cdot \rangle$; $\langle \mathbb{R} \backslash 0, \cdot \rangle$ – мультипликативные абелевы группы;

в) $\langle \{A \in \mathbb{R}^{n \times n} | \det A \neq 0\}, \cdot \rangle$; $\langle \{A \in \mathbb{R}^{n \times n} | \det A = 1\}, \cdot \rangle$ – мультипликативные неабелевы группы;

г) множество $S(X)$ всех биективных отображений $f : X \to X$ – мультипликативная неабелева группа; в частности, если X – конечное множество из n элементов (т.е. $S(X)$ – множество перестановок n-го порядка (§8)), то эта группа называется ***симметрической группой n-го порядка*** и обозначается S_n.

Из аксиом группы и свойств алгебраической операции (§9) вытекают следующие факты.

1°. *В любой группе существует, и притом единственный, нейтральный элемент.*

2°. *В любой группе для каждого элемента существует, и притом единственный, симметричный элемент.*

Аксиомы группы можно несколько ослабить. Убедимся в этом. Пусть G – множество, наделенное алгебраической операцией. Элемент $e_n \in G$, такой, что $ae_n = a$ для любого $a \in G$, назовем *правой единицей* множества G, а элемент $a_n^{-1} \in G$, такой, что $aa_n^{-1} = e_n$, – *правым обратным к элементу* a.

Теорема 37.1. *Множество G с ассоциативной алгебраической операцией является группой, если оно обладает правой единицей e_n и по отношению к ней каждый элемент $a \in G$ обладает правым обратным.*

Доказательство. Пусть a – произвольный элемент множества G и a_n^{-1} – правый обратный к нему элемент. Тогда $aa_n^{-1} = e_n = e_n e_n = e_n(aa_n^{-1}) = (e_n a)a_n^{-1}$, т.е. $aa_n^{-1} = (e_n a)a_n^{-1}$. Умножив обе части последнего равенства справа на правый обратный элемент к элементу a_n^{-1}, получим, что $a = e_n a$. Таким образом, правая единица множества G оказывается и левой единицей, т.е. просто единицей e множества G.

Покажем теперь, что правый обратный элемент к элементу a является и левым обратным, т.е. $a_n^{-1}a = e$. Действительно, $ea_n^{-1} = a_n^{-1}e = a_n^{-1}(aa_n^{-1}) = (a_n^{-1}a)a_n^{-1}$, т.е. $ea_n^{-1} = (a_n^{-1}a)a_n^{-1}$. Умножив обе части последнего равенства справа на правый обратный элемент к элементу a_n^{-1}, получим $e = a_n^{-1}a$. ■

Следствие. *В группе любая правая единица является левой и той единственной единицей, которой эта группа обладает, а любой правый обратный элемент к элементу a группы является левым и тем единственным обратным, которым обладает элемент a.*

Теорема 37.2. *Множество G с ассоциативной алгебраической операцией является группой тогда и только тогда, когда эта операция обладает обратной.*

Доказательство. Необходимость. Пусть G – группа; покажем, что уравнения

$$ax = b, \tag{37.1}$$

$$ya = b \tag{37.2}$$

имеют единственное решение для любых $a, b \in G$. Действительно, элемент $x = a^{-1}b$ является решением уравнения (37.1), так как $ax = a(a^{-1}b) = (aa^{-1})b = eb = b$. Это решение единственно, так как если x' – любое другое решение (37.1), то $ax' = ax$ и $a^{-1}(ax') = a^{-1}(ax)$. Следовательно, $x' = x$. Аналогично доказывается утверждение, касающееся уравнения (37.2).

Достаточность. Пусть теперь уравнения (37.1), (37.2) при любых $a, b \in G$ имеют единственное решение. Покажем, что G – группа. Для этого покажем, что выполнены все условия теоремы 37.1.

Действительно, возьмем произвольный элемент $a \in G$. Уравнение $ax = a$ имеет единственное решение. Очевидно, оно играет роль правой единицы для элемента a. Обозначим его e_n^a. Для любого другого элемента $b \in G$ существует элемент $y \in G$ такой, что $b = ya$ (т.е. y – решение (37.2)). Тогда $b = ya = y(ae_n^a) = (ya)e_n^a = be_n^a$. Следовательно, $be_n^a = b$, $\forall b \in G$, и $e_n^a = e_n$ – правая единица множества G. Далее, уравнение $ax = e_n$ имеет решение, которое, очевидно, является правым обратным к элементу a по отношению к e_n. Отсюда в силу теоремы 37.1 следует, что G – группа. ■

Замечание. Попутно доказано еще одно свойство группы: *в группе любая правая (левая) единица для одного элемента является общей правой (соответственно левой) единицей.*

Итак, групповая операция обладает обратной операцией. В аддитивной группе обратная операция называется *вычитанием (справа и слева)*, а элементы $x = b + (-a)$ и $y = (-a) + b$ – *разностью* (*правосторонней и левосторонней* соответственно). В абелевой группе обе разности совпадают и обозначаются единым символом $b - a$.

§38. Подгруппа

Определение. Подмножество H группы G называется *подгруппой* группы G, если оно само является группой относительно алгебраической операции в G.

Примеры. 1. Простейшими подгруппами любой группы являются ее единичный элемент и сама группа. Эти подгруппы называют *тривиальными*.

2. Множество четных чисел – подгруппа аддитивной группы целых чисел.

3. Множество чисел, кратных p, где $p \in \mathbb{N}$, $p > 1$, – подгруппа аддитивной группы целых чисел.

4. Множество рациональных чисел, отличных от нуля, – подгруппа мультипликативной группы ненулевых действительных чисел.

5. Множество квадратных матриц n-го порядка, определители которых равны единице, – подгруппа мультипликативной группы невырожденных матриц n-го порядка.

Теорема 38.1. *Подмножество H группы G является подгруппой этой группы тогда и только тогда, когда имеют место следующие импликации:*

1) $a, b \in H \;\Rightarrow ab \in H$;
2) $a \in H \;\Rightarrow a^{-1} \in H$.

Доказательство. Необходимость очевидна, она вытекает из того, что H – группа.

Достаточность. Из первого условия следует, что алгебраическая операция в G является алгебраической операцией и в H. Что же касается аксиом, то необходимо проверить только аксиому единицы. Пусть $a \in H$, тогда, согласно условиям 1 и 2, $a^{-1} \in H$ и $aa^{-1} = e \in H$. ■

Произведение подмножеств группы. Пусть G – группа, M и N – два ее подмножества. *Произведением* MN *этих подмножеств* называется множество всевозможных произведений mn, где $m \in M$, $n \in N$. Итак, $MN = \{mn | m \in M, n \in N\}$. Очевидно, что имеет место свойство ассоциативности: $(MN)K = M(NK)$, так как оба этих произведения состоят из элементов $(mn)k = m(nk)$, где $m \in M$, $n \in N$, $k \in K$.

Если одно из подмножеств состоит только из одного элемента, например $M = \{m\}$, то произведение MN обозначается символом mN, а произведние NM – символом Nm, т.е. в этом контексте нет необходимости отличать элемент от состоящего из одного этого элемента множества.

Замечание. Термин "произведение", используемый здесь, весьма условен, он соответствует терминологии мультипликативной группы, принятой в теории групп. В аддитивной группе, очевидно, MN состоит из сумм $m + n$ элементов подмножеств M и N и обозначается в зависимости от контекста символом $M + N$.

Смежные классы. Пусть H – подгруппа группы G, a – элемент группы G. Множество aH называется *левым смежным классом* группы G по подгруппе H, *порожденным элементом* a, а множество Ha – *правым смежным классом.* Из определения вытекают следующие простейшие свойства смежных классов.

1°. $a \in aH$, $a \in Ha$, так как подгруппа H содержит единицу.

2°. *Смежный класс состоит из элементов группы, причем любой элемент группы входит в какой-нибудь смежный класс* (в силу свойства 1°).

3°. *Подгруппа H является одним из смежных классов* (как левых, так и правых), поскольку $H = eH = He$.

4°. *В абелевой группе* $aH = Ha$, $\forall a \in G$.

Таблица 1 содержит примеры смежных классов.

Теорема 38.2. *Смежный класс порождается любым своим элементом.*

Доказательство. Надо показать, что если $g \in aH$, то $aH = gH$. Пусть $g = ah_1$, $h_1 \in H$. Тогда для любого элемента $h \in H$ имеем $ah = (ah_1)(h_1^{-1}h) = gh_2$, где $h_2 = h_1^{-1}h \in H$ в силу теоремы 38.1. Значит, $aH \subset gH$. В то же время $gh = a(h_1h) = ah_2$, где $h_2 = h_1h_2 \in H$. Следовательно, $gH \subset aH$, и с учетом вложения в другую сторону получаем требуемое равенство. Доказательство проведено для левого смежного класса, но очевидно, что оно может быть использовано и для правого. ■

Теорема 38.3. *Любые два левых (правых) смежных класса либо совпадают, либо не пересекаются.*

Доказательство. Утверждение теоремы вытекает из теоремы 38.2, так как если два смежных класса aH и bH имеют общий элемент g, то $aH = bH = gH$. ■

Таблица 1

G	H	aH
$\langle \mathbb{Z}, + \rangle$	$H = \{2k \mid k \in \mathbb{Z}\}$	$mH = Hm = \{m + 2k \mid k \in \mathbb{Z}\} = \{n \in \mathbb{Z} \mid n \equiv m(\mathrm{mod}\,2)\}$. Итак, аддитивная группа целых чисел разбивается на два смежных класса по подгруппе четных чисел: это классы всех четных и всех нечетных чисел
$\langle \mathbb{Z}, + \rangle$	Пусть $p \in \mathbb{N}$, $p > 1$. $H = \{pk \mid k \in \mathbb{Z}\}$	$mH = Hm = \{n \in \mathbb{Z} \mid n \equiv m(\mathrm{mod}\,p)\}$ (см. (7.1)). Итак, аддитивная группа целых чисел разбивается на p смежных классов по подгруппе H чисел, кратных p: C_0, C_1, ... , C_{p-1}, где C_r – класс вычетов по модулю p
$\langle \mathbb{R}^n, + \rangle$	Пусть $A \in \mathbb{R}^{m \times n}$. $H = \{x \in \mathbb{R}^n \mid Ax = 0\}$	Положим $Aa = b$. Тогда $aH = Ha = \{a + x \mid Ax = 0\} = \{x \in \mathbb{R}^n \mid Ax = b\}$. Итак, смежный класс аддитивной группы $\mathbb{R}^n$ по подгруппе решений однородной системы $Ax = 0$, порожденный элементом a, представляет собой множество всех решений неоднородной системы $Ax = b$, частным решением которой является a
$\langle V, + \rangle$, где V – линейное пространство	Пусть L – линейное подпространство V, $H = L$	Пусть $x_0 \in V$. Тогда $x_0H = Hx_0 = \{x_0 + x \mid x \in L\}$ – линейное многообразие $x_0 + H$

Итак, вся группа разбивается на непересекающиеся левые (правые) смежные классы по подгруппе H. Это разбиение называется *левосторонним (соответственно правосторонним) разложением группы G по подгруппе H.*

§39. Конечная группа

Группа, состоящая из конечного числа элементов, называется *конечной группой.* Число элементов конечной группы называется ее *порядком* и обозначается символом $\mathrm{card}\,G$. Очевидно, алгебраическая операция в конечной группе может быть представлена "таблицей умножения" ее элементов. Этим мы воспользуемся в следующих примерах.

1. Группа из одного элемента состоит только из единицы: $G = \{e\}$, при этом алгебраическая операция определяется таблицей

	e
e	e

2. Группа из двух элементов $G = \{e, a\}$ определяется таблицей

	e	a
e	e	a
a	a	e

Здесь $aa \neq a$, так как иначе элемент a будет единицей для a, т.е. (§37, замечание) общей и единственной единицей e. Поэтому $aa = e$.

3. Группа из трех элементов $G = \{e, a, b\}$ определяется таблицей

	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

Здесь, рассуждая так же, как в примере 2, получаем, что $ab \neq a$, $ab \neq b$, $ba \neq a$, $ba \neq b$. Следовательно, $ab = e = ba$. Аналогично $aa \neq a$, $aa \neq e$ (так как единственным обратным элементу a, как выяснено выше, является элемент b). Следовательно, $aa = b$. Аналогично $bb = a$. Заметим, что все три группы абелевы.

Теорема 39.1 (теорема Лагранжа). *Во всякой конечной группе порядок ее подгруппы является делителем порядка самой группы.*

Доказательство. Пусть $\operatorname{card} G = n$, $\operatorname{card} H = k$. Рассмотрим левостороннее разложение группы G по подгруппе H. Оно состоит из всех левых смежных классов aH, где $a \in G$. Каждый класс aH состоит из всех элементов ah, где $h \in H$. Все элементы ah различны (так как если $ah_1 = ah_2$, то $a^{-1}ah_1 = a^{-1}ah_2$ и, значит, $h_1 = h_2$), поэтому каждый класс aH состоит ровно из k элементов. Общее число смежных классов aH равно $m \leq n$. Таким образом, количество всех элементов группы G равно mk, т.е. $n = mk$. ■

§40. Нормальный делитель

Подгруппа H группы G называется ***нормальным делителем***, если для любого элемента $a \in H$

$$aH = Ha, \tag{40.1}$$

т.е. если любой левый (правый) смежный класс одновременно является правым (левым) смежным классом.

Очевидно, что в абелевой группе любая подгруппа является нормальным делителем. Тривиальные подгруппы в любой группе также являются нормальными делителями: оба разложения по единичной подгруппе совпадают с разложением группы на отдельные элементы, оба разложения по самой группе состоят из одного класса – это сама

группа. Укажем критерий нормального делителя для произвольной группы.

Элементы a и b группы G называются ***сопряженными***, если существует элемент $c \in G$ такой, что $a = c^{-1}bc$.

Теорема 40.1. *Подгруппа H группы G является нормальным делителем тогда и только тогда, когда она вместе с каждым элементом содержит все сопряженные с ним элементы.*

Доказательство. Необходимость. Пусть H – нормальный делитель, тогда согласно (40.1) $Hc \subset cH$ для любого элемента $c \in G$. Это означает, что для любого элемента $h \in H$ существует $h_1 \in H$ такой, что $hc = ch_1$, т.е. $c^{-1}hc = h_1 \in H$.

Достаточность. Возьмем произвольные элементы $h \in H$, $c \in G$. Тогда $c^{-1}hc \in H$ и $chc^{-1} \in H$. Это означает, что существуют элементы $h_1, h_2 \in H$ такие, что $c^{-1}hc = h_1$, $chc^{-1} = h_2$, или $hc = ch_1$, $ch = h_2c$. Значит, $Hc \subset cH$ и $cH \subset Hc$, т.е. $cH = Hc$. ■

Примером нормального делителя в неабелевой группе является подгруппа матриц с определителем, равным единице, мультипликативной группы невырожденных матриц n-го порядка, так как для элементов этой подгруппы выполнено условие теоремы 40.1: если $|A| = 1$, то $|C^{-1}AC| = |A| = 1$ для любой невырожденной матрицы C.

Теорема 40.2. *Смежные классы по нормальному делителю образуют группу относительно умножения подмножеств группы.*

Доказательство. Покажем, что произведение смежных классов по нормальному делителю является смежным классом. Действительно, $(aH)(bH) = \{$ в силу ассоциативности произведения $\} = a(Hb)H = \{$ так как H – нормальный делитель $\} = a(bH)H = ab(HH) = abH$. Итак,

$$(aH)(bH) = (ab)H, \tag{40.2}$$

значит, умножение подмножеств группы является алгебраической операцией на множестве смежных классов по нормальному делителю. Ассоциативность этого умножения уже известна (§38). Нейтральным элементом служит сама подгруппа H, так как $(aH)H = (aH)(eH) = (ae)H = aH$, $\forall a \in G$. Обратным элементом к классу aH является класс $a^{-1}H$, так как $(aH)(a^{-1}H) = (aa^{-1})H = eH = H$. ■

Группа смежных классов группы G по нормальному делителю H называется ***фактор-группой группы G по подгруппе H***.

Обозначение: $G|H$.

Группа вычетов по модулю p. Вернемся к смежным классам аддитивной группы целых чисел по подгруппе H чисел, кратных p. Как следует из §38 (пример 2 в таблице), фактор-группа $\mathbb{Z}|H$ состоит ровно из p классов и совпадает с множеством $\mathbb{Z}_p$ классов вычетов по модулю p (§7, пример 3). Алгебраическую операцию в $\mathbb{Z}_p$ (т.е. произведение смежных классов) будем называть сложением, придерживаясь названия исходной операции в группе целых чисел. Согласно (40.2) суммой смежных классов $C_m + C_n$ является класс, который содержит сумму чисел $m + n$:

$$C_m + C_n = C_r, \quad \text{где } r \equiv (m+n)(\operatorname{mod} p). \tag{40.3}$$

Итак, $\mathbb{Z}_p$ – аддитивная группа вычетов по модулю p.

§41. Морфизмы групп

Изоморфизм. Уже примеры конечных групп (§39) приводят к естественному выводу, что все группы (например, третьего порядка) обнаруживают большое сходство: алгебраические операции в них описываются одной и той же таблицей умножения. Общий подход к различению или, напротив, к отождествлению групп предлагает понятие изоморфизма.

Две группы G_1 и G_2 с операциями $*_1$ и $*_2$ называют ***изоморфными***, если существует биективное отображение $f : G_1 \to G_2$, которое сохраняет групповую операцию, т.е. $f(a *_1 b) = f(a) *_2 f(b)$, $\forall a, b \in G_1$. Обозначение: $G_1 \cong G_2$. Само отображение f при этом называют ***изоморфизмом***.

Примеры. 1. Аддитивные группы целых чисел и четных чисел изоморфны, так как отображение $f(n) = 2n$, $n \in \mathbb{Z}$, есть изоморфизм.

2. Мультипликативная группа положительных вещественных чисел и аддитивная группа всех вещественных чисел изоморфны. В качестве изоморфного отображения f можно взять $f(a) = \ln a$. Известное свойство логарифма $\ln ab = \ln a + \ln b$ как раз моделирует условие изоморфизма.

Отметим простейшие свойства изоморфизма.

1°. ***Отношение изоморфизма является отношением эквивалентности на множестве всех групп.***

Действительно, рефлексивность следует из того, что тождественное отображение является изоморфизмом; симметричность – из того, что если f – изоморфизм, то f^{-1} (которое существует в силу биективности f) тоже является изоморфизмом, так как $f^{-1}(a' *_2 b') = \{a' = f(a), b' = f(b)$ в силу сюръективности $f\} = f^{-1}(f(a) *_2 f(b)) = \{$ибо f – изоморфизм $\} = f^{-1}f(a *_1 b) = a *_1 b = f^{-1}(a') *_1 f^{-1}(b')$; транзитивность – из того, что суперпозиция изоморфизмов, как легко проверить, также является изоморфизмом.

2°. ***В изоморфных группах G_1 и G_2 образ (и прообраз) единицы является единицей.***

Действительно, если $a *_1 e = e *_1 a = a$, то $f(a) *_2 f(e) = f(e) *_2 f(a) = f(a)$. Отсюда с учетом того, что элементами $f(a)$ исчерпывается вся группа G_2 (в силу сюръективности f), следует, что $f(e) = e'$ – единица в G_2. Следовательно, образом единицы e является единица e'. Значит, и прообразом единицы e' является e и, в силу инъективности f, только e.

3°. ***В изоморфных группах G_1 и G_2 образ (и прообраз) обратного элемента является обратным элементом, т.е. $f(a^{-1}) = (f(a))^{-1}$.***

Доказательство этого факта аналогично доказательству 2°, оно опирается на определение обратного элемента.

Изоморфное отображение группы G на себя называется *автоморфизмом*. Несложная проверка показывает, что множество всех автоморфизмов группы образует группу относительно суперпозиции отображений.

Гомоморфизм. Две группы G_1 и G_2 с операциями $*_1$ и $*_2$ называют *гомоморфными*, если существует отображение $\varphi : G_1 \to G_2$, сохраняющее групповую операцию. Само отображение φ называется *гомоморфизмом*.

Отметим простейшие свойства гомоморфизма. В гомоморфных группах

1) *образом единицы является единица, т.е.* $\varphi(e) = e'$, где e и e' – единицы групп G_1 и G_2;

2) *образом обратного элемента является обратный элемент к образу:* $\varphi(a^{-1}) = (\varphi(a))^{-1}$.

Доказательства этих свойств повторяют доказательства аналогичных свойств изоморфизма. Отметим лишь, что к гомоморфизму относятся только первые части свойств изоморфизма.

Пусть φ – гомоморфизм группы G_1 на группу G_2. Множество $\ker\varphi = \{a \in G_1 | \varphi(a) = e'\}$ называется *ядром гомоморфизма* φ.

Теорема 41.1. *Нормальные делители группы, и только они, являются ядрами гомоморфизмов этой группы.*

Доказательство. Достаточность. Ядро гомоморфизма φ группы G_1 является подгруппой группы G_1, так как для него выполнены оба условия теоремы 38.1:

1) если $h_1, h_2 \in \ker\varphi$, то $\varphi(h_1 *_1 h_2) = \varphi(h_1) *_2 \varphi(h_2) = e' *_2 e' = e'$, т.е. $h_1 *_1 h_2 \in \ker\varphi$;

2) если $h \in \ker\varphi$, то $\varphi(h) = e'$ и, значит, $\varphi(h^{-1}) = (e')^{-1} = e'$, т.е. $h^{-1} \in \ker\varphi$.

Эта подгруппа является нормальным делителем, так как для нее выполнено условие теоремы 40.1: если $h \in \ker\varphi$ и $c \in G$, то $\varphi(c^{-1} *_1 h *_1 c) = \varphi(c^{-1}) *_2 \varphi(h) *_2 \varphi(c) = \varphi(c^{-1}) *_2 \varphi(c) = \varphi(c^{-1} *_1 c) = \varphi(e) = e'$, т.е. $c^{-1} *_1 h *_1 c \in \ker\varphi$.

Необходимость. Пусть H – подгруппа группы G с операцией $\cdot$, покажем, что существует некоторая группа G' и некоторый гомоморфизм $\varphi : G \to G'$, ядром которого является подгруппа H. В качестве группы G' возьмем фактор-группу $G|H$ группы G по нормальному делителю H. В качестве гомоморфизма возьмем отображение $\varphi : G \to G|H$, которое каждому элементу $a \in G$ ставит в соответствие смежный класс aH, так что $\varphi(a) = aH$. Отображение φ является гомоморфизмом, так как $\varphi(ab) = (ab)H = \{$ согласно (40.2) $\} = (aH)(bH) = \varphi(a)\varphi(b)$, $\forall a, b \in G$. ■

Построенное отображение $\varphi : G \to G|H$ называется *естественным гомоморфизмом*.

§42. Кольцо

Определение, простейшие свойства. Перейдем к рассмотрению множеств, наделенных двумя алгебраическими операциями. Будем называть одну из этих операций сложением (и обозначать символом $+$), а другую – умножением (и обозначать символом $\cdot$).

Непустое множество K, наделенное двумя алгебраическими операциями – сложением и умножением, называется ***кольцом***, если эти операции удовлетворяют следующим аксиомам: $\forall a, b, c \in K$

1) $a + b = b + a$;
2) $(a + b) + c = a + (b + c)$;
3) $\exists 0 \in K : a + 0 = 0 + a$;
4) $\forall a \in K \;\; \exists -a \in K : a + (-a) = (-a) + a = 0$;
5) $(ab)c = a(bc)$;

6) $(a+b)c = ab+bc$, $a(b+c) = ab+ac$.

Кольцо называется ***коммутативным***, если умножение в нем коммутативно, кольцо называется ***кольцом с единицей***, если операция умножения обладает нейтральным элементом. Из аксиом кольца следует, что кольцо является аддитивной абелевой группой.

Очевидно, множества $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ образуют коммутативные кольца с единицей относительно обычных операций сложения и умножения чисел; множество всех четных чисел – коммутативное кольцо без единицы; множество $\mathbb{R}^{n\times n}$ вещественных квадратных матриц n-го порядка – некоммутативное кольцо с единицей относительно известных операций сложения и умножения матриц.

Отметим простейшие свойства кольца, все они опираются на свойства множеств с одной ассоциативной операцией, и в частности на свойства группы.

1°. *Кольцо обладает всеми свойствами аддитивной абелевой группы;* в частности, в кольце:

а) существует, и притом единственный, нулевой элемент 0;

б) для каждого элемента a существует, и притом единственный, противоположный элемент $-a$;

в) для любых элементов $a, b \in K$ существует, и притом единственное, решение уравнения $x + a = b$; это решение называется *разностью* элементов b и a и обозначается символом $b - a$. Итак, $b - a = b + (-a)$.

2°. *В кольце умножение дистрибутивно относительно вычитания*, т.е. $a(b-c) = ab - ac$, $(a-b)c = ac - bc$, $\forall a, b, c \in K$.

Это следует из того, что $b = (b-c) + c$ и $ab = a(b-c) + ac$.

3°. *В кольце для любого элемента a*: $a0 = 0a = 0$.

Это следует из дистрибутивности умножения относительно вычитания: $a0 = a(b-b) = ab - ab = 0$.

4°. *В кольце для любых элементов a, b*: $(-a)b = a(-b) = -ab$.

Это следует из того, что $ab + (-a)b = (a + (-a))b = 0b = 0$.

Следствие. $(-a)(-b) = ab$, $\forall a, b \in K$.

5°. *В кольце с единицей для любого элемента a*:

$$(-1)a = a(-1) = -a.$$

Это следует из того, что $a + (-1)a = 1a + (-1)a = (1 + (-1))a = 0a = 0$.

6°. *В кольце с единицей, содержащем не менее двух элементов*:

$$1 \neq 0.$$

Действительно, если $0 = 1$, то существует $a \in K$: $a \neq 0$, $a \neq 1$. Тогда из равенства $0 = 1$ следует, что $0a = 1a$, т.е. $0 = a$, но $a \neq 0$.

7°. *В кольце с единицей множество обратимых (по умножению) элементов образует мультипликативную группу.*

Это следует из того, что произведение обратимых элементов обратимо, т.е. умножение в кольце является алгебраической операцией на этом множестве.

Делители нуля. В кольце, как мы отмечали выше, алгебраические операции сложения, вычитания и умножения обладают свойствами привычных нам операций над числами. Однако кольцо обладает и специфическим свойством, которого нет в числовых множествах. Так, для чисел из равенства $ab = 0$ следует, что одно из чисел a или b равно нулю. В кольце это может и не выполняться. Например, в кольце матриц второго порядка существуют ненулевые матрицы, произведение которых равно нулю:

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

Ненулевые элементы a и b кольца называются ***делителями нуля***, если $ab = 0$. При этом элемент a называется ***левым*** делителем нуля, а элемент b – ***правым***.

Кольцо вычетов. Рассмотрим аддитивную группу $\mathbb{Z}_p = \{C_0, C_1, \ldots, C_{p-1}\}$ вычетов по модулю p (§40). Напомним, что сложение на $\mathbb{Z}_p$ определено правилом (40.3), т.е. $C_m + C_n$ – это класс, в который входит число $m+n$. Известно, что $\mathbb{Z}_p$ – группа относительно так введенной операции сложения (§40). Отметим, что это абелева группа, так как числа $m+n$ и $n+m$ входят в один и тот же класс. Отметим также, что нулем этой группы является класс C_0 чисел, кратных p, а противоположным к классу C_m – класс

$$-C_m = \begin{cases} C_0, & m = 0, \\ C_{p-m}, & m > 0. \end{cases}$$

Определим на $\mathbb{Z}_p$ операцию умножения. Положим

$$C_m C_n = C_r, \quad \text{где} \quad r \equiv mn \,(\mathrm{mod}\, p),$$

т.е. $C_m C_n$ – это класс, в который входит число mn. Эта операция обладает следующими свойствами:

1) $C_m C_n = C_n C_m$, так как $C_m C_n$ – класс, в который входит nm, а $mn = nm$;

2) $(C_m C_n)C_k = C_m(C_n C_k)$, так как оба произведения представляют собой один и тот же класс, в который входит число $(mn)k = m(nk)$;

3) *существует единица* C_1, так как $C_m C_1 = C_1 C_m = C_m$;

4) $(C_m + C_n)C_k = C_m C_k + C_n C_k$, так как оба произведения представляют собой класс, в который входит число $(m+n)k = mk + nk$;

5) *если p – составное число, то в кольце $\mathbb{Z}_p$ есть делители нуля*, так как если $p = mn$, где $1 < m < p$, $1 < n < p$, то $C_m, C_n \in \mathbb{Z}_p$ и $C_m C_n = C_0$.

Таким образом, $\mathbb{Z}_p$ – конечное коммутативное кольцо с единицей, которое имеет делители нуля, если p – составное число. Оно называется ***кольцом вычетов по модулю p***.

§43. Поле

Определение, простейшие свойства. *Полем* называется коммутативное кольцо с единицей, содержащее не менее двух элементов, в котором каждый отличный от нуля элемент имеет обратный элемент. Из определения вытекают следующие свойства поля.

1°. *Поле обладает всеми свойствами кольца.*

2°. *В поле нет делителей нуля*, так как если $ab = 0$, $a \neq 0$, то $a^{-1}(ab) = b = 0$.

Следствие. Умножение является алгебраической операцией на множестве всех ненулевых элементов поля.

3°. *В поле P множество всех ненулевых элементов образует мультипликативную коммутативную группу*, и поэтому в поле:

а) *существует, и притом единственная, единица, причем* $1 \neq 0$;

б) *для любого элемента $a \neq 0$ существует, и притом единственный, обратный элемент;*

в) *для любых $a, b \in P$, $a \neq 0$, уравнение $ax = b$ имеет единственное решение, при этом $x = a^{-1}b = ba^{-1}$*; этот элемент называется *частным от деления b на a* и обозначается символом $\dfrac{b}{a}$ или b/a.

4°. *В поле сохраняются все обычные правила обращения с дробями:*

а) $\dfrac{a}{b} = \dfrac{c}{d} \Leftrightarrow ad = bc$, так как равенство $b^{-1}a = cd^{-1}$ равносильно равенству $b(b^{-1}a)d = b(cd^{-1})d$, т.е. $ad = bc$;

б) $\dfrac{a}{b} \pm \dfrac{c}{d} = \dfrac{ad \pm bc}{bd}$, $\quad \dfrac{a}{b}\dfrac{c}{d} = \dfrac{ac}{bd}$;

в) $\dfrac{-a}{b} = \dfrac{a}{-b} = -\dfrac{a}{b}$.

Правила "б","в" доказываются так же, как и "а".

Тем самым все правила и формулы элементарной алгебры полностью сохраняются в любом поле, так как в их основе лежат одни и те же свойства операций сложения, вычитания, умножения и деления.

Элементы поля называют ***числами***. Очевидно, множества $\mathbb{Q}$ всех рациональных чисел и $\mathbb{R}$ всех действительных чисел являются полями.

Характеристика поля. Существуют поля, в которых некоторое целое кратное 1, т.е. $n \cdot 1 = \underbrace{1 + 1 + \ldots + 1}_{n}$, равно нулю. Наименьшее натуральное число n, обладающее этим свойством, называется ***характеристикой поля***. Если указанное свойство не имеет места ни для какого натурального числа n, то говорят, что такое поле имеет характеристику 0. Очевидно, $\mathbb{Q}, \mathbb{R}$ – поля характеристики 0.

Теорема 43.1. *Характеристикой поля может быть либо 0, либо простое число.*

Доказательство. Пусть n – характеристика поля и n – составное число, т.е. $n = mk$, где $m \in \mathbb{N}$, $k \in \mathbb{N}$, $m < n$, $k < n$. Тогда в силу дистрибутивности

$$\underbrace{1 + 1 + \ldots + 1}_{n} = \underbrace{(1 + 1 + \ldots + 1)}_{m}\underbrace{(1 + 1 + \ldots + 1)}_{k} = 0.$$

Так как в поле нет делителей нуля, то отсюда следует, что **либо** $\underbrace{1 + 1 + \ldots + 1}_{m} = 0$, либо $\underbrace{1 + 1 + \ldots + 1}_{k} = 0$, т.е. n – не наименьшее натуральное число, обладающее указанным свойством. ■

Поле вычетов. Кольцо вычетов $\mathbb{Z}_p$ не будет полем, если p составное, так как в этом случае в кольце $\mathbb{Z}_p$ есть делители нуля (§42).

Теорема 43.2. *Если p – простое число и $p \geq 2$, то $\mathbb{Z}_p$ – поле характеристики p.*

Доказательство. Покажем, что $\mathbb{Z}_p$ – поле. Так как $\mathbb{Z}_p$ – коммутативное кольцо с единицей (§42), то достаточно показать, что для любого элемента $C_m \in \mathbb{Z}_p$, $m \neq 0$, существует обратный элемент. Для этого рассмотрим все возможные произведения C_mC_k, где $k = \overline{0, p-1}$. Этих произведений ровно p (т.е. столько же, сколько элементов в $\mathbb{Z}_p$), и все они различны, так как из того, что $C_mC_k = C_mC_n$, $k \neq n$, следует, что $mk \equiv mn (\operatorname{mod} p)$ или что $mk - mn$ делится нацело на p. Следовательно, $m(k-n)$ делится на p, причем $0 < m < p$, $0 < k - n < p$. Это означает, что p – составное число. Таким образом, произведения C_mC_k $(k = \overline{0, p-1})$ по одному разу пробегают все множество $\mathbb{Z}_p$ и, значит, одно из них совпадает с C_1. Итак, для любого класса C_m, где $m \neq 0$, существует класс C_k такой, что $C_mC_k = C_1$. Следовательно, $\mathbb{Z}_p$ – поле. Характеристика этого поля равна p, так как $\underbrace{C_1 + C_1 + \ldots + C_1}_{p} = C_0$, $\underbrace{C_1 + C_1 + \ldots + C_1}_{m} = C_m \neq C_0$ при $m < p$. ■

Замечание. До сих пор при изучении алгебраических объектов мы предполагали, что в их основе лежат вещественные числа: вещественные матрицы, вещественные линейные пространства, вещественные системы линейных алгебраических уравнений, многочлены с вещественными коэффициентами. Все факты, касающиеся этих объектов, остаются справедливыми, если вместо поля вещественных чисел рассматривать любое поле. Это легко проверить, так как доказательства всех фактов опирались лишь на свойства алгебраических операций в поле.

Впредь вместо термина "вещественная матрица" будем употреблять термин "матрица над полем P", так же как и термины "линейное пространство над полем P", "система линейных алгебраических уравнений над полем P", "многочлен над полем P".

Глава IX. Комплексные числа

В этой главе рассматривается новая система чисел, которая является расширением поля действительных чисел. Подобно известной из элементарной алгебры цепочке расширений $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$, каждое звено которой связано с необходимостью решения той или иной задачи, расширение действительных чисел связано с проблемой решения квадратных уравнений. Мы расширим поле действительных чисел так, чтобы любое квадратное уравнение, в частности простейшее из них $x^2+1=0$, имело решение. Прежде чем вводить эти числа, отметим, что при каждом известном расширении "старые" числа становились частью "новых", т.е. отождествлялись с некоторым классом новых чисел. Например, натуральное число 5 отождествлялось с целым числом $+5$, или с рациональным числом $\frac{5}{1}$, или с действительным числом $4,9999\ldots$

§44. Поле комплексных чисел

Понятие комплексного числа. *Комплексными числами* называются упорядоченные пары (a,b) вещественных чисел, для которых понятия равенства, суммы, произведения и отождествления с вещественными числами вводятся согласно следующим правилам (аксиомам):

1) $(a,b)=(c,d) \iff a=c,\ b=d$;
2) $(a,b)+(c,d)=(a+c,\ b+d)$;
3) $(a,b)\cdot(c,d)=(ac-bd,\ ad+bc)$;
4) пара $(a,0)$ отождествляется с действительным числом a.

Обозначения: $z=(a,b)$, $\mathbb{C}$ – множество всех комплексных чисел.

Замечание 1. В первых трех аксиомах речь идет об определении разных понятий, поэтому их сопоставление не может привести к каким-либо противоречиям. В несколько другом положении находится аксиома 4. Дело в том, что понятия равенства, суммы и произведения для вещественных чисел уже имеют определенный смысл и они ***не должны противоречить*** правилам 1 – 3. Но, как нетрудно проверить, правила 1 – 3 для вещественных чисел, как пар специального вида, совпадают с обычными правилами равенства, сложения и умножения вещественных чисел (проверьте!).

Теорема 44.1. *Множество $\mathbb{C}$ всех комплексных чисел является полем.*

Доказательство. Действительно, сложение и умножение являются алгебраическими операциями на множестве $\mathbb{C}$, при этом непосредственной проверкой легко установить, что они подчиняются

всем аксиомам поля. Отметим лишь, что $0=(0,0)$; $1=(1,0)$; $-z=(-a,-b)$; $z^{-1}=\left(\frac{a}{a^2+b^2}, -\frac{b}{a^2+b^2}\right)$, если $z=(a,b)\neq 0$. ■

Следствие 1. *Для любой пары комплексных чисел $z_1=(a,b)$, $z_2=(c,d)$ существует, и притом единственная, разность $z_1-z_2=(a-c,b-d)$.*

Следствие 2. *Для любой пары комплексных чисел $z_1=(a,b)$, $z_2=(c,d)\neq 0$ существует, и притом единственное, частное*

$$\frac{z_1}{z_2}=\left(\frac{ac+bd}{c^2+d^2}, \frac{bc-ad}{c^2+d^2}\right).$$

Алгебраическая форма комплексного числа. Введем обозначение: $i=(0,1)$. Из аксиом 1 – 4 непосредственно вытекают следующие свойства комплексных чисел.

1°. $i^2=-1$, так как $i^2=(0,1)(0,1)=(-1,0)=-1$.

2°. Любое комплексное число $z=(a,b)$ может быть записано в виде

$$z=a+bi, \tag{44.1}$$

так как $z=(a,b)=(a,0)+(0,b)=(a,0)+(b,0)(0,1)$.

Форма (44.1) записи комплексного числа $z=(a,b)$ называется ***алгебраической формой числа*** z, при этом число a называется ***действительной частью*** комплексного числа $z=a+bi$ и обозначается символом $\operatorname{Re} z$, а b – ***мнимой частью*** и обозначается $\operatorname{Im} z$. Подчеркнем, что $\operatorname{Re} z$, $\operatorname{Im} z \in \mathbb{R}$. Для вещественных чисел мнимая часть равна нулю. Комплексные числа, у которых действительная часть равна нулю, называются ***чисто мнимыми***. Очевидно, два комплексных числа равны тогда и только тогда, когда отдельно равны их действительные и мнимые части.

Комплексное число $\overline{z}=a-bi$ называется ***сопряженным*** к числу $z=a+bi$.

Теорема 44.2. *Операция сопряжения комплексного числа обладает следующими свойствами:*

1) $\overline{\overline{z}}=z$;
2) $z=\overline{z} \iff z\in\mathbb{R}$;
3) $\overline{z}+z=2a$, $\forall z=a+bi$;
4) $\overline{z}z=a^2+b^2$, $\forall z=a+bi$;
5) $\overline{z_1\pm z_2}=\overline{z_1}\pm\overline{z_2}$; $\overline{z_1z_2}=\overline{z_1}\,\overline{z_2}$; $\overline{(z_1/z_2)}=\overline{z_1}/\overline{z_2}$, $z_2\neq 0$.

Все свойства проверяются непосредственно исходя из определения. ■

Замечание 2. Для комплексных чисел, заданных в алгебраической форме, операции сложения, вычитания, умножения и деления производятся по обычным правилам выполнения этих операций над двучленами $a+bi$ с учетом того, что $i^2=-1$, и последующим приведением подобных членов (т.е. отдельно группируются вещественные числа и чисто мнимые). Особенно это удобно для умножения чисел: если $z_1=a+bi$, $z_2=c+di$, то $z_1z_2=(a+bi)(c+di)=ac+adi+bci-bd=$

$(ac-bd)+(ad+bc)i$. Для деления чисел удобно числитель и знаменатель дроби z_1/z_2 предварительно умножить на $\overline{z_2}$:

$$\frac{z_1}{z_2}=\frac{(a+bi)(c-di)}{z_2\overline{z_2}}=\left\{z_2\overline{z_2}=c^2+d^2\neq 0\right\}=\frac{(ac+bd)}{c^2+d^2}+\frac{(bc-ad)}{c^2+d^2}i.$$

Комплексная плоскость. Пусть на плоскости выбрана прямоугольная декартова система координат. Поставим в соответствие каждому комплексному числу $z=a+bi$ точку M (рис. 1) этой плоскости с координатами (a,b). Очевидно, что это соответствие взаимно однозначно. Вещественные числа изображаются точками оси абсцисс; на оси ординат располагаются изображения чисто мнимых чисел. Началу координат соответствует число ноль; сопряженные комплексные числа изображаются точками, симметричными относительно оси абсцисс (рис.1).

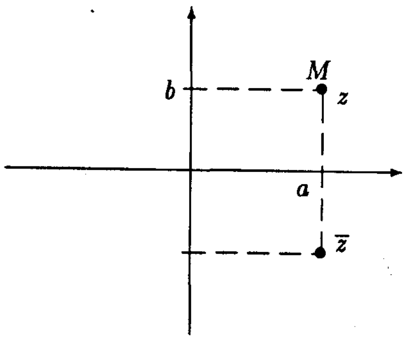


Рис. 1

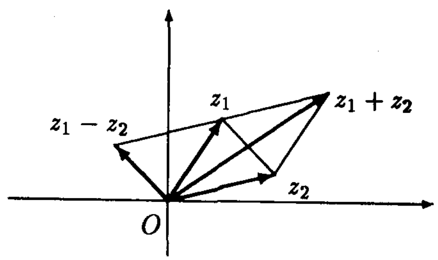


Рис. 2

Плоскость, точками которой изображаются комплексные числа, называется *комплексной плоскостью*, ее ось абсцисс – *вещественной осью*, ось ординат – *мнимой осью* (в соответствии с наименованием чисел, изображения которых лежат на этих осях).

Простое геометрическое истолкование получают на комплексной плоскости операции сложения и вычитания комплексных чисел. Так как каждая точка $M(a,b)$ плоскости связана с ее радиус-вектором $\mathbf{r}=\{a,b\}$, то комплексное число $z=a+bi$ можно изображать на комплексной плоскости радиус-вектором точки $M(a,b)$. При сложении (вычитании) комплексных чисел $z_1=a+bi$ и $z_2=c+di$ складываются (вычитаются) координаты радиус-векторов $\mathbf{r}_1\{a,b\}$ и $\mathbf{r}_2\{c,d\}$ точек, изображающих эти числа. Поэтому сложение (вычитание) комплексных чисел равносильно сложению (вычитанию) радиус-векторов (рис. 2) и может быть выполнено по правилу параллелограмма (§11).

Геометрический смысл умножения и деления комплексных чисел станет ясным из другой формы комплексного числа.

Сопряженная матрица. Пусть $A=(a_{ij})\in\mathbb{C}^{m\times n}$ – матрица размера $m\times n$ над полем комплексных чисел. Матрица $A^H=(a^h_{ij})$ размера $n\times m$ называется ***сопряженной*** к матрице A, если

$$a^h_{ij}=\overline{a_{ji}},\qquad i=\overline{1,n},\ j=\overline{1,m}.$$

Очевидно, что $A^H=(\overline{A})^T=\overline{(A^T)}$, где $\overline{A}=(\overline{a_{ij}})$.

Из определения и теоремы 44.2 вытекают следующие **свойства** сопряженной матрицы:

1) $(A+B)^H=A^H+B^H$,
2) $(\alpha A)^H=\overline{\alpha}A^H$, $\forall\alpha\in\mathbb{C}$,
3) $(AB)^H=B^HA^H$,
4) $(A^H)^H=A$,
5) $\det A^H=\overline{\det A}$,
6) $\operatorname{rg}A^H=\operatorname{rg}A$,

выполненные для всех **матриц, для которых определены левые части** равенств.

§45. Тригонометрическая форма комплексного числа

Алгебраическая форма комплексного числа $z=a+bi$ связана с декартовыми координатами точки $M(a,b)$, изображающей это число на комплексной плоскости. Однако положение точки на плоскости однозначно определяется и ее полярными координатами (§25): расстоянием r от этой точки до начала координат и углом φ между положительным направлением оси абсцисс и радиус-вектором точки $M(a,b)$ (рис. 1). В соответствии с этим вводятся следующие характеристики комплексного числа.

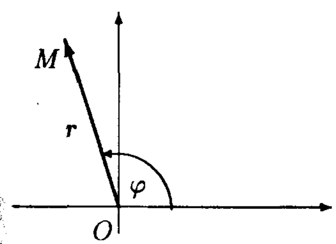


Рис. 1

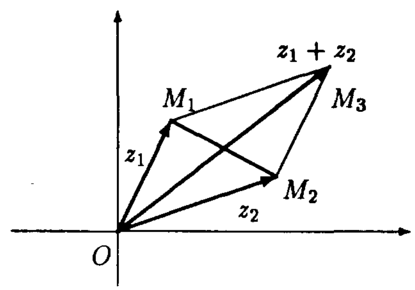


Рис. 2

Модулем комплексного числа $z=a+bi$ называется число $r=\sqrt{a^2+b^2}$. Обозначение: $|z|$. Из определения вытекают следующие факты.

1°. $|z|$ – действительное неотрицательное число, причем

$$|z|=0\iff z=0.$$

2°. $|z|$ совпадает с полярным радиусом точки M, изображающей это число на комплексной плоскости (рис. 1).

3°. $|z| = \sqrt{z\bar{z}}$.

4°. Модуль вещественного числа совпадает с абсолютным значением этого числа.

Аргументом комплексного числа $z \neq 0$ называется угол φ между положительным направлением оси абсцисс и радиус-вектором точки M (рис. 1), отсчитываемый от оси абсцисс в любом направлении, при этом положительным считается направление против часовой стрелки. Обозначение: $\arg z$. Из определения вытекают следующие факты.

1°. $\arg z$ не определен для $z = 0$, а для $z \neq 0$ определен с точностью до слагаемого, кратного 2π, так как при одном и том же r углы, отличающиеся на слагаемое $2\pi k$, где $k \in \mathbb{Z}$, определяют одну и ту же точку комплексной плоскости, т.е. одно и то же комплексное число.

2°. $\arg z$ отличается от полярного угла точки M тем, что $\arg z$ имеет бесконечно много значений.

3°. Два комплексных числа z_1 и z_2 равны тогда и только тогда, когда

$$\begin{cases} |z_1| = |z_2| \quad \text{и, если} \quad |z_1| \neq 0, \\ \arg z_1 = \arg z_2 + 2\pi k, k \in \mathbb{Z}. \end{cases} \tag{45.1}$$

Теорема 45.1. *Любое комплексное число $z \neq 0$ может быть записано в виде*

$$z = r(\cos\varphi + i\sin\varphi), \tag{45.2}$$

где $r = |z|$, $\varphi = \arg z$.

Доказательство. Пусть $z = a + bi$. Тогда из соотношений (25.1), связывающих прямоугольные координаты с полярными, следует, что $a = r\cos\varphi$, $b = r\sin\varphi$ и $z = r(\cos\varphi + i\sin\varphi)$. ■

Форма (45.2) записи комплексного числа называется *тригонометрической формой* этого числа.

Теорема 45.2. *Для любых комплексных чисел z_1 и z_2 имеют место неравенства*

$$||z_1| - |z_2|| \leq |z_1 \pm z_2| \leq |z_1| + |z_2|. \tag{45.3}$$

Доказательство. Неравенства (45.3) вытекают из правила параллелограмма сложения и вычитания комплексных чисел (§44) и неравенств треугольника (рис. 2: ΔOM_2M_3, ΔOM_1M_2). Это обоснование относится к случаю, когда точки O, M_1, M_2 не лежат на одной прямой (очевидно, при этом неравенства (45.3) будут строгими). Если же точки O, M_1, M_2 лежат на одной прямой, то несложный перебор вариантов расположения этих точек приводит к неравенствам (45.3), которые в некоторых из этих вариантов могут обратиться в равенства. ■

Неравенства (45.3) называют *неравенствами треугольника на комплексной плоскости*. Отметим, что теорема 45.2 относится только к модулям; для самих комплексных чисел неравенство не определено.

Теорема 45.3. *При умножении комплексных чисел их модули умножаются, а аргументы складываются; при делении комплексных чисел их модули делятся, а аргументы вычитаются:*

$$|z_1 z_2| = |z_1||z_2|, \quad \arg z_1 z_2 = \arg z_1 + \arg z_2; \tag{45.4}$$

$$\left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|}, \quad \arg\frac{z_1}{z_2} = \arg z_1 - \arg z_2, \quad z_2 \neq 0. \tag{45.5}$$

Доказательство. Пусть $z_k = r_k(\cos\varphi_k + i\sin\varphi_k)$, где $k = 1, 2$. Тогда $z_1 z_2 = r_1 r_2((\cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2) + i(\sin\varphi_1\cos\varphi_2 + \cos\varphi_1\sin\varphi_2)) = r_1 r_2(\cos(\varphi_1 + \varphi_2) + i\sin(\varphi_1 + \varphi_2))$. Отсюда следует (45.4). Если $z_2 \neq 0$, то $\dfrac{z_1}{z_2} = \dfrac{r_1(\cos\varphi_1 + i\sin\varphi_1)}{r_2(\cos\varphi_2 + i\sin\varphi_2)}\,\dfrac{\cos\varphi_2 - i\sin\varphi_2}{\cos\varphi_2 - i\sin\varphi_2} = \dfrac{r_1}{r_2}(\cos(\varphi_1 - \varphi_2) + i\sin(\varphi_1 - \varphi_2))$. Отсюда следует (45.5). ■

Замечание 1. Если один из сомножителей в (45.4) или z_1 в (45.5) равен нулю, то теорема относится только к модулям.

Замечание 2. Соотношения (45.4) переносятся и на любое число сомножителей (достаточно применить метод математической индукции).

§46. Возведение в степень и извлечение корня

Пусть $n \in \mathbb{Z}$ и $z \in \mathbb{C}$. Число z^n, определенное равенствами $z^n = \underbrace{z \cdot z \cdot \ldots \cdot z}_{n}$ при $n \in \mathbb{N}$; $z^n = 1$ при $n = 0$ (в этом случае предполагается, что $z \neq 0$); $z^n = 1/z^m$ при $m = -n \in \mathbb{N}$, называется *n-й степенью числа z*.

Теорема 46.1. *Если $z = r(\cos\varphi + i\sin\varphi)$, $n \in \mathbb{Z}$, то*

$$z^n = r^n(\cos n\varphi + i\sin n\varphi). \tag{46.1}$$

Доказательство. Для $n \in \mathbb{N}$ равенство (46.1) вытекает из (45.4). Для $n = 0$ равенство (46.1) следует из определения: $z^0 = 1 = r^0(\cos 0 + i\sin 0)$. Если $n < 0$ и $n = -m$, $m \in \mathbb{N}$, то

$$z^n = \frac{1}{z^m} = \frac{1}{r^m(\cos m\varphi + i\sin m\varphi)} = r^n(\cos m\varphi - i\sin m\varphi) = \\ = r^n(\cos(-m\varphi) + i\sin(-m\varphi)) = r^n(\cos n\varphi + i\sin n\varphi). \quad ■$$

Формула (46.1) называется *формулой Муавра*.

Пусть $n \in \mathbb{N}$. ***Корнем n-й степени из комплексного числа z*** называется число $\alpha \in \mathbb{C}$ такое, что $\alpha^n = z$.

Очевидно, что для $z = 0$ существует единственный корень n-й степени, равный нулю (так как $\alpha^n = 0 \Leftrightarrow |\alpha|^n = 0 \Leftrightarrow \alpha = 0$).

Теорема 46.2. *Для ненулевого числа $z = r(\cos\varphi + i\sin\varphi)$ существует ровно n различных корней $\alpha_1, \alpha_2, \ldots, \alpha_n$ n-й степени:*

$$\alpha_k = \sqrt[n]{r}\left(\cos\frac{\varphi+2\pi k}{n} + i\sin\frac{\varphi+2\pi k}{n}\right), \quad k = \overline{0, n-1}. \tag{46.2}$$

Доказательство. Пусть α – корень n-й степени из числа z и $\alpha = \rho(\cos\theta + i\sin\theta)$. Тогда $\alpha^n = z$ или, в силу (46.1), $\rho^n(\cos n\theta + i\sin n\theta) = r(\cos\varphi + i\sin\varphi)$. Согласно (45.1) это означает, что

$$\rho^n = r, \qquad n\theta = \varphi + 2\pi k, \quad k \in \mathbb{Z}.$$

Так как $\rho > 0$, $r > 0$, то существует единственный положительный корень n-й степени ρ из положительного числа r – это арифметический корень $\rho = \sqrt[n]{r}$. Таким образом, число $\alpha = \rho(\cos\theta + i\sin\theta)$ является корнем n-й степени из числа z тогда и только тогда, когда

$$\rho = \sqrt[n]{r}, \qquad \theta = \frac{\varphi+2\pi k}{n}, \; k \in \mathbb{Z},$$

т.е. числа $\alpha_k = \sqrt[n]{r}\left(\cos\frac{\varphi+2\pi k}{n} + i\sin\frac{\varphi+2\pi k}{n}\right)$, $k \in \mathbb{Z}$, и только они, являются корнями n-й степени из числа z. Числа α_k для $k = \overline{0, n-1}$ различны, так как их аргументы отличаются самое большее на $\frac{2\pi(n-1)}{n} < 2\pi$. Числа α_k при $k \geq n$ совпадают с одним из корней $\alpha_0, \alpha_1, \ldots, \alpha_{n-1}$, так как если $k = nq + r$, $0 \leq r \leq n-1$, то

$$\frac{\varphi+2\pi k}{n} = \frac{\varphi+2\pi nq+2\pi r}{n} = \frac{\varphi+2\pi r}{n} + 2\pi q, \; q \in \mathbb{Z}.$$

Следовательно, $\alpha_k = \alpha_r$. ■

Геометрическая интерпретация корней. На комплексной плоскости все корни n-й степени из ненулевого числа z расположены на окружности радиуса $\rho = \sqrt[n]{|z|}$ и делят эту окружность на n равных частей (так как $\arg\alpha_k = \arg\alpha_{k-1} + 2\pi/n$).

В частности, все корни n-й степени из единицы имеют вид

$$\varepsilon_k = \cos\frac{2\pi k}{n} + i\sin\frac{2\pi k}{n}, \quad k = \overline{0, n-1}. \tag{46.3}$$

Они расположены на единичной окружности и делят ее на n равных частей, начиная с $\varepsilon_0 = 1$. При этом действительных корней

может быть либо два, если n четно (ε_0 и $\varepsilon_{n/2}$), либо один, если n нечетно (ε_0). В любом случае недействительных корней четное число, они расположены симметрично относительно действительной оси, т.е. попарно сопряжены.

Теорема 46.3. ***Все корни n-й степени из комплексного числа $z \neq 0$ получаются умножением одного из этих корней на все корни n-й степени из единицы.***

Доказательство. Пусть α_k, $k = \overline{0, n-1}$, – корни n-й степени из $z = r(\cos\varphi + i\sin\varphi)$, определяемые равенством (46.2). Тогда, согласно (45.4) и (46.3), $\alpha_k\varepsilon_0 = \alpha_k$, $\alpha_k\varepsilon_1 = \alpha_{k+1}, \ldots, \alpha_k\varepsilon_{n-k} = \alpha_0, \ldots, \alpha_k\varepsilon_{n-1} = \alpha_{k-1}$. ■

Теореме 46.3 можно придать и другой смысл. Действительно, нетрудно проверить, что все корни n-й степени из единицы образуют мультипликативную группу. Тогда смежный класс мультипликативной группы ненулевых комплексных чисел по подгруппе $H = \{\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{n-1}\}$ корней n-й степени из единицы, порожденный корнем α_k, состоит из произведений $\alpha_k\varepsilon_m$, где $m = \overline{0, n-1}$. Но $(\alpha_k\varepsilon_m)^n = \alpha_k^n\varepsilon_m^n = z \cdot 1 = z$. Таким образом, ***все корни n-й степени из числа z образуют смежный класс мультипликативной группы ненулевых комплексных чисел по подгруппе корней n-й степени из единицы, порожденный одним из корней n-й степени из z.***

Глава X. Многочлены над произвольным полем

В §13, 31 уже упоминались многочлены с вещественными коэффициентами в связи с линейными пространствами и в связи с алгебраическими линиями и поверхностями. В этой главе многочлены будут основным объектом изучения. Начнем с алгебраических (формальных) свойств многочленов без учета того, что многочлен является не только формальным выражением, но также и функцией.

§47. Кольцо многочленов

Пусть P – поле. *Многочленом (полиномом) n-й степени от переменной x над полем P* называется выражение

$$a_0 + a_1x + a_2x^2 + \ldots + a_nx^n, \tag{47.1}$$

где a_i, $i = \overline{0,n}$, — фиксированные числа из поля P и $a_n \neq 0$. Эти числа называются *коэффициентами многочлена*, а число a_n — *старшим коэффициентом*. Число $0 \in P$ по определению считается многочленом с нулевыми коэффициентами и называется ***нулевым многочленом.*** Степень нулевого многочлена не определена.

Многочлен обозначается символом $f(x)$ или $f_n(x)$, степень многочлена $f(x)$ – символом $\deg f$, множество всех многочленов от переменной x над полем P – символом $P[x]$. Итак,

$$f(x) = \sum_{k=0}^{n} a_kx^k \in P[x], \quad \deg f = n.$$

Замечание 1. Выражения $x, x^2, \ldots, x^n$ не несут смысловой нагрузки (пока), они должны восприниматься как символы. Таким образом, на многочлен (47.1) следует смотреть как на некоторое формальное выражение, вполне определяемое набором своих коэффициентов $a_0, a_1, \ldots, a_n$, где $a_n \neq 0$.

Два многочлена $f(x), g(x) \in P[x]$ называются ***равными***, если равны их коэффициенты при одинаковых степенях переменной x.

Обозначение: $f(x) = g(x)$. Из определения следует, что нулевой многочлен равен только нулевому многочлену, а для ненулевых многочленов $f(x) = \sum_{k=0}^{n} a_kx^k$ и $g(x) = \sum_{k=0}^{s} b_kx^k$ равенство $f(x) = g(x)$ означает, что

$$\deg f = \deg g = n; \qquad a_i = b_i,\ i = \overline{0,n}. \tag{47.2}$$

Равенство многочленов, определенное здесь, означает тождественное равенство, и мы будем называть его *формальным* в отличие от равенства многочленов как функций. В §49 будет доказана равносильность этих определений равенства.

Введем на множестве $P[x]$ алгебраические операции.

Суммой многочленов $f(x) = \sum_{k=0}^{n} a_kx^k$ и $g(x) = \sum_{k=0}^{s} b_kx^k$ называется многочлен

$$h(x) = \sum_{k=0}^{\max(n,s)} c_kx^k, \quad \text{где} \quad c_k = a_k + b_k. \tag{47.3}$$

Здесь недостающие коэффициенты a_k или b_k заменяются нулями. Обозначение: $f(x) + g(x)$.

Из определения вытекают следующие факты.

1°. Для любого многочлена $f(x) \in P[x]$

$$f(x) + 0 = 0 + f(x) = f(x). \tag{47.4}$$

2°. Для ненулевых многочленов $f(x)$, $g(x)$, $f(x) + g(x)$

$$\deg(f + g) \leq \max(\deg f, \deg g). \tag{47.5}$$

3°. Если $f(x), g(x) \in P[x]$, то $f(x) + g(x) \in P[x]$, т.е. сложение многочленов является алгебраической операцией на множестве $P[x]$.

Произведением многочленов $f(x) = \sum_{k=0}^{n} a_kx^k$ и $g(x) = \sum_{k=0}^{s} b_kx^k$ называется многочлен

$$h(x) = \sum_{k=0}^{n+s} c_kx^k, \quad \text{где} \quad c_k = \sum_{i+j=k} a_ib_j,\ k = \overline{0, n+s}. \tag{47.6}$$

Здесь суммирование в $\sum_{i+j=k} a_ib_j$ ведется по всевозможным индексам i и j, для которых $i + j = k$. Обозначение: $f(x)g(x)$.

Из определения вытекают следующие факты.

1°. Произведение ненулевых многочленов не может быть нулевым, при этом

$$\deg fg = \deg f + \deg g. \tag{47.7}$$

2°. Если $f(x), g(x) \in P[x]$, то $f(x)g(x) \in P[x]$, т.е. умножение многочленов является алгебраической операцией на множестве $P[x]$.

3°. Операция умножения многочленов порождает операцию умножения многочлена на число из поля P как частный случай умножения многочленов: если $f(x) = \sum_{k=0}^{n} a_kx^k$ и $\alpha \in P$, то

$$\alpha f(x) = \sum_{k=0}^{n} \alpha a_kx^k; \tag{47.8}$$

таким образом, на множестве $P[x]$ определен и внешний закон композиции.

Теорема 47.1. *Множество $P[x]$ всех многочленов над полем P является коммутативным кольцом с единицей и без делителей нуля.*

Доказательство. Проверим все аксиомы кольца (§42). Прежде всего отметим, что $P[x]$ – аддитивная абелева группа: коммутативность и ассоциативность сложения очевидны (в силу (47.3)), нулем является нулевой многочлен (как отмечено в (47.4)), противоположным к многочлену $f(x) = \sum_{k=0}^{n} a_k x^k$, как легко проверить, является многочлен $-f(x) = \sum_{k=0}^{n}(-a_k)x^k$. Коммутативность умножения следует из определения. Докажем ассоциативность умножения. Пусть $f_n(x) = \sum_{k=0}^{n} a_k x^k$, $g_s(x) = \sum_{k=0}^{s} b_k x^k$, $h_p(x) = \sum_{k=0}^{p} c_k x^k$. Обозначим через α_k, β_k, γ_k и δ_k коэффициенты при x^k у многочленов $f(x)g(x)$, $g(x)h(x)$, $(f(x)g(x))h(x)$ и $f(x)(g(x)h(x))$ соответственно. Тогда в силу (47.6)

$$\gamma_k = \sum_{i+j=k} \alpha_i c_j = \sum_{i+j=k} \Big(\sum_{r+t=i} a_r b_t \Big) c_j = \sum_{r+t+j=k} a_r b_t c_j,$$

$$\delta_k = \sum_{r+i=k} a_r \beta_i = \sum_{r+i=k} a_r \Big(\sum_{t+j=i} b_t c_j \Big) = \sum_{r+t+j=k} a_r b_t c_j,$$

т.е. $\gamma_k = \delta_k$. Отсюда, если учесть, что $\deg(fg)h = \deg f(gh) = n+s+p$, следует равенство $(f(x)g(x))h(x) = f(x)(g(x)h(x))$.

Роль единицы при умножении многочленов играет число 1, рассматриваемое как многочлен нулевой степени.

Справедливость аксиомы дистрибутивности вытекает из равенства $\sum_{i+j=k}(a_i + b_i)c_j = \sum_{i+j=k} a_i c_j + \sum_{i+j=k} b_i c_j$, так как левая часть этого равенства является коэффициентом при x^k в многочлене $(f(x)+g(x))h(x)$, а правая часть – коэффициентом при той же степени x в многочлене $f(x)h(x) + g(x)h(x)$.

Наконец, из (47.7) следует, что в $P[x]$ нет делителей нуля. ■

Следствие. Множество $P[x]$ является линейным пространством над полем P. Это следует из того, что любое кольцо по сложению – абелева группа; что же касается внешнего закона композиции (47.8), то его дистрибутивность относительно сложения является частным проявлением общего закона дистрибутивности в кольце $P[x]$, а справедливость других аксиом умножения на число очевидна.

Замечание 2. Произведение (47.6) многочленов $f(x)$ и $g(x)$ может быть получено обычным для элементарной алгебры перемножением двух сумм $(a_0 + a_1 x + \ldots + a_n x^n)(b_0 + b_1 x + \ldots + b_s x^s)$ с последующей группировкой одночленов одинаковой степени. Это следует из общего закона дистрибутивности и из того, что многочлены $f(x)$ и $g(x)$ можно рассматривать как суммы многочленов.

Замечание 3. Кольцо $P[x]$ не является полем, так как не всякий многочлен $f(x) \in P[x]$ обладает обратным многочленом $f^{-1}(x)$. Действительно, равенство $f(x)f^{-1}(x) = 1$ с учетом (47.7) означает, что *многочлены нулевой степени, и только они, обладают обратными.*

§48. Деление многочленов

Кольцо $P[x]$ всех многочленов над полем P по своим свойствам близко к кольцу $\mathbb{Z}$ всех целых чисел. Эта аналогия проявляется и в том, что для многочленов, как и для целых чисел, имеют место понятия деления нацело, деления с остатком, делителя, наибольшего общего делителя и др.

Теорема 48.1. *Для любых двух многочленов $f(x), g(x) \in P[x]$, где $g(x) \neq 0$, существует, и притом единственная, пара многочленов $q(x), r(x) \in P[x]$ такая, что:*

$$\begin{aligned} &f(x) = g(x)q(x) + r(x), \\ &\textit{где либо} \quad r(x) = 0, \quad \textit{либо} \quad \deg r < \deg g. \end{aligned} \tag{48.1}$$

Доказательство. Докажем первую часть теоремы. Будем считать, что $f(x) \neq 0$, так как в противном случае можно положить $q(x) = 0$, $r(x) = 0$. Пусть $f(x) = \sum_{k=0}^{n} a_k x^k$, $g(x) = \sum_{k=0}^{s} b_k x^k$, $\deg f = n$, $\deg g = s$. Без ограничения общности считаем, что $n \geq s$, так как в противном случае можно взять $q(x) = 0$ и $r(x) = f(x)$. Применим метод математической индукции по степени n многочлена $f(x)$, считая $g(x)$ фиксированным.

1. Пусть $n = 0$. Тогда $s = 0$ и $q(x) = a_0 b_0^{-1}$, $r(x) = 0$.

2. Пусть теперь теорема верна для любого многочлена степени меньшей n. Докажем ее для многочлена $f(x)$ степени $n \geq s$. Воспроизведем первый шаг известного из элементарной алгебры алгоритма деления многочленов с действительными коэффициентами, т.е. построим одночлен $\dfrac{a_n}{b_s}x^{n-s}$ и составим разность

$$f_1(x) = f(x) - \frac{a_n}{b_s}x^{n-s}g(x). \tag{48.2}$$

Либо многочлен $f_1(x)$ равен 0, либо $\deg f_1 < n$. В первом случае можно положить $q(x) = \dfrac{a_n}{b_s}x^{n-s}$, $r(x) = 0$. Во втором случае для многочлена $f_1(x)$ по индуктивному предположению найдутся многочлены $q_1(x)$ и $r(x)$ такие, что $f_1(x) = g(x)q_1(x) + r(x)$, где либо $r(x) = 0$, либо $\deg r < \deg g$. Тогда, согласно (48.2), $f(x) = \dfrac{a_n}{b_s}x^{n-s}g(x) + g(x)q_1(x) + r(x)$. Положив $q(x) = \dfrac{a_n}{b_s}x^{n-s} + q_1(x)$, приходим к паре многочленов $q(x), r(x) \in P[x]$, удовлетворяющей условиям (48.1).

Остается доказать единственность. Пусть существует еще одна пара многочленов $q_1(x), r_1(x) \in P[x]$, удовлетворяющая условиям (48.1). Тогда $f(x) = g(x)q(x) + r(x)$ и $f(x) = g(x)q_1(x) + r_1(x)$, т.е.

$$g(x)(q(x) - q_1(x)) = r(x) - r_1(x). \tag{48.3}$$

Если $r(x) - r_1(x) \neq 0$, то и $q(x) - q_1(x) \neq 0$, при этом степень правой части равенства (48.3) меньше s (в силу (47.5) и (48.1)), а левой части – не меньше s (в силу (47.7)), что невозможно. Следовательно, $r(x) - r_1(x) = 0$, а так как в кольце $P[x]$ нет делителей нуля, то и $q(x) - q_1(x) = 0$. Таким образом, $r(x) = r_1(x)$ и $q(x) = q_1(x)$. ■

Заметим, что теорема очень похожа на соответствующую теорему о делимости целых чисел. По аналогии с целыми числами многочлен $q(x)$ называется *частным* (или *неполным частным*) от деления $f(x)$ на $g(x)$, а $r(x)$ – *остатком* от этого деления. Если $r(x) = 0$, то говорят, что $f(x)$ *делится* (или *нацело делится*) на $g(x)$, при этом $g(x)$ называется *делителем* $f(x)$. Очевидно, делителями любого многочлена $f(x)$ будут все многочлены нулевой степени и все многочлены вида $\alpha f(x)$, где $\alpha \neq 0$.

Многочлен $\varphi(x)$ называется *общим делителем* многочленов $f(x)$ и $g(x)$, если он является делителем каждого из них. Очевидно, все многочлены нулевой степени – общие делители любой пары многочленов. Многочлены, не имеющие других общих делителей, кроме многочленов нулевой степени, называются *взаимно простыми.*

Многочлен $d(x)$ называется *наибольшим общим делителем* ненулевых многочленов $f(x)$ и $g(x)$, если:

1) $d(x)$ – общий делитель многочленов $f(x)$ и $g(x)$;
2) $d(x)$ делится на любой общий делитель многочленов $f(x)$ и $g(x)$.

Обозначение: $\mathrm{HOD}(f, g)$. Очевидно, если $d(x) = \mathrm{HOD}(f, g)$, то $\alpha d(x) = \mathrm{HOD}(f, g)$ для любого $\alpha \neq 0$.

Теорема 48.2. *Для любой пары ненулевых многочленов $f(x)$, $g(x) \in P[x]$ существует наибольший общий делитель. Он определен однозначно с точностью до множителя нулевой степени.*

Доказательство. Приведем описание алгоритма построения $\mathrm{HOD}(f, g)$, называемого *алгоритмом Евклида* или *алгоритмом последовательного деления*. Он состоит в следующем.

Выполним цепочку делений с остатком согласно теореме 48.1:

$$\left\{\begin{array}{rcll} f(x) & = & g(x)q_1(x) + r_1(x), & \deg r_1 < \deg g, \\ g(x) & = & r_1(x)q_2(x) + r_2(x), & \deg r_2 < \deg r_1, \\ r_1(x) & = & r_2(x)q_3(x) + r_3(x), & \deg r_3 < \deg r_2, \\ \multicolumn{4}{c}{\dotfill} \\ r_{k-3}(x) & = & r_{k-2}(x)q_{k-1}(x) + r_{k-1}(x), & \deg r_{k-1} < \deg r_{k-2}, \\ r_{k-2}(x) & = & r_{k-1}(x)q_k(x) + r_k(x), & \deg r_k < \deg r_{k-1}, \\ r_{k-1}(x) & = & r_k(x)q_{k+1}(x). & \end{array}\right. \tag{48.4}$$

Степени остатков понижаются, поэтому процесс оборвется в тот момент, когда деление выполнится нацело. Пусть $r_k(x)$ – последний отличный от нуля остаток. Покажем, что $r_k(x) = \mathrm{HOD}(f, g)$. Действительно, просмотр равенств (48.4) снизу вверх показывает, что $r_k(x)$ является делителем $r_{k-1}(x)$, $r_{k-2}(x)$, ..., $r_1(x)$, $g(x)$, $f(x)$, т.е. общим делителем $f(x)$ и $g(x)$. Просмотр равенств (48.4) сверху вниз

показывает, что все остатки $r_1(x)$, $r_2(x)$, ..., $r_k(x)$ делятся на любой общий делитель $f(x)$ и $g(x)$.

Докажем вторую часть теоремы. Положим $d_1(x) = \mathrm{HOD}(f, g)$, $d_2(x) = \mathrm{HOD}(f, g)$. Тогда, согласно определению $\mathrm{HOD}(f, g)$, $d_1(x) = d_2(x)q_1(x)$, $d_2(x) = d_1(x)q_2(x)$ и, с учетом (47.7), $\deg d_1 \geq \deg d_2$, $\deg d_2 \geq \deg d_1$. Следовательно, $\deg d_1 = \deg d_2$ и многочлены $d_1(x)$ и $d_2(x)$ отличаются лишь множителем нулевой степени. ■

§49. Корни многочленов

До сих пор мы рассматривали многочлен как формально-алгебраическое выражение. Однако многочлен $f(x)$ можно рассматривать и как функцию от переменной x. Если $f(x) = \sum_{k=0}^{n} a_k x^k$ – многочлен над полем P, c – некоторое число из поля P, то число $f(c) = \sum_{k=0}^{n} a_k c^k$ называется *значением многочлена $f(x)$ при $x = c$.*

Известно (из теории функций), что две функции называются равными, если их значения равны при всех значениях переменной. Очевидно, что если многочлены $f(x)$ и $g(x)$ равны как многочлены (см. (47.2)), то они равны и как функции. Обратное, однако, потребует дополнительных исследований и будет доказано позже. А пока, забегая вперед, будем иметь в виду, что оба подхода к понятию равенства многочленов совпадают.

То же относится и к операциям над многочленами: сложение и умножение многочленов, определенные в §47, превращаются в сложение и умножение функций, так как если $\varphi(x) = f(x) + g(x)$, $\psi(x) = f(x)g(x)$, то $\varphi(c) = f(c) + g(c)$, $\psi(c) = f(c)g(c)$, $\forall c \in \mathbb{R}$.

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

Теорема 49.1 (теорема Безу). *Остаток от деления многочлена $f(x)$ на $x - c$ равен $f(c)$.*

Доказательство. Разделим согласно теореме 48.1 многочлен $f(x)$ на многочлен $x - c$. Тогда $f(x) = (x - c)q(x) + r(x)$, где $\deg r < \deg(x - c) = 1$, так что $r(x) = r$ – константа. Беря значения обеих частей этого равенства при $x = c$, получим, что $r = f(c)$. ■

Следствие. Число $c \in P$ является корнем многочлена $f(x) \in P[x]$ тогда и только тогда, когда многочлен $f(x)$ делится на $x - c$ в кольце $P[x]$.

Алгебраическая замкнутость поля $\mathbb{C}$. До сих пор все поля были равноправны. Однако в вопросе существования корней многочлена это далеко не так. Поле P называется *алгебраически замкнутым*, если любой многочлен $f(x) \in P[x]$ степени $n \geq 1$ обладает в P хотя бы одним корнем. Очевидно, что поле действительных чисел $\mathbb{R}$ не является алгебраически замкнутым, так как многочлен $x^2 + 1$ не имеет действительных корней.

Теорема 49.2 (основная теорема алгебры). *Поле $\mathbb{C}$ комплексных чисел алгебраически замкнуто.*

Эта теорема называется "основной" по традиции, установившейся с тех времен, когда проблема решения алгебраических уравнений была главной проблемой в алгебре (впервые теорема доказана Гауссом в 1799 г. для частного случая). Теперь она относится к числу рядовых утверждений, хотя и очень важных. Во всяком случае, на ней основана вся дальнейшая теория многочленов. И тем не менее эта теорема не является чисто алгебраической. Все ее доказательства (а их после Гаусса найдено довольно много) в той или иной мере опираются на другие разделы математики. Мы приведем одно из них, наиболее алгебраическое из доступных нам. Для этого потребуются дополнительные понятия и факты.

В приводимых ниже леммах, если не оговаривается особо, $f(z) = a_0 + a_1 z + \ldots + a_n z^n$ – многочлен над полем комплексных чисел от комплексной переменной z.

Лемма 1. *Пусть $f(z) = a_1 z + a_2 z^2 + \ldots + a_n z^n$ – многочлен с нулевым свободным членом. Тогда для любого $\varepsilon > 0$ найдется $\delta > 0$ такое, что для всех z, для которых $|z| < \delta$, выполняется неравенство $|f(z)| < \varepsilon$.*

Доказательство. Пусть $|z| < 1$. Тогда в силу (45.4) и (45.3)

$$|f(z)| = |z|\left|a_1 + a_2 z + \ldots + a_n z^{n-1}\right| \le |z|\left(|a_1| + |a_2| + \ldots + |a_n|\right).$$

Положим $M = |a_1| + |a_2| + \ldots + |a_n|$ и возьмем $\delta = \min(1, \varepsilon/M)$. Тогда для всех z, для которых $|z| < \delta$, выполняется неравенство $|f(z)| \le |z| \cdot M < \varepsilon/M \cdot M = \varepsilon$. ■

Комплексная функция $f(z)$ от комплексной переменной z называется *непрерывной в точке* z_0, если для любого $\varepsilon > 0$ существует $\delta > 0$ такое, что для всех z, для которых $|z - z_0| < \delta$, выполняется неравенство $|f(z) - f(z_0)| < \varepsilon$.

Лемма 2. *Многочлен $f(z) = a_0 + a_1 z + \ldots + a_n z^n$ есть непрерывная функция во всех точках комплексной плоскости.*

Доказательство. Пусть z_0 – произвольное комплексное число. Разложим многочлен $f(z)$ по степеням $z - z_0$: $f(z) = c_0 + c_1(z - z_0) + \ldots + c_n(z - z_0)^n$. Тогда $c_0 = f(z_0)$, так что $f(z) - f(z_0) = c_1(z - z_0) + \ldots + c_n(z - z_0)^n$.

Правая часть представляет собой многочлен от $z - z_0$ с нулевым свободным членом. По лемме 1 для любого $\varepsilon > 0$ найдется $\delta > 0$ такое, что $|f(z) - f(z_0)| < \varepsilon$ для всех z, для которых $|z - z_0| < \delta$. ■

Лемма 3. *Модуль многочлена есть непрерывная функция.*

Доказательство. Утверждение вытекает из леммы 2 и свойств модуля комплексных чисел (45.3): $|f(z) - f(z_0)| \ge \big||f(z)| - |f(z_0)|\big|$. ■

Лемма 4. *Если $f(z)$ – многочлен степени $n \ge 1$, то для любого $M > 0$ существует $R > 0$ такое, что для всех z, для которых $|z| > R$, выполняется неравенство $|f(z)| > M$.*

Доказательство. Пусть $f(z) = a_0 + a_1 z + \ldots + a_n z^n$. Запишем его в виде

$$f(z) = a_n z^n \left(1 + \frac{a_{n-1}}{a_n} z^{-1} + \ldots + \frac{a_0}{a_n} z^{-n}\right) = a_n z^n \left(1 + g(z^{-1})\right), \tag{49.1}$$

где $g(z^{-1})$ – многочлен от z^{-1} с нулевым свободным членом. В силу леммы 1 для $\varepsilon = 1/2$ найдется $\delta > 0$ такое, что при $|z^{-1}| < \delta$ имеет место неравенство $|g(z^{-1})| < 1/2$. Модуль $a_n z^n$ может быть сделан сколь угодно большим, именно при $|z| > \sqrt[n]{2M/|a_n|}$ будет $|a_n z^n| > 2M$. Возьмем $R = \max\left(\sqrt[n]{2M/|a_n|}, \frac{1}{\delta}\right)$. Тогда если $|z| > R$, то $|z^{-1}| < \delta$ и $|z| > \sqrt[n]{2M/|a_n|}$, так что согласно (49.1)
$|f(z)| = |a_n z^n|\left|1 + g(z^{-1})\right| \ge |a_n z^n|\left|1 - |g(z^{-1})|\right| > 2M(1 - \frac{1}{2}) = M$. ■

Число $z_0 = x_0 + iy_0$ называется *пределом последовательности* $z_n = x_n + iy_n$, если для любого $\varepsilon > 0$ существует натуральное число N такое, что $|z_n - z_0| < \varepsilon$ для всех $n > N$. Обозначение: $\lim\limits_{n\to\infty} z_n = z_0$. Очевидно, что сходимость последовательности комплексных чисел z_n равносильна сходимости двух последовательностей действительных чисел: $\lim\limits_{n\to\infty} x_n = x_0$, $\lim\limits_{n\to\infty} y_n = y_0$.

Последовательность z_n называется *ограниченной*, если существует число $R > 0$ такое, что $|z_n| \le R$.

Лемма 5. *Из любой ограниченной последовательности z_n можно выделить сходящуюся последовательность.*

Доказательство. Пусть $z_n = x_n + iy_n$ и $|z_n| \le R$, тогда $|x_n| \le R$, так что x_n – ограниченная последовательность действительных чисел. Из нее согласно теореме Больцано–Вейерштрасса [7] можно выделить сходящуюся подпоследовательность $x_{n_k} \to x_0$. Рассмотрим соответствующую подпоследовательность мнимых частей y_{n_k}. Она ограничена, и из нее также можно выделить сходящуюся подпоследовательность $y_{n_{k_m}} \to y_0$. Тогда соответствующая подпоследовательность $z_{n_{k_m}}$ сходится к $z_0 = x_0 + iy_0$. ■

Лемма 6. *Точная нижняя грань модуля многочлена достигается, т.е. существует число z_0 такое, что $|f(z_0)| \le |f(z)|$ при всех комплексных z.*

Доказательство. Рассмотрим множество всевозможных значений модуля многочлена $f(z)$. Так как $|f(z)| \ge 0$, то это множество ограничено снизу и, следовательно, имеет точную нижнюю грань. Обозначим ее через m. Тогда для любого натурального числа n можно найти комплексное число z_n такое, что

$$|f(z_n)| \le m + \frac{1}{n}. \tag{49.2}$$

Воспользуемся леммой 4: для $M = m + 1$ найдем R такое, что при $|z| > R$ будет $|f(z)| > M \ge m + \frac{1}{n}$. Отсюда и из (49.2) следует, что

$|z_n| \leq R$. Последовательность z_n оказалась ограниченной, и из нее согласно лемме 5 можно выделить сходящуюся подпоследовательность $z_{n_k} \to z_0$. Тогда в силу непрерывности $|f(z)|$ (лемма 3)

$$\lim_{k\to\infty} |f(z_{n_k})| = |f(z_0)|. \tag{49.3}$$

С другой стороны, из (49.2) и определения нижней грани имеем $m \leq |f(z_{n_k})| \leq m + \frac{1}{n_k}$, поэтому

$$\lim_{k\to\infty} |f(z_{n_k})| = m. \tag{49.4}$$

Сопоставление (49.3) и (49.4) приводит к требуемому равенству

$$|f(z_0)| = m. \blacksquare$$

Лемма 7 (лемма Даламбера). *Если $f(z)$ - многочлен степени $n \geq 1$ и $f(z_0) \neq 0$, то найдется число z_1 такое, что $|f(z_1)| < |f(z_0)|$.*

Доказательство. Разложим многочлен $f(z)$ по степеням $z - z_0$:

$$f(z) = c_0 + c_1(z-z_0) + \ldots + c_n(z-z_0)^n. \tag{49.5}$$

Очевидно, что $c_0 = f(z_0) \neq 0$. Пусть c_k - первый ненулевой коэффициент в (49.5) после c_0 (такой коэффициент имеется, так как $f(z)$ не константа). Тогда

$$f(z) = c_0 + c_k(z-z_0)^k + c_{k+1}(z-z_0)^{k+1} + \ldots + c_n(z-z_0)^n = c_0\Big(1+$$
$$+\tfrac{c_k}{c_0}(z-z_0)^k + \tfrac{c_k}{c_0}(z-z_0)^k\big(\tfrac{c_{k+1}}{c_k}(z-z_0) + \ldots + \tfrac{c_n}{c_k}(z-z_0)^{n-k}\big)\Big) =$$
$$= c_0\Big(1 + \frac{c_k}{c_0}(z-z_0)^k + \frac{c_k}{c_0}(z-z_0)^k g(z-z_0)\Big), \tag{49.6}$$

где $g(z-z_0) = \dfrac{c_{k+1}}{c_k}(z-z_0) + \ldots + \dfrac{c_n}{c_k}(z-z_0)^{n-k}$ - многочлен от $z - z_0$ с нулевым свободным членом. По лемме 1 для $\varepsilon = 1/2$ найдется такое δ, что если $|z - z_0| < \delta$, то

$$|g(z-z_0)| < 1/2. \tag{49.7}$$

Оценим правую часть (49.6). Пусть $\dfrac{c_k}{c_0} = R(\cos\theta + i\sin\theta)$, $z - z_0 = r(\cos\varphi + i\sin\varphi)$. Выберем r так, чтобы $Rr^k < 1$. Для этого нужно взять $r < \sqrt[k]{1/R}$. Далее положим $\theta + k\varphi = \pi$, т.е. возьмем $\varphi = (\pi - \theta)/k$. При таком выборе $\dfrac{c_k}{c_0}(z-z_0)^k = -Rr^k$. Теперь положим $z_1 = z_0 + r(\cos\varphi + i\sin\varphi)$ при $r < \min(\delta, \sqrt[k]{1/R})$ и $\varphi = (\pi-\theta)/k$.

Тогда из (49.6) следует, что $f(z_1) = c_0\left(1 - Rr^k - Rr^k g(z_1 - z_0)\right)$, и тем самым

$$|f(z_1)| = |c_0|\,|1 - Rr^k - Rr^k g(z_1-z_0)| \leq$$
$$\leq |c_0|\left(|1 - Rr^k| + Rr^k|g(z_1-z_0)|\right) \leq \{\text{в силу выбора } r \text{ и } (49.7)\} \leq$$
$$\leq |c_0|\left(1 - Rr^k + Rr^k/2\right) = |c_0|\left(1 - Rr^k/2\right) < |c_0| = |f(z_0)|. \blacksquare$$

Доказательство основной теоремы. Пусть $f(z)$ - произвольный многочлен над полем $\mathbb{C}$ от комплексной переменной z степени $n \geq 1$. Согласно лемме 6 множество всевозможных значений $|f(z)|$ имеет точную нижнюю грань m, которая достигается в некоторой точке z_0, так что $|f(z_0)| = m$. Тогда $f(z_0) = 0$, так как в противном случае, если $|f(z_0)| \neq 0$, то согласно лемме 7 найдется точка z_1, для которой $|f(z_1)| < |f(z_0)| = \inf|f(z)|$, что невозможно. Таким образом, z_0 - корень $f(z)$ и поле $\mathbb{C}$ комплексных чисел алгебраически замкнуто. $\blacksquare$

§50. Каноническое разложение многочлена над полем комплексных чисел

Теорема 50.1. *Для любого многочлена $f(z) = \sum_{k=0}^n a_k z^k \in \mathbb{C}[z]$ степени $n \geq 1$ существуют числа $c_1, c_2, \ldots, c_n \in \mathbb{C}$ такие, что*

$$f(z) = a_n(z-c_1)(z-c_2)\ldots(z-c_n). \tag{50.1}$$

Это разложение единственно с точностью до порядка сомножителей.

Доказательство. Из алгебраической замкнутости поля $\mathbb{C}$ следует существование корня $c_1 \in \mathbb{C}$ многочлена $f(z)$. Тогда в кольце $\mathbb{C}[z]$ многочлен $f(z)$ делится (теорема 49.1) на многочлен $z - c_1$, так что $f(z) = (z-c_1)f_1(z)$, где $f_1(z) \in \mathbb{C}[z]$, $\deg f_1 = n - 1$. Если $n - 1 \geq 1$, то к многочлену $f_1(z)$ также применима основная теорема алгебры и, следовательно, $f(z) = (z-c_1)(z-c_2)f_2(z)$, где $f_2(z) \in \mathbb{C}[z]$, $\deg f_2 = n-2$, $c_2 \in \mathbb{C}$. Применив эти рассуждения n раз, найдем числа $c_1, c_2, \ldots, c_n \in \mathbb{C}$ такие, что

$$f(z) = (z-c_1)(z-c_2)\ldots(z-c_n)f_n, \tag{50.2}$$

где $\deg f_n = 0$ и, следовательно, f_n - константа. Сравнив коэффициенты при z^n в обеих частях равенства (50.2), получим, что $f_n = a_n$. Тем самым доказано существование разложения (50.1).

Докажем его единственность. Пусть существует другое разложение:

$$f(z) = a_n(z-d_1)(z-d_2)\ldots(z-d_n). \tag{50.3}$$

Каждое число c_i из первого разложения встречается среди чисел $d_1, \ldots, d_n$ второго разложения, так как в противном случае для $c_i \neq$

d_j, $j=\overline{1,n}$, из (50.1) получим, что $f(c_i)=0$, а из (50.3) – что $f(c_i)\neq 0$. Аналогично каждое число d_j встречается в первом разложении.

Покажем теперь, что если $c_i=d_j$, то c_i встречается в (50.1) столько же раз, сколько d_j в (50.3). Пусть c_i равно d_j и встречается в (50.1) k раз, а в (50.3) – m раз и пусть $k>m$. Положим $(z-c_i)^m=\varphi(z)$. Тогда

$$\varphi(z)(z-c_i)^{k-m}\prod_{c_s\neq c_i}(z-c_s)=\varphi(z)\prod_{d_s\neq d_j}(z-d_s),$$

откуда следует, что

$$\varphi(z)\left((z-c_i)^{k-m}\prod_{c_s\neq c_i}(z-c_s)-\prod_{d_s\neq d_j}(z-d_s)\right)=0.$$

Так как в кольце $\mathbb{C}[z]$ нет делителей нуля и $\varphi(z)\neq 0$, то

$$(z-c_i)^{k-m}\prod_{c_s\neq c_i}(z-c_s)=\prod_{d_s\neq d_j}(z-d_s).$$

Положив в этом равенстве $z=c_i$, придем к противоречию. Итак, $k\leq m$. Аналогично показывается, что $m\leq k$. Значит, $k=m$. ■

Теорема 50.2 (о каноническом разложении многочлена над полем $\mathbb{C}$). *Для любого многочлена $f(z)=\sum_{k=0}^{n}a_kz^k\in\mathbb{C}[z]$ степени $n\geq 1$ существуют числа $c_1,c_2,\ldots,c_m\in\mathbb{C}$, где $c_i\neq c_j$ при $i\neq j$, и числа $k_1,k_2,\ldots,k_m\in\mathbb{N}$, где $k_1+k_2+\ldots+k_m=n$, такие, что*

$$f(z)=a_n(z-c_1)^{k_1}(z-c_2)^{k_2}\ldots(z-c_m)^{k_m}. \tag{50.4}$$

Это разложение единственно с точностью до порядка сомножителей.

Доказательство. Рассмотрим разложение (50.1) для многочлена $f(z)$. Среди чисел $c_1,c_2,\ldots,c_n$ могут быть одинаковые. Будем считать, что $c_1,c_2,\ldots,c_m$ различны, а каждое из $c_{m+1},\ldots,c_n$ равно одному из $c_1,c_2,\ldots,c_m$. Объединив одинаковые сомножители в этом разложении, получим

$$f(z)=a_n(z-c_1)^{k_1}(z-c_2)^{k_2}\ldots(z-c_m)^{k_m},$$

где $c_i\neq c_j$ при $i\neq j$ по построению, а $k_1+k_2+\ldots+k_m=n$, так как $k_1+k_2+\ldots+k_m$ – это число всех сомножителей в (50.1). Единственность разложения следует из теоремы 50.1. ■

Разложение (50.4) называется ***каноническим разложением многочлена над полем комплексных чисел.***

Заметим, что числа c_i, $i=\overline{1,m}$, в каноническом разложении (50.4) являются корнями многочлена $f(z)$ (так как $f(c_i)=0$, $i=\overline{1,m}$) и других корней этот многочлен не имеет (так как $f(\alpha)\neq 0$ для любого числа $\alpha\neq c_i$, $i=\overline{1,m}$). Число k_i в каноническом разложении многочлена называется ***кратностью корня*** c_i. Если $k_i=1$, то корень c_i называется ***простым***, если $k_i>1$ – ***кратным***.

Следствие 1. *Любой многочлен $f(z)\in\mathbb{C}[z]$ степени $n\geq 1$ имеет ровно n корней, если каждый корень считать столько раз, какова его кратность.*

Замечание. Это утверждение применимо и к многочлену нулевой степени, неприменимо лишь к нулевому многочлену, для которого любое число $z\in\mathbb{C}$ является корнем.

Вернемся к вопросу о равносильности двух определений равенства многочленов, о котором шла речь в начале §49.

Теорема 50.3. *Если два многочлена $f(z),g(z)\in\mathbb{C}[z]$, степени которых не превосходят $n\geq 1$, имеют равные значения при более чем n различных значениях переменной, то они равны.*

Доказательство. Действительно, многочлен $h(z)=f(z)-g(z)\in\mathbb{C}[z]$ имеет более n различных корней. Но $\deg h\leq n$, поэтому $h(z)=0$. ■

Следствие 2. *Два многочлена $f(z),g(z)\in\mathbb{C}[z]$ равны тогда и только тогда, когда совпадают их значения при всех значениях переменной z.*

Следствие 3. *Многочлен $f(z)\in\mathbb{C}[z]$ степени $n\geq 1$ однозначно определяется своими значениями в $n+1$ различных точках.*

Нетрудно показать, что если многочлен $f(z)$ в точках $\alpha_1,\alpha_2,\ldots,\alpha_{n+1}$, где $\alpha_i\neq\alpha_j$ при $i\neq j$, принимает значения $f(\alpha_1),f(\alpha_2),\ldots,f(\alpha_{n+1})$, то

$$f(z)=\sum_{i=1}^{n+1}f(\alpha_i)\frac{(z-\alpha_1)\ldots(z-\alpha_{i-1})(z-\alpha_{i+1})\ldots(z-\alpha_{n+1})}{(\alpha_i-\alpha_1)\ldots(\alpha_i-\alpha_{i-1})(\alpha_i-\alpha_{i+1})\ldots(\alpha_i-\alpha_{n+1})}.$$

Этот многочлен называется *интерполяционным многочленом Лагранжа.*

Теорема 50.4 (формулы Виета). *Если $c_1,c_2,\ldots,c_n$ – корни многочлена $f(z)=\sum_{k=0}^{n}a_kz^k$, то*

$$\begin{array}{rcl} a_{n-1}/a_n &=& -(c_1+c_2+\ldots+c_n),\\ a_{n-2}/a_n &=& +(c_1c_2+c_1c_3+\ldots+c_1c_n+c_2c_3+\ldots+c_{n-1}c_n),\\ &\ldots&\\ a_{n-k}/a_n &=& (-1)^k\sum\limits_{i_1<i_2<\ldots<i_k}c_{i_1}c_{i_2}\ldots c_{i_k},\\ &\ldots&\\ a_0/a_n &=& (-1)^nc_1c_2\ldots c_n. \end{array} \tag{50.5}$$

Доказательство. Достаточно рассмотреть случай, когда $a_n=1$, так как многочлены $f(z)$ и $(1/a_n)f(z)$ имеют одинаковые корни. Применим индукцию по n. Для $n=2$ формулы (50.5) являются известными из элементарной алгебры формулами Виета. Пусть для многочленов $(n-1)$-й степени формулы (50.5) имеют место, докажем их для многочлена $f(z)=a_0+a_1z+\ldots+a_{n-1}z^{n-1}+z^n$. Согласно (50.1)

$$f(z) = ((z-c_1)(z-c_2)\dots(z-c_{n-1}))(z-c_n) = g(z)(z-c_n),$$

где $g(z) = (z-c_1)(z-c_2)\dots(z-c_{n-1})$ – многочлен степени $n-1$. Пусть $g(z) = b_0 + b_1 z + \dots + b_{n-2}z^{n-2} + z^{n-1}$, тогда

$$f(z) = (b_0 + b_1 z + \dots + b_{n-2}z^{n-2} + z^{n-1})(z-c_n) =$$
$$= b_0 z + b_1 z^2 + \dots + b_{n-2}z^{n-1} + z^n - (b_0 c_n + b_1 c_n z + \dots + b_{n-2}c_n z^{n-2} + c_n z^{n-1}) = -b_0 c_n + (b_0 - b_1 c_n)z + \dots + (b_{n-k-1} - b_{n-k}c_n)z^{n-k} + \dots + z^n$$

и $f(z) = a_0 + a_1 z + \dots + a_{n-k}z^{n-k} + \dots + z^n$. Сравним коэффициенты при z^{n-k} и учтем, что для коэффициентов b_i, $i = \overline{0, n-2}$, верны соотношения (50.5), т.е. $b_{(n-1)-k} = (-1)^k \sum' c_{i_1}c_{i_2}\dots c_{i_k}$, где суммирование в $\sum'$ ведется по всем $1 \le i_1 < i_2 < \dots < i_k \le n-1$. Тогда

$$a_{n-k} = b_{n-k-1} - b_{n-k}c_n = b_{(n-1)-k} - b_{(n-1)-(k-1)}c_n =$$
$$= (-1)^k \sum{}' c_{i_1}c_{i_2}\dots c_{i_k} - (-1)^{k-1}c_n \sum{}' c_{i_1}c_{i_2}\dots c_{i_{k-1}} =$$
$$= (-1)^k \left(\sum{}' c_{i_1}c_{i_2}\dots c_{i_k} + c_n \sum{}' c_{i_1}c_{i_2}\dots c_{i_{k-1}}\right) =$$
$$= (-1)^k \sum c_{i_1}c_{i_2}\dots c_{i_k},$$

где суммирование ведется по всем $1 \le i_1 < i_2 < \dots < i_k \le n$. ■

Замечание. Все утверждения этого параграфа имеют место в любом алгебраически замкнутом поле.

§51. Многочлены над полем вещественных чисел

Пусть $f(x) = \sum_{k=0}^n a_k x^k$ – многочлен над полем $\mathbb{R}$ степени n, т.е. $a_k \in \mathbb{R}$, $k = \overline{0, n}$, и $a_n \neq 0$. Так как $\mathbb{R} \subset \mathbb{C}$, то многочлен $f(x)$ можно рассматривать и как многочлен над полем $\mathbb{C}$. Тогда, как показано в §50, этот многочлен с учетом кратностей имеет n корней (вообще говоря, комплексных).

Теорема 51.1. *Если c – комплексный (и не действительный) корень многочлена $f(x) \in \mathbb{R}[x]$, то корнем многочлена будет и $\bar{c}$.*

Доказательство. Если c является корнем многочлена $f(x) = \sum_{k=0}^n a_k x^k$, то $a_0 + a_1 c + \dots + a_n c^n = 0$. Взяв сопряжение от обеих частей этого равенства, получим с учетом теоремы 44.2 и соотношения $\bar{a}_k = a_k$, $k = \overline{0, n}$, что $a_0 + a_1\bar{c} + \dots + a_n\bar{c}^n = 0$. ■

Положим $\varphi(x) = (x-c)(x-\bar{c}) = x^2 - (c+\bar{c})x + |c|^2$. Очевидно, $\varphi(x) \in \mathbb{R}[x]$.

Теорема 51.2. *Если c – комплексный (но не действительный) корень многочлена $f(x) \in \mathbb{R}[x]$, то в кольце $\mathbb{R}[x]$ многочлен $f(x)$ делится на $\varphi(x)$.*

Доказательство. Разделим многочлен $f(x)$ на $\varphi(x)$ в кольце $\mathbb{R}[x]$ согласно (48.1), тогда $f(x) = \varphi(x)q(x) + r(x)$, где $q(x), r(x) \in \mathbb{R}[x]$. Это же разложение можно рассматривать как результат деления $f(x)$ на $\varphi(x)$ в кольце $\mathbb{C}[x]$. Но в кольце $\mathbb{C}[x]$ многочлен $f(x)$ делится нацело на $\varphi(x)$, так как c и $\bar{c}$ являются корнями $f(x)$. Отсюда и

из единственности результата деления (теорема 48.1) следует, что $f(x) = \varphi(x)q(x)$, где $q(x) \in \mathbb{R}[x]$. ■

Теорема 51.3. *Кратности корней c и $\bar{c}$ многочлена с вещественными коэффициентами совпадают.*

Доказательство. Пусть c и $\bar{c}$ – корни многочлена $f(x) \in \mathbb{R}[x]$ кратностей m и k соответственно. Покажем, что $k \le m$. Пусть $k > m$. Тогда, согласно теореме 51.2, $f(x) = (\varphi(x))^m q(x)$, где $q(x) \in \mathbb{R}[x]$. Но $q(c) = 0$, а $q(\bar{c}) \neq 0$, что невозможно в силу теоремы 51.1. Значит, $k \le m$. Аналогично можно показать, что $m \le k$. Следовательно, $k = m$. ■

Следствие. Многочлен с вещественными коэффициентами нечетной степени имеет хотя бы один вещественный корень.

Теорема 51.4 (о каноническом разложении многочлена над полем $\mathbb{R}$). *Для любого многочлена $f(x) = \sum_{k=0}^n a_k x^k \in \mathbb{R}[x]$ степени $n \ge 1$ существуют числа $c_1, \dots, c_r \in \mathbb{R}$, где $c_i \neq c_j$ при $i \neq j$; числа $p_1, q_1, \dots, p_s, q_s \in \mathbb{R}$, где $(p_i, q_i) \neq (p_j, q_j)$ при $i \neq j$; числа $k_1, \dots, k_r \in \mathbb{N}$ и $l_1, \dots, l_s \in \mathbb{N}$, где $\sum_{i=1}^r k_i + 2\sum_{j=1}^s l_j = n$, такие, что*

$$f(x) = a_n \prod_{i=1}^r (x-c_i)^{k_i} \prod_{j=1}^s (x^2 + p_j x + q_j)^{l_j}. \tag{51.1}$$

Доказательство. Используем каноническое разложение (50.4) многочлена $f(x)$, рассматривая $f(x)$ как многочлен с комплексными коэффициентами:

$$f(x) = a_n (x-c_1)^{k_1} \dots (x-c_m)^{k_m}, \tag{51.2}$$

где $c_i \neq c_j$ при $i \neq j$ и $k_1 + \dots + k_m = n$. Числа $c_1, \dots, c_m$ являются комплексными корнями $f(x)$. Среди этих корней могут быть и действительные. Пусть $c_1, \dots, c_r$ – действительные корни, а $c_{r+1}, \dots, c_m$ – комплексные (не действительные). Тогда среди $c_{r+1}, \dots, c_m$ вместе с каждым корнем c_j находится и сопряженный ему корень $c_i = \bar{c}_j$, причем той же кратности, так что $k_i = k_j = l_j$ и, следовательно, $2l_j = k_i + k_j$. Объединим в (51.2) пару сомножителей:

$$(x-c_j)^{k_j}(x-c_i)^{k_i} = ((x-c_j)(x-\bar{c}_j))^{l_j} = (x^2 + p_j x + q_j)^{l_j},$$

где $p_j = -(c_j + \overline{c_j}) \in \mathbb{R}$, $q_j = c_j\overline{c_j} \in \mathbb{R}$. Тогда

$$(x-c_{r+1})\dots(x-c_m) = (x^2 + p_1 x + q_1)^{l_1} \dots (x^2 + p_s x + q_s)^{l_s}$$

и разложение (51.2) перейдет в (51.1), причем $k_1 + \dots + k_r + 2(l_1 + \dots + l_s) = k_1 + \dots + k_r + k_{r+1} + \dots + k_m = n$. ■

Разложение (51.1) называется ***каноническим разложением многочлена над полем вещественных чисел.***

Замечание. Теорема 50.3 и ее следствия справедливы и для многочленов с вещественными коэффициентами. В этом можно убедиться, если рассматривать их как многочлены с комплексными коэффициентами. Формулы Виета справедливы лишь для тех многочленов из $\mathbb{R}[x]$, которые имеют полный набор действительных корней.

Глава XI. Алгебраические линии второго порядка на плоскости

§52. Эллипс

Эллипсом называется геометрическое место точек M плоскости, для которых сумма расстояний от двух фиксированных точек F_1 и F_2 плоскости есть постоянное число, большее, чем расстояние между F_1 и F_2. Это число мы обозначим через $2a$. Точки F_1, F_2 называются ***фокусами эллипса***, расстояние между ними называется ***фокусным расстоянием*** и обозначается через $2c$. Числа $r_1 = \rho(M, F_1)$ и $r_2 = \rho(M, F_2)$ называются ***фокальными радиусами*** точки M.

Из определения эллипса следует, что точка M плоскости является точкой эллипса тогда и только тогда, когда

$$r_1 + r_2 = 2a, \quad a > c. \tag{52.1}$$

Каноническое уравнение. Запишем условие (52.1) в координатной форме. Для этого введем на плоскости прямоугольную декартову систему координат Oxy, приняв за начало O середину отрезка F_1F_2, за ось Ox – прямую F_1F_2, ориентированную от F_1 к F_2; ориентацию на оси Oy выбираем произвольно (рис. 1). Эта система координат называется ***канонической системой координат*** (для данного эллипса).

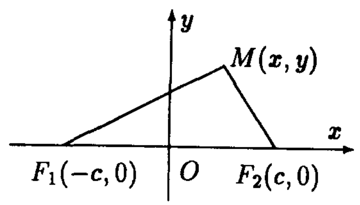


Рис. 1

Фокусы F_1 и F_2 в канонической системе координат имеют координаты $(-c, 0)$ и $(c, 0)$ соответственно.

Теорема 52.1. *Уравнение эллипса в канонической системе координат имеет вид*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad a \ge b > 0. \tag{52.2}$$

Доказательство. Пусть $M(x, y)$ – точка эллипса (52.1). Тогда условие (52.1) может быть записано через координаты точек F_1, F_2, M в виде

$$\sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2} = 2a, \quad a > c.$$

После переноса одного слагаемого в правую часть и возведения в квадрат получим

$$x^2 + 2cx + c^2 + y^2 = 4a^2 - 4a\sqrt{(x-c)^2 + y^2} + x^2 - 2cx + c^2 + y^2$$

или, после очевидных преобразований,

$$a\sqrt{(x-c)^2 + y^2} = a^2 - cx.$$

Еще раз возведем в квадрат: $a^2x^2 - 2a^2xc + a^2c^2 + a^2y^2 = a^4 - 2a^2cx + c^2x^2$ или $x^2(a^2 - c^2) + a^2y^2 = a^2(a^2 - c^2)$, где $a^2 - c^2 > 0$.

Разделив обе части уравнения на $a^2(a^2 - c^2)$, придем к уравнению

$$\frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1, \qquad \text{или} \qquad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

где $b^2 = a^2 - c^2$, $a \ge b > 0$. Таким образом, любая точка $M(x, y)$ эллипса удовлетворяет уравнению (52.2). Обратное, вообще говоря, не очевидно, так как при переходе от (52.1) к (52.2) мы дважды возводили в квадрат обе части уравнений и поэтому на линии, определяемой уравнением (52.2), могут быть и "лишние" точки. Покажем, что уравнению (52.2) отвечают только точки эллипса (52.1). Пусть точка $M(x, y)$ удовлетворяет уравнению (52.2), тогда

$$y^2 = b^2\left(1 - \frac{x^2}{a^2}\right) \quad \text{и} \quad |x| \le a. \tag{52.3}$$

Найдем фокальные радиусы точки M. Имеем

$$r_1 = \sqrt{(x+c)^2 + y^2} = \sqrt{x^2 + 2cx + c^2 + b^2 - \frac{b^2}{a^2}x^2} =$$

$$= \sqrt{\frac{x^2(a^2 - b^2)}{a^2} + 2cx + c^2 + b^2} = \sqrt{\frac{x^2c^2}{a^2} + 2cx + a^2} =$$

$$= \sqrt{\left(\frac{xc}{a} + a\right)^2} = \left|\frac{xc}{a} + a\right| = a + \frac{cx}{a},$$

так как $\frac{|cx|}{a} \le \frac{ca}{a} = c < a$ в силу (52.3). Итак,

$$r_1 = a + \frac{cx}{a}. \tag{52.4}$$

Аналогично доказывается, что

$$r_2 = a - \frac{cx}{a}. \tag{52.5}$$

Из (52.4) и (52.5) следует, что $r_1 + r_2 = 2a$. ■

Уравнение (52.2) называется ***каноническим уравнением эллипса.*** В случае когда $a = b$ (т.е. фокусы совпадают), эллипс превращается в окружность радиуса a с центром в точке $O = F_1 = F_2$.

Из канонического уравнения вытекают следующие свойства эллипса.

1°. Координатные оси канонической системы координат (рис. 2) являются осями симметрии эллипса, так как вместе с точкой $M(x, y)$ эллипса точки $M_1(-x, y)$ и $M_2(x, -y)$ также принадлежат эллипсу. Начало координат канонической системы координат является центром симметрии эллипса, так как точки $M(x, y)$ и $M_3(-x, -y)$ одновременно принадлежат эллипсу. Начало координат канонической системы координат называется ***центром*** эллипса, числа $2a$ и $2b$ – ***большой*** и ***малой осями*** эллипса, а числа a и b – ***большой*** и ***малой полуосями.***

2°. Все точки эллипса лежат в прямоугольнике, ограниченном прямыми $x = \pm a$ и $y = \pm b$, так как для координат (x, y) точек эллипса $|x| \le a$, $|y| \le b$. Точки $A_1(-a, 0)$, $A_2(a, 0)$, $B_1(0, -b)$, $B_2(0, b)$ пересечения эллипса с осями координат называются ***вершинами эллипса.***

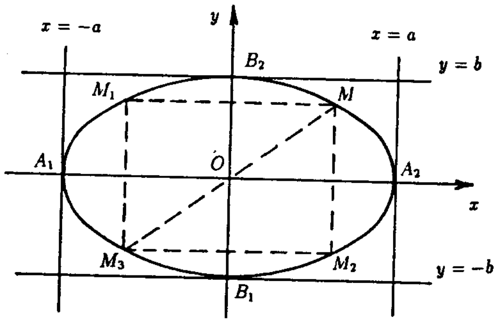


Рис. 2

3°. Исследуем форму эллипса. В силу симметрии достаточно провести это исследование лишь для точек I четверти. Здесь уравнение эллипса имеет вид $y = \dfrac{b}{a}\sqrt{a^2 - x^2}$, $x \in [0, a]$. Легко показать, что:

а) $y(x)$ непрерывна и убывает от $y = b$ до $y = 0$;

б) $y'(0) = 0$, $y'(a) = \infty$, следовательно, касательная к эллипсу в точке $x = 0$ параллельна оси Ox, а в точке $x = a$ – оси Oy;

в) $y''(x) < 0$, следовательно, линия выпукла вверх (рис. 2).

Простой *способ построения эллипса* дается определением эллипса: достаточно нить длиной $2a$ закрепить концами в точках F_1 и F_2, натянуть нить карандашом и передвигать карандаш, держа нить натянутой.

Директориальное свойство. Число $\varepsilon = c/a$ называется ***эксцентриситетом эллипса.*** Из определения следует, что $0 \le \varepsilon < 1$, при

этом

$$\varepsilon = \sqrt{1 - \frac{b^2}{a^2}} \quad \text{или} \quad \frac{b}{a} = \sqrt{1 - \varepsilon^2} \tag{52.6}$$

и, с учетом (52.4), (52.5),

$$r_1 = a + \varepsilon x, \quad r_2 = a - \varepsilon x. \tag{52.7}$$

Соотношения (52.6) означают, что чем меньше эксцентриситет, тем эллипс "ближе" к окружности (для окружности $\varepsilon = 0$).

Для эллипса, не являющегося окружностью, две прямые d_1 и d_2, заданные в канонической системе координат уравнениями

$$d_1: \ x = -\frac{a}{\varepsilon}, \qquad d_2: \ x = \frac{a}{\varepsilon},$$

называются ***директрисами эллипса*** (рис. 3).

Директриса d_i называется ***соответствующей фокусу*** F_i, $i = 1, 2$.

Теорема 52.2. ***Эллипс, не являющийся окружностью, есть геометрическое место точек M плоскости, для которых отношение расстояния от данной точки F к расстоянию до данной прямой d, не проходящей через эту точку, равно данному положительному числу, меньшему*** 1, ***т.е.***

$$\frac{\rho(M, F)}{\rho(M, d)} = \varepsilon, \quad 0 < \varepsilon < 1. \tag{52.8}$$

Доказательство. Пусть $\rho(F, d) = m$. Найдем число $a > 0$, для которого

$$m = \frac{a}{\varepsilon} - a\varepsilon. \tag{52.9}$$

Очевидно, $a = \dfrac{m\varepsilon}{1 - \varepsilon^2}$. Положим

$$c = a\varepsilon, \quad b^2 = a^2(1 - \varepsilon^2). \tag{52.10}$$

Введем прямоугольную декартову систему координат Oxy, принимая за ось Ox прямую, проходящую через точку F, перпендикулярную прямой d (рис. 4) и ориентированную от F к d; за начало O – точку на оси Ox, расположенную от точки F на расстоянии c по другую сторону от прямой d; ориентация на оси Oy выбирается произвольно.

В этой системе координат точка F имеет координаты $(c, 0)$, а прямая d определяется уравнением $x = a/\varepsilon$, так как $c + m = a\varepsilon + m = a/\varepsilon$ (рис. 4). Условие (52.8) для точки $M(x, y)$ означает, что

$$\sqrt{(x - c)^2 + y^2} = \left| x - \frac{a}{\varepsilon} \right| \varepsilon.$$

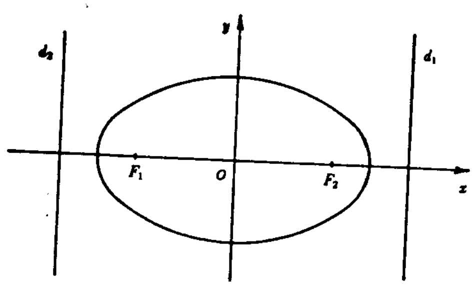


Рис. 3

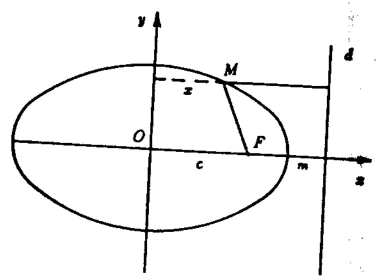


Рис. 4

Возведя обе части в квадрат, получим равносильное уравнение
$$x^2 - 2cx + c^2 + y^2 = x^2\varepsilon^2 - 2xa\varepsilon + a^2,$$
которое с учетом того, что $c = a\varepsilon$, может быть записано в виде
$$(1 - \varepsilon^2)x^2 + y^2 = a^2(1 - \varepsilon^2).$$
Поделив обе части этого уравнения на правую часть, приходим в обозначениях (52.10) к каноническому уравнению эллипса, эквивалентному (52.8):
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad a > b > 0.$$

Итак, точка $M(x, y)$ принадлежит геометрическому месту точек (52.8) тогда и только тогда, когда она является точкой эллипса, для которого данная точка F является фокусом, прямая d – соответствующей директрисой, а число ε – эксцентриситетом. Отметим, что доказательство теоремы как раз и состоит в построении этого эллипса. Его полуоси a и b определены равенствами (52.9), (52.10). ■

§53. Гипербола

Гиперболой называется геометрическое место точек M плоскости, для которых абсолютная величина разности расстояний от двух фиксированных точек F_1 и F_2 плоскости есть постоянное число, меньшее, чем расстояние между F_1 и F_2. Это число мы обозначим через $2a$. Точки F_1, F_2 называются *фокусами* гиперболы, расстояние между ними называется *фокальным расстоянием* и обозначается $2c$. Согласно определению $a < c$. Числа $r_i = \rho(M, F_i)$, $i = 1, 2$, называются *фокальными радиусами* точки M.

Из определения гиперболы следует, что точка M плоскости является точкой гиперболы тогда и только тогда, когда
$$|r_1 - r_2| = 2a, \quad a < c. \tag{53.1}$$

Каноническое уравнение. Запишем условие (53.1) в координатной форме. Поступая так же, как и в случае эллипса, построим каноническую систему координат Oxy для гиперболы (53.1).

Теорема 53.1. *Уравнение гиперболы в канонической системе координат имеет вид*
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{53.2}$$

Доказательство. Если $M(x, y)$ – точка гиперболы (53.1), то
$$\left|\sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2}\right| = 2a, \quad a < c,$$
или
$$\sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2} = \pm 2a, \quad a < c.$$
Используя те же преобразования, что и для эллипса, придем к уравнению
$$x^2(a^2 - c^2) + a^2y^2 = a^2(a^2 - c^2), \quad a^2 - c^2 < 0,$$
или, в обозначениях $-b^2 = a^2 - c^2$, к уравнению (53.2).

Таким образом, любая точка $M(x, y)$ гиперболы удовлетворяет уравнению (53.2). Обратное доказывается так же, как и для эллипса. Здесь $|x| \geq a$, поэтому
$$\begin{cases} r_1 = \dfrac{cx}{a} + a, \\ r_2 = \dfrac{cx}{a} - a, \end{cases} \text{если } x \geq a, \quad \begin{cases} r_1 = -\left(\dfrac{cx}{a} + a\right), \\ r_2 = -\left(\dfrac{cx}{a} - a\right), \end{cases} \text{если } x \leq -a. \tag{53.3}$$
Таким образом, в обоих случаях $|r_1 - r_2| = 2a$. ■

Уравнение (53.2) называется *каноническим уравнением гиперболы*. Отметим простейшие свойства гиперболы, вытекающие из канонического уравнения.

1°. Как и в случае эллипса, координатные оси канонической системы координат являются осями симметрии гиперболы, а начало координат – центром симметрии. Ось Ox пересекает гиперболу в точках $A_1(-a, 0)$, $A_2(a, 0)$ и называется *вещественной осью гиперболы*. Ось Oy не пересекает гиперболу и называется *мнимой осью*. Начало координат называется *центром гиперболы*, числа $a > 0$, $b > 0$ – *вещественной* и *мнимой полуосями гиперболы*.

2°. Все точки гиперболы лежат вне полосы, определяемой прямыми $x = \pm a$, так как $|x| \geq a$ (рис. 1). Это означает, что гипербола состоит из двух отдельных частей, называемых *ветвями гиперболы*.

3°. Все точки гиперболы лежат в тех вертикальных углах, образованных прямыми $y = \pm\frac{b}{a}x$, которые содержат вещественную ось, так как для точек $M(x, y)$ гиперболы $y = \pm\frac{b}{a}\sqrt{x^2 - a^2}$, т.е. $|y| \leq \frac{b}{a}|x|$.

4°. Исследуем форму гиперболы. В силу симметрии достаточно провести это исследование лишь в I четверти. Здесь уравнение гиперболы имеет вид $y = \frac{b}{a}\sqrt{x^2 - a^2}$, $x \geq a$. Легко проверить, что (рис. 1):

а) $y(x)$ непрерывна и возрастает на $[a, +\infty)$, при этом $y(a) = 0$ и $\lim\limits_{x\to+\infty} y(x) = +\infty$;

б) $y'(a) = \infty$, так что касательная к гиперболе в точке $x = a$ параллельна оси Oy;

в) $y''(x) < 0$, следовательно, линия выпукла вверх;

г) если l – прямая, заданная уравнением $y = \frac{b}{a}x$, то для всех точек $M(x, y)$ гиперболы $\lim\limits_{x\to+\infty} \rho(M, l) = 0$, так как

$$\rho(M,l) = \frac{|bx/a - y|}{\sqrt{1 + b^2/a^2}} = \frac{|bx - ay|}{\sqrt{a^2+b^2}} = \frac{|bx - b\sqrt{x^2 - a^2}|}{\sqrt{a^2+b^2}} = \\ = \frac{a^2 b}{\sqrt{a^2+b^2}(x + \sqrt{x^2 - a^2})} \to 0 \quad \text{при } x \to +\infty.$$

Итак, все точки гиперболы с ростом x сколь угодно **близко приближаются** к прямым

$$l_1 : \ y = -\tfrac{b}{a}x \quad \text{и} \quad l_2 : \ y = \tfrac{b}{a}x,$$

не пересекая их. Прямые l_1, l_2 называются ***асимптотами гиперболы.***

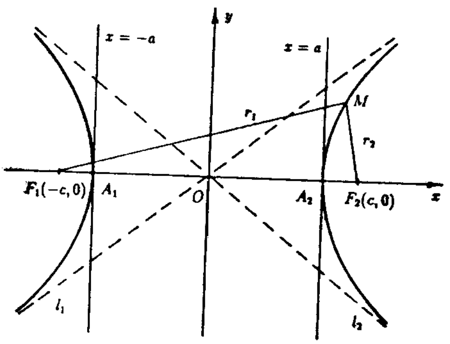


Рис. 1

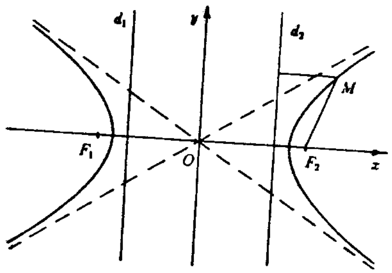


Рис. 2

Директориальное свойство. Число $\varepsilon = c/a$ называется *эксцентриситетом гиперболы*. Из определения следует, что $\varepsilon > 1$, при этом

$$\varepsilon = \sqrt{1 + \frac{b^2}{a^2}}\,, \quad \frac{b}{a} = \sqrt{\varepsilon^2 - 1}$$

и, с учетом (53.3),

$$\begin{cases} r_1 = -(\varepsilon x + a), \\ r_2 = -(\varepsilon x - a) \end{cases} \quad \text{и} \quad \begin{cases} r_1 = \varepsilon x + a, \\ r_2 = \varepsilon x - a \end{cases}$$

для левой и правой ветвей гиперболы соответственно.

Прямые d_1 и d_2, заданные в канонической системе координат уравнениями

$$d_1 : \ x = -\frac{a}{\varepsilon}, \qquad d_2 : \ x = \frac{a}{\varepsilon},$$

называются *директрисами гиперболы* (рис. 2).

Директриса d_i называется *соответствующей* фокусу F_i, $i = 1, 2$.

Теорема 53.2. ***Гипербола есть геометрическое место точек M плоскости, для которых отношение расстояния от данной точки F к расстоянию до данной прямой d, не проходящей через эту точку, равно данному числу $\varepsilon > 1$, т.е.***

$$\frac{\rho(M,F)}{\rho(M,d)} = \varepsilon, \quad \varepsilon > 1.$$

Доказательство. Так же как и в случае эллипса (теорема 52.2), для данных точки F, прямой d и числа $\varepsilon > 1$ строится гипербола, для которой точка F является фокусом, прямая d – соответствующей директрисой, число ε – эксцентриситетом. Если $m = \rho(F, d)$, то полуоси гиперболы определяются соотношениями

$$m = a\varepsilon - \frac{a}{\varepsilon}, \quad c = a\varepsilon \tag{53.4}$$

или $a = \dfrac{m\varepsilon}{\varepsilon^2 - 1}$, $c = a\varepsilon$, $b^2 = a^2(\varepsilon^2 - 1)$. Использование тех же преобразований, что и при доказательстве теоремы 52.2, завершает доказательство. ■

§54. Парабола

Параболой называется геометрическое место точек плоскости, для которых расстояние от некоторой фиксированной точки F плоскости равно расстоянию до некоторой фиксированной прямой d, не проходящей через точку F. Точка F называется *фокусом параболы*, прямая d – ее *директрисой*. Расстояние от фокуса параболы до ее директрисы называется ***фокальным параметром*** параболы и обозначается через p. *Эксцентриситет параболы* по определению считается равным 1.

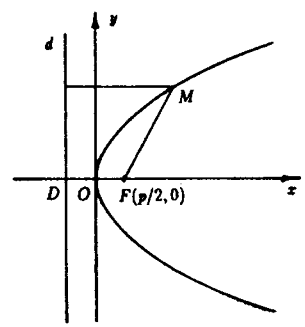


Рис. 1

Каноническое уравнение. Введем на плоскости прямоугольную декартову систему координат Oxy, принимая за ось Ox прямую, проходящую через точку F перпендикулярно прямой d, ориентированную от прямой d к точке F, за начало O – середину отрезка FD, где D – точка пересечения оси Ox с прямой d; ориентацию на оси Oy выбираем произвольно (рис. 1). Эта система координат называется ***канонической системой координат*** для данной параболы.

Теорема 54.1. *Уравнение параболы в канонической системе координат имеет вид*

$$y^2 = 2px, \quad p > 0. \tag{54.1}$$

Доказательство. В канонической системе координат Oxy фокус F имеет координаты $(p/2, 0)$, директриса d определяется уравнением $x = -p/2$. Точка $M(x,y)$ является точкой параболы тогда и только тогда, когда $\rho(M,F) = \rho(M,d)$, или, что то же самое,

$$\sqrt{(x-\frac{p}{2})^2 + y^2} = \left|x + \frac{p}{2}\right|.$$

Возведя в квадрат обе части уравнения, получим равносильное уравнение $x^2 - px + p^2/4 + y^2 = x^2 + px + p^2/4$ или уравнение (54.1). ■

Уравнение (54.1) называется ***каноническим уравнением параболы.***

Следующие свойства параболы (рис. 1) непосредственно вытекают из канонического уравнения.

1°. Ось Ox канонической системы координат является осью симметрии параболы. Она называется ***осью параболы.*** Начало координат лежит на параболе и называется ***вершиной параболы.***

2°. Все точки параболы расположены в правой полуплоскости от оси Oy, так как $x \geq 0$.

3°. Исследуем форму параболы. В силу симметрии достаточно рассмотреть I четверть. Здесь уравнение параболы имеет вид

$$y = \sqrt{2px}, \ x \geq 0.$$

Легко проверить, что:

а) $y(x)$ непрерывна и возрастает на $[0, +\infty)$, причем $y(0) = 0$ и $\lim\limits_{x\to+\infty} y(x) = +\infty$;

б) $y'(0) = \infty$, так что касательная к параболе в точке $x = 0$ параллельна оси Oy;

в) $y''(0) < 0$, следовательно, парабола выпукла вверх (рис. 1).

Замечание. Определение параболы, по существу, означает ее директориальное свойство

$$\frac{\rho(M,F)}{\rho(M,d)} = \varepsilon, \tag{54.2}$$

где F – фокус, d – директриса, ε – эксцентриситет, которое, как мы видели в §52, 53, имеет место для эллипсов и гипербол. Отличие лишь

в том, что для эллипсов $0 < \varepsilon < 1$, для парабол $\varepsilon = 1$, для гипербол $\varepsilon > 1$. Итак, если на плоскости даны прямая d и не принадлежащая ей точка F, то геометрическое место точек M плоскости, удовлетворяющих условию (54.2), является либо эллипсом (если $0 < \varepsilon < 1$), либо параболой (если $\varepsilon = 1$), либо гиперболой (если $\varepsilon > 1$).

Эти линии обладают целым рядом других свойств, имеющих "родственные" формулировки.

§55. Касательные к эллипсу, гиперболе и параболе

Будем считать, что эллипс, гипербола и парабола заданы своими каноническими уравнениями (52.2), (53.2), (54.1). Найдем уравнения касательных к этим линиям.

Теорема 55.1. *В канонической системе координат уравнения касательных к эллипсу, гиперболе и параболе в точке* (x_0, y_0) *линии имеют вид*

$$\frac{xx_0}{a^2} + \frac{yy_0}{b^2} = 1, \tag{55.1}$$

$$\frac{xx_0}{a^2} - \frac{yy_0}{b^2} = 1, \tag{55.2}$$

$$yy_0 = p(x + x_0) \tag{55.3}$$

соответственно.

Доказательство. Все три уравнения выводятся по одной схеме. Приведем ее для эллипса. Воспользуемся известным из курса математического анализа [7] уравнением касательной в неособой точке (x_0, y_0) линии, заданной неявно уравнением $F(x,y) = 0$. Это уравнение имеет вид

$$F'_x(x_0,y_0)(x - x_0) + F'_y(x_0,y_0)(y - y_0) = 0, \tag{55.4}$$

где $F'_x(x_0,y_0)$, $F'_y(x_0,y_0)$ – частные производные функции $F(x,y)$ в точке (x_0,y_0). Для эллипса (52.2) уравнение касательной в точке (x_0,y_0) этого эллипса имеет согласно (55.4) вид

$$\frac{x_0}{a^2}(x - x_0) + \frac{y_0}{b^2}(y - y_0) = 0$$

или, с учетом того, что $\frac{x_0^2}{a^2} + \frac{y_0^2}{b^2} = 1$, вид (55.1). ■

§56. Оптические свойства эллипса, гиперболы и параболы

Здесь речь идет о биссекториальных свойствах касательных к этим линиям.

Теорема 56.1. *Касательная к эллипсу в произвольной точке M_0 эллипса есть биссектриса внешнего угла M_0 треугольника $F_1F_2M_0$, где F_1, F_2 – фокусы эллипса.*

Доказательство. Рассмотрим касательную (55.1) к эллипсу (52.2) в точке $M_0(x_0, y_0)$ (рис. 1).

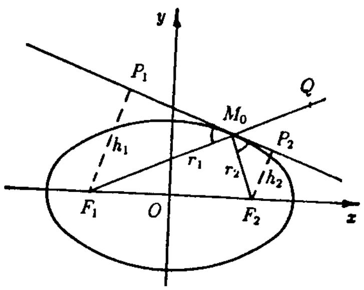


Рис. 1

Фокусы $F_1(-c, 0)$, $F_2(c, 0)$ расположены по одну сторону от касательной, так как в (34.2), с учетом условий $0 \le \varepsilon < 1$ и $|x_0| \le a$, имеем

$$\frac{-cx_0}{a^2} - 1 = -\frac{\varepsilon x_0}{a} - 1 = -\frac{\varepsilon x_0 + a}{a} < 0,$$
$$\frac{cx_0}{a^2} - 1 = \frac{\varepsilon x_0}{a} - 1 = \frac{\varepsilon x_0 - a}{a} < 0.$$

Обозначим через P_1 и P_2 основания перпендикуляров, опущенных из фокусов эллипса на касательную. Утверждение теоремы будет доказано, если мы покажем, что $\angle F_2M_0P_2 = \angle P_2M_0Q$ или, что то же самое, $\angle F_1M_0P_1 = \angle F_2M_0P_2$ (рис. 1). Пусть h_1, h_2 – расстояния от фокусов до касательной. Тогда

$$\frac{h_1}{h_2} = \frac{-(cx_0/a^2) - 1}{(cx_0/a^2) - 1} = \frac{a + \varepsilon x_0}{a - \varepsilon x_0} = \frac{r_1}{r_2},$$

следовательно, $\Delta F_2M_0P_2 \sim \Delta F_1M_0P_1$ (так как оба они прямоугольные), поэтому $\angle F_1M_0P_1 = \angle F_2M_0P_2$. ■

Доказанной теореме можно дать следующую оптическую интерпретацию: если поместить в один из фокусов источник света, то лучи после отражения от эллипса соберутся в другом фокусе, так как световой луч отразится от эллипса как от касательной, проведенной к эллипсу в точке падения луча.

Теорема 56.2. *Касательная к гиперболе в произвольной ее точке является биссектрисой внутреннего угла M_0 треугольника $F_1F_2M_0$, где F_1, F_2 – фокусы гиперболы.*

Доказательство. Эта теорема доказывается точно так же, как и предыдущая. Отметим лишь одно отличие (не затрагивая очевидные): фокусы F_1 и F_2 расположены по разные стороны от любой касательной (рис. 2). ■

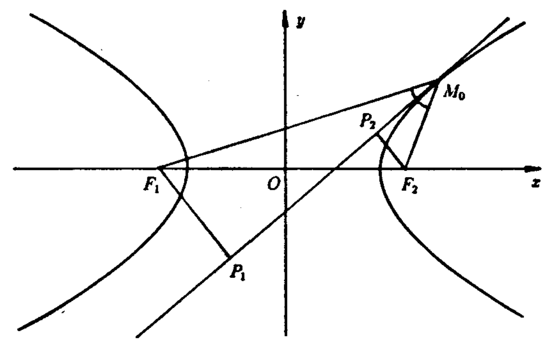


Рис. 2

Теореме 56.2 можно дать оптическое истолкование, аналогичное тому, которое было дано для эллипса.

Теорема 56.3. *Касательная к параболе есть биссектриса угла между фокальным радиусом M_0F точки касания M_0 и перпендикуляром M_0D, опущенным из точки M_0 на директрису.*

Доказательство. Из уравнения (55.3) касательной к параболе $y^2 = 2px$ в точке $M_0(x_0, y_0)$ этой параболы следует, что точка A (рис. 3) пересечения касательной с осью Ox имеет кординаты $(-x_0, 0)$. Следовательно, $AO = x_0$ и $AF = AO + OF = x_0 + p/2$. С другой стороны, $M_0F = M_0D = M_0C + CD = x_0 + p/2$. Таким образом, $AF = M_0F$ и, значит, $\angle FAM_0 = \angle AM_0F$. Но $\angle FAM_0 = \angle DM_0A$, следовательно, $\angle DM_0A = \angle AM_0F$. ■

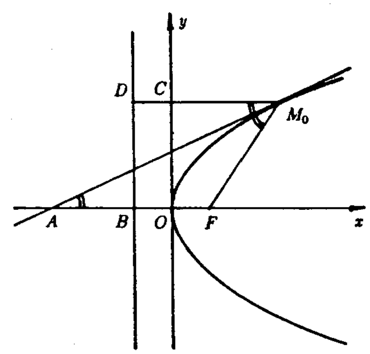


Рис. 3

Эта теорема имеет следующее оптическое истолкование: если в фокусе параболического зеркала поместить источник света, то лучи, отразившись от зеркала, образуют пучок параллельных лучей. Это свойство используется в конструкции зеркальных прожекторов.

§57. Полярные уравнения эллипса, гиперболы и параболы

Пусть $\mathcal{L}$ – какая-нибудь из трех линий: эллипс, гипербола или парабола, заданная своим каноническим уравнением.

Введем полярную систему координат, приняв за полюс фокус F линии (для эллипса F – левый фокус, для правой ветви гиперболы F – правый фокус, а для левой ветви – левый фокус), полярную ось направим в положительную сторону оси Ox канонической системы координат Oxy во всех случаях, кроме левой ветви гиперболы, а в этом случае – в отрицательную сторону оси Ox (рис. 1).

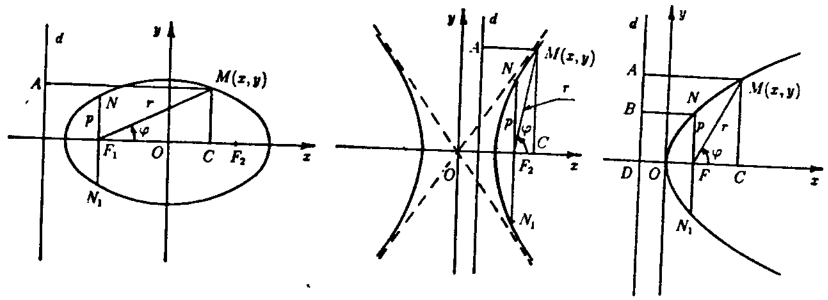


Рис. 1

Тогда для любой точки M плоскости ее полярный радиус r совпадает с фокальным радиусом $\rho(M,F)$. Согласно директориальному свойству линии $\mathcal{L}$ точка $M(x,y)$ принадлежит $\mathcal{L}$ тогда и только тогда, когда

$$\frac{r}{\rho(M,d)} = \varepsilon, \tag{57.1}$$

где ε – эксцентриситет линии $\mathcal{L}$. Пусть $\rho(F,d) = m$, тогда $\rho(M,d) = MA = r\cos\varphi + m$ и уравнение (57.1) линии может быть записано в виде

$$r = (r\cos\varphi + m)\varepsilon$$

или

$$r = \frac{m\varepsilon}{1-\varepsilon\cos\varphi}. \tag{57.2}$$

Для параболы $\varepsilon = 1$, $m = p$, где p – фокальный параметр параболы. Следовательно, в выбранной полярной системе координат парабола определяется уравнением

$$r = \frac{p}{1-\cos\varphi}. \tag{57.3}$$

Для эллипса и гиперболы число $p = b^2/a$ называется *фокальным параметром*. Из (52.9) и (53.4) следует, что число $m\varepsilon$ в (57.2) совпадает с фокальным параметром p, так как

$$m\varepsilon = a - a\varepsilon^2 = a\left(1 - \frac{c^2}{a^2}\right) = \frac{ab^2}{a^2} = \frac{b^2}{a} = p \quad \text{для эллипса и}$$

$$m\varepsilon = a\varepsilon^2 - a = a\left(\frac{c^2}{a^2} - 1\right) = \frac{ab^2}{a^2} = \frac{b^2}{a} = p \quad \text{для гиперболы.}$$

Это позволяет записать уравнение (57.2) в виде

$$\cdot - \frac{p}{1-\varepsilon\cos\varphi}, \tag{57.4}$$

где p – фокальный параметр линии, а ε – ее эксцентриситет . Поскольку для параболы $\varepsilon = 1$, то уравнение (57.4) совпадает с уравнением (57.3) параболы. Таким образом, эллипс, гипербола и парабола описываются в полярной системе координат единым уравнением (57.4).

Замечание. Проведем из фокуса F (рис. 1) прямую, перпендикулярную оси Ox. Эта прямая пересечет линию $\mathcal{L}$ в двух точках N и N_1. Для точки N, как и для любой точки $\mathcal{L}$, справедливо соотношение (57.1), так что

$$\frac{\rho(N,F)}{\rho(N,d)} = \varepsilon.$$

Но $\rho(N,d) = \rho(F,d) = m$, следовательно, $\rho(N,F) = m\varepsilon = p$. Таким образом, фокальный параметр линии $\mathcal{L}$ совпадает с половиной длины хорды NN_1.

§58. Общее уравнение линии второго порядка

Пусть Oxy – аффинная система координат на плоскости. Алгебраическая линия второго порядка определяется уравнением

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

где $F(x,y)$ – алгебраический многочлен второй степени от переменных x, y с вещественными коэффициентами (§31). Этот многочлен принято записывать в виде

$$F(x,y) = a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33},$$

где $a_{11}^2 + a_{12}^2 + a_{22}^2 \neq 0$. В соответствии с этим алгебраическая линия второго порядка определяется уравнением

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0, \quad a_{11}^2 + a_{12}^2 + a_{22}^2 \neq 0. \tag{58.1}$$

Уравнение (58.1) называется *общим уравнением алгебраической линии второго порядка на плоскости.* Группа слагаемых $a_{11}x^2 + 2a_{12}xy + a_{22}y^2$ (т.е. одночленов второй степени) называется *квадратичной частью* уравнения (58.1) (или *группой старших членов*), группа слагаемых $2a_{13}x + 2a_{23}y$ – ***линейной частью***, a_{33} – ***свободным членом***.

Рассмотренные в предыдущих параграфах эллипс, гипербола и парабола, очевидно, представляют собой алгебраические линии второго порядка. Естественно поставить вопрос, какие еще линии являются линиями второго порядка. Чтобы ответить на него, нужно найти такую систему координат, в которой уравнение (58.1) принимает наиболее простой вид (как это было в случае эллипса, гиперболы и параболы). Поиском такой системы координат мы и займемся.

Компактная запись общего уравнения. Положим

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}; \quad b = \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}; \quad X = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Матрица A называется ***матрицей квадратичной части***. В этих обозначениях уравнение (58.1) может быть записано в компактной форме:

$$X^T AX + 2b^T X + a_{33} = 0, \quad A = A^T, \ A \neq O. \tag{58.2}$$

В этом нетрудно убедиться, выполнив все умножения в левой части (58.2). Введем новую матрицу

$$B = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{21} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} = \left[\begin{array}{c|c} A & b \\ \hline b^T & a_{33} \end{array}\right].$$

Числа $I_1 = \operatorname{tr} A$, $I_2 = |A|$, $K_3 = |B|$ называются ***инвариантами линии второго порядка***, число $K_2 = \begin{vmatrix} a_{11} & a_{13} \\ a_{13} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{23} & a_{33} \end{vmatrix}$ – ***полуинвариантом***. Далее нам потребуется несколько дополнительных понятий, относящихся к матрицам.

Характеристический многочлен. *Характеристическим многочленом* матрицы $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ называется функция $f(\lambda)$, определенная равенством

$$f(\lambda) = |A - \lambda I|. \tag{58.3}$$

Легко проверить, что:

а) если $n = 2$, то $f(\lambda) = (-\lambda)^2 + a_1(-\lambda) + a_0$, где

$$a_1 = \operatorname{tr} A, \quad a_0 = |A|; \tag{58.4}$$

б) если $n = 3$, то $f(\lambda) = (-\lambda)^3 + a_2(-\lambda)^2 + a_1(-\lambda) + a_0$, где

$$a_2 = \operatorname{tr} A, \quad a_1 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \quad a_0 = |A|. \tag{58.5}$$

Главным минором матрицы называется минор, расположенный в строках и столбцах с одинаковыми номерами. Нетрудно заметить (см. (58.4), (58.5)), что коэффициенты характеристического многочлена связаны с главными минорами матрицы: если $n = 2$, то a_1 – сумма главных миноров первого порядка, a_0 – единственный главный минор второго порядка; если $n = 3$, то a_2, a_1, a_0 – суммы главных миноров первого, второго, третьего порядков соответственно.

Матрицы $A, B \in \mathbb{R}^{n\times n}$ называются ***подобными***, если существует невырожденная матрица Q такая, что

$$A = Q^{-1}BQ. \tag{58.6}$$

Теорема 58.1. *Характеристические многочлены подобных матриц совпадают.*

Доказательство. Действительно, если $A = Q^{-1}BQ$, то $|A - \lambda I| = |Q^{-1}BQ - \lambda I| = |Q^{-1}BQ - Q^{-1}\lambda IQ| = |Q^{-1}(B - \lambda I)Q| = |B - \lambda I|$. Таким образом, при любом значении λ характеристические многочлены матриц A и B принимают одинаковые значения и, следовательно, совпадают. ■

Следствие. У подобных матриц второго порядка совпадают следы и определители. У подобных матриц третьего порядка совпадают следы, суммы главных миноров второго порядка и определители.

Преобразования общего уравнения. Пусть исходная аффинная система координат Oxy соответствует началу O и базису $e = (e_1, e_2)$. Переход к новой системе координат $O'x'y'$ означает (§24) перенос начала в точку $O'(\alpha, \beta)$ и преобразование базиса $eQ = e'$ с матрицей перехода Q. При этом старые координаты $X = (x, y)^T$ связаны с новыми $X' = (x', y')^T$ формулами преобразования координат:

1) $X = a + X'$, где $a = (\alpha, \beta)^T$, в случае переноса начала;
2) $X = QX'$ в случае преобразования базиса.

Исследуем особенности преобразования уравнения линии в каждом из этих случаев. Пусть линия $\mathcal{L}$ в системе координат Oxy задана своим общим уравнением (58.2).

Теорема 58.2. *При переходе к новому базису $e' = eQ$ общее уравнение (58.2) преобразуется в уравнение*

$$X'^T A' X' + 2b'^T X' + a_{33} = 0, \tag{58.7}$$

где $A' = Q^T AQ$, $b' = Q^T b$, при этом:

1) знаки инвариантов I_2, K_3 не изменяются;

2) в случае когда e и e' – ортонормированные базисы, инварианты I_1, I_2, K_3 и полуинвариант K_2 не изменяются.

Доказательство. Подставим в уравнение (58.2) вместо старых координат X их выражения через новые координаты X': $X = QX'$. Тогда в новой системе координат $Ox'y'$ линия $\mathcal{L}$ определяется уравнением

$$X'^T Q^T AQX' + 2b^T QX' + a_{33} = 0$$

или

$$X'^T(Q^TAQ)X' + 2(Q^Tb)^TX' + a_{33} = 0.$$

Это означает, что в новом уравнении квадратичная часть определяется матрицей

$$A' = Q^TAQ, \tag{58.8}$$

а линейная часть – столбцом $b' = Q^Tb$. Таким образом, общее уравнение (58.2) линии $\mathcal{L}$ преобразуется в (58.7).

Докажем п.1. Из (58.8) следует, что $|A'| = |Q^TAQ| = |A|\cdot|Q|^2$, т.е. $I_2' = I_2|Q|^2$, где $|Q| \neq 0$. Следовательно, $\operatorname{sgn} I_2' = \operatorname{sgn} I_2$. Далее, если

$$\widetilde{Q} = \left[\begin{array}{c|c} Q & O \\ \hline O & 1 \end{array}\right], \text{ то } \widetilde{Q}^T = \left[\begin{array}{c|c} Q^T & O \\ \hline O & 1 \end{array}\right] \text{ и } |\widetilde{Q}| = |\widetilde{Q}^T| = |Q|.$$

Матрица B', соответствующая уравнению (58.7), имеет вид

$$B' = \left[\begin{array}{c|c} A' & b' \\ \hline b'^T & a_{33} \end{array}\right] = \left[\begin{array}{c|c} Q^TAQ & Q^Tb \\ \hline b^TQ & a_{33} \end{array}\right] = \widetilde{Q}^TB\widetilde{Q}.$$

Следовательно, $|B'| = |B|\cdot|\widetilde{Q}|^2 = |B|\cdot|Q|^2$ и $K_3' = K_3|Q|^2$. Таким образом, $\operatorname{sgn} K_3' = \operatorname{sgn} K_3$.

Докажем п.2. Если оба базиса e и e' ортонормированы, то матрица перехода Q будет ортогональной матрицей (§24) и $Q^T = Q^{-1}$. При этом матрица $\widetilde{Q}$ также будет ортогональной, ибо $\widetilde{Q}^T\widetilde{Q} = \widetilde{Q}\widetilde{Q}^T = I$, и, следовательно, $\widetilde{Q}^T = \widetilde{Q}^{-1}$. Таким образом, $A' = Q^{-1}AQ$ и $B' = \widetilde{Q}^{-1}B\widetilde{Q}$, т.е. пары матриц A' и A, B' и B подобны. Отсюда и из теоремы 58.1 (и ее следствия) вытекает, что числа $I_1 = \operatorname{tr} A$, $I_2 = |A|$, $K_3 = |B|$, $K_2 = a_1 - I_2$ (где a_1 – сумма главных миноров второго порядка матрицы B) при переходе к новому базису не изменяются. ■

Замечание. Отметим, что при переходе к новому базису свободный член не изменяется.

Теорема 58.3. *При переносе начала координат в точку $O'(\alpha,\beta)$ общее уравнение (58.2) преобразуется в уравнение*

$$X'^TAX' + 2b'^TX' + a_{33}' = 0, \tag{58.9}$$

где $b' = b + Aa$, $a_{33}' = a^TAa + 2b^Ta + a_{33}$, $a = (\alpha,\beta)^T$, при этом инварианты I_1, I_2, K_3 не изменяются.

Доказательство. Подставим в уравнение (58.2) $X = X' + a$. Тогда $(X'^T + a^T)A(X' + a) + 2b^T(X' + a) + a_{33} = 0$, или

$$X'^TAX' + X'^TAa + a^TAX' + a^TAa + 2b^TX' + 2b^Ta + a_{33} = 0. \tag{58.10}$$

Заметим, что произведение X'^TAa есть вещественное число и его можно заменить результатом транспонирования, так что $X'^TAa = (X'^TAa)^T = a^TA^TX' = (Aa)^TX'$. Так как $A = A^T$, то $a^TAX' = a^TA^TX' = (Aa)^TX'$. С учетом этих соотношений уравнение (58.10) может быть записано в виде

$$X'^TAX' + 2(Aa + b)^TX' + a^TAa + 2b^Ta + a_{33} = 0.$$

Это означает, что в новом уравнении квадратичная часть остается прежней, линейная часть определяется столбцом $b' = Aa + b$, свободный член a_{33}' равен $a^TAa + 2b^Ta + a_{33}$. Таким образом, уравнение (58.2) линии $\mathcal{L}$ преобразуется в (58.9).

Что касается инвариантов, то неизменность I_1, I_2 очевидна, так как матрица квадратичной части A осталась прежней. Докажем, что K_3 не изменяется. Имеем

$$B = \left[\begin{array}{cc|c} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ \hline a_{13} & a_{23} & a_{33} \end{array}\right],$$

$$B' = \left[\begin{array}{cc|c} a_{11} & a_{12} & a_{13} + \alpha a_{11} + \beta a_{12} \\ a_{12} & a_{22} & a_{23} + \alpha a_{12} + \beta a_{22} \\ \hline a_{13} + \alpha a_{11} + \beta a_{12} & a_{23} + \alpha a_{12} + \beta a_{22} & a_{33}' \end{array}\right],$$

где $a_{33}' = a_{11}\alpha^2 + 2a_{12}\alpha\beta + a_{22}\beta^2 + 2a_{13}\alpha + 2a_{23}\beta + a_{33} = a_{33} + \alpha a_{13} + \beta a_{23} + \alpha(a_{13} + \alpha a_{11} + \beta a_{12}) + \beta(a_{23} + \alpha a_{12} + \beta a_{22})$. Последнее соотношение означает, что матрица B' получена из матрицы B с помощью элементарных преобразований: если к третьей строке матрицы B прибавить линейную комбинацию первых строк с коэффициентами α, β, а затем к третьему столбцу прибавить такую же линейную комбинацию первых столбцов, то получится матрица B'. Следовательно, $K_3' = |B'| = |B| = K_3$. ■

Теорема 58.4. *Общее уравнение линии второго порядка, заданное в прямоугольной декартовой системе координат, переходом к другой прямоугольной декартовой системе координат приводится к одному и только одному из следующих типов уравнений:*

$$\begin{array}{ll} \text{I.} & \lambda_1x^2 + \lambda_2y^2 + a_0 = 0, \quad \textit{где} \quad \lambda_1\lambda_2 \neq 0; \\ \text{II.} & \lambda_2y^2 + 2b_0x = 0, \quad \textit{где} \quad \lambda_2b_0 \neq 0; \\ \text{III.} & \lambda_2y^2 + c_0 = 0, \quad \textit{где} \quad \lambda_2 \neq 0. \end{array} \tag{58.11}$$

Доказательство. Пусть Oxy – прямоугольная декартова система координат и линия $\mathcal{L}$ задана в этой системе координат общим уравнением (58.1).

Шаг 1 (преобразование базиса). **Метод вращений.** Покажем, что если $a_{12} \neq 0$, то поворотом осей можно привести квадратичную часть уравнения (58.1) к сумме квадратов. Действительно, поворот осей на угол φ приводит к новому базису $e' = eQ$ с матрицей перехода (§24) $Q = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$. Очевидно, $Q^TQ = QQ^T = I$, т.е. Q – ортогональная матрица. Согласно теореме 58.2 при переходе к системе координат $\{O; \mathbf{e}'_1, \mathbf{e}'_2\}$ матрица квадратичной части A преобразуется в матрицу

$$A' = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix},$$

при этом

$$a'_{12} = -a_{11}\cos\varphi\sin\varphi - a_{12}\sin^2\varphi + a_{12}\cos^2\varphi + a_{22}\cos\varphi\sin\varphi = \\ = a_{12}\cos 2\varphi - \frac{1}{2}(a_{11} - a_{22})\sin 2\varphi.$$

Если $\operatorname{ctg} 2\varphi = (a_{11} - a_{22})/(2a_{12})$, то $a'_{12} = 0$. Следовательно, при повороте осей на такой угол φ квадратичная часть уравнения преобразуется в сумму квадратов и уравнение (58.1) в новой системе координат $Ox'y'$ будет иметь вид

$$a'_{11}x'^2 + a'_{22}y'^2 + 2a'_{13}x' + 2a'_{23}y' + a_{33} = 0. \tag{58.12}$$

При этом в силу теоремы 58.2 инварианты I_1, I_2, K_3 и полуинвариант K_2 останутся прежними.

Метод, использованный здесь, называется *методом вращений*.

Шаг 2 (перенос начала). Дальнейшее упрощение уравнения (58.12) основано на том, что если в нем содержится ненулевой квадрат какой-либо переменной, то переносом начала можно освободиться от этой переменной в первой степени. Действительно, если $a'_{11} \neq 0$, то

$$a'_{11}x'^2 + 2a'_{13}x' = a'_{11}\left(x'^2 + \frac{2a'_{13}}{a'_{11}}x'\right) = \\ = a'_{11}\left(x' + \frac{a'_{13}}{a'_{11}}\right)^2 - \frac{{a'_{13}}^2}{a'_{11}} = \left\{x'' = x' + \frac{a'_{13}}{a'_{11}}\right\} = a'_{11}x''^2 - \frac{{a'_{13}}^2}{a'_{11}}.$$

Таким образом, если $a'_{11} \neq 0$, $a'_{22} \neq 0$, то переносом начала

$$x'' = x' + \frac{a'_{13}}{a'_{11}}, \quad y'' = y' + \frac{a'_{23}}{a'_{22}}$$

уравнение (58.12) преобразуется в уравнение

$$a'_{11}x''^2 + a'_{22}y''^2 + a'_{33} = 0, \quad a'_{11}a'_{22} \neq 0,$$

где $a'_{33} = a_{33} - \dfrac{{a'_{13}}^2}{a'_{11}} - \dfrac{{a'_{23}}^2}{a'_{22}}$, т.е. в уравнение типа I.

Пусть один из коэффициентов a'_{11} и a'_{22} равен нулю. **Если** $a'_{11} = 0$, $a'_{22} \neq 0$, то переносом начала

$$x'' = x', \quad y'' = y' + \frac{a'_{23}}{a'_{22}}$$

уравнение (58.12) преобразуется в уравнение

$$a'_{22}y''^2 + 2a'_{13}x'' + a'_{33} = 0, \quad a'_{22} \neq 0, \tag{58.13}$$

где $a'_{33} = a_{33} - {a'_{23}}^2/a'_{22}$.

Случай, когда $a'_{11} \neq 0$, $a'_{22} = 0$, сводится к **предыдущему** переименованием переменных

$$x'' = y', \quad y'' = x', \tag{58.14}$$

что соответствует переходу к новому базису с матрицей перехода $Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Нетрудно проверить, что Q – ортогональная матрица, поэтому числа I_1, I_2, K_3, K_2 при таком переходе не изменяются. Очевидно, что новая система координат получена поворотом осей с последующим отражением одной из осей относительно другой.

Итак, дальнейшее преобразование общего уравнения (58.1) сводится к преобразованию уравнения (58.13).

Если в этом уравнении $a'_{13} = 0$, то уравнение (58.13) относится к уравнению типа III.

Если же $a'_{13} \neq 0$, то $2a'_{13}x'' + a'_{33} = 2a'_{13}(x'' + a'_{33}/(2a'_{13}))$ и переносом начала

$$x''' = x'' + a'_{33}(2a'_{13}), \quad y''' = y''$$

уравнение (58.13) приводится к уравнению

$$a'_{22}y'''^2 + 2a'_{13}x''' = 0, \quad a'_{22}a'_{13} \neq 0,$$

которое относится к типу II.

Отметим, что все промежуточные и окончательная системы координат оставались прямоугольными, так как преобразования базиса с помощью ортогональной матрицы перехода сохраняют свойство ортонормированности (§24). Итак, переходом к новой прямоугольной системе координат общее уравнение (58.1) приводится к одному из трех указанных типов уравнений.

Перейдем к вопросу о единственности. Для этого найдем инварианты I_2, K_3 для каждого из уравнений (58.11). Имеем для уравнения типа I

$$A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad B = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & a_0 \end{bmatrix}; \tag{58.15}$$

для уравнения типа II

$$A = \begin{bmatrix} 0 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & b_0 \\ 0 & \lambda_2 & 0 \\ b_0 & 0 & 0 \end{bmatrix}; \tag{58.16}$$

для уравнения типа III

$$A = \begin{bmatrix} 0 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & c_0 \end{bmatrix}. \qquad \textbf{(58.17)}$$

Следовательно,
1) $I_2 \neq 0$ для уравнения типа I;
2) $I_2 = 0$, $K_3 \neq 0$ для уравнения типа II;
3) $I_2 = 0$, $K_3 = 0$ для уравнения типа III.

Эти условия взаимно исключают друг друга, и так как общее уравнение и уравнения (58.11) имеют одинаковые инварианты I_2, K_3, то общее уравнение (58.1) приводится только к одному из трех указанных типов уравнений. ■

Уравнения (58.11) называются ***приведенными уравнениями линии второго порядка.***

Замечание. Особо отметим, что в ***прямоугольных*** координатах коэффициенты λ_1, λ_2 приведенных уравнений являются инвариантами линии, так как

$$\lambda_1 + \lambda_2 = I_1, \quad \lambda_1\lambda_2 = I_2 \qquad (58.18)$$

и, следовательно, λ_1 и λ_2 являются корнями характеристического многочлена матрицы A

$$\lambda^2 - I_1\lambda + I_2 = 0. \qquad (58.19)$$

§59. Классификация линий второго порядка на плоскости

Канонические уравнения.

Теорема 59.1. *Общее уравнение* (58.1) *линии второго порядка, заданное в прямоугольной декартовой системе координат, определяет одну и только одну из девяти линий, для каждой из которых существует прямоугольная система координат, в которой уравнение этой линии имеет вид:*

1) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1,\ a \geq b > 0,$ ***эллипс;***

2) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = -1,$ ***мнимый эллипс;***

3) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 0,$ ***пара мнимых пересекающихся прямых;***

4) $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1,$ ***гипербола;***

5) $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 0,$ ***пара пересекающихся прямых;***

6) $y^2 = 2px, \quad p > 0,$ ***парабола;***
7) $y^2 = a^2, \quad a \neq 0,$ ***пара параллельных прямых;***
8) $y^2 = -a^2, \quad a \neq 0,$ ***пара мнимых параллельных прямых;***
9) $y^2 = 0,$ ***пара совпадающих прямых.***

Доказательство. Пусть общее уравнение (58.1) переходом к новой прямоугольной декартовой системе координат преобразовалось в приведенное уравнение. Рассмотрим все возможные при этом варианты.

Если $I_2 \neq 0$, то уравнение (58.1) преобразуется в приведенное уравнение типа I:

$$\lambda_1 x^2 + \lambda_2 y^2 + a_0 = 0, \qquad (59.1)$$

где $\lambda_1\lambda_2 \neq 0$. Для этого уравнения согласно (58.15)

$$I_1 = \lambda_1 + \lambda_2, \quad I_2 = \lambda_1\lambda_2, \quad K_3 = \lambda_1\lambda_2 a_0. \qquad (59.2)$$

В зависимости от знаков λ_1, λ_2, a_0 уравнение (59.1) может быть записано по-разному.

1. Если $\lambda_1\lambda_2 > 0$, $\lambda_1 a_0 < 0$, т.е.

$$I_2 > 0, \quad I_1 K_3 < 0, \qquad (59.3)$$

то простейшие преобразования приводят уравнение (59.1) к равносильному уравнению $\dfrac{x^2}{-a_0/\lambda_1} + \dfrac{y^2}{-a_0/\lambda_2} = 1$ или $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$, где $a^2 = -\dfrac{a_0}{\lambda_1}$, $b^2 = -\dfrac{a_0}{\lambda_2}$. В этом уравнении можно считать, что $a \geq b > 0$, так как в противном случае достаточно переименовать переменные, как это было сделано в (58.14). Мы получили уравнение 1, известное нам как каноническое уравнение *эллипса*.

2. Если $\lambda_1\lambda_2 > 0$, $\lambda_1 a_0 > 0$, т.е.

$$I_2 > 0, \quad I_1 K_3 > 0, \qquad (59.4)$$

то, поступая аналогично, приходим к уравнению 2.

Ясно, что нет ни одной точки плоскости, удовлетворяющей этому уравнению. Принято говорить о нем как об уравнении ***мнимого эллипса.***

3. Если $\lambda_1\lambda_2 > 0$, $a_0 = 0$, т.е.

$$I_2 > 0, \quad K_3 = 0, \tag{59.5}$$

то уравнение (59.1) эквивалентно уравнению 3, где $a^2 = |\lambda_1|^{-1}$, $b^2 = |\lambda_2|^{-1}$.

Ясно, что только начало координат удовлетворяет этому уравнению. Принято говорить о нем как об уравнении ***пары мнимых пересекающихся прямых*** (или *вырожденного эллипса*).

4. Если $\lambda_1\lambda_2 < 0$, $a_0 \neq 0$, т.е.

$$I_2 < 0, \quad K_3 \neq 0, \tag{59.6}$$

то уравнение (59.1) эквивалентно уравнению 4. Это известное нам каноническое уравнение *гиперболы*.

5. Если $\lambda_1\lambda_2 < 0$, $a_0 = 0$, т.е.

$$I_2 < 0, \quad K_3 = 0, \tag{59.7}$$

то уравнение (59.1) эквивалентно уравнению **5**. Оно определяет ***пару пересекающихся прямых*** $y = \pm\dfrac{b}{a}x$.

Рассмотренные линии образуют первую группу линий второго порядка на плоскости. Ими исчерпываются все линии, которые определяются приведенными уравнениями типа I, т.е. случаем, когда $I_2 \neq 0$.

6. Если $I_2 = 0$, $K_3 \neq 0$, то уравнение (59.1) преобразуется в приведенное уравнение типа II:

$$\lambda_2 y^2 + 2b_0 x = 0, \tag{59.8}$$

где $\lambda_2 b_0 \neq 0$. Для этого уравнения согласно (58.16)

$$I_1 = \lambda_2, \quad I_2 = 0, \quad K_3 = -\lambda_2 b_0^2 \neq 0. \tag{59.9}$$

Уравнение (59.8) эквивалентно уравнению

$$y^2 = 2px,$$

где $p = -b_0/\lambda_2$. В этом уравнении можно считать, что $p > 0$, так как в противном случае достаточно выполнить отражение оси Ox относительно оси Oy: $x' = -x$, $y' = y$, что соответствует переходу к новому базису с матрицей перехода $Q = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$, которая в силу ортогональности приводит к ортонормированному базису и не меняет инвариантов. Итак, мы получили известное каноническое уравнение 6, известное нам как каноническое уравнение ***параболы***.

Если $I_2 = 0$, $K_3 = 0$, то уравнение (59.1) преобразуется в приведенное уравнение типа III:

$$\lambda_2 y^2 + c_0 = 0, \tag{59.10}$$

где $\lambda_2 \neq 0$. Для этого уравнения согласно (58.17)

$$I_1 = \lambda_2, \quad I_2 = 0, \quad K_3 = 0, \quad K_2 = \lambda_2 c_0. \tag{59.11}$$

Линии этой группы определяются полуинвариантом K_2. Как следует из теоремы 58.2, преобразования базиса с помощью ортогональной матрицы перехода не изменяют K_2.

Лемма. *Если $I_2 = K_3 = 0$, то полуинвариант K_2 не изменяется при параллельном переносе.*

Доказательство. Не изменяя числа K_2, преобразуем общее уравнение (58.1) в уравнение (58.12), в котором $a'_{12} = 0$. Для этого уравнения

$$A' = \begin{bmatrix} a'_{11} & 0 \\ 0 & a'_{22} \end{bmatrix}.$$

Так как $I_2 = 0$, то $a'_{11}a'_{22} = 0$. Поэтому одно из чисел a'_{11} или a'_{22} равно нулю. Не изменяя инвариантов, можно считать (см. обсуждение (58.14)), что $a'_{11} = 0$. Тогда

$$B' = \begin{bmatrix} 0 & 0 & a'_{13} \\ 0 & a'_{22} & a'_{23} \\ a'_{13} & a'_{23} & a_{33} \end{bmatrix}.$$

Но $K_3 = 0$, поэтому $a'_{13} = 0$ и уравнение (58.12) имеет вид

$$a'_{22}y'^2 + 2a'_{23}y' + a_{33} = 0, \tag{59.12}$$

при этом

$$B' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & a'_{22} & a'_{23} \\ 0 & a'_{23} & a_{33} \end{bmatrix}, \quad K_2 = K'_2 = a'_{22}a_{33} - a'^2_{23}. \tag{59.13}$$

При параллельном переносе $x' = x'' + \alpha$, $y' = y'' + \beta$ уравнение (59.12) преобразуется в уравнение

$$a'_{22}y''^2 + 2(\beta a'_{22} + a'_{23})y'' + 2\beta a'_{23} + a_{33} = 0,$$

для которого

$$B'' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & a'_{22} & \beta a'_{22} + a'_{23} \\ 0 & \beta a'_{22} + a'_{23} & a_{33} + a'_{22}\beta^2 + 2a'_{23}\beta \end{bmatrix},$$

$$K''_2 = a'_{22}(a_{33} + a'_{22}\beta^2 + 2a'_{23}\beta) - (\beta a'_{22} + a'_{23})^2 = a'_{22}a_{33} - a'^2_{23}.$$

Из (59.13) следует, что $K''_2 = K_2$. Лемма доказана.

Вернемся к уравнению (59.10). Отметим, что согласно лемме полуинварианты K_2 в уравнениях (58.1) и (59.10) совпадают.

7. Если $\lambda_2 c_0 < 0$, т.е.

$$K_2 < 0, \tag{59.14}$$

то уравнение (59.10) эквивалентно уравнению 7 (где $a^2 = -c_0/\lambda_2 > 0$), которое определяет ***пару параллельных прямых***: $y = a$ и $y = -a$.

8. Если $\lambda_2 c_0 > 0$, т.е.

$$K_2 > 0, \tag{59.15}$$

то уравнение (59.10) эквивалентно уравнению 8 (где $a^2 = c_0/\lambda_2 > 0$). Ясно, что ни одна точка плоскости не удовлетворяет этому уравнению. Принято говорить о нем как об уравнении ***пары мнимых параллельных прямых***.

9. Если $c_0 = 0$, т.е.

$$K_2 = 0, \tag{59.16}$$

то уравнение (59.10) эквивалентно уравнению 9, которое определяет *пару совпадающих прямых*: $y = 0$ и $y = 0$.

Условия (59.3)–(59.7), (59.9), (59.11), (59.14)–(59.16) исчерпывают все варианты линий второго порядка на плоскости и взаимно исключают друг друга. Следовательно, общее уравнение (58.1) определяет одну и только одну из девяти перечисленных линий. ■

Для каждой из этих линий найдена прямоугольная декартова система координат, в которой уравнение линии имеет вид 1–9. Эти уравнения называют ***каноническими уравнениями линий второго порядка***.

Линии, для которых $I_2 > 0$, называются ***линиями эллиптического типа***, $I_2 < 0$ – ***гиперболического типа***, $I_2 = 0$ – ***параболического типа***.

Замечание. Если общее уравнение (58.1) линии второго порядка задано в прямоугольной декартовой системе координат, то каноническое уравнение линии может быть найдено по инвариантам, так как согласно (58.18) и (58.19) коэффициенты λ_1, λ_2 приведенных уравнений являются корнями характеристического многочлена матрицы A:

$$\lambda^2 - I_1\lambda + I_2 = 0.$$

При этом если $\lambda_1 \leq \lambda_2$, то в силу (59.2), (59.9), (59.11)

$$a_0 = \frac{K_3}{I_2}, \quad b_0^2 = -\frac{K_3}{I_1}, \quad c_0 = \frac{K_2}{I_1}.$$

Результаты проведенных исследований **сведены в следующую таблицу.**

Приведенное уравнение	Номер	Каноническое уравнение линии	Название линии	Признак линии
$\lambda_1 x^2 + \lambda_2 y^2 + \frac{K_3}{I_2} = 0$, $\lambda_1\lambda_2 \neq 0$	1	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, $a \geq b > 0$	Эллипс	$I_2 > 0$, $I_1 K_3 < 0$
	2	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1$	Мнимый эллипс	$I_2 > 0$, $I_1 K_3 > 0$
	3	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0$	Пара мнимых пересекающихся прямых	$I_2 > 0$, $K_3 = 0$
	4	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$	Гипербола	$I_2 < 0$, $K_3 \neq 0$
	5	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$	Пара пересекающихся прямых	$I_2 < 0$, $K_3 = 0$
$I_1 y^2 \pm 2\sqrt{-\frac{K_3}{I_1}}x = 0$, $I_1 K_3 \neq 0$	6	$y^2 = 2px$, $p > 0$	Парабола	$I_2 = 0$, $K_3 \neq 0$
$I_1 y^2 + \frac{K_2}{I_1} = 0$, $I_1 \neq 0$	7	$y^2 = a^2$, $a \neq 0$	Пара параллельных прямых	$I_2 = 0$, $K_3 = 0$, $K_2 < 0$
	8	$y^2 = -a^2$, $a \neq 0$	Пара мнимых параллельных прямых	$I_2 = 0$, $K_3 = 0$, $K_2 > 0$
	9	$y^2 = 0$	Пара совпадающих прямых	$I_2 = 0$, $K_3 = 0$, $K_2 = 0$

Метод Лагранжа. Теорема 58.4 о приведенных уравнениях остается справедливой и в аффинной системе координат, так как тип приведенного уравнения определяется только ***знаками*** инвариантов I_2, K_3, которые в силу теоремы 58.2 сохраняются при преобразованиях аффинной системы координат. Вместо вычисления инвариантов или метода вращений можно использовать ***метод выделения полных квадратов*** (***метод Лагранжа***), который состоит в следующем.

1. Пусть в общем уравнении (58.1) $a_{11} \neq 0$. Выделим полный квадрат в группе членов, содержащих переменную x:

$$a_{11}x^2 + 2a_{12}xy + 2a_{13}x = a_{11}\left(x + \frac{a_{12}}{a_{11}}y + a_{13}\right)^2 - a_{11}\left(\frac{a_{12}}{a_{11}}y + a_{13}\right)^2.$$

Положим

$$x' = x + \frac{a_{12}}{a_{11}}y + a_{13}. \tag{59.17}$$

Тогда уравнение (58.1) примет вид

$$a_{11}x'^2 + a'_{22}y^2 + 2a'_{23}y + a'_{33} = 0. \tag{59.18}$$

Если $a'_{22} \neq 0$, то, положив

$$y' = y + \frac{a'_{23}}{a'_{22}}, \tag{59.19}$$

приводим уравнение (59.18) к уравнению

$$a_{11}x'^2 + a'_{22}y^2 + a''_{33} = 0, \quad a_{11}a'_{22} \neq 0, \tag{59.20}$$

которое является приведенным уравнением типа I. Преобразования координат (59.17), (59.19) определяют переход к новому базису с помощью матрицы перехода

$$Q = \begin{bmatrix} 1 & -\dfrac{a_{12}}{a_{11}} \\ 0 & 1 \end{bmatrix}$$

и переноса начала. Согласно теоремам 58.2 и 58.3 знаки инвариант I_2, K_3 в уравнениях (58.1) и (59.20) совпадают.

Если в уравнении (59.18) $a'_{22} = 0$, то:

а) в случае когда $a'_{23} = 0$, уравнение (59.18) является приведенны уравнением типа III;

б) в случае когда $a'_{23} \neq 0$, положив

$$y' = y + \frac{a'_{33}}{2a'_{23}},$$

приводим уравнение (59.18) к уравнению

$$a_{11}x'^2 + 2a'_{13}y' = 0, \quad a_{11}a'_{13} \neq 0,$$

которое является приведенным уравнением типа II.

2. Пусть в общем уравнении (58.1) $a_{11} = 0$, $a_{22} \neq 0$. Поступая так же, как и в п.1 (с точностью до названия переменных), освобождаемся от слагаемого $2a_{12}xy$ и линейной части (если это возможно) и приходим к приведенным уравнениям, не изменив знаков инвариантов I_2, K_3.

3. Пусть в общем уравнении (58.1) $a_{11} = 0$, $a_{22} = 0$. Тогда $a_{12} \neq 0$. Положив

$$x = x' + y', \quad y = x' - y', \tag{59.21}$$

сводим этот случай к уже рассмотренным. Отметим, что преобразования координат (59.21) отвечают переходу к новому базису с матрицей перехода $Q = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

Так как приведенные уравнения (58.11) определяются лишь знаком инвариантов I_2, K_3, то в результате выполненных преобразовани

общее уравнение (58.1) приведется к одному и только одному из уравнений (58.11). С помощью метода Лагранжа легко устанавливается тип линии, однако не может быть получено каноническое уравнение.

Глава XII. **Линейное пространство над произвольным полем**

В главе IV мы рассматривали вещественные линейные пространства. Вещественными они назывались потому, что операция умножения на число предусматривала умножение вектора на вещественные числа. В §43 мы уже отмечали, что все свойства линейного пространства остаются в силе, если поле вещественных чисел $\mathbb{R}$ заменить произвольным полем P. В этой главе воспроизводится известное определение линейного пространства для случая произвольного поля, напоминаются уже известные факты и рассматриваются новые.

§60. Определение и основные свойства

Пусть дано поле P. Непустое множество V называется ***линейным*** или ***векторным пространством над полем*** P, если на этом множестве определены внутренний закон композиции $V \times V \to V$, называемый сложением, и внешний закон композиции $P \times V \to V$, называемый умножением на число из поля P, удовлетворяющие следующим аксиомам: для любых $a, b, c \in V$ и $\alpha, \beta \in P$

1) $a + b = b + a$;
2) $(a + b) + c = a + (b + c)$;
3) существует элемент $\theta \in V$ такой, что $a + \theta = \theta + a = a$;
4) для любого элемента $a \in V$ существует элемент $-a \in V$ **такой,** что $a + (-a) = (-a) + a = \theta$;
5) $1 \cdot a = a$;
6) $\alpha(\beta a) = (\alpha\beta)a$;
7) $(\alpha + \beta)a = \alpha a + \beta a$;
8) $\alpha(a + b) = \alpha a + \alpha b$.

Линейное пространство над полем $\mathbb{R}$ называется ***вещественным линейным пространством***, а над полем $\mathbb{C}$ – ***комплексным***.

Примеры. 1. Арифметическое пространство P^n – линейное пространство над полем P. Как и в случае $\mathbb{R}^n$, здесь $P^n = \{x = (x_1, \ldots, x_n) \mid x_k \in P,\ k = \overline{1, n}\}$ с операциями: если $x = (x_1, \ldots, x_n)$ и $y = (y_1, \ldots, y_n)$, $\alpha \in P$, то

$$x + y = (x_1 + y_1, \ldots, x_n + y_n), \quad \alpha x = (\alpha x_1, \ldots, \alpha x_n).$$

2. В частности, арифметическое пространство $\mathbb{C}^n$ есть комплексное линейное пространство. В таком пространстве арифметические векторы умножаются на комплексные числа. Однако можно было бы ограничиться умножением только на вещественные числа, при этом результат умножения, несомненно, останется арифметическим вектором из $\mathbb{C}^n$. Это означает, что внешний закон композиции можно

вводить и как умножение на вещественные числа, т.е. как отображение $\mathbb{R} \times \mathbb{C}^n \to \mathbb{C}^n$. Символом $\mathbb{C}^n_{\mathbb{R}}$ будем обозначать множество $\mathbb{C}^n$, в котором внешний закон композиции определен как $\mathbb{R} \times \mathbb{C}^n \to \mathbb{C}^n$. Нетрудно проверить, что $\mathbb{C}^n_{\mathbb{R}}$ – вещественное линейное пространство. В частности, само поле $\mathbb{C}$ можно рассматривать как вещественное линейное пространство $\mathbb{C}_{\mathbb{R}}$.

3. Аксиомы линейного пространства выявляют алгебраические свойства многих классов функций, часто встречающихся в математическом анализе: например, $C[0,1]$ – множество непрерывных функций на отрезке $[0,1]$; $D[0,1]$ – множество всех дифференцируемых функций на отрезке $[0,1]$; $C^1[0,1]$ – множество всех функций, непрерывных вместе с первой производной на отрезке $[0,1]$. Хотя у каждого из перечисленных множеств имеются и другие замечательные свойства, все они охватываются единой алгебраической схемой, аксиоматизируемой определением линейного пространства. Нетрудно проверить, что эти множества являются вещественными линейными пространствами относительно обычных операций сложения функций и умножения функции на число. Линейные пространства функций принято называть ***функциональными пространствами***.

§61. Линейная зависимость. Ранг и база системы векторов

Понятие линейной зависимости и связанные с ней утверждения рассматривались в главе IV (§ 14–16) и относились к вещественным линейным пространствам. Все определения, а также теоремы и их доказательства остаются в силе и в произвольном линейном пространстве. Введем новые понятия. В дальнейшем, употребляя термин "линейное пространство", будем иметь в виду линейное пространство над произвольным полем P, не оговаривая это дополнительно.

Ранг и база системы векторов. Будем рассматривать конечные системы $a_1, \ldots, a_k$ векторов линейного пространства.

Линейно независимая подсистема системы векторов, через которую линейно выражается любой вектор системы, называется *базой* этой *системы векторов*.

Примеры. 1. В системе из нулевых векторов нет ни одной базы, так как любая ее подсистема линейно зависима.

2. Базисные строки матрицы согласно теореме о базисном миноре образуют базу системы строк, рассматриваемых как векторы арифметического пространства. Это же относится и к базисным столбцам.

3. В системе трех неколлинеарных векторов плоскости V_2 любая пара векторов образует базу.

Обратим внимание на то, что в этих примерах система векторов может обладать не единственной базой, но при этом все базы одной системы состоят из одинакового числа векторов.

Теорема 61.1. *Подсистема системы векторов является базой системы векторов тогда и только тогда, когда образует максимальную линейно независимую подсистему.*

Доказательство. Необходимость. Пусть в системе векторов $a_1, \ldots, a_r, \ldots, a_k$ подсистема $a_1, \ldots, a_r$ образует базу (здесь для упрощения записи в качестве подсистемы взяты первые r векторов, что, как будет видно из доказательства, не нарушает общности рассуждений). Тогда любая бо́льшая подсистема будет линейно зависимой в силу теоремы 16.2, так как любой ее вектор линейно выражается через базу $a_1, \ldots, a_r$. Таким образом, база образует максимальную линейно независимую подсистему системы векторов.

Достаточность. Пусть $a_1, \ldots, a_r$ – максимальная линейно независимая подсистема системы векторов $a_1, \ldots, a_r, \ldots, a_k$. Тогда для любого вектора a_i, $i = \overline{1, k}$, подсистема $a_1, \ldots, a_r, a_i$ линейно зависима (если $i \le r$, то как подсистема, содержащая два одинаковых вектора; если же $i > r$, то как подсистема из $r + 1 > r$ векторов). В силу теоремы 14.6 вектор a_i линейно выражается через $a_1, \ldots, a_r$. Таким образом, $a_1, \ldots, a_r$ – база. ■

Следствие 1. *Все базы одной системы векторов состоят из одинакового числа векторов, равного максимальному числу линейно независимых векторов системы.*

Число векторов базы называется ***рангом системы векторов***. Очевидно, ранг системы векторов равен максимальному числу линейно независимых векторов системы. Обозначение: $\mathrm{rg}(a_1, \ldots, a_n)$.

Две системы векторов линейного пространства называются ***эквивалентными***, если каждая из этих систем выражается через другую. Из определения следует, что ***база системы векторов эквивалентна самой системе.***

Теорема 61.2. *Если система векторов $a_1, \ldots, a_k$ линейно выражается через $b_1, \ldots, b_m$, то $\mathrm{rg}(a_1, \ldots, a_k) \le \mathrm{rg}(b_1, \ldots, b_m)$.*

Доказательство. Пусть $a_1, \ldots, a_r$ и $b_1, \ldots, b_s$ – базы рассматриваемых систем. Из условия теоремы и транзитивности свойства "линейной выражаемости" следует, что база $a_1, \ldots, a_r$ первой системы линейно выражается через базу $b_1, \ldots, b_s$ второй. Тогда $r \le s$, так как в противном случае, если $r > s$, система $a_1, \ldots, a_r$ была бы линейно зависимой в силу теоремы 16.2. ■

Следствие 2. *Ранги эквивалентных систем совпадают.*

Следствие 3. *Эквивалентные линейно независимые системы векторов состоят из одинакового числа векторов.*

§62. Базис и размерность

Говорят, что *система векторов* $e_1, \ldots, e_n$ линейного пространства V ***порождает пространство*** V, если любой вектор $x \in V$ является линейной комбинацией $e_1, \ldots, e_n$.

Упорядоченная система векторов $e_1, \ldots, e_n$ линейного пространства V называется ***базисом*** V, если она линейно независима и порождает V. Очевидно, что это определение совпадает с определением

базиса, данным в §17 для вещественного линейного пространства.

Теорема 62.1. ***Любые два базиса линейного пространства состоят из одинакового числа векторов.***

Это утверждение вытекает из эквивалентности двух базисов линейного пространства и следствия 2 (§61). ■

Итак, число векторов базиса не зависит от самого базиса и однозначно определяется самим пространством. Ответ на вопрос, что представляет собой это число, дается теоремой 17.1, справедливой и в пространстве над произвольным полем.

Напомним, что число векторов базиса линейного пространства V называется ***размерностью пространства*** V и обозначается символом $\dim V$. Размерность нулевого пространства по определению считается равной нулю. Из теоремы 17.1 следует, что размерность линейного пространства совпадает с максимальным числом линейно независимых векторов этого пространства. Линейное пространство размерности n, где $n \in \mathbb{N}$, называется ***n-мерным***. Нулевое пространство и n-мерные пространства называются ***конечномерными***.

Линейное пространство называется ***бесконечномерным***, если для любого числа $k \in \mathbb{N}$ в пространстве существует k линейно независимых векторов. Пример бесконечномерного пространства дан в §17 (пример 5). В курсе линейной алгебры изучаются лишь конечномерные линейные пространства.

Теорема 62.2 (о неполном базисе). *В n-мерном пространстве любую линейно независимую систему из k, где $k < n$, векторов можно дополнить до базиса.*

Доказательство. Пусть $e_1, \ldots, e_k$ – линейно независимая система векторов пространства V. Так как $k < n$, то в силу теоремы 17.1 система векторов $e_1, \ldots, e_k$ не является базисом V и, следовательно, не порождает всего пространства V. Пусть вектор $e_{k+1} \in V$ не является линейной комбинацией $e_1, \ldots, e_k$, тогда система векторов $e_1, \ldots, e_k, e_{k+1}$ линейно независима (в силу теоремы 14.6). Если $k+1 = n$, то эта система векторов образует базис V; если же $k+1 < n$, то аналогичным образом построим линейно независимую систему из $k + 2$ векторов. За $n - k$ таких шагов мы построим искомый базис $e_1, \ldots, e_k, \ldots, e_n$. ■

Точно так же, как и в §17, определяются координаты вектора в базисе и матрица перехода от одного базиса к другому. При этом все факты, связанные с этими понятиями, остаются в силе и в общем случае, так как их доказательства опираются на те свойства вещественных чисел, которые имеют место в любом поле. Напомним эти факты.

1°. ***Разложение вектора по базису единственно.***

2°. ***Координаты вектора обладают свойством линейности.***

3°. ***При переходе от базиса e к базису $f = eQ$ координаты вектора x изменяются по следующему закону: $x_e = Qx_f$.***

Примеры. 1. В арифметических пространствах P^n, $\mathbb{R}^n$, $\mathbb{C}^n$ единичные векторы (14.6) образуют базис (§17, пример 2). Этот базис принято называть *естественным*. Координаты любого вектора $a = (a_1, \dots, a_n)$ в естественном базисе совпадают с компонентами $a_1, \dots, a_n$ этого вектора.

2. В арифметическом пространстве $\mathbb{C}^n_{\mathbb{R}}$ векторы (14.6), очевидно, не могут быть базисом, так как умножением этих векторов только на вещественные числа нельзя получить любой комплексный арифметический вектор $a = (\alpha_1 + i\beta_1, \dots, \alpha_n + i\beta_n)$, где $\alpha_k, \beta_k \in \mathbb{R}$, $k = \overline{1,n}$. Добавим к векторам (14.6) еще n векторов:

$$f_1 = (i, 0, \dots, 0),\ f_2 = (0, i, \dots, 0),\ \dots,\ f_n = (0, 0, \dots, i).$$

Покажем, что система векторов $e_1, \dots, e_n, f_1, \dots, f_n$ образует базис $\mathbb{C}^n_{\mathbb{R}}$. Действительно, линейная комбинация этих векторов с вещественными коэффициентами α_k, β_k имеет вид

$$\sum_{i=1}^{n} \alpha_k e_k + \sum_{i=1}^{n} \beta_k f_k = (\alpha_1 + i\beta_1, \dots, \alpha_n + i\beta_n) \qquad (62.1)$$

и равна нулевому вектору $\theta = (0, \dots, 0)$ тогда и только тогда, когда $\alpha_k + i\beta_k = 0$, $k = \overline{1,n}$, т.е. $\alpha_k = \beta_k = 0$, $k = \overline{1,n}$. Следовательно, векторы $e_1, \dots, e_n, f_1, \dots, f_n$ линейно независимы. Они порождают все пространство $\mathbb{C}^n_{\mathbb{R}}$, так как любой вектор $(\alpha_1 + i\beta_1, \dots, \alpha_n + i\beta_n)$ из $\mathbb{C}^n_{\mathbb{R}}$ согласно (62.1) является линейной комбинацией этих векторов с вещественными коэффициентами $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n$.

Итак, $\dim \mathbb{C}^n_{\mathbb{R}} = 2n$.

§63. Изоморфизм линейных пространств

Два линейных пространства V_1 и V_2 над общим полем P называются ***изоморфными***, если существует биективное отображение $\varphi : V_1 \to V_2$, которое сохраняет законы композиции, т.е. если для любых векторов $x, y \in V_1$ и любого числа $\alpha \in P$

1) $\varphi(x + y) = \varphi(x) + \varphi(y)$;
2) $\varphi(\alpha x) = \alpha\varphi(x)$.

Обозначение: $V_1 \cong V_2$. Само отображение φ называется ***изоморфизмом линейных пространств***.

Примеры. 1. Геометрические пространства V_1, V_2, V_3 векторов на прямой, на плоскости и в пространстве изоморфны арифметическим пространствам $\mathbb{R}^1$, $\mathbb{R}^2$ и $\mathbb{R}^3$ соответственно. Действительно, поставим в соответствие каждому вектору $\mathbf{x} \in V_n$ набор его координат в каком-либо базисе e, т.е. арифметический вектор $x_e \in \mathbb{R}^n$, где $n = 1, 2, 3$. Это соответствие взаимно однозначно и сохраняет законы композиции, так как координаты вектора обладают свойством линейности (§17).

2. $V_2 \cong C_{\mathbb{R}}$.

3. $P^{m\times n} \cong P^{mn}$, и в частности $\mathbb{R}^{m\times n} \cong \mathbb{R}^{mn}$.

Отметим простейшие свойства изоморфных пространств.

1°. *Отношение изоморфизма есть отношение эквивалентности на множестве всех линейных пространств над полем P.*

2°. *В изоморфных пространствах*

а) образ (и прообраз) линейной комбинации векторов есть линейная комбинация образов (прообразов) с теми же коэффициентами;

б) образ (и прообраз) нулевого вектора есть нулевой вектор;

в) образ (и прообраз) линейно независимой системы векторов образует линейно независимую систему;

г) образ (и прообраз) базиса есть базис.

Доказательства этих свойств опираются лишь на определение и элементарные свойства рассматриваемых объектов. Мы предоставляем их читателю.

Несмотря на простоту формулировок, уже эти свойства говорят о том, что изоморфные линейные пространства, даже самой различной природы, с точки зрения свойств линейного пространства неразличимы. Это позволяет многие свойства линейных пространств изучать на линейных пространствах простой структуры. К этой возможности мы будем неоднократно прибегать, проводя все вычисления в арифметических пространствах, а "рисуя" – в геометрических.

Теорема 63.1 (критерий изоморфизма). *Два линейных пространства над общим полем изоморфны тогда и только тогда, когда их размерности совпадают.*

Доказательство. Необходимость вытекает из свойства "г" изоморфных пространств.

Достаточность. Пусть V_1 и V_2 – линейные пространства над полем P и пусть $\dim V_1 = \dim V_2 = n$. Выберем в этих пространствах по базису: пусть $e_1, \dots, e_n$ – базис V_1, $f_1, \dots, f_n$ – базис V_2. Построим отображение $\varphi : V_1 \to V_2$, поставив в соответствие каждому вектору $x = \sum_{i=1}^n \alpha_i e_i \in V_1$ вектор $y = \sum_{i=1}^n \alpha_i f_i \in V_2$ (т.е. вектор, который имеет те же координаты, что и вектор x). В силу единственности разложения вектора по базису отображение φ биективно. При этом φ – изоморфизм, так как координаты вектора обладают свойством линейности. ∎

Следствие. Любое n-мерное вещественное пространство изоморфно арифметическому пространству $\mathbb{R}^n$, а любое n-мерное комплексное пространство – арифметическому пространству $\mathbb{C}^n$.

§64. Линейные подпространства

Напомним определение линейного подпространства, данное в §18 (оно остается неизменным и в случае линейного пространства над произвольным полем). Подмножество L линейного пространства V над полем P называется ***линейным подпространством пространства*** V, если оно само является линейным пространством относительно законов композиции в V. Другое определение линейного подпространства, эквивалентное этому, устанавливается известной теоремой 18.1, которая имеет место и в общем случае.

Как указывалось в §30, множество всех решений однородной системы уравнений

$$Ax = 0 \tag{64.1}$$

с n неизвестными образует линейное подпространство арифметического пространства $\mathbb{R}^n$ (или P^n в общем случае, когда $A \in P^{m\times n}$). Про это подпространство L говорят, что оно задано однородной системой (64.1), и записывают в виде

$$L: \quad Ax = 0 . \tag{64.2}$$

Другой способ задания линейного подпространства дается понятием линейной оболочки.

Линейная оболочка. Пусть $a_1, \ldots, a_k$ – система векторов линейного пространства V над полем P. *Линейной оболочкой системы векторов* $a_1, \ldots, a_k$ называется множество всевозможных линейных комбинаций этих векторов. Говорят также, что линейная оболочка *натянута на векторы* $a_1, \ldots, a_k$. Обозначение: $\mathcal{L}(a_1, \ldots, a_k)$. Итак,

$$\mathcal{L}(a_1, \ldots, a_k) = \left\{ a = \sum_{i=1}^{k} \alpha_i a_i \mid \alpha_i \in P\,,\ i = \overline{1,k} \right\}.$$

Из определения следует, что каждое конечномерное пространство является линейной оболочкой векторов своего базиса. В частности, линейное подпространство, заданное однородной системой (64.2), является линейной оболочкой фундаментальной системы решений (§30).

Теорема 64.1. *Если $a_1, \ldots, a_k$ – векторы линейного пространства V, то $\mathcal{L}(a_1, \ldots, a_k)$ является линейным подпространством пространства V.*

Это утверждение вытекает из теоремы 18.1, так как для линейной оболочки $\mathcal{L}(a_1, \ldots, a_k)$ обе импликации имеют место. ■

Итак, линейная оболочка $\mathcal{L}(a_1, \ldots, a_k)$ – это линейное подпространство, которое порождается векторами $a_1, \ldots, a_k$.

Теорема 64.2. *Две системы векторов линейного пространства эквивалентны тогда и только тогда, когда их линейные оболочки совпадают.*

Доказательство. Необходимость. Пусть системы векторов $a_1, \ldots, a_k$ и $b_1, \ldots, b_m$ эквивалентны. Тогда $\mathcal{L}(a_1, \ldots, a_k) = \mathcal{L}(b_1, \ldots, b_m)$, так как для этих множеств, как легко проверить, имеет место двустороннее вложение. Достаточность очевидна. ■

Следствие 1. *Линейная оболочка системы векторов совпадает с линейной оболочкой своей базы.*

Следствие 2. *Размерность линейной оболочки системы векторов равна рангу этой системы:*

$$\dim \mathcal{L}(a_1, \ldots, a_k) = \operatorname{rg}(a_1, \ldots, a_k). \tag{64.3}$$

Теорема 64.3 (о монотонности размерности). *Размерность линейного подпространства не превосходит размерности пространства. Подпространство той же размерности, что и все пространство, совпадает с пространством.*

Доказательство. Пусть L – линейное подпространство V, $\dim L = k$, $\dim V = n$. Тогда $k \le n$, так как в противном случае в n-мерном пространстве V существует k, где $k > n$, линейно независимых векторов (например, векторы базиса L). Пусть $k = n$ и $e_1, \ldots, e_n$ – базис L. Так как $\dim V = n$, то векторы $e_1, \ldots, e_n$ образуют базис V (§17, утверждение 1). Таким образом, $L = V = \mathcal{L}(e_1, \ldots, e_n)$. ■

§65. Сумма и пересечение подпространств

Пусть $L_1, \ldots, L_k$ – линейные подпространства линейного пространства V. *Суммой подпространств* $L_1, \ldots, L_k$ называется множество всевозможных векторов x, представимых в виде

$$x = x_1 + \ldots + x_k\,, \tag{65.1}$$

где $x_i \in L_i$, $i = \overline{1,k}$. Обозначение: $L_1 + \ldots + L_k$ или $\sum_{i=1}^{k} L_i$.

Итак,

$$\sum_{i=1}^{k} L_i = \{x = x_1 + \ldots + x_k \mid x_i \in L_i\,,\ i = \overline{1,k}\}.$$

Представление (65.1) вектора x называется *разложением вектора x по подпространствам* $L_1, \ldots, L_k$.

Пересечением подпространств $L_1, \ldots, L_k$ называется множество

$$L_1 \cap \ldots \cap L_k = \{x \in V \mid x_i \in L_i\,,\ i = \overline{1,k}\}.$$

Заметим, что пересечение подпространств не пусто, так как всегда содержит нулевой вектор θ пространства.

Теорема 65.1. *Сумма и пересечение подпространств линейного пространства V являются линейными подпространствами пространства V.*

Это утверждение вытекает из теоремы 18.1, так как для $\sum_{i=1}^{k} L_i$ и $\bigcap_{i=1}^{k} L_i$ справедливы обе импликации. ■

Заметим, что каждое из подпространств L_i, $i = \overline{1,k}$, а также их пересечение являются линейными подпространствами суммы $L_1 + \ldots + L_k$, а пересечение $L_1 \cap \ldots \cap L_k$ – линейным подпространством каждого из L_i, $i = \overline{1,k}$.

Теорема 65.2. *Сумма линейных подпространств есть линейная оболочка совокупности базисов слагаемых подпространств.*

Доказательство. Пусть $e_1, \ldots, e_m, f_1, \ldots, f_s, \ldots, g_1, \ldots, g_t$ – базисы подпространств $L_1, L_2, \ldots, L_k$ соответственно. Положим $W = \mathcal{L}(e_1, \ldots, e_m, f_1, \ldots, f_s, \ldots, g_1, \ldots, g_t)$. Тогда $W = L_1 + \ldots + L_k$, так как для этих множеств, как нетрудно проверить, имеет место двустороннее вложение. ■

Следствие. Размерность суммы линейных подпространств равна рангу совокупности слагаемых подпространств:

$$\dim\sum_{i=1}^{k} L_i = \mathrm{rg}(e_1,\ldots,e_m,f_1,\ldots,f_s,\ldots,g_1,\ldots,g_t).$$

Это утверждение вытекает из теоремы 65.2 (с учетом 64.3).

Теорема 65.3. *Для любых двух линейных подпространств L_1 и L_2 линейного пространства V имеет место соотношение*

$$\dim(L_1+L_2) = \dim L_1 + \dim L_2 - \dim L_1\cap L_2 . \tag{65.2}$$

Доказательство. Рассмотрим сначала случай, когда $L_1\cap L_2\neq\{\theta\}$. Выполним следующее построение. Пусть $f_1,\ldots,f_s$ – базис $L_1\cap L_2$. Так как $L_1\cap L_2\subset L_1$, $L_1\cap L_2\subset L_2$, то согласно теореме о неполном базисе векторы $f_1,\ldots,f_s$ можно дополнить как до базиса L_1, так и до базиса L_2. Пусть $e_1,\ldots,e_m,f_1,\ldots,f_s$ – базис L_1, а $f_1,\ldots,f_s,g_1,\ldots,g_t$ – базис L_2. Покажем, что система векторов

$$e_1,\ldots,e_m,f_1,\ldots,f_s,g_1,\ldots,g_t \tag{65.3}$$

образует базис L_1+L_2. Действительно, она порождает L_1+L_2 в силу теоремы 65.2 и, кроме того, она линейно независима, так как если

$$\sum_{i=1}^{m}\alpha_i e_i+\sum_{i=1}^{s}\beta_i f_i+\sum_{i=1}^{t}\gamma_i g_i=\theta , \tag{65.4}$$

то

$$g=\sum_{i=1}^{t}\gamma_i g_i=-\sum_{i=1}^{m}\alpha_i e_i-\sum_{i=1}^{s}\beta_i f_i. \tag{65.5}$$

Отсюда следует, что вектор g является вектором как L_1, так и L_2, т.е. $g\in L_1\cap L_2$ и, следовательно, вектор g является линейной комбинацией векторов $f_1,\ldots,f_s$. Из единственности разложения вектора g по линейно независимой системе векторов $e_1,\ldots,e_m,f_1,\ldots,f_s$ следует, что в (65.5) $\alpha_1=\ldots=\alpha_m=0$. Тогда в (65.4) в силу линейной независимости $f_1,\ldots,f_s,g_1,\ldots,g_t$ получим, что $\beta_1=\ldots=\beta_s=0$, $\gamma_1=\ldots=\gamma_t=0$. Таким образом, только тривиальная линейная комбинация векторов (65.3) равна нулевому вектору. Итак, векторы (65.3) образуют базис L_1+L_2 и, следовательно,

$$\dim(L_1+L_2)=m+s+t. \tag{65.6}$$

С другой стороны, согласно схеме построения векторов (65.3), $\dim L_1+\dim L_2-\dim L_1\cap L_2=(m+s)+(s+t)-s=m+s+t$. Сравнение этого соотношения с (65.6) дает (65.2). В случае когда $L_1\cap L_2=\{\theta\}$, доказательство проводится аналогично, только в нем не участвуют векторы $f_1,\ldots,f_s$. ■

§66. Прямая сумма подпространств

Сумма подпространств линейного пространства называется ***прямой суммой***, если разложение каждого вектора в ней по слагаемым подпространствам единственно. Обозначение: $L_1\oplus\ldots\oplus L_k$.

Теорема 66.1 (критерии прямой суммы). *Для подпространств $L_1,\ldots,L_k$ конечномерного линейного пространства V следующие утверждения равносильны:*

1) *сумма подпространств $L_1,\ldots,L_k$ – прямая;*

2) *совокупность базисов подпространств $L_1,\ldots,L_k$ линейно независима;*

3) *совокупность базисов подпространств $L_1,\ldots,L_k$ образует базис суммы $\sum_{i=1}^{k} L_i$;*

4) $\dim\sum_{i=1}^{k} L_i=\sum_{i=1}^{k}\dim L_i$;

5) *существует вектор $a\in\sum_{i=1}^{k} L_i$, для которого разложение по подпространствам $L_1,\ldots,L_k$ единственно;*

6) *произвольная система ненулевых векторов $a_1,\ldots,a_k$, взятых по одному из каждого подпространства L_i, $i=\overline{1,k}$, линейно независима;*

7) $L_1\cap L_2=\{\theta\}$ *(для $k=2$).*

Доказательство. $1\Rightarrow 2$. Пусть совокупность $e_1,\ldots,e_m$, $f_1,\ldots,f_s,\ldots,g_1,\ldots,g_t$ базисов подпространств $L_1,\ldots,L_k$ линейно зависима и

$$\sum_{i=1}^{m}\alpha_i e_i+\sum_{i=1}^{s}\beta_i f_i+\ldots+\sum_{i=1}^{t}\gamma_i g_i=\theta , \tag{66.1}$$

где

$$\sum_{i=1}^{m}\alpha_i^2+\sum_{i=1}^{s}\beta_i^2+\ldots+\sum_{i=1}^{t}\gamma_i^2\neq 0 . \tag{66.2}$$

Положим

$$x_1=\sum_{i=1}^{m}\alpha_i e_i ,\quad x_2=\sum_{i=1}^{s}\beta_i f_i ,\quad \ldots ,\quad x_k=\sum_{i=1}^{t}\gamma_i g_i .$$

Отметим, что $x_i\in L_i$, $i=\overline{1,k}$, причем среди $x_1,\ldots,x_k$ в силу (66.2) и линейной независимости векторов каждого базиса существует вектор $x_i\neq\theta$. Тогда соотношение (66.1) может быть записано в виде

$$\theta=x_1+\ldots+x_i+\ldots+x_k ,\ x_i\neq\theta . \tag{66.3}$$

Это дает второе разложение нулевого вектора θ (после известного: $\theta=\theta+\ldots+\theta+\ldots+\theta$) по подпространствам $L_1,\ldots,L_i,\ldots,L_k$.

$2\Rightarrow 1$. Пусть $L_1,\ldots,L_k$ – не прямая сумма. Тогда существует вектор b из этой суммы, который имеет два разложения по подпространствам $L_1,\ldots,L_k$:

$$b=b_1+\ldots+b_i+\ldots+b_k , \tag{66.4}$$

$$b = b'_1 + \ldots + b'_i + \ldots + b'_k\,, \qquad (66.5)$$

которые отличаются хотя бы одним слагаемым $b_i \neq b'_i$. Если вычесть из первого равенства второе и разложить каждое слагаемое $b_j - b'_j$, $j = \overline{1,n}$, по базису L_j, получим нетривиальную линейную комбинацию базисных векторов пространств $L_1, \ldots, L_k$, равную нулевому вектору. Это противоречит линейной независимости совокупности базисов $L_1, \ldots, L_k$.

$2 \iff 3$. Это следует из теоремы 65.2.

$3 \iff 4$. Эти утверждения отличаются только терминологией.

$1 \Rightarrow 5$. Очевидно.

$5 \Rightarrow 1$. Пусть $L_1 + \ldots + L_k$ – не прямая сумма. Тогда существует вектор b из этой суммы, для которого имеют место два различных разложения (66.4) и (66.5). Вычитая (66.5) из (66.4), получим нетривиальное разложение нулевого вектора (66.3). Если его сложить с разложением вектора a, то получим еще одно разложение вектора a.

$1 \Rightarrow 6$. Пусть система векторов $a_1, \ldots, a_k$, где $a_i \in L_i$, $a_i \neq \theta$, $i = \overline{1,k}$, линейно зависима. Тогда существуют числа $\alpha_1, \ldots, \alpha_k \in P$, одновременно не равные нулю (пусть, например, $\alpha_i \neq 0$) и такие, что $\alpha_1 a_1 + \ldots + \alpha_i a_i + \ldots + \alpha_k a_k = \theta$. Это равенство дает второе разложение нулевого вектора, отличное от тривиального (так как $\alpha_i a_i \neq \theta$), что противоречит утверждению 1.

$6 \Rightarrow 1$. Пусть $L_1 + \ldots + L_k$ – не прямая сумма, тогда существует вектор b, для которого имеют место два разложения (66.4) и (66.5). Вычитая одно из другого, получим, что $a_{i_1} + a_{i_2} + \ldots + a_{i_q} = \theta$, где $a_{i_m} = b_{i_m} - b'_{i_m} \neq \theta$, $a_{i_m} \in L_{i_m}$, $m = \overline{1,q}$. Значит, векторы $a_{i_1}, a_{i_2}, \ldots, a_{i_q}$ линейно зависимы и, следовательно, любая система ненулевых векторов, взятых по одному из каждого L_i, $i = \overline{1,k}$, содержащая эти векторы, линейно зависима (в силу теоремы 14.3). Это противоречит утверждению 6.

$4 \iff 7$. Это следует из теоремы 65.3. ■

Теорема 66.2. *Линейное пространство V является прямой суммой своих подпространств L_1 и L_2 тогда и только тогда, когда:*

1) $\dim V = \dim L_1 + \dim L_2$;
2) $L_1 \cap L_2 = \{\theta\}$.

Доказательство. Необходимость следует из теоремы 66.1 (утверждения 4,7).

Достаточность. Из условия 2 следует, что $L_1 + L_2$ – прямая сумма. Положим $L = L_1 \oplus L_2$. Согласно теореме 66.1 (утверждение 4) $\dim L = \dim L_1 + \dim L_2$. Отсюда в силу условия 1 вытекает, что $\dim L = \dim V$. Это означает, что $L = V$ (теорема 64.3), т.е. $V = L_1 \oplus L_2$. ■

Дополнительное подпространство. Пусть L – линейное подпространство пространства V. Подпространство L^δ называется ***дополнительным подпространством к** L* (рис. 1), если $L \oplus L^\delta = V$.

Очевидно, что при этом L – дополнительное подпространство к L^δ.

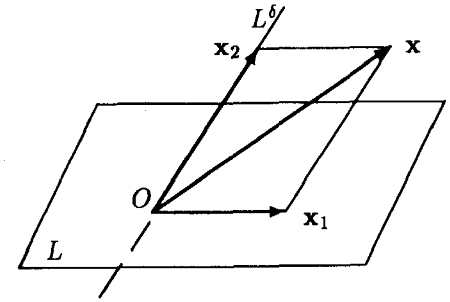


Рис. 1

Теорема 66.3. *Для любого подпространства L линейного пространства V существует дополнительное подпространство.*

Доказательство. Будем считать, что L – нетривиальное подпространство (если $L = \{\theta\}$, то $L^\delta = V$, а если $L = V$, то $L^\delta = \{\theta\}$). Пусть $e_1, \ldots, e_k$ – базис L. Дополним его до базиса $e_1, \ldots, e_k, e_{k+1}, \ldots, e_n$ всего пространства V. Тогда $\mathcal{L}(e_{k+1}, \ldots, e_n) = L^\delta$, так как для подпространства $\mathcal{L}(e_{k+1}, \ldots, e_n)$ выполнены все условия теоремы 66.2. ■

Замечание. В случае если L – нетривиальное подпространство, дополнительное подпространство определено неоднозначно.

§67. Линейное многообразие

Линейное многообразие и связанные с ним понятия были определены в §18. Там же приведены некоторые элементарные утверждения, относящиеся к вещественному линейному пространству. Все они остаются в силе и в линейном пространстве над произвольным полем P (с очевидной заменой поля $\mathbb{R}$ на поле P).

Линейные многообразия $H_1 = x_1 + L_1$ и $H_2 = x_2 + L_2$ в линейном пространстве V называются ***параллельными***, если либо $L_1 \subset L_2$, либо $L_2 \subset L_1$ (рис. 1).

Теорема 67.1. *Если два линейных многообразия с непустым пересечением параллельны, то одно из них содержит другое.*

Доказательство. Пусть линейные аффинные многообразия $H_1 = x_1 + L_1$ и $H_2 = x_2 + L_2$ параллельны и пусть

$$L_1 \subset L_2\,. \qquad (67.1)$$

По условию существует вектор $x_0 \in H_1 \cap H_2$. Так как вектором сдвига может быть любой вектор линейного многообразия (§18), то $H_1 = x_0 + L_1$, $H_2 = x_0 + L_2$. Отсюда с учетом (67.1) следует, что $H_1 \subset H_2$. ■

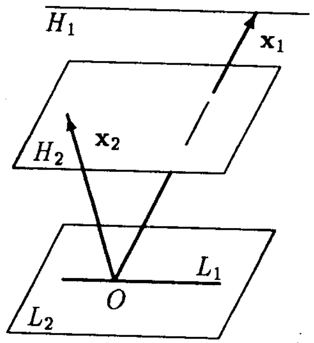


Рис. 1

Следствие 1. *Если линейные многообразия параллельны, то либо они не пересекаются, либо одно из них содержится в другом.*

Теорема 67.2. *Непустое пересечение линейных многообразий $H_1 = x_1 + L_1$ и $H_2 = x_2 + L_2$ является линейным многообразием с направляющим подпространством $L_1 \cap L_2$.*

Доказательство. По условию существует вектор $x_0 \in H_1 \cap H_2$, следовательно, $H_1 = x_0 + L_1$, $H_2 = x_0 + L_2$. Тогда $H_1 \cap H_2 = x_0 + L_1 \cap L_2$, так как для этих множеств имеет место двустороннее вложение. ■

Теорема 67.3. *Всякое k-мерное линейное многообразие в n-мерном пространстве можно задать в виде пересечения $n-k$ гиперплоскостей.*

Доказательство. Пусть $H = x_0 + L$ – k-мерное линейное многообразие и $e_1, \dots, e_k$ – базис L. Дополним его до базиса $e_1, \dots, e_k, e_{k+1}, \dots, e_n$ пространства V. В качестве искомых гиперплоскостей можно взять линейные многообразия $H_i = x_0 + L_i$, $i = \overline{1, n-k}$, где

$$\begin{gathered} L_1 = \mathcal{L}(e_1, \dots, e_k, e_{k+2}, e_{k+3}, \dots, e_n), \\ L_2 = \mathcal{L}(e_1, \dots, e_k, e_{k+1}, e_{k+3}, \dots, e_n), \\ \dots \quad \dots \quad \dots \quad \dots \\ L_{n-k} = \mathcal{L}(e_1, \dots, e_k, e_{k+1}, e_{k+2}, \dots, e_{n-1}), \end{gathered}$$

т.е. L_i – линейная оболочка всех векторов базиса V, кроме e_{k+i}.

Очевидно, что $L_1 \cap L_2 \cap \ldots \cap L_{n-k} = \mathcal{L}(e_1, \dots, e_k) = L$. В силу теоремы 67.2 отсюда следует, что $H = H_1 \cap H_2 \cap \ldots \cap H_{n-k}$. Так как $\dim L_i = n-1$, $i = \overline{1, n-k}$, то все H_i – гиперплоскости. ■

Теорема 67.4. *Пересечение линейного многообразия $H = x_0 + L$ с любым подпространством, дополнительным к L, состоит ровно из одного вектора.*

Доказательство. Пусть L^δ – дополнительное подпространство к L. Так как $L \oplus L^\delta = V$, то для вектора x_0 имеет место разложение $x_0 = y + z$, где $y \in L$, $z \in L^\delta$. Тогда $z = x_0 - y \in H$, ибо $-y \in L$. Следовательно, z – общий вектор L^δ и H, т.е. $z \in L^\delta \cap H$.

Покажем, что z – единственный вектор пересечения $L^\delta \cap H$. Пусть $z' \in L^\delta \cap H$, тогда так как $z \in H$, $z' \in H$, то (§18) $z - z' \in L$, а так как $z \in L^\delta$, $z' \in L^\delta$, то $z - z' \in L^\delta$. Из того, что $L \cap L^\delta = \{\theta\}$, следует, что $z - z' = \theta$, т.е. $z = z'$. ■

Фактор-пространство. Пусть V – линейное пространство над полем P, L – некоторое его линейное подпространство. Рассмотрим множество $V|L$ всех линейных многообразий с направляющим подпространством L: $V|L = \{H = x + L \mid x \in V\}$. Введем на этом множестве операции сложения и умножения на число из поля P.

Определим сумму многообразий $H_1 = x_1 + L$, $H_2 = x_2 + L$ по следующему правилу:

$$H_1 + H_2 = (x_1 + x_2) + L. \tag{67.2}$$

Это определение корректно, так как результат не зависит от выбора векторов сдвига многообразий H_1, H_2. Фактически эта операция совпадает с известной операцией сложения смежных классов по нормальному делителю, т.е. с алгебраической операцией в фактор-группе (§40). Таким образом, правило (67.3) определяет внутренний закон композиции на $V|L$ (рис. 2).

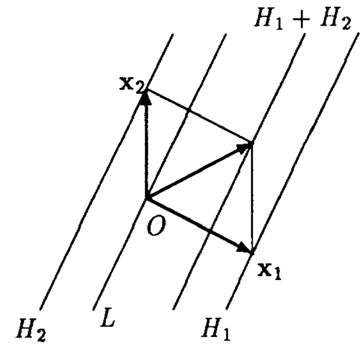


Рис. 2

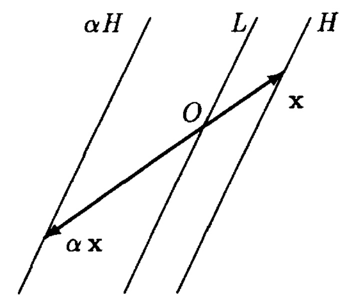


Рис. 3

Определим умножение многообразия $H = x_0 + L$ на число $\alpha \in P$ по следующему правилу:

$$\alpha H = \alpha x_0 + L. \tag{67.3}$$

Проверим корректность этого определения, т.е. что результат не зависит от выбора вектора сдвига x_0. Пусть $H = x_0' + L$. Покажем, что

$$\alpha x_0 + L = \alpha x_0' + L. \tag{67.4}$$

Для любого элемента $\alpha x_0 + y$, $y \in L$, из левого множества имеем $\alpha x_0 + y = \alpha(x_0' + x_0 - x_0') + y = \alpha x_0' + \alpha(x_0 - x_0') + y = \alpha x_0' + y_1$, где

$y_1 = \alpha(x_0 - x_0') + y \in L$, так как $x_0 - x_0' \in L$, $y \in L$. Следовательно, $\alpha x_0 + L \subset \alpha x_0' + L$. С другой стороны, для элемента из правого множества $\alpha x_0' + y$, $y \in L$, имеем $\alpha x_0' + y = \{x_0' = x_0 + y_1$, где $y_1 \in L\} = \alpha(x_0+y_1)+y = \alpha x_0+\alpha y_1+y = \alpha x_0+y_2$, где $y_2 = \alpha y_1+y \in L$. Следовательно, $\alpha x_0' + L \subset \alpha x_0' + L$. Оба вложения доказывают (67.4).

Таким образом, правило (67.3) определяет внешний закон композиции на $V|L$ (рис. 3).

Теорема 67.5. *Множество $V|L$ всех линейных многообразий с направляющим подпространством L есть линейное пространство относительно законов композиции* (67.2), (67.3).

Доказательство. Справедливость аксиом группы сложения вытекает из того, что $V|L$ – фактор-группа. Впрочем, непосредственная проверка этих аксиом так же проста, как и проверка остальных аксиом линейного пространства, ее мы предоставляем читателю. ■

Линейное пространство $V|L$ называется *фактор-пространством линейного пространства V по подпространству L*.

Теорема 67.6. *Фактор-пространство $V|L$ изоморфно L^δ.*

Доказательство. Если $L = \{\theta\}$, то $L^\delta = V$, $V|L = V$ и, следовательно, $V|L \cong L^\delta$. Если $L = V$, то $L^\delta = \{\theta\}$, $V|L = \{\theta\}$ и $V|L \cong \{\theta\}$. Пусть $0 < \dim L < \dim V$ и L^δ – какое-нибудь дополнительное подпространство к L. Согласно теореме 67.4 в каждом многообразии $H = x + L$ существует, и притом единственный, вектор $z \in L^\delta$, так что $H = z+L$. Таким образом, $V|L = \{H = z+L \,|\, z \in L^\delta\}$. Отображение $\varphi : L^\delta \to V|L$, которое каждому вектору $z \in L^\delta$ ставит в соответствие линейное аффинное многообразие $H = z + L \in V|L$, биективно в силу теоремы 67.4, оно сохраняет законы композиции в силу (67.2), (67.3). Таким образом, φ – изоморфизм. ■

Следствие 2. $\dim V|L = \dim V - \dim L$.

Глава XIII. Евклидовы и унитарные пространства

Линейные пространства, изучавшиеся нами до сих пор, несмотря на абстрактный характер, по своим свойствам очень близки к геометрическим пространствам. Однако в них еще не нашли отражение понятия, связанные с измерениями, такие, как длина вектора и угол между векторами. В §22 мы видели, что эти величины тесно связаны с понятием скалярного произведения векторов. Они просто выражаются через скалярное произведение, а само скалярное произведение векторов оказывается при этом весьма простой с точки зрения алгебры функцией: оно линейно по каждой из двух переменных. Это дает основание при введении метрических понятий в теории линейных пространств отталкиваться от понятия скалярного произведения.

В этой главе рассматриваются только вещественные и комплексные пространства. Итак, V – линейное пространство над полем P, где P – либо $\mathbb{R}$, либо $\mathbb{C}$.

§68. Скалярное произведение

Пусть V – вещественное или комплексное линейное пространство. Отображение

$$(\,,\,): \quad V \times V \to P$$

называется ***скалярным произведением***, если оно удовлетворяет следующим аксиомам: для любых $x, y, z \in V$ и любого $\alpha \in P$

$$\begin{aligned} &1)\ \ (x,y) = \overline{(y,x)}\,;\\ &2)\ \ (\alpha x, y) = \alpha(x,y)\,;\\ &3)\ \ (x+y, z) = (x,z) + (y,z)\,;\\ &4)\ \ (x,x) \ge 0 \text{ для любого } x \in V,\\ &(x,x) = 0 \text{ тогда и только тогда, когда } x = \theta\,. \end{aligned} \tag{68.1}$$

Число (x,y) называется ***скалярным произведением векторов x и y***, аксиомы 1 – 4 называются ***аксиомами скалярного произведения***.

Замечание 1. В вещественном случае черта в первой аксиоме может быть опущена.

Замечание 2. Аксиома 4 в комплексном случае на первый взгляд кажется парадоксальной, так как для комплексных чисел знак не определен. Однако из первой аксиомы следует, что скалярный квадрат (x,x) есть вещественное число.

Вещественное линейное пространство со скалярным произведением называется ***евклидовым пространством***, а комплексное – ***унитарным***. Обозначение: E и U соответственно.

Примеры. 1. В геометрических пространствах V_n, где $n = 1, 2, 3$, было введено скалярное произведение (22.1) на основании метрики, имеющейся на прямой, на плоскости и в пространстве. Оно удовлетворяет всем аксиомам скалярного произведения (теорема 22.2) и, следовательно, является скалярным произведением и в принятом выше смысле. Аксиоматическое определение скалярного произведения говорит о том, что геометрическое правило (22.1) не является единственным скалярным произведением в пространствах V_n, $n = 1, 2, 3$.

2. В арифметическом пространстве $\mathbb{R}^n$ скалярное произведение векторов $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ может быть введено следующим образом:

$$(x, y) = \sum_{i=1}^{n} x_i y_i\,, \tag{68.2}$$

а в арифметическом пространстве $\mathbb{C}^n$ –

$$(x, y) = \sum_{i=1}^{n} x_i \overline{y_i}\,. \tag{68.3}$$

Для (68.2), (68.3) аксиомы (68.1) легко проверяются непосредственно.

3. В функциональном пространстве $C[0,1]$ скалярное произведение функций $f(x)$ и $g(x)$ может быть задано так:

$$(f, g) = \int_0^1 f(x)g(x)dx\,. \tag{68.4}$$

Справедливость аксиом (68.1) вытекает из свойств определенного интеграла [7].

Итак, операции (68.2)–(68.4) определены корректно и превращают пространства V_n, $\mathbb{R}^n$, $C[0,1]$ в евклидовы пространства, а пространство $\mathbb{C}^n$ – в унитарное.

Из определения скалярного произведения вытекают следующие *простейшие свойства* этой операции.

1°. $(x, y+z) = (x, y) + (x, z)$, $\forall x, y, z \in E\ (U)$.

2°. $(x, \alpha x) = \overline{\alpha}(x, y)$, $\forall x, y \in E\ (U)$, $\forall \alpha \in \mathbb{R}$ (соответственно $\mathbb{C}$).

3°. $(\theta, x) = (x, \theta) = 0$, $\forall x \in E\ (U)$.

4°. $(x, y) = 0$ для любого вектора $y \in E\ (U)$ тогда и только тогда, когда $x = \theta$.

5°. Любое подпространство L евклидова (унитарного) пространства является евклидовым (соответственно унитарным) пространством.

Отметим, что скалярное произведение (x, y) векторов x, y линейно по первому аргументу, а в евклидовом пространстве – линейно по обоим аргументам (свойства 1°, 2°).

Теорема 68.1. *Для любых векторов $x, y \in E\ (U)$ имеет место неравенство*

$$|(x, y)|^2 \le (x, x)(y, y) \tag{68.5}$$

или, в другой форме,

$$\begin{vmatrix} (x,x) & (x,y) \\ (y,x) & (y,y) \end{vmatrix} \ge 0\,.$$

Доказательство. Будем считать, что $x \ne \theta$ (для $x = \theta$ неравенство (68.5), очевидно, обращается в равенство). Для любых векторов $x, y \in E\ (U)$ и любого числа $\alpha \in \mathbb{R}\ (\mathbb{C})$, согласно (68.1) и свойствам 1°–4°, имеем

$$0 \le (\alpha x - y, \alpha x - y) = (\alpha x, \alpha x) - (\alpha x, y) - (y, \alpha x) + (y, y) =$$
$$= |\alpha|^2 (x, x) - \alpha(x, y) - \overline{\alpha}(y, x) + (y, y)\,. \tag{68.6}$$

Так как $x \ne \theta$, то $(x, x) \ne 0$. Возьмем $\alpha = (y, x)/(x, x)$. Тогда из (68.6) имеем $0 \le \dfrac{|(y,x)|^2}{(x,x)^2}(x,x) - \dfrac{(y,x)(x,y)}{(x,x)} - \dfrac{(x,y)(y,x)}{(x,x)} + (y,y)$, откуда после очевидных преобразований получим (68.5). ■

Неравенство (68.5) называется *неравенством Коши–Буняковского.*

Замечание 3. В евклидовом пространстве неравенство Коши–Буняковского может быть записано в виде $(x, y)^2 \le (x, x)(y, y)$.

Теорема 68.2. *Неравенство Коши–Буняковского обращается в равенство тогда и только тогда, когда векторы x и y коллинеарны.*

Доказательство этой теоремы фактически содержится в доказательстве теоремы 68.1. Действительно, если $x = \theta$, то $x = 0y$. Если $x \ne \theta$, то неравенство (68.6) обращается в равенство тогда и только тогда, когда $0 = (\alpha x - y, \alpha x - y)$, т.е. когда $y = \alpha x$. ■

§69. Основные метрические понятия

В евклидовом и унитарном пространствах *длиной* вектора x называется арифметическое значение квадратного корня из его скалярного квадрата. Обозначение: $|x|$. Итак, $|x| = \sqrt{(x,x)}$.

Из аксиом скалярного произведения вытекают следующие факты.

1°. Любой вектор x евклидова (и унитарного) пространства имеет длину, при этом $|x| \ge 0$, $\forall x \in E\ (U)$ и $|x| = 0 \Leftrightarrow x = \theta$.

2°. $|\alpha x| = |\alpha||x|$, $\forall x \in E\ (U)$, $\forall \alpha \in \mathbb{R}$ (соответственно $\mathbb{C}$).

В новой терминологии неравенство Коши–Буняковского может быть записано следующим образом:

$$|(x, y)| \le |x| \cdot |y|\,. \tag{69.1}$$

Вектор единичной длины называется ***нормированным***. Любой ненулевой вектор можно нормировать, поделив его на его длину.

Теорема 69.1. *В евклидовом (унитарном) пространстве для любых векторов x, y имеют место неравенства*

$$||x|-|y|| \le |x+y| \le |x|+|y|. \qquad (69.2)$$

Доказательство. Имеем $|x+y|^2 = (x+y, x+y) = (x,x) + (x,y)+(y,x)+(y,y)$. Применив к этой сумме числовые неравенства треугольника (которые имеют место как для вещественных, так и для комплексных чисел), с учетом неравенства (69.1) получим, что

$$|x+y|^2 \le |x|^2 + 2|x|\cdot|y| + |y|^2 = (|x|+|y|)^2,$$
$$|x+y|^2 \ge |x|^2 - 2|x|\cdot|y| + |y|^2 = (|x|-|y|)^2.$$

Отсюда следует (69.2). ■

Неравенства (69.2) называются ***неравенствами треугольника*** *в евклидовом (унитарном) пространстве.*

В евклидовом пространстве ***углом*** между ненулевыми векторами x и y называется угол φ, $0 \le \varphi \le \pi$, для которого

$$\cos\varphi = \frac{(x,y)}{|x|\cdot|y|}. \qquad (69.3)$$

Корректность определения следует из неравенства Коши–Буняковского (69.1).

В унитарном пространстве понятие угла между векторами не определено. Однако и в евклидовом и в унитарном пространстве можно ввести обобщение понятия прямого угла.

§70. Ортогональные векторы

Два вектора $x, y \in E\,(U)$ называются ***ортогональными***, если $(x,y) = 0$. Из свойства 4° скалярного произведения (§68) следует, что нулевой вектор θ, и только нулевой, ортогонален любому вектору пространства.

В евклидовом пространстве вследствие (69.3) ортогональность векторов x и y означает, что либо один из них нулевой, либо угол между ними равен $\pi/2$.

Система векторов $x_1,\dots,x_k \in E\,(U)$ называется ***ортогональной системой***, если

$$(x_i, x_j) = 0 \quad \text{при} \quad i \ne j. \qquad (70.1)$$

Система векторов $x_1,\dots,x_k \in E\,(U)$ называется ***ортонормированной системой***, если $(x_i,x_j) = \delta_{ij}$, где δ_{ij} – символ Кронекера, т.е.

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \ne j. \end{cases} \qquad (70.2)$$

Теорема 70.1. *Ортогональная система ненулевых векторов линейно независима.*

Доказательство. Пусть $x_1,\dots,x_k \in E\,(U)$ – ортогональная система ненулевых векторов. Умножая обе части равенства

$$\alpha_1 x_1 + \ldots + \alpha_i x_i + \ldots + \alpha_k x_k = \theta \qquad (70.3)$$

скалярно на x_i, получаем в силу (70.1)

$$\alpha_i (x_i, x_i) = 0, \quad i = \overline{1,k}. \qquad (70.4)$$

По условию $x_i \ne \theta$, значит, $(x_i,x_i) \ne 0$, и в силу (70.4) все коэффициенты линейной комбинации в (70.3) равны нулю. Следовательно, векторы рассматриваемой системы линейно независимы. ■

Следствие 1. *Ортонормированная система векторов линейно независима.*

Следствие 2. *В n-мерном евклидовом (унитарном) пространстве любая ортонормированная система из n векторов образует базис.*

Базис, векторы которого образуют ортонормированную систему, называется ***ортонормированным базисом***. В соответствии с (70.2) $e_1,\dots,e_k$ – ортонормированный базис $E\,(U)$, если

$$(e_i, e_j) = \begin{cases} 1, & i = j, \\ 0, & i \ne j. \end{cases} \qquad (70.5)$$

Теорема 70.2. *В евклидовом (унитарном) пространстве координаты $x_1,\dots,x_n$ вектора x в базисе $e = (e_1,\dots,e_n)$ вычисляются по правилу*

$$x_i = (x, e_i), \quad i = \overline{1,n}, \qquad (70.6)$$

тогда и только тогда, когда e – ортонормированный базис.

Доказательство. Необходимость. Пусть для любого вектора $x \in E\,(U)$ координаты в базисе e вычисляются согласно (70.6). Тогда по этому же правилу вычисляются координаты и базисных векторов, которые известны, так как

$$e_i = 0e_1 + \ldots + 0e_{i-1} + 1e_i + 0e_{i+1} + \ldots + 0e_n, \quad i = \overline{1,n}.$$

Сравнение координат вектора e_i с правилом (70.6) приводит к требуемым равенствам (70.5).

Достаточность. Если e – ортонормированный базис и $x = \sum_{j=1}^n x_j e_j$, то свойство линейности скалярного произведения с учетом (70.5) приводит к (70.6). ■

Теорема 70.3. ***В евклидовом (унитарном) пространстве скалярное произведение векторов $x = \sum_{i=1}^n x_i e_i$, $y = \sum_{i=1}^n y_i e_i$, заданных своими координатами в базисе e, вычисляется по правилу***

$$(x,y) = \sum_{i=1}^n x_i \overline{y_i} \qquad (70.7)$$

тогда и только тогда, когда e – ортонормированный базис.

Доказательство. Необходимость. Если скалярное произведение вычисляется согласно (70.7) для любой пары векторов, то это же верно для пары базисных векторов e_i и e_j, координаты которых известны. Применив правило (70.7) для вычисления скалярного произведения векторов e_i и e_j, получим требуемые равенства (70.5).

Достаточность. Если e – ортонормированный базис и $x = \sum_{i=1}^n x_i e_i$, $y = \sum_{i=1}^n y_i e_i$, то в силу свойства линейности скалярного произведения имеем

$$(x,y) = (\sum_{i=1}^n x_i e_i\,,\ \sum_{j=1}^n y_j e_j\,) = \sum_{i=1}^n \sum_{j=1}^n x_i \overline{y_j}(e_i,e_j) = \sum_{i=1}^n x_i \overline{y_i}\,. \ \blacksquare$$

Замечание. В евклидовом пространстве черта в равенстве (70.7) может быть опущена: $(x,y) = \sum_{i=1}^n x_i y_i$.

До сих пор все базисы линейного пространства были равноправны. С помощью любого базиса можно сводить операции сложения векторов и умножения вектора на число к операциям над числами. В евклидовом и унитарном пространствах теорема 70.3 отводит ортонормированному базису особую роль: с помощью такого базиса и третья операция – скалярное произведение векторов – сводится к операциям над числами. Возникает естественный вопрос, всегда ли существует ортонормированный базис.

Теорема 70.4. *В конечномерном евклидовом (унитарном) пространстве существует ортонормированный базис.*

Доказательство. Пусть $\dim V = n$. Используем индукцию по n. При $n = 1$ утверждение очевидно: достаточно взять любой вектор $f \neq \theta$ и положить $e_1 = f/|f|$.

Пусть в любом $(n-1)$-мерном евклидовом (унитарном) пространстве существует ортонормированный базис; покажем, что ортонормированный базис существует и в n-мерном пространстве $E\ (U)$. Пусть $f_1,\ldots,f_n$ – базис $E\ (U)$. Линейная оболочка $\mathcal{L}(f_1,\ldots,f_{n-1})$ является $(n-1)$-мерным пространством, и в нем по индуктивному предположению существует ортонормированный базис $e_1,\ldots,e_{n-1}$. Так как $f_n \notin \mathcal{L}(f_1,\ldots,f_{n-1}) = \mathcal{L}(e_1,\ldots,e_{n-1})$, то вектор $g_n = f_n - \alpha_1 e_1 - \alpha_2 e_2 - \ldots - \alpha_{n-1}e_{n-1}$ отличен от нулевого вектора при любых $\alpha_i \in \mathbb{R}(\mathbb{C})$. Выберем коэффициенты α_i, $i = \overline{1,n-1}$, из условия ортогональности вектора g_n всем векторам $e_1,\ldots,e_{n-1}$: $0 = (g_n,e_i) = (f_n,e_i) - \alpha_i$, $i = \overline{1,n-1}$, или $\alpha_i = (f_n,e_i)$, $i = \overline{1,n-1}$. Тогда, положив $e_n = g_n/|g_n|$, получим ортонормированный базис $e_1,\ldots,e_n$ пространства $E\ (U)$. ■

Процесс ортогонализации. Доказательство теоремы 70.4, по существу, представляет собой алгоритм последовательного построения ортонормированного базиса по заданному базису $f_1,\ldots,f_n$.

Первый шаг. Полагая $g_1 = f_1$, находим $e_1 = g_1/|g_1|$.

k-й шаг **$(k \geq 2)$.** Полагаем

$$g_k = f_k - \alpha_1 e_1 - \alpha_2 e_2 - \ldots - \alpha_{k-1}e_{k-1}\,, \qquad (70.8)$$

где $\alpha_i = (f_k, e_i)$, $i = \overline{1,k-1}$, и находим $e_k = g_k/|g_k|$.

Через n шагов получим ортонормированный базис $e_1,\ldots,e_n$ пространства. Описанный алгоритм называется ***процессом ортогонализации Грама–Шмидта*** и состоит в следующем.

Ортогональная (унитарная) матрица. Матрица $U \in \mathbb{C}^{n\times n}$ называется *унитарной*, если

$$UU^H = U^H U = I\,, \qquad (70.9)$$

матрица $Q \in \mathbb{R}^{n\times n}$ называется *ортогональной*, если

$$QQ^T = Q^T Q = I\,. \qquad (70.10)$$

Напомним, что в §24 уже рассматривались ортогональные матрицы в связи с ортонормированным базисом геометрических пространств. В произвольном евклидовом (унитарном) пространстве имеет место то же утверждение.

Теорема 70.5. *Матрица ортогональна (унитарна) тогда и только тогда, когда она является матрицей перехода от одного ортонормированного базиса к другому ортонормированному базису евклидова (унитарного) пространства.*

Доказательство повторяет доказательство теоремы 24.1 с учетом соотношений (70.5), (70.7), (70.9), (70.10). ■

§71. Матрица Грама

Матрицей Грама системы векторов $a_1,\ldots,a_k$ евклидова (унитарного) пространства называется матрица

$$G(a_1,\ldots,a_k) = \begin{bmatrix} (a_1,a_1) & (a_2,a_1) & \ldots & (a_k,a_1) \\ \ldots & \ldots & \ldots & \ldots \\ (a_1,a_k) & (a_2,a_k) & \ldots & (a_k,a_k) \end{bmatrix}. \qquad (71.1)$$

Определитель матрицы Грама называется *определителем Грама.*

Теорема 71.1. *Система векторов $a_1,\ldots,a_k$ евклидова (унитарного) пространства линейно зависима тогда и только тогда, когда* $\det G(a_1,\ldots,a_k) = 0$.

Доказательство. Необходимость. Пусть $a_1,\ldots,a_k$ – линейно зависимая система векторов. Последовательно умножая нетривиальную линейную комбинацию

$$\alpha_1 a_1 + \ldots + \alpha_k a_k = \theta \qquad (71.2)$$

скалярно на векторы $a_1,\ldots,a_k$, получим однородную **систему** уравнений относительно неизвестных $\alpha_1,\ldots,\alpha_k$:

$$\begin{cases} \alpha_1(a_1,a_1) + & \ldots & +\alpha_k(a_k,a_1) = 0, \\ \ldots & \ldots & \ldots \\ \alpha_k(a_1,a_k) + & \ldots & +\alpha_k(a_k,a_k) = 0 \end{cases} \qquad (71.3)$$

с матрицей коэффициентов $G(a_1,\dots,a_k)$. Из существования нетривиального решения полученной системы уравнений следует (теорема 28.6), что $\det G(a_1,\dots,a_k)=0$.

Достаточность. Пусть $\det G(a_1,\dots,a_k)=0$. Тогда система (71.3) имеет нетривиальное решение $\alpha_1,\dots,\alpha_k$. Перепишем систему (71.3) в виде

$$\left\{\begin{array}{l} \left(\sum_{i=1}^k \alpha_i a_i, a_1\right)=0, \\ \dots\ \dots\ \dots\ \dots \\ \left(\sum_{i=1}^k \alpha_i a_i, a_k\right)=0. \end{array}\right. \tag{71.4}$$

Это означает, что вектор $g=\sum_{i=1}^k \alpha_i a_i$, с одной стороны, принадлежит $\mathcal{L}(a_1,\dots,a_k)$, а с другой стороны, ортогонален $\mathcal{L}(a_1,\dots,a_k)$. Согласно аксиоме 4 скалярного произведения вектор g может быть только нулевым. Значит, для векторов $a_1,\dots,a_k$ имеет место соотношение (71.2), откуда с учетом нетривиальности набора $\alpha_1,\dots,\alpha_k$ следует линейная зависимость $a_1,\dots,a_k$. ■

Матрица $A\in\mathbb{C}^{n\times n}$ называется *эрмитовой матрицей*, если

$$A^H=A; \tag{71.5}$$

матрица $A\in\mathbb{R}^{n\times n}$ называется *симметрической* (или *вещественной эрмитовой*) *матрицей*, если

$$A^T=A. \tag{71.6}$$

Из (71.5) следует, что $|A|\in\mathbb{R}$. Таким образом, для эрмитовых матриц (комплексных и вещественных) можно говорить о знаке определителя.

Теорема 71.2. *Матрица Грама системы векторов евклидова (унитарного) пространства эрмитова.*

Доказательство. Пусть $G(a_1,\dots,a_n)=G=(g_{ij})$. Из (71.1) следует, что $g_{ij}=(a_j,a_i)$, $g_{ji}=(a_i,a_j)$, т.е. $g_{ij}=\overline{g_{ji}}$. Это означает, что $G^H=G$ или, в вещественном случае, $G^T=G$. ■

Теорема 71.3. *Определитель Грама линейно независимой системы векторов в евклидовом (унитарном) пространстве положителен.*

Доказательство. Пусть $a_1,\dots,a_k$ – линейно независимая система векторов евклидова пространства. Тогда $\dim\mathcal{L}(a_1,\dots,a_k)=k$. Выберем ортонормированный базис $e_1,\dots,e_k$ линейной оболочки $\mathcal{L}(a_1,\dots,a_k)$. Составим матрицу

$$A=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{kk}\end{bmatrix}\in\mathbb{C}^{k\times k},$$

столбцами которой являются координаты векторов $a_1,\dots,a_k$ в базисе $e=(e_1,\dots,e_k)$. Тогда, согласно (70.7), $(a_i,a_j)=\sum_{s=1}^k a_{sj}\overline{a_{si}}=$

$\sum_{s=1}^k a_{is}^h a_{sj}=\{A^HA\}_{ij}$, $i=\overline{1,k}$, $j=\overline{1,k}$. Следовательно,

$$G(a_1,\dots,a_k)=A^HA \tag{71.7}$$

и $\det G(a_1,\dots,a_k)=|\det A|^2$. Таким образом,

$$\det G(a_1,\dots,a_k)\geq 0. \tag{71.8}$$

По условию система $a_1,\dots,a_k$ линейно независима, поэтому, согласно теореме 71.1, $\det G(a_1,\dots,a_k)\neq 0$. Отсюда с учетом (71.8) следует, что $\det G(a_1,\dots,a_k)>0$. ■

Замечание. Равенство (71.7) дает компактную форму записи матрицы Грама, в частности в вещественном случае

$$G(a_1,\dots,a_k)=A^TA.$$

§72. Ортогональное дополнение

Задача о перпендикуляре. Пусть L – линейное подпространство евклидова (унитарного) пространства E (U). Вектор x называется *ортогональным к подпространству* L, если он ортогонален любому вектору $y\in L$. Обозначение: $x\perp L$.

Очевидно, что $x\perp\mathcal{L}(a_1,\dots,a_k)$ тогда и только тогда, когда $x\perp a_i$, $i=\overline{1,k}$. Совокупность всех векторов $x\in E$ (U), ортогональных подпространству L, называется *ортогональным дополнением* к L. Обозначение: $L^\perp$.

Теорема 72.1. *Ортогональное дополнение к подпространству является линейным подпространством.*

Доказательство. Пусть $y_1,y_2\in L^\perp$, тогда $(y_1,x)=(y_2,x)=0$ для любого вектора $x\in L$. Складывая эти равенства, получим, что $(y_1+y_2,x)=0$, $\forall x\in L$, т.е. $y_1+y_2\in L^\perp$. Аналогично если $(y,x)=0$, $\forall x\in L$, то $(\alpha y,x)=0$, т.е. $\alpha y\in L$. Отсюда на основании теоремы 18.1 следует, что $L^\perp$ – линейное подпространство. ■

Теорема 72.2. *Если L – линейное подпространство $E(U)$, то*

$$L\oplus L^\perp=E\,(U). \tag{72.1}$$

Доказательство. Утверждение очевидно, если L – тривиальное подпространство. Пусть L – нетривиальное подпространство. Возьмем $e_1,\dots,e_k$ – ортонормированный базис L, $e_{k+1},\dots,e_n$ – ортонормированный базис $L^\perp$. Система векторов $e_1,\dots,e_k,e_{k+1},\dots,e_n$ ортонормирована и, следовательно, линейно независима (теорема 70.1). Покажем, что она образует базис всего пространства E (U). Пусть это не так. Тогда существует вектор f пространства, который не является линейной комбинацией $e_1, \dots, e_n$. Система векторов $e_1, \dots, e_n, f$ линейно независима, и применение к ней процесса ортогонализации приводит к вектору e_{n+1}, который ортогонален $e_1, \dots, e_n$ и, значит, $e_{n+1} \in L^\perp$. С другой стороны, $e_{n+1} \perp L^\perp$, так как e_{n+1} ортогонален $e_{k+1}, \dots, e_n$. Следовательно, $e_{n+1} = \theta$. Отсюда согласно (70.8) вытекает линейная зависимость $e_1, \dots, e_n, f$, что противоречит допущению. Таким образом, система векторов $e_1, \dots, e_n$ является базисом $E\ (U)$ и $\dim L + \dim L^\perp = \dim E\ (\dim U)$. Так как $L \cap L^\perp = \{\theta\}$, то согласно теореме 66.2 получаем (72.1). ■

Следствие. Если L – линейное подпространство $E\ (U)$, то для любого вектора $f \in E\ (U)$ существует, и притом единственное, разложение

$$f = g + h, \tag{72.2}$$

где $g \in L$, $h \perp L$.

Вектор g в разложении (72.2) называется *ортогональной проекцией вектора f на подпространство L*, а вектор h – *ортогональной составляющей вектора f*.

Задачу разложения (72.2) вектора на ортогональную проекцию и ортогональную составляющую называют *задачей о перпендикуляре*. Этот термин заимствован из геометрии. С разложением (72.2) геометрического вектора мы уже встречались (рис. 1): чтобы получить это разложение, достаточно опустить перпендикуляр из конца вектора f на плоскость L. Имея в виду эту аналогию, ортогональную составляющую h в разложении (72.2) называют *перпендикуляром, опущенным из вектора f на подпространство L*, а сам вектор f – *наклонной к подпространству L*.

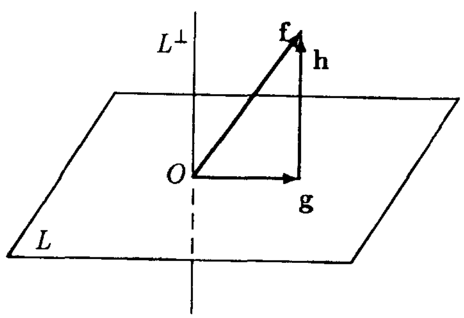


Рис. 1

Аналогия с геометрическими векторами состоит не только в названии. Отметим несколько тех свойств составляющих g и h в разложении (72.2), которые имеют место и в геометрии: так как $|f|^2 = (g+h, g+h) = (g,g) + (h,h)$, то

$$|f|^2 = |g|^2 + |h|^2, \tag{72.3}$$

$$|h| \le |f|.$$

Последнее неравенство свидетельствует о том, что *длина перпендикуляра не превосходит длины наклонной*. Равенство (72.3) называется *теоремой Пифагора в евклидовом (унитарном) пространстве*.

Решение задачи о перпендикуляре. Из теоремы 72.2 следует, что задача о перпендикуляре имеет, и притом единственное, решение. Укажем один из способов построения этого решения.

Пусть $L = \mathcal{L}(a_1, \dots, a_k)$. Разложение (72.2) равносильно задаче построения векторов g и h таких, что

$$\begin{cases} f = g + h, \\ g \in L, \\ h \perp L \end{cases} \Leftrightarrow \begin{cases} g = \sum_{j=1}^k x_j a_j, \\ f - g \perp L, \\ h = f - g \end{cases} \Leftrightarrow \begin{cases} (f - g, a_i) = 0, \ i = \overline{1,k}, \\ g = \sum_{j=1}^k x_j a_j, \\ h = f - g \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} (g, a_i) = (f, a_i), \ i = \overline{1,k}, \\ g = \sum_{j=1}^k x_j a_j, \\ h = f - g \end{cases} \Leftrightarrow \begin{cases} (\sum_{j=1}^k x_j a_j, a_i) = (f, a_i), \ i = \overline{1,k}, \\ g = \sum_{j=1}^k x_j a_j, \\ h = f - g. \end{cases}$$

Итак, задача о перпендикуляре сводится к поиску чисел $x_1, \dots, x_k$, для которых

$$(\sum_{j=1}^k x_j a_j, a_i) = (f, a_i), \ i = \overline{1,k}. \tag{72.4}$$

Сравнение этих соотношений с (71.4) и (71.3) говорит о том, что числа $x_1, \dots, x_k$ являются решением системы уравнений с матрицей Грама $G(a_1, \dots, a_k)$ и правой частью (f, a_i), $i = \overline{1,k}$. Эта система всегда имеет решение, так как она равносильна задаче о перпендикуляре.

Отметим возможные при этом варианты.

1°. Если $a_1, \dots, a_k$ – ортонормированный базис L, то $G(a_1, \dots, a_k) = I$ и система (72.4) имеет единственное решение $x_i = (f, a_i)$, $i = \overline{1,k}$, при этом $g = \sum_{i=1}^k x_i a_i$, $h = f - g$.

2°. Если $a_1, \dots, a_k$ линейно независимы, то $\det G(a_1, \dots, a_k) \ne 0$ и система (72.4) имеет единственное решение; решив систему (72.4), найдем $g = \sum_{i=1}^k x_i a_i$, $h = f - g$.

3°. Если $a_1, \dots, a_k$ линейно зависимы, то $\det G(a_1, \dots, a_k) = 0$ и система (72.4) имеет бесконечно много решений, которые соответствуют бесконечному числу разложений вектора g по линейно зависимой системе векторов $a_1, \dots, a_k$.

§73. Линейное многообразие в евклидовом (унитарном) пространстве

Пусть $H = x_0 + L$ – линейное многообразие в евклидовом (унитарном) пространстве. Вектор $n \in H$, ортогональный L, называется ***нормальным вектором линейного многообразия H*** (рис. 1).

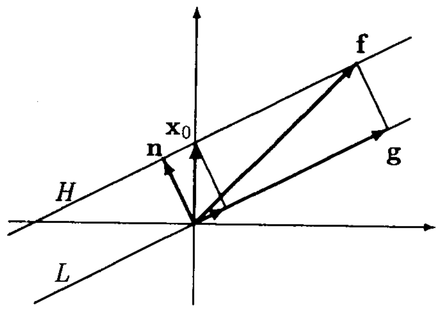


Рис. 1

Теорема 73.1. *Для любого линейного многообразия в евклидовом (унитарном) пространстве существует, и притом единственный, нормальный вектор.*

Доказательство. Рассмотрим линейное многообразие $H = x_0 + L$ (рис. 1). Все векторы из H, ортогональные L, находятся в $H \cap L^\perp$, но пересечение $H \cap L^\perp$ состоит ровно из одного вектора n, так как $L^\perp$ – дополнительное подпространство к L (теоремы 72.2 и 67.4). Этот вектор n и будет единственным нормальным вектором H. ■

Теорема 73.2. *Нормальный вектор линейного многообразия совпадает с перпендикуляром, опущенным из любого вектора линейного многообразия на направляющее подпространство.*

Доказательство. Пусть n – нормальный вектор линейного многообразия $H = x_0 + L$, тогда $H = n + L$ (§18). Следовательно, любой вектор $f \in H$ (рис. 1) может быть представлен в виде

$$f = n + g, \quad g \in L. \tag{73.1}$$

Так как $n \perp L$, то соотношение (73.1) совпадает с разложением вектора f на ортогональную проекцию g и перпендикуляр n. ■

Следствие. Среди всех векторов линейного многообразия нормальный вектор имеет наименьшую длину.

Уравнение гиперплоскости. Пусть $H = x_0 + L$ – гиперплоскость в E (U), т.е. $\dim L = n - 1$, где $n = \dim E$ $(\dim U)$. Тогда $L^\perp$ – одномерное подпространство и его базис состоит из одного вектора n.

Вектор $x \in H$ тогда и только тогда, когда разность $x - x_0 \in L$, т.е. когда

$$(x - x_0, n) = 0. \tag{73.2}$$

Таким образом, уравнению (73.2) удовлетворяют все векторы x гиперплоскости H, и только они. Уравнение (73.2) может быть записано в равносильной форме:

$$(x, n) = p, \tag{73.3}$$

где $p = (x_0, n)$ – фиксированное число для данной гиперплоскости.

Уравнению (73.3) можно придать координатную форму: если $(x_1, \ldots, x_n)$ и $(A_1, \ldots, A_n)$ – координаты векторов x и n в некотором ортонормированном базисе пространства, то уравнение (73.3) равносильно уравнению $A_1 x_1 + \ldots + A_n x_n = p$, которое в геометрических пространствах совпадает с общим уравнением прямой на плоскости и плоскости в пространстве.

§74. Расстояние в евклидовом (унитарном) пространстве

Множество M называется ***метрическим пространством***, если задано отображение

$$\rho : M \times M \to \mathbb{R},$$

которое каждой упорядоченной паре элементов $x, y \in M$ ставит в соответствие число $\rho(x,y) \in \mathbb{R}$ такое, что:

1) $\rho(x,y) \geq 0, \quad \forall x, y \in M,$
$\rho(x,y) = 0 \iff x = y;$
2) $\rho(x,y) = \rho(y,x), \quad \forall x, y \in M;$
3) $\rho(x,z) \leq \rho(x,y) + \rho(y,z), \quad \forall x, y, z \in M.$

Число $\rho(x,y)$ называется ***расстоянием между x и y***; отображение ρ – ***метрикой***, аксиомы 1–3 – ***аксиомами метрики (расстояния)***.

Расстоянием между множествами X и Y в метрическом пространстве называется число

$$\rho(X,Y) = \inf_{x \in X, y \in Y} \rho(x,y). \tag{74.1}$$

Теорема 74.1. *В евклидовом (унитарном) пространстве V правило*

$$\rho(x,y) = |x - y| \tag{74.2}$$

задает метрику.

Доказательство. Действительно, правило (74.2) определяет отображение $\rho : V \times V \to \mathbb{R}$, которое отвечает всем аксиомам метрики. Проверка аксиом тривиальна, отметим лишь одну из них: $\rho(x,z) = |x - z| = |(x-y) + (y-z)| \leq |x-y| + |y-z| = \rho(x,y) + \rho(y,z)$, $\forall x, y, z \in V$. ■

Итак, евклидово (унитарное) пространство является метрическим пространством относительно метрики (74.2). В дальнейшем, говоря об евклидовом (унитарном) пространстве как о метрическом, будем иметь в виду именно эту метрику.

Теорема 74.2 (о кратчайшем расстоянии). *Расстояние между вектором f и линейным подпространством L в евклидовом (унитарном) пространстве равно длине перпендикуляра, опущенного из вектора f на L.*

Доказательство. Пусть $f = g + h$, где $g \in L$, $h \in L^\perp$, и y – произвольный вектор из L. Тогда $\rho(f,y) = |f - y| = |(g+h) - y| =$

$|h+(g-y)| = \{$ в силу (72.3) $\} = |h|^2 + |g-y|^2$. Отсюда следует, что $\rho(f,y) \geq |h|$, $\forall y \in L$ и $\rho(f,y) = |h|$, если $y = g$. Это означает, что $|h| = \inf\limits_{y\in L} \rho(f,y) = \rho(f,L)$. ■

Эта теорема может быть сформулирована и в других терминах, а именно:

1) *расстояние между вектором f и подпространством L равно расстоянию между вектором f и его ортогональной проекцией на L;*

2) *среди всех векторов подпространства L ближе всего к вектору f расположена его ортогональная проекция на L.*

Теорема 74.3. *Пусть $H = x_0 + L$ – линейное аффинное многообразие в евклидовом (унитарном) пространстве. Тогда*

$$\rho(f,H) = \rho(f - x_0, L).$$

Это утверждение следует из теоремы 74.2 и того, что для любого вектора $z = x_0 + y \in H$: $\rho(f,z) = |f-z| = |(f-x_0)-y| = \rho(f-x_0,y)$, где $y \in L$. ■

Теорема 74.4. *Пусть $H_1 = x_1 + L_1$ и $H_2 = x_2 + L_2$ – линейные аффинные многообразия в евклидовом (унитарном) пространстве. Тогда*

$$\rho(H_1, H_2) = \rho(x_1 - x_2, L_1 + L_2).$$

Это утверждение следует из того, что для любых векторов $z_1 = x_1 + y_1 \in H_1$ и $z_2 = x_2 + y_2 \in H_2$: $\rho(z_1,z_2) = |z_1 - z_2| = |(x_1-x_2) - (y_1-y_2)| = \rho(x_1 - x_2, y)$, где $y = y_1 - y_2 \in (L_1 + L_2)$. ■

§75. Изометрия

Евклидовы пространства E_1 и E_2 называются *изоморфными*, если существует биективное отображение $\varphi : E_1 \to E_2$, которое сохраняет законы композиции и скалярное произведение, т.е. если:

1) $\varphi(x+y) = \varphi(x) + \varphi(y)$, $\quad \forall x,y \in E_1$;
2) $\varphi(\alpha x) = \alpha\varphi(x)$, $\quad \forall x \in E_1$, $\forall \alpha \in \mathbb{R}$;
3) $(\varphi(x), \varphi(y)) = (x,y)$, $\quad \forall x,y \in E_1$.

Само отображение φ при этом называется *изоморфизмом евклидовых пространств* или *изометрией*.

Точно так же определяется изоморфизм унитарных пространств U_1 и U_2 (с очевидным отличием в свойстве 2: $\alpha \in \mathbb{C}$). Из определения следует, что изоморфные евклидовы (унитарные) пространства изоморфны как линейные пространства.

Теорема 75.1. *Два евклидовых (унитарных) пространства изоморфны тогда и только тогда, когда равны их размерности.*

Доказательство. Необходимость вытекает из изоморфизма евклидовых (унитарных) пространств как линейных пространств.

Достаточность. Пусть V_1 и V_2 – оба евклидовы (оба унитарные) пространства и пусть $\dim V_1 = \dim V_2 = n$. Выберем в V_1 и V_2 ортонормированные базисы $e_1, \ldots, e_n$ и $e'_1, \ldots, e'_n$ и построим отображение $\varphi : V_1 \to V_2$, положив для каждого вектора $x = \sum_{i=1}^n x_i e_i \in V_1$ вектор $\varphi(x) = \sum_{i=1}^n x_i e'_i$. Из доказательства теоремы 63.1 следует, что отображение φ – изоморфизм линейных пространств V_1 и V_2. Оно сохраняет скалярное произведение, так как если $x = \sum_{i=1}^n x_i e_i$, $y = \sum_{i=1}^n y_i e'_i$, то согласно (70.7)

$$(x,y) = \sum_{i=1}^n x_i \overline{y_i}, \quad (\varphi(x), \varphi(y)) = \sum_{i=1}^n x_i \overline{y_i}. \quad ■$$

Глава XIV. **Линейные операторы**

В этой главе рассматриваются отображения (§8), которые действуют в линейных пространствах. Отображения в отличие от матриц будем обозначать рукописными латинскими буквами (например, $\mathcal{A}$), образ вектора x – символом $\mathcal{A}x$.

§76. Определение и простейшие свойства

Пусть V и W – линейные пространства над общим **полем** P. Отображение

$$\mathcal{A} : V \to W \tag{76.1}$$

называется ***линейным отображением пространства** V **в пространство*** W, если для любых $x, y \in V$, $\alpha \in P$

1) $\mathcal{A}(x+y) = \mathcal{A}x + \mathcal{A}y$;
2) $\mathcal{A}(\alpha x) = \alpha \mathcal{A}x$.

Линейное отображение (76.1) называют также *линейным преобразованием пространства V в пространство W* или *линейным оператором, действующим из пространства V в пространство W*.

Если $V = W$, то линейное отображение $\mathcal{A} : V \to V$ называют *линейным отображением (преобразованием) пространства V в себя*, а чаще – *линейным оператором, действующим в V*.

Если $W = P$, то линейное отображение (76.1) называют *линейной формой* или *линейным функционалом в пространстве V*. Линейный функционал обозначают строчными латинскими буквами (например, f), при этом образ вектора x – символом $f(x)$.

Множество всех линейных операторов, действующих из пространства V в пространство W, будем обозначать символом $\mathcal{L}(V, W)$.

В соответствии с определением равенства отображений (§8) операторы $\mathcal{A}, \mathcal{B} \in \mathcal{L}(V, W)$ равны, если $\mathcal{A}x = \mathcal{B}x$, $\forall x \in V$.

Примеры. 1. Пусть M_n – пространство вещественных многочленов степени не выше n. Отображение $\mathcal{D} : M_n \to M_n$, определенное правилом

$$\mathcal{D}p(t) = p'(t), \tag{76.2}$$

является линейным и называется *оператором дифференцирования*.

2. Пусть $V = L_1 \oplus L_2$. Отображение $\mathcal{P} : V \to V$, определенное правилом

$$\mathcal{P}x = x_1 \tag{76.3}$$

для вектора $x \in V$ с разложением $x = x_1 + x_2$, где $x_1 \in L_1$, $x_2 \in L_2$, является линейным и называется *оператором проектирования пространства V на L_1 параллельно L_2*.

Отображение $\mathcal{R} : V \to V$, определенное правилом

$$\mathcal{R}x = x_1 - x_2, \tag{76.4}$$

также является линейным и называется *оператором отражения пространства V относительно L_1 параллельно L_2*.

3. Отображение $\mathcal{O} : V \to W$, которое каждый вектор $x \in V$ переводит в нулевой вектор $\theta \in W$, является линейным и называется *нулевым оператором*.

4. Отображение $\mathcal{I} : V \to V$, которое каждый вектор $x \in V$ переводит в x, является линейным и называется *тождественным (единичным) оператором*.

5. Изоморфизм φ линейных пространств V и W, очевидно, является линейным оператором, действующим из V в W.

Простейшие свойства. Из определения вытекают следующие свойства линейных операторов.

1°. *Линейный оператор переводит нулевой вектор в нулевой вектор*, так как $\mathcal{A}\theta_1 = \mathcal{A}(0x) = 0\mathcal{A}x = \theta_2$ (здесь θ_1, θ_2 – нулевые векторы V и W соответственно).

2°. *Линейный оператор сохраняет линейные комбинации*, т.е. переводит линейную комбинацию векторов в линейную комбинацию образов с теми же коэффициентами: $\mathcal{A}\left(\sum_{i=1}^k \alpha_i x_i\right) = \sum_{i=1}^k \alpha_i \mathcal{A}x_i$.

3°. *Линейный оператор сохраняет линейную зависимость*, т.е. переводит линейно зависимую систему векторов в линейно зависимую.

Задание линейного оператора. Свойство 2° говорит о том, что для задания линейного оператора $\mathcal{A} \in \mathcal{L}(V, W)$ достаточно определить его только на векторах $e_1, \ldots, e_n$ некоторого базиса пространства V. Зная векторы $\mathcal{A}e_1, \ldots, \mathcal{A}e_n$, можно однозначно найти образ любого вектора $x = \sum_{i=1}^n x_i e_i \in V$: $\mathcal{A}x = \sum_{i=1}^n x_i \mathcal{A}e_i \in W$. Формализуем это рассуждение.

Теорема 76.1. *Пусть $e_1, \ldots, e_n$ – базис пространства V, а $g_1, \ldots, g_n$ – произвольные векторы пространства W. Тогда существует, и притом единственный, линейный оператор $\mathcal{A} \in \mathcal{L}(V, W)$, который переводит векторы $e_1, \ldots, e_n$ в векторы $g_1, \ldots, g_n$ соответственно.*

Доказательство. Построим искомый оператор, положив для каждого вектора $x = \sum_{i=1}^n x_i e_i \in V$

$$\mathcal{A}x = \sum_{i=1}^n x_i g_i. \tag{76.5}$$

Из единственности разложения вектора x по базису следует, что правило (76.5) однозначно определяет образ вектора x, при этом, как легко проверить, $\mathcal{A}e_i = g_i$, $i = \overline{1,n}$. Линейность построенного оператора вытекает из линейности координат.

Оператор $\mathcal{A}$ единствен, так как если $\mathcal{B}$ – любой другой линейный оператор, переводящий векторы $e_1, \ldots, e_n$ в $g_1, \ldots, g_n$, то

$$\mathcal{B}x = \mathcal{B}\Big(\sum_{i=1}^{n} x_i e_i\Big) = \sum_{i=1}^{n} x_i \mathcal{B}e_i = \sum_{i=1}^{n} x_i g_i = \mathcal{A}x\,, \quad \forall x \in V\,.$$

Следовательно, $\mathcal{B} = \mathcal{A}$. ■

Следствие. Линейные операторы $\mathcal{A}, \mathcal{B} \in \mathcal{L}(V,W)$ равны тогда и только тогда, когда они совпадают на векторах базиса V.

§77. Матрица линейного оператора

Пусть $e = (e_1, \ldots, e_n)$ и $f = (f_1, \ldots, f_m)$ – базисы пространств V и W. Как следует из теоремы 76.1, линейный оператор $\mathcal{A} \in \mathcal{L}(V,W)$ однозначно определяется заданием векторов $\mathcal{A}e_1, \ldots, \mathcal{A}e_n$. В свою очередь, векторы $\mathcal{A}e_i$, $i = \overline{1,n}$, однозначно определяются своими координатами в базисе f, т.е. коэффициентами разложений

$$\begin{cases} \mathcal{A}e_1 = a_{11}f_1 + a_{21}f_2 + \ldots + a_{m1}f_m\,, \\ \mathcal{A}e_2 = a_{12}f_1 + a_{22}f_2 + \ldots + a_{m2}f_m\,, \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ \mathcal{A}e_n = a_{1n}f_1 + a_{2n}f_2 + \ldots + a_{mn}f_m\,. \end{cases} \tag{77.1}$$

Матрица

$$A_{fe} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \tag{77.2}$$

называется ***матрицей оператора $\mathcal{A}$ в паре базисов e и f***. Для обозначения этой матрицы используется также символ $(\mathcal{A})_{fe}$.

Из единственности разложения вектора по базису следует, что при фиксированных базисах e и f матрица линейного оператора определена однозначно.

Теорема 77.1. *Пусть $\dim V = n$, $\dim W = m$. Тогда существует взаимно однозначное соответствие между линейными операторами из $\mathcal{L}(V,W)$ и матрицами из $P^{m\times n}$.*

Доказательство. Построим это соответствие. Зафиксируем базисы $e = (e_1, \ldots, e_n)$ и $f = (f_1, \ldots, f_m)$ пространств V и W. Поставим в соответствие каждому линейному оператору $\mathcal{A} \in \mathcal{L}(V,W)$ его матрицу A_{fe} в паре базисов e и f. Очевидно, что матрица $A_{fe} \in P^{m\times n}$ определена однозначно. Докажем биективность построенного таким образом отображения. Действительно, оно:

1) сюръективно, так как любая матрица $B = (b_{ij}) \in P^{m\times n}$ является матрицей линейного оператора из $\mathcal{L}(V,W)$, переводящего векторы

e_j в векторы $\sum_{i=1}^{m} b_{ij}f_i$, $j = \overline{1,n}$ (в силу теоремы 76.1 такой оператор существует);

2) инъективно, ибо различные операторы из $\mathcal{L}(V,W)$ не совпадают на базисных векторах и, значит, имеют разные матрицы. ■

Доказанная теорема играет важную роль в теории линейных операторов. Она позволяет описывать свойства линейных операторов через известный нам аппарат теории матриц. В дальнейшем мы увидим, что связь между операторами и матрицами более тесная, чем взаимно однозначное соответствие.

Координаты вектора и его образа. Пусть $\mathcal{A} \in \mathcal{L}(V,W)$, e и f – базисы пространств V и W.

Теорема 77.2. *Если $y = \mathcal{A}x$, то*

$$y_f = A_{fe}x_e\,. \tag{77.3}$$

Доказательство. Пусть $x = \sum_{i=1}^{n} x_i e_i$, $y = \sum_{i=1}^{m} y_i f_i$ и $A_{fe} = (a_{ij})$. Утверждение (77.3) равносильно соотношениям

$$y_i = \sum_{j=1}^{n} a_{ij}x_j\,, \quad i = \overline{1,m}\,. \tag{77.4}$$

Докажем их. Имеем $y = \mathcal{A}x = \mathcal{A}\big(\sum_{j=1}^{n} x_j e_j\big) = \sum_{j=1}^{n} x_j \mathcal{A}e_j = \{$в силу (77.1)$\} = \sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij}f_i = \sum_{i=1}^{m}\big(\sum_{j=1}^{n} a_{ij}x_j\big)f_i$. Из единственности разложения вектора y по базису f следует (77.4). ■

Матрицы оператора в различных базисах. Пусть e и $t = eC$ – два базиса пространства V с матрицей перехода C, а f и $s = fD$ – два базиса пространства W с матрицей перехода D. Одному и тому же линейному оператору $\mathcal{A} \in \mathcal{L}(V,W)$ в паре базисов e и f соответствует матрица A_{fe}, а в паре базисов t и s – матрица A_{st}.

Теорема 77.3. *Матрицы A_{fe} и A_{st} линейного оператора $\mathcal{A} \in \mathcal{L}(V,W)$ в различных парах базисов связаны соотношением*

$$A_{st} = D^{-1}A_{fe}C\,. \tag{77.5}$$

Доказательство. Для произвольного вектора $x \in V$ и его образа $y = \mathcal{A}x$ в силу (77.3) имеем

$$y_f = A_{fe}x_e\,, \quad y_s = A_{st}x_t\,. \tag{77.6}$$

В свою очередь, согласно (17.8), $x_e = Cx_t$, $y_f = Dy_s$. Подставив эти соотношения в (77.6), получим, что $Dy_s = A_{fe}Cx_t$ или $DA_{st}x_t = A_{fe}Cx_t$. Так как это соотношение имеет место для любых x_t, то $DA_{st} = A_{fe}C$. В силу невырожденности матрицы перехода отсюда следует (77.5). ■

Следствие 1. Матрицы линейного оператора в различных парах базисов эквивалентны (§16).

Следствие 2. *Ранг матрицы линейного оператора не зависит от выбора базисов.*

Имеет место более общее утверждение.

Теорема 77.4. *Две матрицы A и B над полем P одинакового размера $m \times n$ эквивалентны тогда и только тогда, когда они являются матрицами одного и того же линейного оператора $\mathcal{A} \in \mathcal{L}(V,W)$, где V и W – линейные пространства над полем P размерностей n и m соответственно.*

Доказательство. Необходимость. Пусть $A, B \in P^{m\times n}$ и $B = D^{-1}AC$. Рассмотрим любые линейные пространства V и W над полем P такие, что $\dim V = n$, $\dim W = m$. Возьмем в пространстве V произвольный базис e, а в пространстве W – базис f. В силу взаимно однозначного соответствия между $P^{m\times n}$ и $\mathcal{L}(V,W)$ (теорема 77.1) существует единственный оператор $\mathcal{A} \in \mathcal{L}(V,W)$, который в паре базисов e и f имеет матрицу A. Тогда согласно (77.5) матрица B будет матрицей этого же оператора в паре базисов $t = eC$ и $s = fD$.

Достаточность рассмотрена в теореме 77.3. ■

§78. Линейное пространство операторов

На множестве $\mathcal{L}(V,W)$ введем операции сложения операторов и умножения оператора на число $\alpha \in P$.

Суммой линейных операторов $\mathcal{A}, \mathcal{B} \in \mathcal{L}(V,W)$ называется отображение $\mathcal{C} : V \to W$, выполняемое по правилу $\mathcal{C}x = \mathcal{A}x + \mathcal{B}x$, $\forall x \in V$. Обозначение: $\mathcal{A}+\mathcal{B}$. Итак,

$$(\mathcal{A}+\mathcal{B})x = \mathcal{A}x + \mathcal{B}x\,, \quad \forall x \in V\,. \tag{78.1}$$

Произведением линейного оператора $\mathcal{A} \in \mathcal{L}(V,W)$ *на число* $\alpha \in P$ называется отображение $\mathcal{C} : V \to W$, выполняемое по правилу $\mathcal{C}x = \alpha\mathcal{A}x$, $\forall x \in V$. Обозначение: $\alpha\mathcal{A}$. Итак,

$$(\alpha\mathcal{A})x = \alpha\mathcal{A}x\,, \quad \forall x \in V\,.$$

Теорема 78.1. *Для любых операторов $\mathcal{A}, \mathcal{B} \in \mathcal{L}(V,W)$ и числа $\alpha \in P$*

$$\mathcal{A}+\mathcal{B} \in \mathcal{L}(V,W)\,, \quad \alpha\mathcal{A} \in \mathcal{L}(V,W).$$

Доказательство. Действительно, для любых $x, y \in V$ согласно (78.1) имеем $(\mathcal{A}+\mathcal{B})(x+y) = \mathcal{A}(x+y) + \mathcal{B}(x+y)$. В силу линейности $\mathcal{A}$, $\mathcal{B}$ и аксиом линейного пространства $(\mathcal{A}+\mathcal{B})(x+y) = (\mathcal{A}x+\mathcal{A}y)+(\mathcal{B}x+\mathcal{B}y) = (\mathcal{A}x+\mathcal{B}x)+(\mathcal{A}y+\mathcal{B}y) = (\mathcal{A}+\mathcal{B})x+(\mathcal{A}+\mathcal{B})y$.

Аналогично показывается, что $(\mathcal{A}+\mathcal{B})(\lambda x) = \lambda((\mathcal{A}+\mathcal{B})x)$ для любых $x \in V$, $\lambda \in P$. Следовательно, $\mathcal{A}+\mathcal{B} \in \mathcal{L}(V,W)$.

Точно так же доказывается, что $\alpha\mathcal{A} \in \mathcal{L}(V,W)$. ■

Следствие 1. *Сложение операторов и умножение оператора на число являются внутренним и внешним законами композиции на множестве $\mathcal{L}(V,W)$.*

Теорема 78.2. *Множество $\mathcal{L}(V,W)$ – линейное пространство над полем P относительно введенных выше операций.*

Доказательство. Достаточно проверить аксиомы линейного пространства, взяв в качестве нулевого элемента нулевое отображение $\mathcal{O} \in \mathcal{L}(V,W)$, а в качестве противоположного к оператору $\mathcal{A}$ отображение $-\mathcal{A} \in \mathcal{L}(V,W)$, выполняемое по правилу $(-\mathcal{A})x = -\mathcal{A}x$, $\forall x \in V$. Все аксиомы вытекают из соответствующих аксиом линейного пространства, примененных к V и W, и проверяются по единой схеме. Проверим, например, ассоциативность. Для любых $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathcal{L}(V,W)$ и любого $x \in V$

$$((\mathcal{A}+\mathcal{B})+\mathcal{C})x = (\mathcal{A}+\mathcal{B})x + \mathcal{C}x = (\mathcal{A}x+\mathcal{B}x)+\mathcal{C}x = \mathcal{A}x + (\mathcal{B}x+\mathcal{C}x)\,,$$
$$(\mathcal{A}+(\mathcal{B}+\mathcal{C}))x = \mathcal{A}x + (\mathcal{B}+\mathcal{C})x = \mathcal{A}x + (\mathcal{B}x+\mathcal{C}x)\,.$$

Таким образом, $(\mathcal{A}+\mathcal{B})+\mathcal{C} = \mathcal{A}+(\mathcal{B}+\mathcal{C})$. ■

Теорема 78.3. *Если $\dim V = n$, $\dim W = m$, то линейное пространство $\mathcal{L}(V,W)$ изоморфно пространству матриц $P^{m\times n}$.*

Доказательство. Зафиксируем базисы e и f пространств V и W. Построим отображение $\varphi : \mathcal{L}(V,W) \to P^{m\times n}$, положив $\varphi(\mathcal{A}) = A_{fe}$. Это отображение биективно в силу теоремы 77.1. Покажем, что оно сохраняет законы композиции, т.е. что

$$(\mathcal{A}+\mathcal{B})_{fe} = A_{fe} + B_{fe}\,, \quad (\alpha\mathcal{A})_{fe} = \alpha A_{fe}\,. \tag{78.2}$$

Пусть $A_{fe} = (a_{ij})$, $B_{fe} = (b_{ij})$. Тогда, согласно (77.1), $\mathcal{A}e_j = \sum_{i=1}^m a_{ij}f_i$, $\mathcal{B}e_j = \sum_{i=1}^m b_{ij}f_i$, поэтому $(\mathcal{A}+\mathcal{B})e_j = \mathcal{A}e_j + \mathcal{B}e_j = \sum_{i=1}^m (a_{ij}+b_{ij})f_i$. По определению (77.2) матрицы линейного оператора отсюда следует первое из соотношений (78.2).

Аналогично проверяется второе соотношение. ■

Следствие 2. $\dim \mathcal{L}(V,W) = \dim V \cdot \dim W$.

Замечание. Соотношения (78.2) означают, что при сложении операторов их матрицы складываются, при умножении оператора на число его матрица умножается на это же число.

§79. Умножение линейных операторов

Тот факт, что линейные операторы образуют линейное пространство, используется редко. Дело в том, что о линейных операторах можно сказать гораздо больше, используя еще одну операцию – умножение операторов, знакомую по §8 как суперпозиция или композиция отображений. Напомним определение.

Пусть V, W, Z – линейные пространства над полем P. *Произведением линейных операторов* $\mathcal{A} \in \mathcal{L}(V,W)$, $\mathcal{B} \in \mathcal{L}(W,Z)$ называется отображение $\mathcal{C} : V \to Z$, выполняемое по правилу $\mathcal{C}x = \mathcal{B}(\mathcal{A}x)$, $\forall x \in V$. Обозначение: $\mathcal{B}\mathcal{A}$. Итак, $(\mathcal{B}\mathcal{A})x = \mathcal{B}(\mathcal{A}x)$, $\forall x \in V$.

Теорема 79.1. *Если $\mathcal{A} \in \mathcal{L}(V,W)$, $\mathcal{B} \in \mathcal{L}(W,Z)$, то $\mathcal{B}\mathcal{A} \in \mathcal{L}(V,Z)$.*

Доказательство. Линейность оператора $\mathcal{B}\mathcal{A}$ проверяется непосредственно: для любых $x, y \in V$ и $\alpha \in P$

$(\mathcal{BA})(x+y) = \mathcal{B}(\mathcal{A}(x+y)) = \mathcal{B}(\mathcal{A}x+\mathcal{A}y) = \mathcal{B}(\mathcal{A}x)+\mathcal{B}(\mathcal{A}y) = \mathcal{BA}x+\mathcal{BA}y$,
$(\mathcal{BA})(\alpha x) = \mathcal{B}(\mathcal{A}(\alpha x)) = \mathcal{B}(\alpha(\mathcal{A}x)) = \alpha\mathcal{B}(\mathcal{A}x) = \alpha(\mathcal{BA}x) = (\alpha\mathcal{BA})x$. ■

Произведение линейных операторов определено не для любой пары линейных операторов. Однако если это произведение имеет смысл, то:

$$\begin{aligned} &1)\ (\mathcal{AB})\mathcal{C} = \mathcal{A}(\mathcal{BC}) \quad \text{(ассоциативность)};\\ &2)\ \alpha(\mathcal{AB}) = (\alpha\mathcal{A})\mathcal{B} = \mathcal{A}(\alpha\mathcal{B});\\ &3)\ (\mathcal{A}+\mathcal{B})\mathcal{C} = \mathcal{AC}+\mathcal{BC},\\ &\quad\ \mathcal{A}(\mathcal{B}+\mathcal{C}) = \mathcal{AB}+\mathcal{AC} \quad \text{(дистрибутивность)}. \end{aligned} \tag{79.1}$$

Эти свойства легко проверяются непосредственно.

Умножение линейных операторов не обладает свойством коммутативности. В самом деле, о коммутативности можно говорить лишь для операторов, действующих в одном пространстве, т.е. для операторов $\mathcal{A}, \mathcal{B} \in \mathcal{L}(V, V)$. Но и в этом случае умножение не коммутативно. В этом можно убедиться на простом примере операторов ортогонального проектирования плоскости V_2 на ось Ox и поворота на угол $\pi/2$.

Теорема 79.2. *При умножении линейных операторов их матрицы умножаются, т.е. если e, f, g – базисы пространств V, W, Z, то*

$$(\mathcal{BA})_{ge} = B_{gf}A_{fe}. \tag{79.2}$$

Доказательство. Пусть $A_{fe} = (a_{ij})$, $B_{gf} = (b_{ij})$, $(\mathcal{BA})_{ge} = (c_{ij})$, $\dim V = n$, $\dim W = m$, $\dim Z = k$. Тогда в силу (77.1)

$$\mathcal{BA}e_j = \sum_{i=1}^{k} c_{ij}g_i. \tag{79.3}$$

В то же время $\mathcal{BA}e_j = \mathcal{B}(\mathcal{A}e_j) = \mathcal{B}(\sum_{s=1}^m a_{sj}f_s) = \sum_{s=1}^m a_{sj}(\mathcal{B}f_s) = \sum_{s=1}^m a_{sj}\sum_{i=1}^k b_{is}g_i = \sum_{s=1}^m\sum_{i=1}^k a_{sj}b_{is}g_i = \sum_{i=1}^k(\sum_{s=1}^m b_{is}a_{sj})g_i$.

Сравнение этого разложения с (79.3) приводит к равенству $c_{ij} = \sum_{s=1}^m b_{is}a_{sj}$, которое означает (79.2). ■

§80. Образ и ядро линейного оператора

Образом линейного оператора $\mathcal{A} \in \mathcal{L}(V, W)$ называется множество

$$\operatorname{im}\mathcal{A} = \{y \in W \mid \exists x \in V : \mathcal{A}x = y\},$$

ядром оператора $\mathcal{A}$ – множество

$$\ker\mathcal{A} = \{x \in V \mid \mathcal{A}x = \theta\}.$$

Примеры. 1. В пространстве многочленов M_n для оператора дифференцирования (76.2): $\operatorname{im}\mathcal{D} = M_{n-1}$, $\ker\mathcal{D} = M_0$.

2. Для оператора проектирования (76.3): $\operatorname{im}\mathcal{P} = L_1$, $\ker\mathcal{P} = L_2$.

3. Для оператора отражения (76.4): $\operatorname{im}\mathcal{R} = V$, $\ker\mathcal{R} = \{\theta\}$.

Теорема 80.1. *Если $\mathcal{A} \in \mathcal{L}(V, W)$, то $\ker\mathcal{A}$ – линейное подпространство пространства V, $\operatorname{im}\mathcal{A}$ – линейное подпространство пространства W.*

Эти утверждения вытекают из теоремы 18.1, условия которой легко проверяются. ■

Рангом линейного оператора называется размерность его образа, а *дефектом* – размерность ядра. Обозначения: $\operatorname{rg}\mathcal{A}$, $\operatorname{def}\mathcal{A}$. Итак, $\operatorname{rg}\mathcal{A} = \dim\operatorname{im}\mathcal{A}$, $\operatorname{def}\mathcal{A} = \dim\ker\mathcal{A}$.

Теорема 80.2. *Если $e_1, \dots, e_n$ – базис пространства V, то*

$$\operatorname{im}\mathcal{A} = \mathcal{L}(\mathcal{A}e_1, \dots, \mathcal{A}e_n). \tag{80.1}$$

Доказательство. Достаточно показать, что для множеств (80.1) имеет место двустороннее вложение: с одной стороны, если $y \in \operatorname{im}\mathcal{A}$, то $y = \mathcal{A}x$ для некоторого вектора $x \in V$, т.е. $y = \mathcal{A}(\sum_{i=1}^n x_ie_i) = \sum_{i=1}^n x_i\mathcal{A}e_i \in \mathcal{L}(\mathcal{A}e_1, \dots, \mathcal{A}e_n)$; с другой стороны, если $y \in \mathcal{L}(\mathcal{A}e_1, \dots, \mathcal{A}e_n)$, то $y = \sum_{i=1}^n x_i\mathcal{A}e_i = \mathcal{A}(\sum_{i=1}^n x_ie_i) = \mathcal{A}x$, где $x = \sum_{i=1}^n x_ie_i$, т.е. $y \in \operatorname{im}\mathcal{A}$. ■

Теорема 80.3. *Ранг линейного оператора равен рангу его матрицы в произвольной паре базисов.*

Доказательство. Из теоремы 80.2 и (64.3) следует, что $\operatorname{rg}\mathcal{A} = \dim\operatorname{im}\mathcal{A} = \dim\mathcal{L}(\mathcal{A}e_1, \dots, \mathcal{A}e_n) = \operatorname{rg}(\mathcal{A}e_1, \dots, \mathcal{A}e_n)$. Ранг системы векторов $\mathcal{A}e_1, \dots, \mathcal{A}e_n$ совпадает с рангом системы арифметических векторов, составленных из координат этих векторов в базисе f пространства W, т.е. с рангом системы столбцов матрицы A_{fe}. ■

Теорема 80.4 (о ранге и дефекте). *Если $\mathcal{A} \in \mathcal{L}(V, W)$, то*

$$\operatorname{rg}\mathcal{A} + \operatorname{def}\mathcal{A} = \dim V. \tag{80.2}$$

Доказательство. Пусть $e_1, \dots, e_k$ – базис $\ker\mathcal{A}$. Дополним его до базиса $e_1, \dots, e_k, e_{k+1}, \dots, e_n$ пространства V. Согласно теореме 80.2, $\operatorname{im}\mathcal{A} = \mathcal{L}(\mathcal{A}e_1, \dots, \mathcal{A}e_k, \mathcal{A}e_{k+1}, \dots, \mathcal{A}e_n) = \mathcal{L}(\mathcal{A}e_{k+1}, \dots, \mathcal{A}e_n)$. Докажем, что векторы $\mathcal{A}e_{k+1}, \dots, \mathcal{A}e_n$ линейно независимы. Пусть это не так. Тогда для нетривиальной линейной комбинации этих векторов имеют место соотношения $\alpha_{k+1}\mathcal{A}e_{k+1} + \dots + \alpha_n\mathcal{A}e_n = \theta$ или $\mathcal{A}(\alpha_{k+1}e_{k+1} + \dots + \alpha_ne_n) = \theta$. Следовательно, $\alpha_{k+1}e_{k+1} + \dots + \alpha_ne_n \in \ker\mathcal{A}$. Это означает, что вектор $\alpha_{k+1}e_{k+1} + \dots + \alpha_ne_n$ линейно выражается через $e_1, \dots, e_k$, что невозможно в силу линейной независимости $e_1, \dots, e_k, e_{k+1}, \dots, e_n$. Таким образом, $\dim\operatorname{im}\mathcal{A} = n - k$, $\dim\ker\mathcal{A} = k$. Отсюда следует (80.2). ■

Теорема 80.5. *Пусть $\mathcal{A} \in \mathcal{L}(V, W)$, $\operatorname{rg}\mathcal{A} = r$, $\dim V = n$, $\dim W = m$. Тогда существуют базисы e и f пространств V и W,* *в которых оператор $\mathcal{A}$ имеет матрицу $I_r \in P^{m\times n}$ вида*

$$I_r = \left[\begin{array}{ccc|c} 1 & & 0 & \\ & \ddots & & O \\ 0 & & 1 & \\ \hline & O & & O \end{array}\right] \Big\} r,$$

в которой все элементы равны нулю, кроме первых r диагональных элементов, равных 1.

Доказательство. Возьмем произвольные базисы t и s пространств V и W. Пусть A_{st} – матрица оператора $\mathcal{A}$ в паре базисов t и s. Тогда, согласно теореме 80.3, $\operatorname{rg} A_{st} = r$. В силу теоремы 16.10 отсюда следует, что матрицы A_{st} и I_r эквивалентны. Следовательно (теорема 77.4), они являются матрицами одного линейного оператора. Отсюда вытекает утверждение теоремы. ■

Базисы e и f, в которых оператор $\mathcal{A}$ имеет матрицу I_r, называют *канонической парой базисов.*

Теоремой 80.5 решается вопрос о наиболее простой форме матрицы линейного оператора, действующего в различных пространствах V и W. Очевидно, этот вопрос далеко не праздный, так как одному и тому же линейному оператору соответствует целый класс эквивалентных друг другу матриц и выбор самой простой из них упрощает исследование свойств оператора.

Остановимся более подробно на одном из частных случаев линейных операторов – случае, когда пространство W совпадает с основным полем P.

§81. Линейные формы

Как уже отмечалось в §76, линейное отображение $f : V \to P$ линейного пространства V над полем P в это поле называется ***линейной формой*** (или *линейным функционалом*) *в пространстве* V.

Примеры. 1. Простейшей линейной формой в пространстве V является отображение $f : V \to P$, определенное равенством $f(x) = 0$, где $0 \in P$.

2. Линейной формой является отображение $f : \mathbb{R}^n \to \mathbb{R}$, определенное следующим правилом: если $(x_1, \ldots, x_n) \in \mathbb{R}^n$, то $f(x) = x_1$.

3. В пространстве M_n вещественных многочленов степени не выше n линейной формой является отображение $f : M_n \to \mathbb{R}$, определенное правилом: если $p(t) \in M_n$, то $f(p) = p(1)$.

Из общей теории линейных операторов следует, что если $e_1, \ldots, e_n$ – базис пространства V, то линейная форма $f \in \mathcal{L}(V, P)$ однозначно определяется числами

$$\alpha_1 = f(e_1), \ \ldots, \ \alpha_n = f(e_n).$$

При этом для произвольного вектора $x = \sum_{i=1}^n x_i e_i$ в силу линейности f имеем

$$f(x) = \alpha_1 x_1 + \ldots + \alpha_n x_n . \tag{81.1}$$

Представление (81.1) называется ***общим видом линейной формы f в базисе $e_1, \ldots, e_n$***. Числа $\alpha_1, \ldots, \alpha_n$ называются ***коэффициентами линейной формы f в базисе $e_1, \ldots, e_n$***.

Сопряженное пространство. Известно (теорема 78.2), что множество $\mathcal{L}(V, P)$ всех линейных форм в линейном пространстве V образует линейное пространство относительно операций сложения и умножения на число, введенных в §78 правилами

$$(f_1 + f_2)(x) = f_1(x) + f_2(x), \quad (\alpha f)(x) = \alpha f(x).$$

Линейное пространство всех линейных форм на пространстве V называется *сопряженным пространством к пространству* V. Оно обозначается символом V^*.

Теорема 81.1. $\dim V^* = \dim V$.

Это равенство вытекает из теоремы 78.3 и ее следствия, так как $\dim P = 1$. ■

Следствие. Всякое конечномерное линейное пространство изоморфно своему сопряженному. Это вытекает из теоремы 63.1 об изоморфизме линейных пространств одинаковой размерности.

Специальное представление линейной формы в евклидовом (унитарном) пространстве.

Теорема 81.2. *Для любой линейной формы f в евклидовом (унитарном) пространстве V существует, и притом единственный, вектор $h \in V$ такой, что*

$$f(x) = (x, h), \quad \forall x \in V.$$

Доказательство. Пусть $e_1, \ldots, e_n$ – ортонормированный базис V и $\alpha_1 \ldots, \alpha_n$ – коэффициенты линейной формы f в этом базисе. Тогда вектор $h = \sum_{i=1}^n \overline{\alpha_i} e_i$ будет искомым в силу (81.1) и (70.7). Единственность вектора h следует из того, что если вектор h_1 удовлетворяет требованиям теоремы, то $(x, h_1) = (x, h)$, $\forall x \in V$, т.е. $(x, h_1 - h) = 0$, $\forall x \in V$. Так как $h_1 - h \in V$, то отсюда следует, что $h_1 - h = \theta$. ■

§82. Алгебра линейных операторов, действующих в одном пространстве

Обратимся теперь к линейным операторам, действующим в одном пространстве. Основной целью наших исследований будет поиск матрицы такого оператора, имеющей наиболее простую форму. Однако об этом речь пойдет позже. В ближайших параграфах будут рассматриваться общие вопросы, относящиеся к операторам, действующим в одном пространстве.

Пусть V – линейное пространство над полем P. Рассмотрим множество $\mathcal{L}(V,V)$ всех линейных операторов, действующих в пространстве V. В этом множестве для любых операторов выполнимы не только сложение и умножение на число, но и умножение операторов друг на друга. Из общей теории линейных операторов вытекают следующие факты.

1°. $\mathcal{L}(V,V)$ – ***линейное пространство над полем*** P (теорема 78.2).

2°. $\mathcal{L}(V,V)$ – ***некоммутативное кольцо с единицей*** (см. (79.1)).

3°. Линейное пространство над полем P, которое является кольцом и удовлетворяет условию 2 из (79.1), называется ***алгеброй*** (или ***линейной алгеброй***) ***над полем*** P. Размерность линейного пространства при этом называется также ***размерностью алгебры***. Из свойства 1° и (79.1) следует, что $\mathcal{L}(V,V)$ – ***алгебра над полем*** P.

Отметим, что множество $P^{n\times n}$ квадратных матриц n-го порядка над полем P также является алгеброй. Примером бесконечномерной алгебры служит кольцо многочленов $P[x]$.

Может быть построена другая алгебра, в которой нарушена как коммутативность, так и ассоциативность произведения. Примером такой алгебры является множество всех векторов геометрического пространства V_3 относительно линейных операций над векторами и векторного произведения (§23).

4°. При построении матрицы линейного оператора $\mathcal{A} \in \mathcal{L}(V,V)$ используется один базис e пространства V: столбцами этой матрицы являются коэффициенты разложений векторов $\mathcal{A}e_1,\ldots,\mathcal{A}e_n$ по базису $e_1,\ldots,e_n$ (§77); матрица обозначается символом A_e или $(\mathcal{A})_e$ и называется ***матрицей оператора*** $\mathcal{A}$ ***в базисе*** e. Очевидно, $A_e \in P^{n\times n}$.

5°. *Линейное пространство* $\mathcal{L}(V,V)$ *изоморфно пространству* $P^{n\times n}$ и $\dim \mathcal{L}(V,V) = n^2$ (теорема 78.3 и ее следствие).

6°. Как и в любом кольце, оператор $\mathcal{A} \in \mathcal{L}(V,V)$ можно возводить в степень $n \in \mathbb{N}$, и если $p(t) = a_0 + a_1 t + \ldots + a_n t^n$ – произвольный многочлен над полем P от переменной t, то однозначно определен оператор

$$p(\mathcal{A}) = a_0 I + a_1 \mathcal{A} + \ldots + a_n \mathcal{A}^n\,, \qquad (82.1)$$

называемый ***многочленом от оператора*** $\mathcal{A}$. Из свойств матрицы линейного оператора следует, что матрицей многочлена (82.1) от оператора $\mathcal{A}$ является тот же многочлен от матрицы A_e: $(p(\mathcal{A}))_e = p(A_e)$.

7°. При переходе от базиса e к базису $f = eQ$ матрица оператора изменяется согласно (77.5) по следующему закону:

$$A_f = Q^{-1} A_e Q\,. \qquad (82.2)$$

Из (82.2) и (58.6) следует, что *одному и тому же линейному оператору* $\mathcal{A} \in \mathcal{L}(V,V)$ *соответствует целый класс матриц, подобных друг другу.*

8°. Очевидно, что *две матрицы* $A, B \in P^{n\times n}$ ***подобны тогда и только тогда, когда они являются матрицами одного и того же линейного оператора, действующего в n-мерном линейном пространстве над полем*** P (теорема 77.4).

9°. Из (82.2) следует, что *все матрицы одного и того же линей го оператора имеют одинаковый определитель.* Это означает, определитель матрицы линейного оператора не зависит от матри и определяется самим оператором. *Определителем линейного опе тора* $\mathcal{A} \in \mathcal{L}(V,V)$ называется определитель матрицы этого операт в произвольном базисе. Обозначение: $\det \mathcal{A}$. Итак,

$$\det \mathcal{A} = \det A_e\,. \qquad (82$$

Заметим, что кольцо $\mathcal{L}(V,V)$ не является полем хотя бы пото что оно некоммутативно. К тому же вопрос об обратимости линейн операторов пока остается открытым.

§83. Обратный оператор

Напомним определение обратного отображения и некоторые фак (§8) применительно к линейному оператору, действующему в одн пространстве.

Пусть $\mathcal{A} \in \mathcal{L}(V,V)$. Отображение $\mathcal{A}^{-1} : V \to V$ называет ***обратным оператором*** к оператору $\mathcal{A}$, если

$$\mathcal{A}\mathcal{A}^{-1} = \mathcal{A}^{-1}\mathcal{A} = \mathcal{I}\,. \qquad (83.$$

Теорема 83.1. *Линейный оператор* $\mathcal{A} \in \mathcal{L}(V,V)$ *обрат тогда и только тогда, когда он биективен.*

Теорема 83.2. *Обратный оператор единствен.*

Исследуем теперь свойства обратимости, относящиеся непосре ственно к линейным операторам.

Теорема 83.3. *Обратный оператор линеен.*

Доказательство. Пусть $\mathcal{A} \in \mathcal{L}(V,V)$. Покажем, что обратны оператор $\mathcal{A}^{-1}$, если он существует, является линейным операторо действующим в пространстве V. Действительно, если $\mathcal{A}$ обратим, т он биективен и, значит, сюръективен. Это означает, что для любы векторов $y_1, y_2 \in V$ существуют $x_1, x_2 \in V$ такие, что $y_1 = \mathcal{A}x$ $y_2 = \mathcal{A}x_2$. При этом $x_1 = \mathcal{A}^{-1}y_1$, $x_2 = \mathcal{A}^{-1}y_2$. Отсюда получим что $\mathcal{A}^{-1}(y_1 + y_2) = \mathcal{A}^{-1}(\mathcal{A}x_1 + \mathcal{A}x_2) = \mathcal{A}^{-1}\mathcal{A}(x_1 + x_2) = x_1 + x_2 =$ $\mathcal{A}^{-1}y_1 + \mathcal{A}^{-1}y_2$. Аналогично, $\mathcal{A}^{-1}(\alpha y_1) = \mathcal{A}^{-1}(\alpha\mathcal{A}x_1) = \mathcal{A}^{-1}\mathcal{A}(\alpha x_1) =$ $\alpha x_1 = \alpha\mathcal{A}^{-1}y_1$, $\forall \alpha \in P$. ■

Теорема 83.4. *Матрица обратного оператора* $\mathcal{A}^{-1}$ *в про извольном базисе является обратной к матрице оператора* $\mathcal{A}$ *этом же базисе.*

Доказательство. Пусть e – произвольный базис пространств V и для оператора $\mathcal{A} \in \mathcal{L}(V,V)$ существует обратный оператор $\mathcal{A}^{-1}$

Перейдем в равенствах (83.1) к матрицам операторов в базисе e Согласно (79.2) получим, что $A_e(\mathcal{A}^{-1})_e = (\mathcal{A}^{-1})_e A_e = I$. Эти равен ства совпадают с определением обратной матрицы для матрицы A,

(§5). Таким образом,

$$(A_e)^{-1} = (\mathcal{A}^{-1})_e. \quad \blacksquare \tag{83.2}$$

Оператор $\mathcal{A} \in \mathcal{L}(V,V)$ называется *невырожденным*, если его ядро состоит только из нулевого вектора, т.е.

$$\ker \mathcal{A} = \{\theta\}, \tag{83.3}$$

и *вырожденным* в противном случае.

Теорема 83.5. *В конечномерном пространстве V следующие утверждения равносильны: для $\mathcal{A} \in \mathcal{L}(V,V)$*

$$\begin{array}{lll} 1)\ \mathcal{A}\mathcal{A}^{-1} = \mathcal{I}; & 4)\ \operatorname{im}\mathcal{A} = V; & 6)\ \mathcal{A} \text{ обратим}; \\ 2)\ \mathcal{A}^{-1}\mathcal{A} = \mathcal{I}; & 5)\ \det\mathcal{A} \neq 0; & 7)\ \mathcal{A} \text{ биективен}. \\ 3)\ \mathcal{A} \text{ не вырожден}; & & \end{array} \tag{83.4}$$

Доказательство. $1 \iff 2 \iff 5 \iff 6 \iff 7$. Пусть $e = (e_1,\ldots,e_n)$ – произвольный базис пространства V. В силу свойства (79.2) утверждение 1 равносильно матричному равенству $A_e(\mathcal{A}^{-1})_e = I$ или, с учетом теоремы 5.3, соотношениям $A_e(A_e)^{-1} = (A_e)^{-1}A_e = I$, эквивалентным (в силу (79.2)) равенствам (83.1) и (в силу теоремы 5.2) условию

$$\det A_e \neq 0. \tag{83.5}$$

Это доказывает импликации $1 \iff 5$, $1 \iff 2$, $1 \iff 6$ и, согласно теореме 83.1, $1 \iff 7$.

$1 \iff 3 \iff 4$. Утверждение 1 в силу (83.5) равносильно равенству

$$\operatorname{rg}\mathcal{A} = n, \tag{83.6}$$

которое согласно (80.2) эквивалентно тому, что

$$\operatorname{def}\mathcal{A} = 0. \tag{83.7}$$

Равенство (83.6) означает, что $\dim\operatorname{im}\mathcal{A} = \dim V$ или, в силу теоремы 64.3 о монотонности размерности, что $\operatorname{im}\mathcal{A} = V$. Это доказывает импликацию $1 \iff 4$. Равенство (83.7) означает, что $\dim\ker\mathcal{A} = 0$. Это доказывает импликацию $1 \iff 3$. ■

Теорема 83.6. *Произведение обратимых операторов обратимо, при этом*

$$(\mathcal{A}\mathcal{B})^{-1} = \mathcal{B}^{-1}\mathcal{A}^{-1}.$$

Это утверждение уже было доказано в §8 применительно к отображениям общего вида. ■

Следствие 1. *Умножение линейных операторов является алгебраической операцией на множестве всех обратимых операторов, действующих в пространстве V.*

Следствие 2. *Множество всех обратимых линейных операторов из $\mathcal{L}(V,V)$ образует неабелеву мультипликативную группу.*

Глава XV. Структура линейного оператора в комплексном пространстве

Приступим к вопросу о поиске в классе подобных матриц, соответствующих линейному оператору $\mathcal{A} \in \mathcal{L}(V,V)$, матрицы наиболее простого вида. Ответ на этот вопрос совсем прост, когда линейный оператор действует в пространствах V и W, никак не связанных между собой (теорема 80.5). Гораздо интереснее (и сложнее), когда $V = W$. В первом случае базисы пространств V и W можно было выбирать независимо друг от друга, во втором – необходимо строить базис одного пространства $V = W$. Мы увидим, к какому разнообразию ответов приводит эта меньшая свобода выбора.

Наиболее полные результаты этой главы относятся к случаю, когда V – комплексное пространство.

§84. Инвариантные подпространства

Пусть V – линейное пространство над полем P и $\mathcal{A} \in \mathcal{L}(V,V)$. Линейное подпространство L пространства V называется *инвариантным подпространством относительно оператора $\mathcal{A}$*, если для любого вектора x из L его образ $\mathcal{A}x$ также лежит в L.

Примеры. 1. Тривиальные подпространства $\{\theta\}$ и V инварианты относительно любого оператора $\mathcal{A} \in \mathcal{L}(V,V)$.

2. Для любого линейного оператора $\mathcal{A}$ инвариантными подпространствами будут $\ker\mathcal{A}$ и $\operatorname{im}\mathcal{A}$, так как если $\mathcal{A}x = \theta$, то $\mathcal{A}(\mathcal{A}x) = \mathcal{A}\theta = \theta$, и если $y = \mathcal{A}x$, то $\mathcal{A}y = \mathcal{A}(\mathcal{A}x) = \mathcal{A}x_1$, где $x_1 = \mathcal{A}x$.

3. Для оператора дифференцирования (76.2) в пространстве M_n вещественных многочленов инвариантными подпространствами являются все подпространства $M_0, M_1, \ldots, M_{n-1}$.

Для линейного оператора наличие инвариантных подпространств означает возможность построить базис, в котором его матрица имеет несколько более простую форму, чем матрица общего вида.

Теорема 84.1. *Пусть $\mathcal{A} \in \mathcal{L}(V,V)$ и L – инвариантное подпространство относительно $\mathcal{A}$. Тогда существует базис пространства V, в котором матрица оператора $\mathcal{A}$ имеет квазитреугольную форму.*

Доказательство. Пусть $e_1,\ldots,e_k$ – базис L. Дополним его до базиса $e_1,\ldots,e_k,e_{k+1},\ldots,e_n$ пространства V. Построим матрицу оператора $\mathcal{A}$ в этом базисе. Из инвариантности L вытекает, что $\mathcal{A}e_1,\ldots,\mathcal{A}e_k \in L$ и, следовательно, векторы $\mathcal{A}e_1,\ldots,\mathcal{A}e_k$ линейно выражаются только через $e_1, \dots, e_k$. Таким образом,

$$\begin{cases} \mathcal{A}e_1 &= a_{11}e_1 + \ldots + a_{k1}e_k, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \mathcal{A}e_k &= a_{1k}e_1 + \ldots + a_{kk}e_k, \\ \mathcal{A}e_{k+1} &= a_{1,k+1}e_1 + \ldots + a_{k,k+1}e_k + \ldots + a_{n,k+1}e_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \mathcal{A}e_n &= a_{1n}e_1 + \ldots + a_{kn}e_k + \ldots + a_{nn}e_n\,. \end{cases} \tag{84.1}$$

Это означает, что матрица A_e имеет вид

$$A_e = \begin{bmatrix} a_{11} & \ldots & a_{1k} & a_{1,k+1} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{k1} & \ldots & a_{kk} & a_{k,k+1} & \ldots & a_{kn} \\ 0 & \ldots & 0 & a_{k+1,k+1} & \ldots & a_{k+1,n} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & a_{n,k+1} & \ldots & a_{nn} \end{bmatrix}$$

и, следовательно, имеет квазитреугольную форму:

$$A_e = \left[\begin{array}{c|c} A_1 & B \\ \hline O & C \end{array}\right]. \blacksquare \tag{84.2}$$

Замечание 1. Верна и теорема, обратная доказанной: переход от (84.2) к (84.1) очевиден и приводит к тому, что $L = \mathcal{L}(e_1, \dots, e_k)$ инвариантно относительно $\mathcal{A}$.

Теорема 84.2. *Если пространство V является прямой суммой подпространств $L_1, \dots, L_k$, инвариантных относительно оператора $\mathcal{A} \in \mathcal{L}(V, V)$, то в пространстве V существует базис, в котором матрица оператора $\mathcal{A}$ имеет квазидиагональную форму.*

Доказательство аналогично доказательству теоремы 84.1. В качестве искомого базиса берется базис e, составленный из базисов слагаемых подпространств (теорема 66.1). Тогда в силу инвариантности подпространств $L_1, \dots, L_k$ матрица A_e имеет вид

$$A_e = \begin{bmatrix} \boxed{A_1} & & & O \\ & \boxed{A_2} & & \\ & & \ddots & \\ O & & & \boxed{A_k} \end{bmatrix}. \blacksquare \tag{84.3}$$

Замечание 2. Верна и теорема, обратная доказанной.

Индуцированный оператор. Рассматривая линейный оператор только на его инвариантном подпространстве, можно получить новый оператор. Пусть L – подпространство, инвариантное относительно оператора $\mathcal{A} \in \mathcal{L}(V, V)$. Отображение $\mathcal{A}|L : L \to L$, определенное равенством

$$(\mathcal{A}|L)x = \mathcal{A}x\,, \quad \forall x \in L,$$

называется ***индуцированным оператором, порожденным оператором*** $\mathcal{A}$. В силу линейности оператора $\mathcal{A}$ индуцированный оператор также будет линейным. Он совпадает с оператором $\mathcal{A}$ на подпространстве L и не определен вне его. Итак, $\mathcal{A}|L \in \mathcal{L}(L, L)$.

Замечание 3. Из разложений (84.1) следует, что матрицы $A_1, \dots, A_k$ в (84.3) являются матрицами индуцированных операторов $\mathcal{A}|L_1, \dots, \mathcal{A}|L_k$ в базисах инвариантных подпространств $L_1, \dots, L_k$.

§85. Собственные значения и собственные векторы

Пусть V – линейное пространство над полем P. Ненулевой вектор $x \in X$ называется ***собственным вектором оператора*** $\mathcal{A} \in \mathcal{L}(V, V)$, если существует такое число $\lambda \in P$, что

$$\mathcal{A}x = \lambda x\,. \tag{85.1}$$

Число λ называется ***собственным значением оператора*** $\mathcal{A}$, ***соответствующим собственному вектору*** x. Множество всех собственных значений оператора $\mathcal{A}$ называется *спектром* этого оператора.

Примеры. 1. В пространстве вещественных многочленов M_n любой многочлен нулевой степени будет собственным вектором оператора дифференцирования (76.2), ему соответствует собственное значение $\lambda = 0$.

2. Для оператора проектирования (76.3) любой ненулевой вектор из L_1 будет собственным вектором, отвечающим собственному значению $\lambda = 1$, так как $\mathcal{P}x = x$, $\forall x \in L_1$, а любой ненулевой вектор из L_2 будет собственным вектором, отвечающим собственному значению $\lambda = 0$, так как $\mathcal{P}x = 0x$, $\forall x \in L_2$.

Из определения следует, что если x – собственный вектор оператора $\mathcal{A}$, отвечающий собственному значению λ, то любой вектор αx, где $\alpha \neq 0$, также будет собственным вектором оператора $\mathcal{A}$, отвечающим тому же собственному значению λ. Это означает, что любой собственный вектор порождает целое одномерное подпространство собственных векторов, из которого исключен вектор θ. Таким образом, в отношении собственных векторов понятие "различные" не имеет содержательного смысла, существеннее ставить вопрос об их линейной независимости.

Теорема 85.1. *Собственные векторы $x_1, \dots, x_k$ оператора, отвечающие различным собственным значениям $\lambda_1, \dots, \lambda_k$, линейно независимы.*

Доказательство. Применим индукцию по k. Для $k = 1$ утверждение заведомо верно, так как собственный вектор является ненулевым по определению. Пусть оно верно для любой системы из $k - 1$

векторов. Докажем его для k векторов $x_1, \ldots, x_k$. Приравняем нулевому вектору линейную комбинацию этих векторов:

$$\alpha_1 x_1 + \ldots + \alpha_k x_k = \theta . \qquad (85.2)$$

Под действием оператора $\mathcal{A}$ это равенство перейдет в равенство

$$\alpha_1 \lambda_1 x_1 + \ldots + \alpha_k \lambda_k x_k = \theta . \qquad (85.3)$$

Умножая обе части (85.2) на λ_k и вычитая полученное равенство из (85.3), получаем, что $\alpha_1(\lambda_1 - \lambda_k)x_1 + \ldots + \alpha_{k-1}(\lambda_{k-1} - \lambda_k)x_{k-1} = \theta$. В силу индуктивного предположения отсюда следует, что $\alpha_1 = \ldots = \alpha_{k-1} = 0$. Тогда (85.2) перейдет в равенство $\alpha_k x_k = \theta$. Так как $x_k \neq \theta$, то $\alpha_k = 0$. Итак, в (85.2) все коэффициенты равны нулю. Следовательно, $x_1, \ldots, x_k$ линейно независимы. ■

Следствие. Линейный оператор, действующий в n-мерном пространстве, не может иметь более чем n различных собственных значений.

Не всякий линейный оператор обладает собственными векторами. Например, в геометрической плоскости V_2 оператор поворота на угол φ, не кратный π, не имеет ни одного собственного вектора, так как ни один ненулевой вектор после такого поворота не останется коллинеарным самому себе. Выясним вопрос о существовании собственных векторов.

§86. Характеристический многочлен

Это понятие уже встречалось в §58 в связи с матрицами второго и третьего порядка. Остановимся теперь на нем более подробно.

Характеристическим многочленом матрицы $A \in P^{n\times n}$ называется функция

$$f(\lambda) = \det(A - \lambda I), \quad \lambda \in P . \qquad (86.1)$$

Теорема 86.1. *Характеристический многочлен (86.1) матрицы $A \in P^{n\times n}$ является многочленом n-й степени от переменной λ над полем P.*

Доказательство. Пусть $A = (a_{ij}) \in P^{n\times n}$. Тогда

$$f(\lambda) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} - \lambda & \ldots & a_{1n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} - \lambda \end{vmatrix} .$$

Каждый элемент матрицы $A - \lambda I$ представляет собой многочлен от λ степени не выше 1, значит, каждый член $\det(A - \lambda I)$ является многочленом от λ степени не выше n. Отсюда следует, что $f(\lambda)$ – многочлен от λ, степень которого не превосходит n. Осталось показать, что степень этого многочлена в точности равна n. Действительно, все члены $\det(A - \lambda I)$, отличные от $(a_{11} - \lambda)(a_{22} - \lambda) \ldots (a_{nn} - \lambda)$, имеют степень, не превосходящую $n - 2$, следовательно, в многочлене $f(\lambda)$ слагаемые, содержащие λ^{n-1} и λ^n, определяются только членом

$(a_{11} - \lambda)(a_{22} - \lambda) \ldots (a_{nn} - \lambda)$, который согласно формулам Виета имеет вид

$$(-\lambda)^n + (a_{11} + a_{22} + \ldots + a_{nn})(-\lambda)^{n-1} + g_{n-2}(\lambda),$$

где $g_{n-2}(\lambda)$ – многочлен от λ степени не выше $n - 2$. Таким образом,

$$f(\lambda) = a_0 + a_1(-\lambda) + a_2(-\lambda)^2 + \ldots + a_{n-1}(-\lambda)^{n-1} + (-\lambda)^n \qquad (86.2)$$

является многочленом n-й степени от λ над полем P. ■

Замечание 1. Попутно найден коэффициент a_{n-1} характеристического многочлена (86.2): $a_{n-1} = a_{11} + \ldots + a_{nn}$. Кроме того, $a_0 = f(0) = \det A$. Итак, в (86.2)

$$a_0 = \det A, \quad a_{n-1} = \operatorname{tr} A . \qquad (86.3)$$

Теорема 86.2. *Характеристические многочлены подобных матриц совпадают.*

Доказательство теоремы и сама теорема приводятся в §58 (теорема 58.1). ■

Следствие. Все матрицы одного и того же линейного оператора имеют одинаковые характеристические многочлены.

Таким образом, характеристический многочлен матрицы линейного оператора не зависит от базиса, а определяется самим оператором.

Характеристическим многочленом оператора называется функция

$$f(\lambda) = \det(\mathcal{A} - \lambda\mathcal{I}), \quad \lambda \in P.$$

Из (82.3) следует, что характеристический многочлен оператора совпадает с характеристическим многочленом матрицы этого оператора в произвольном базисе.

Из следствия и второго равенства (86.3) вытекает, что след матрицы A_e линейного оператора $\mathcal{A}$ не зависит от выбора базиса e. Поэтому (как и в §82, п.9) можно ввести понятие *следа оператора* равенством $\operatorname{tr} \mathcal{A} = \operatorname{tr} A_e$.

Теорема 86.3. *Характеристический многочлен индуцированного оператора является делителем характеристического многочлена порождающего оператора.*

Это утверждение следует из (84.2) и замечания 3 из §84. ■

Теорема 86.4. *Если $V = L_1 \oplus \ldots \oplus L_k$ – прямая сумма подпространств $L_1, \ldots, L_k$, инвариантных относительно оператора $\mathcal{A} \in \mathcal{L}(V, V)$, то характеристический многочлен $f(\lambda)$ оператора $\mathcal{A}$ равен произведению характеристических многочленов $f_1(\lambda), \ldots, f_k(\lambda)$ индуцированных операторов $\mathcal{A}|L_1, \ldots, \mathcal{A}|L_k$:*

$$f(\lambda) = f_1(\lambda) \ldots f_k(\lambda) . \qquad (86.4)$$

Утверждение является следствием теоремы 84.2. ■

Условие существования собственных векторов линейного оператора определяется его характеристическим многочленом.

Теорема 86.5. *Пусть V – линейное пространство над полем P. Число $\lambda \in P$ является собственным значением оператора $\mathcal{A} \in \mathcal{L}(V, V)$ тогда и только тогда, когда λ – корень его характеристического многочлена.*

Доказательство. Согласно определению число λ является собственным значением оператора $\mathcal{A}$ тогда и только тогда, когда существует вектор x, удовлетворяющий условиям

$$\begin{cases} \mathcal{A}x = \lambda x, \\ x \neq \theta, \\ \lambda \in P, \end{cases} \iff \begin{cases} (\mathcal{A} - \lambda\mathcal{I})x = \theta, \\ x \neq \theta, \\ \lambda \in P. \end{cases}$$

Это равносильно вырожденности оператора $\mathcal{A} - \lambda\mathcal{I}$ при некотором $\lambda \in P$ или, в силу (83.4), условию

$$\det(\mathcal{A} - \lambda\mathcal{I}) = 0 \tag{86.5}$$

при некотором $\lambda \in P$. Таким образом, число λ является собственным значением оператора $\mathcal{A}$ тогда и только тогда, когда оно является корнем характеристического многочлена оператора $\mathcal{A}$, принадлежащим полю P. ■

Уравнение (86.5) называется *характеристическим уравнением* для оператора $\mathcal{A}$.

Собственные векторы линейного оператора в комплексном пространстве. Итак, вопрос о существовании собственных векторов сводится к вопросу о существовании корней характеристического многочлена, принадлежащих основному полю. Известно, что не во всяком поле многочлены имеют корни. Однако, как показано в §49, в алгебраически замкнутом поле $\mathbb{C}$ комплексных чисел любой многочлен степени $n \geq 1$ имеет n корней, если каждый корень считать столько раз, какова его кратность. Отсюда в соответствии с теоремой 86.5 вытекают следующие утверждения.

Теорема 86.6. *Произвольный линейный оператор, действующий в n-мерном комплексном пространстве, имеет:*

1) n собственных значений, если каждое собственное значение считать столько раз, какова его кратность как корня характеристического многочлена;

2) хотя бы один собственный вектор;

3) на любом своем инвариантном подпространстве хотя бы один собственный вектор.

Последнее утверждение следует из того, что индуцированный оператор, как и любой оператор, действующий в комплексном пространстве, имеет собственный вектор, который, очевидно, является собственным вектором основного оператора.

Замечание 2. Теорема 86.6 остается справедливой в вещественном пространстве для тех операторов, чьи характеристические многочлены имеют только вещественные корни.

Способ нахождения собственных векторов определен теоремой 86.5 и состоит в следующем.

1. Строится характеристический многочлен оператора и находятся его корни. Те корни, которые принадлежат полю P, будут собственными значениями оператора.

2. Для каждого найденного собственного значения λ_0 находятся ненулевые векторы ядра оператора $\mathcal{A} - \lambda_0\mathcal{I}$. Они и будут собственными векторами, отвечающими λ_0. Если $e_1, \ldots, e_n$ – базис пространства V, то поиск собственных векторов сводится к решению однородной системы уравнений $(A_e - \lambda_0 I)x_e = 0$, фундаментальная система решений которой дает полный набор линейно независимых собственных векторов, отвечающих λ_0, в координатах относительно базиса e.

Замечание 3. Прием, описанный здесь, представляет собой лишь теоретическую основу решения проблемы собственных значений и собственных векторов. Для реальных вычислений он нуждается в специальных приемах, так как поиск корней многочлена в общем случае (исключая разве что многочлены второй степени) сопряжен с дополнительными трудностями. Уже для многочленов степени $n \geq 5$ задача нахождения корней неразрешима в радикалах (доказано Э.Галуа в 30-х гг. прошлого столетия). Этот результат оказал решающее влияние на развитие вычислительных методов решения проблемы собственных значений. Их изложение не входит в нашу задачу, однако отметим, что вычислительные алгоритмы поиска собственных значений, как правило, не связаны с многочленами.

В ручных вычислениях иногда удается найти собственные значения, минуя прямое построение характеристического многочлена, если вычислять определитель матрицы $A - \lambda I$ *методом выделения линейных множителей*. Этот метод основан на выполнении элементарных преобразований, формирующих в какой-либо строке (столбце) матрицы $A - \lambda I$ общий множитель вида $\lambda - a$, который затем выносится за знак определителя. Это дает корень характеристического многочлена $\lambda = a$.

§87. Собственное подпространство

Пусть λ_0 – собственное значение оператора $\mathcal{A}$. Множество

$$W_{\lambda_0} = \{\, x \in V \mid \mathcal{A}x = \lambda_0 x \,\} \tag{87.1}$$

называется *собственным подпространством* оператора $\mathcal{A}$, отвечающим собственному значению λ_0.

Очевидно, что $W_{\lambda_0} = \ker(\mathcal{A} - \lambda_0\mathcal{I})$, поэтому собственное подпространство является линейным подпространством пространства V.

Из (87.1) следует, что собственное подпространство W_{λ_0} состоит из нулевого вектора θ и всех собственных векторов, отвечающих λ_0. Легко проверить, что собственное подпространство инвариантно относительно оператора $\mathcal{A}$. Размерность собственного подпространства W_{λ_0} называется ***геометрической кратностью собственного значения*** λ_0, а кратность λ_0 как корня характеристического многочлена называется его ***алгебраической кратностью***.

Теорема 87.1. *Геометрическая кратность собственного значения не превосходит его алгебраической кратности.*

Доказательство. Пусть m и s – алгебраическая и геометрическая кратности собственного значения λ_0 оператора $\mathcal{A} \in \mathcal{L}(V, V)$.

Собственное подпространство W_{λ_0} инвариантно относительно оператора $\mathcal{A}$, следовательно, можно рассматривать индуцированный оператор $\mathcal{A}|W_{\lambda_0}$. Найдем его характеристический многочлен $f_1(\lambda)$. Пусть $e_1, \ldots, e_s$ – базис W_{λ_0}. Тогда согласно (87.1) матрицей оператора $\mathcal{A}|W_{\lambda_0}$ в этом базисе будет диагональная матрица s-го порядка с элементами λ_0 на главной диагонали. Следовательно, $f_1(\lambda) = (\lambda_0 - \lambda)^s$. Согласно теореме 86.3, $(\lambda_0 - \lambda)^s$ является делителем характеристического многочлена $f(\lambda)$ оператора $\mathcal{A}$, но $(\lambda_0 - \lambda)$ входит в характеристический многочлен $f(\lambda)$ ровно m раз. Значит, $s \leq m$. ■

Теорема 87.2. *Сумма собственных подпространств оператора, отвечающих различным собственным значениям, является прямой суммой.*

Доказательство. Пусть $\lambda_1, \ldots, \lambda_p$ – попарно различные собственные значения оператора $\mathcal{A}$. Тогда для собственных подпространств $W_1, \ldots, W_{\lambda_p}$ выполнено условие прямой суммы (теорема 66.1): любая система ненулевых векторов, взятых по одному из каждого W_{λ_k}, линейно независима как система собственных векторов, отвечающих различным собственным значениям (теорема 85.1). ■

§88. Операторы простой структуры

Критерии простой структуры. Линейный оператор $\mathcal{A} \in \mathcal{L}(V, V)$ называется ***оператором простой структуры***, если в пространстве V существует базис из собственных векторов оператора $\mathcal{A}$.

Теорема 88.1. *Линейный оператор $\mathcal{A} \in \mathcal{L}(V, V)$ имеет простую структуру тогда и только тогда, когда в пространстве V существует базис, в котором он имеет диагональную матрицу.*

Доказательство. Пусть $\dim V = n$. Согласно определению оператор $\mathcal{A}$ имеет простую структуру тогда и только тогда, когда он имеет n линейно независимых собственных векторов $e_1, \ldots, e_n$. Это равносильно существованию базиса $e_1, \ldots, e_n$, в котором матрица A_e оператора $\mathcal{A}$ имеет вид

$$A_e = \begin{bmatrix} \lambda_1 & & & O \\ & \lambda_2 & & \\ & & \ddots & \\ O & & & \lambda_n \end{bmatrix}, \tag{88.1}$$

где $\lambda_1, \ldots, \lambda_n$ – собственные значения, соответствующие собственным векторам $e_1, \ldots, e_n$. ■

Следствие. В n-мерном пространстве линейный оператор, имеющий n различных собственных значений, является оператором простой структуры.

Обратное не верно. Тождественный оператор имеет простую структуру, однако все его собственные значения равны 1.

В соответствии с (88.1) оператор простой структуры называют также *диагонализуемым оператором*.

Теорема 88.2. *Линейный оператор $\mathcal{A} \in \mathcal{L}(V, V)$ имеет простую структуру тогда и только тогда, когда все его собственные подпространства в прямой сумме дают все пространство V:*

$$W_{\lambda_1} \oplus \ldots \oplus W_{\lambda_p} = V. \tag{88.2}$$

Доказательство. Необходимость. Пусть $\mathcal{A}$ имеет простую структуру. Тогда в пространстве V существует базис $e_1, \ldots, e_n$, состоящий из собственных векторов оператора $\mathcal{A}$. Каждый вектор этого базиса принадлежит некоторому собственному пространству W_{λ_k}. Следовательно, $W_{\lambda_1} + \ldots + W_{\lambda_p} = V$. Но эта сумма является прямой в силу теоремы 87.2.

Достаточность вытекает из критерия прямой суммы (теорема 66.1): совокупность базисов собственных подпространств W_{λ_k}, $k = \overline{1, p}$, образует базис V, тем самым пространство V имеет базис из собственных векторов оператора $\mathcal{A}$. ■

Замечание 1. На основании теоремы 66.1 условие (88.2) может быть заменено равносильным условием

$$\dim W_{\lambda_1} + \ldots + \dim W_{\lambda_p} = \dim V. \tag{88.3}$$

Теорема 88.3. *Линейный оператор, действующий в комплексном пространстве, имеет простую структуру тогда и только тогда, когда для каждого собственного значения этого оператора геометрическая кратность совпадает с алгебраической.*

Доказательство. Пусть $\lambda_1, \ldots, \lambda_p$, где $\lambda_i \neq \lambda_j$ при $i \neq j$, – все различные собственные значения оператора $\mathcal{A} \in \mathcal{L}(V, V)$ и пусть m_k и s_k, $k = \overline{1, p}$, – алгебраические и геометрические кратности λ_k.

Так как V – комплексное пространство, то, согласно теореме 86.6, $\dim V = m_1 + \ldots + m_p$. С другой стороны, $0 < \dim W_{\lambda_k} = s_k \le m_k$, $k = \overline{1,p}$, поэтому равенство (88.3) возможно тогда и только тогда, когда

$$s_k = m_k\,, \quad k = \overline{1,p}. \quad \blacksquare \tag{88.4}$$

Замечание 2. Теорема 88.3 в вещественном пространстве верна для тех операторов, чьи характеристические многочлены имеют только вещественные корни.

Матричная формулировка операторных свойств. В точности так же, как и для линейных операторов $\mathcal{A} \in \mathcal{L}(V,V)$, определяются собственные значения и собственные векторы квадратной матрицы $A \in P^{n\times n}$. Ненулевой вектор-столбец $x \in P^n$ называется ***собственным вектором матрицы*** $A \in P^{n\times n}$, если существует число $\lambda \in P$ такое, что

$$Ax = \lambda x\,.$$

При этом число λ называется ***собственным значением матрицы*** A, соответствующим собственному вектору x.

Если $e = (e_1, \ldots, e_n)$ – произвольный базис пространства V, то для оператора $\mathcal{A} \in \mathcal{L}(V,V)$ имеет место импликация

$$\mathcal{A}x = \lambda x \iff A_e x_e = \lambda x_e\,. \tag{88.5}$$

Это означает, что собственные значения оператора $\mathcal{A}$ и его матрицы в любом базисе $e = (e_1, \ldots, e_n)$ совпадают, а собственные векторы матрицы A_e являются координатными столбцами собственных векторов оператора $\mathcal{A}$ в этом базисе.

Характеристические многочлены оператора и его матрицы совпадают по определению, поэтому для квадратных матриц $A \in P^{n\times n}$ имеют место теоремы 86.5, 86.6.

Эта идентичность свойств оператора и его матрицы, замеченная уже в связи с первыми свойствами оператора (§77), продолжает наблюдаться.

Квадратная матрица $A \in P^{n\times n}$ называется ***матрицей простой структуры***, если она имеет n линейно независимых собственных векторов.

Из (88.5) следует, что линейный оператор является оператором простой структуры тогда и только тогда, когда его матрица в любом базисе имеет простую структуру.

Теорема 88.4 (матричная формулировка теоремы 88.1). *Квадратная матрица является матрицей простой структуры тогда и только тогда, когда она подобна диагональной.*

Доказательство. Пусть $A \in P^{n\times n}$ – заданная матрица. Рассмотрим произвольное линейное пространство V над полем P размерности n. Зафиксируем в пространстве V произвольный базис f. Пусть $\mathcal{A} \in \mathcal{L}(V,V)$ – линейный оператор, матрица которого в базисе

f совпадает с матрицей A, так что $A = A_f$ (согласно теореме 77 такой оператор существует). Матрица $A = A_f$ имеет простую структуру тогда и только тогда, когда $\mathcal{A}$ – оператор простой структуры. В силу теоремы 88.1 это равносильно существованию базиса e, в котором оператор $\mathcal{A}$ имеет диагональную матрицу (88.1). Таким образом, матрица A имеет простую структуру тогда и только тогда, когда она и диагональная матрица (88.1) являются матрицами одного оператора. Это равносильно их подобию (§82). $\blacksquare$

Замечание 3. Доказательство теоремы можно схематически изобразить в виде лаконичной диаграммы, которой мы неоднократно будем пользоваться:

$$\begin{array}{ccc} & \mathcal{A} \in \mathcal{L}(V,V) & \\ \overset{f}{\nearrow} & & \overset{e}{\searrow} \\ A = A_f & \cong & A_e = [\,\ddots\,]\,. \end{array} \tag{88.6}$$

Замечание 4. Теорема может быть доказана и на матричном языке. Действительно, матрица A является матрицей простой структуры тогда и только тогда, когда она имеет n линейно независимых собственных векторов $s_1, \ldots, s_n \in P^n$. Это равносильно следующей цепочке равенств: если $S = [s_1 \ldots s_n]$ – матрица, столбцами которой являются собственные векторы $s_1, \ldots, s_n$, Λ – диагональная матрица с собственными значениями $\lambda_1, \ldots, \lambda_n$ на главной диагонали, то

$$\left\{\begin{array}{l} As_1 = \lambda_1 s_1 \\ \quad\ldots \\ As_n = \lambda_n s_n \end{array}\right. \iff A[s_1 \ldots s_n] = [s_1 \ldots s_n]\Lambda \iff AS = S\Lambda \iff$$

$$\iff A = S\Lambda S^{-1}.$$

Жорданова клетка. Итак, самой простой формой матрицы обладают только операторы простой структуры, т.е. операторы, имеющие полный набор линейно независимых собственных векторов. Как мы уже отмечали, в вещественном пространстве существуют операторы, которые не имеют ни одного собственного вектора. И в комплексном пространстве не каждый линейный оператор обладает необходимым для базиса числом линейно независимых собственных векторов. Рассмотрим важный пример.

Матрица

$$J_k(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \lambda_0 & 1 \\ 0 & 0 & 0 & \ldots & 0 & \lambda_0 \end{bmatrix} \tag{88.7}$$

называется ***жордановой клеткой*** ***k-го порядка***. Эта матрица имеет:

1) характеристический многочлен $f(\lambda) = (\lambda_0 - \lambda)^k$;
2) собственное значение $\lambda = \lambda_0$ алгебраической кратности k;
3) собственные векторы, которые являются нетривиальными решениями однородной системы уравнений с матрицей

$$B = J_k(\lambda_0) - \lambda_0 I = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix},$$

ранг которой, очевидно, равен $k-1$. Таким образом, матрица $J_k(\lambda_0)$ (или оператор, задаваемый этой матрицей) имеет (теорема 30.2) один линейно независимый собственный вектор $(k-(k-1)=1)$ и не может быть матрицей простой структуры.

§89. Треугольная форма матрицы линейного оператора

Теорема 89.1. *В n-мерном комплексном пространстве V для любого линейного оператора $\mathcal{A} \in \mathcal{L}(V, V)$ существует система n вложенных друг в друга инвариантных подпространств $L_1, \dots, L_n$ всех размерностей от 1 до n, т.е. таких, что*

$$L_1 \subset L_2 \subset \ldots \subset L_n = V, \tag{89.1}$$

где $\dim L_k = k, \quad k = \overline{1, n}$.

Доказательство. Используем индукцию по n. Для $n = 1$ утверждение теоремы очевидно. Пусть теорема верна для всех линейных пространств размерности $n-1$. Докажем ее для n-мерного пространства V, воспользовавшись следующей леммой.

Лемма. *Линейный оператор, действующий в n-мерном комплексном пространстве, обладает инвариантным подпространством размерности $n-1$.*

Доказательство леммы. Линейный оператор $\mathcal{A}$, действующий в комплексном пространстве V, имеет собственное значение λ (теорема 86.6). Значит, $\det(\mathcal{A} - \lambda\mathcal{I}) = 0$ и $\operatorname{rg}(\mathcal{A} - \lambda\mathcal{I}) \le n-1$. Следовательно, $\dim \operatorname{im}(\mathcal{A} - \lambda\mathcal{I}) \le n-1$ и в пространстве V существует подпространство L размерности $n-1$, которое содержит $\operatorname{im}(\mathcal{A} - \lambda\mathcal{I})$. Очевидно, L инвариантно относительно оператора $\mathcal{A} - \lambda\mathcal{I}$. Покажем, что оно инвариантно и относительно $\mathcal{A}$. Пусть $x \in L$, тогда $(\mathcal{A} - \lambda\mathcal{I})x = y \in L$. Отсюда следует, что $\mathcal{A}x = \lambda x + y \in L$. Лемма доказана.

Вернемся к доказательству теоремы. Согласно лемме оператор $\mathcal{A}$, действующий в n-мерном комплексном пространстве V, имеет инвариантное подпространство L_{n-1} размерности $n-1$. Тогда индуцированный оператор $\mathcal{A}|L_{n-1}$ действует в $(n-1)$-мерном комплексном пространстве L_{n-1} и согласно индуктивному предположению для него существует система вложенных инвариантных подпространств $L_1 \subset \ldots \subset L_{n-1}$ таких, что $\dim L_k = k$, $k = \overline{1, n-1}$. Так как действия операторов $\mathcal{A}$ и $\mathcal{A}\,|\,L_{n-1}$ совпадают, то подпространства $L_1, \dots, L_{n-2} \subset L_{n-1}$ инвариантны относительно оператора $\mathcal{A}$. Остается добавить, что $L_{n-1} \subset L_n = V$. ■

Теорема 89.2. *Для любого линейного оператора $\mathcal{A}$, действующего в комплексном пространстве, существует базис, в котором матрица линейного оператора имеет треугольную форму.*

Доказательство. Согласно теореме 89.1 для оператора $\mathcal{A}$ найдется система инвариантных подпространств $L_1, \dots, L_n$ таких, что $\dim L_k = k$ и $L_1 \subset \ldots \subset L_{n-1} \subset L_n = V$. Искомый базис $e_1, \dots, e_n$ строим так: в качестве e_1 берем любой базис L_1, в качестве e_k, где $k > 1$, – вектор, дополняющий базис L_{k-1} до базиса L_k. В силу инвариантности подпространств L_k, $k = \overline{1, n}$, матрица A_e имеет верхнюю треугольную форму. ■

Замечание 1. Перенумерацией построенного базиса в обратном порядке можно получить нижнюю треугольную форму матрицы линейного оператора.

Замечание 2. На главной диагонали матрицы A_e расположены собственные значения оператора $\mathcal{A}$.

Теорема 89.3. *Любая квадратная комплексная матрица подобна матрице, имеющей треугольную форму.*

Доказательство теоремы повторяет доказательство теоремы 88.4 и ведется по схеме (88.6). ■

§90. Нильпотентный оператор

Линейный оператор $\mathcal{A} \in \mathcal{L}(V, V)$ называется ***нильпотентным***, если существует число $q \in \mathbb{N}$ такое, что $\mathcal{A}^q = \mathcal{O}$. Наименьшее число q, обладающее этим свойством, называется ***индексом нильпотентности (высотой)*** оператора $\mathcal{A}$. Очевидно, что индекс нильпотентности ненулевого оператора $q \ge 2$. Аналогично определяется нильпотентная матрица $A \in P^{n \times n}$ и ее индекс.

Примеры. 1. В пространстве вещественных многочленов M_n оператор дифференцирования (76.3) является нильпотентным оператором индекса $n+1$.

2. Жорданова клетка

$$J_k(0) = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

является нильпотентной матрицей индекса k.

Теорема 90.1. *Если $\mathcal{A} \in \mathcal{L}(V, V)$ – нильпотентный оператор индекса q и $x_0 \in V$ – вектор, для которого $\mathcal{A}^{q-1}x_0 \ne \theta$, то* *векторы*

$$x_0, \mathcal{A}x_0, \ldots, \mathcal{A}^{q-1}x_0 \qquad (90.1)$$

линейно независимы.

Доказательство. Рассмотрим равенство $\alpha_0 x_0 + \alpha_1 \mathcal{A}x_0 + \ldots + \alpha_{q-1}\mathcal{A}^{q-1}x_0 = \theta$. Применяя последовательно операторы $\mathcal{A}^{q-1}$, $\mathcal{A}^{q-2}$, $\ldots, \mathcal{A}$ к обеим частям этого равенства, получим, что $\alpha_0 = \alpha_1 = \ldots = \alpha_{q-1} = 0$. Это доказывает линейную независимость системы векторов (90.1). ■

Следствие 1. *Индекс нильпотентности не превосходит размерности пространства.*

Теорема 90.2. *В комплексном пространстве линейный оператор нильпотентен тогда и только тогда, когда все его собственные значения равны нулю.*

Доказательство. Необходимость. Если λ – собственное значение нильпотентного оператора $\mathcal{A} \in \mathcal{L}(V,V)$ индекса q и x – соответствующий собственный вектор, то, как нетрудно проверить, имеет место цепочка импликаций

$$\mathcal{A}x = \lambda x \;\Rightarrow\; \mathcal{A}^2 x = \lambda^2 x \;\Rightarrow\; \ldots \;\Rightarrow\; \mathcal{A}^q x = \lambda^q x\,.$$

Отсюда следует, что $\lambda^p x = \theta$. Так как $x \neq \theta$, то $\lambda = 0$.

Достаточность. Рассмотрим базис e комплексного пространства V, в котором оператор $\mathcal{A}$ имеет верхнюю треугольную матрицу (§89), главная диагональ которой состоит целиком из нулей (§89, замечание 2). Итак,

$$A_e = \begin{bmatrix} 0 & a_{12} & a_{13} & \ldots & a_{1n} \\ 0 & 0 & a_{23} & \ldots & a_{2n} \\ & & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & a_{n-1,n} \\ 0 & 0 & 0 & \ldots & 0 \end{bmatrix}.$$

Как нетрудно проверить, при последовательном возведении этой матрицы в степени $q = 2, 3, \ldots, n$ нетривиальный треугольник, расположенный над главной диагональю, перемещается каждый раз на одну диагональ выше, так что $(A_e)^n = O$. Значит, $\mathcal{A}^n = \mathcal{O}$. ■

Замечание. Необходимость этого утверждения, как видно из самого доказательства, имеет место и в вещественном пространстве.

Прямая сумма операторов. Если $V = L_1 \oplus L_2 \oplus \ldots \oplus L_p$ – прямая сумма подпространств $L_1, L_2, \ldots, L_p$, инвариантных относительно линейного оператора $\mathcal{A} \in \mathcal{L}(V,V)$, то оператор $\mathcal{A}$ называется ***прямой суммой индуцированных операторов*** $\mathcal{A}|L_1, \ldots, \mathcal{A}|L_p$. Этот термин мотивирован тем обстоятельством, что для любого вектора $x \in V$ с разложением $x = x_1 + \ldots + x_p$, где $x_k \in L_k$, $k = \overline{1,p}$, имеет место равенство $\mathcal{A}x = \mathcal{A}x_1 + \ldots + \mathcal{A}x_p = (\mathcal{A}|L_1)x_1 + \ldots + (\mathcal{A}|L_p)x_p$.

Теорема 90.3. *Произвольный линейный оператор $\mathcal{A} \in \mathcal{L}(V,V)$ является прямой суммой нильпотентного и обратимого операторов, причем это разложение единственно.*

Доказательство. Для доказательства теоремы необходимо показать, что существует единственная пара подпространств L_1, L_2, инвариантных относительно линейного оператора $\mathcal{A}$ и таких, что $V = L_1 \oplus L_2$, $\mathcal{A}|L_1$ нильпотентен, $\mathcal{A}|L_2$ обратим. Будем считать, что оператор $\mathcal{A}$ вырожден (для невырожденного $\mathcal{A}$: $L_1 = \{\theta\}$, $L_2 = V$) и не нильпотентен (для нильпотентного $\mathcal{A}$: $L_1 = V$, $L_2 = \{\theta\}$).

Существование. Обозначим для $k \in \mathbb{N}$

$$N_k = \ker \mathcal{A}^k\,, \quad T_k = \operatorname{im} \mathcal{A}^k.$$

1. Покажем, что *подпространства N_k строго вложены друг в друга до некоторого момента q, начиная с которого все N_k совпадают*, т.е.

$$N_1 \subset N_2 \subset \ldots \subset N_q = N_{q+1} = \ldots \qquad (90.2)$$

а) Вложение $N_k \subseteq N_{k+1}$ очевидно, так как если $\mathcal{A}^k x = \theta$, то $\mathcal{A}^{k+1}x = \mathcal{A}(\mathcal{A}^k) = \mathcal{A}\theta = \theta$.

б) Пусть $N_k = N_{k+1}$, тогда $N_{k+1} = N_{k+2}$, так как $N_{k+1} \subseteq N_{k+2}$, $N_{k+2} \subseteq N_{k+1}$. Второе из этих вложений следует из того, что если $x \in N_{k+2}$, то $\mathcal{A}^{k+2}x = \theta$, т.е. $\mathcal{A}^{k+1}(\mathcal{A}x) = \theta$. Значит, $\mathcal{A}x \in N_{k+1} = N_k$, следовательно, $\mathcal{A}^k(\mathcal{A}x) = \theta$, т.е. $\mathcal{A}^{k+1}x = \theta$.

Из "а" и "б" следует, что подпространство N_k либо строго вложено в N_{k+1}, либо совпадает со всеми последующими ядрами. Так как в конечномерном пространстве размерности подпространств N_k не могут бесконечно возрастать, то наступит момент q, начиная с которого все ядра N_k будут совпадать с N_q.

2. Зафиксируем этот момент q и покажем, что

$$V = N_q \oplus T_q.$$

Действительно, $\dim V = \dim N_q + \dim T_q$ в силу теоремы 80.4 о ранге и дефекте, при этом $N_q \cap T_q = \{\theta\}$, так как если $y \in N_q$, $y \in T_q$, то $\mathcal{A}^q y = \theta$, $y = \mathcal{A}^q x$, т.е. $\mathcal{A}^{2q}x = \theta$. Значит, $x \in N_{2q} = N_q$ и $\mathcal{A}^q x = \theta$, откуда следует, что $y = \theta$.

3. Подпространства N_q, T_q *инвариантны относительно* $\mathcal{A}$, ибо:

а) если $x \in N_q$, то $x \in N_{q+1} = N_q$, поэтому $\mathcal{A}^{q+1}x = \theta$, т.е. $\mathcal{A}^q(\mathcal{A}x) = \theta$, следовательно, $\mathcal{A}x \in N_q$;

б) если $y \in T_q$, то $y = \mathcal{A}^q x$ и $\mathcal{A}y = \mathcal{A}^{q+1}x = \mathcal{A}^q(\mathcal{A}x) = \mathcal{A}^q x_1$, где $x_1 = \mathcal{A}x$, следовательно, $\mathcal{A}y \in T_q$.

4. Оператор $\mathcal{A}|N_q$ – *нильпотентный оператор индекса* q, так как:

а) $\mathcal{A}^q x = \theta$ для любого вектора $x \in N_q$;

б) существует вектор $x_0 \in N_q$ такой, что $\mathcal{A}^{q-1}x_0 \neq \theta$, ибо $N_{q-1} \neq N_q$.

5. Оператор $\mathcal{A}|T_q$ *обратим*, так как его ядро состоит только из нулевого вектора. Действительно, если $y \in \ker \mathcal{A}|T_q$, то $y \in T_q$, $\mathcal{A}y = \theta$, т.е. $y = \mathcal{A}^q x$ и $\mathcal{A}^{q+1}x = \theta$. Отсюда следует, что $x \in N_{q+1} = N_q$, т.е. $\mathcal{A}^q x = \theta$ и $y = \theta$.

Утверждения пунктов 2 – 5 доказывают существование искомого разложения: $L_1 = N_q$, $L_2 = T_q$.

Единственность. Пусть существует другое разложение $V = N \oplus T$, обладающее всеми свойствами первого.

1. Нильпотентность оператора $\mathcal{A}|N$ означает, что $\mathcal{A}^k x = \theta$, $\forall x \in N$, при некотором $k \in \mathbb{N}$. Следовательно, $N \subseteq N_k \subseteq N_q$ и

$$\dim N \leq \dim N_q\,. \qquad (90.3)$$

2. Обратимость оператора $\mathcal{A}|T$ означает, что $\operatorname{im}\mathcal{A}|T = T$. Следовательно, для любого вектора $y \in T$ имеет место представление $y = \mathcal{A}y_1$, где $y_1 \in T$. Используя такое же представление для вектора y_1 и всех последующих, получаем, что $y = \mathcal{A}y_1 = \mathcal{A}^2 y_2 = \ldots = \mathcal{A}^q y_q$. Таким образом, $T \subseteq T_q$ и

$$\dim T \leq \dim T_q\,. \qquad (90.4)$$

Так как $\dim N + \dim T = \dim V = \dim N_q + \dim T_q$, то из (90.3), (90.4) следует, что $N = N_q$, $T = T_q$. ■

Следствие 2. *Оператор $\mathcal{A}$ на подпространстве N_q имеет только нулевые собственные значения, а на подпространстве T_q не имеет нулевых собственных значений.* Это вытекает из доказанной теоремы с учетом теоремы 90.2 и того, что обратимый оператор не вырожден.

Следствие 3. *Для оператора $\mathcal{A}$, действующего в комплексном пространстве V, с характеристическим многочленом $f(\lambda) = \det(\mathcal{A} - \lambda\mathcal{I}) = (-\lambda)^{m_1}(\lambda_2 - \lambda)^{m_2}\ldots(\lambda_p - \lambda)^{m_p}$:*

а) характеристические многочлены $f_1(\lambda)$ и $f_2(\lambda)$ операторов $\mathcal{A}|N_q$ и $\mathcal{A}|T_q$ имеют вид

$$f_1(\lambda) = (-\lambda)^{m_1}, \; f_2(\lambda) = (\lambda_2 - \lambda)^{m_2}\ldots(\lambda_p - \lambda)^{m_p}; \qquad (90.5)$$

б) при этом

$$\dim N_q = m_1, \quad \dim T_q = m_2 + \ldots + m_p\,. \qquad (90.6)$$

Соотношение (90.5) вытекает из (86.4), если учесть, что $f_2(0) \neq 0$. Соотношение (90.6) следует из (90.5), так как размерность пространства совпадает со степенью характеристического многочлена.

§91. Корневые подпространства

Теоремой 90.3 осуществляется одновременное расщепление пространства $V = N_q \oplus T_q$, характеристического многочлена $f(\lambda) = f_1(\lambda)f_2(\lambda)$ и самого оператора $\mathcal{A}$ на индуцированные операторы $\mathcal{A}|N_q$ и $\mathcal{A}|T_q$. Такое расщепление направлено на выделение во всем пространстве V максимального подпространства N_q, на котором оператор $\mathcal{A}$ имеет только нулевые собственные значения. Этот же прием может быть использован и для других собственных значений.

Теорема 91.1 (**о расщеплении линейного оператора**). *Для любого линейного оператора $\mathcal{A}$, действующего в комплексном пространстве V, с характеристическим многочленом*

$$f(\lambda) = a_n(\lambda_1 - \lambda)^{m_1}\ldots(\lambda_p - \lambda)^{m_p}, \quad \text{где} \quad \lambda_i \neq \lambda_j \; \text{при} \; i \neq j\,, \qquad (91.1)$$

существуют инвариантные подпространства $K_{\lambda_1}, \ldots, K_{\lambda_p}$ такие, что

$$V = K_{\lambda_1} \oplus \ldots \oplus K_{\lambda_p}\,; \qquad (91.2)$$

$$\dim K_{\lambda_j} = m_j\,, \quad j = \overline{1,p}\,; \qquad (91.3)$$

$$f_j(\lambda) = \det(\mathcal{A}|K_{\lambda_j} - \lambda\mathcal{I}) = (\lambda_j - \lambda)^{m_j}\,, \quad j = \overline{1,p}\,. \qquad (91.4)$$

Доказательство. Заметим предварительно, что:

а) если $\lambda_{\mathcal{A}}$ и $\lambda_{\mathcal{B}}$ – собственные значения операторов $\mathcal{A}$ и $\mathcal{B} = \mathcal{A} - \lambda_0\mathcal{I}$, то, как нетрудно проверить,

$$\lambda_{\mathcal{B}} = \lambda_{\mathcal{A}} - \lambda_0\,; \qquad (91.5)$$

б) если подпространство L инвариантно относительно оператора $\mathcal{B} = \mathcal{A} - \lambda_0\mathcal{I}$, то L инвариантно относительно $\mathcal{A}$, так как если $x \in L$, то $\mathcal{B}x = y \in L$, т.е. $(\mathcal{A} - \lambda_0\mathcal{I})x = y \in L$, и, следовательно, $\mathcal{A}x = \lambda_0 x + y \in L$.

Доказательство теоремы состоит из $p-1$ однотипных шагов.

Первый шаг. Рассмотрим оператор $\mathcal{B}_1 = \mathcal{A} - \lambda_1\mathcal{I}$. Применив к нему теорему 90.3, получим подпространства N_{q_1} и T_{q_1}, которые с учетом (91.5) обладают следующими свойствами:

$V = N_{q_1} \oplus T_{q_1}$;
N_{q_1}, T_{q_1} инвариантны относительно $\mathcal{B}_1$;
$\dim N_{q_1} = m_1$, $\dim T_{q_1} = m_2 + \ldots + m_p$;
$\det(\mathcal{B}_1|N_{q_1} - \lambda\mathcal{I}) = (-\lambda)^{m_1}$;
$\det(\mathcal{B}_1|T_{q_1} - \lambda\mathcal{I}) = (\lambda_2 - \lambda_1 - \lambda)^{m_2}\ldots(\lambda_p - \lambda_1 - \lambda)^{m_p}$.

Положив $K_{\lambda_1} = N_{q_1}$, $V_1 = T_{q_1}$, получим подпространства K_{λ_1} и V_1, инвариантные относительно оператора $\mathcal{A}$ (согласно п."б") и такие, что:

$$\begin{aligned} &V = K_{\lambda_1} \oplus V_1; \\ &\dim K_{\lambda_1} = m_1, \quad \dim V_1 = m_2 + \ldots + m_p; \\ &\det(\mathcal{A}|K_{\lambda_1} - \lambda\mathcal{I}) = (\lambda_1 - \lambda)^{m_1}; \\ &\det(\mathcal{A}|V_1 - \lambda\mathcal{I}) = (\lambda_2 - \lambda)^{m_2}\ldots(\lambda_p - \lambda)^{m_p}. \end{aligned} \qquad (91.6)$$

Второй шаг состоит в применении первого шага к оператору $\mathcal{A}_1 = \mathcal{A}|V_1$. В результате второго шага получим подпространства K_{λ_2} и V_2, обладающие свойствами (91.6) с очевидными изменениями. Через $p-1$ шагов получим все подпространства $K_{\lambda_1}, \ldots, K_{\lambda_p} = V_{p-1}$, удовлетворяющие требованиям теоремы. ■

Следствие. Для любого линейного оператора, действующего в комплексном пространстве, существует базис, в котором его матрица имеет квазидиагональную форму (согласно теореме 84.2), *у которой число диагональных клеток совпадает с числом различных собственных значений, а их размеры – с алгебраическими кратностями собственных значений*, или, в матричной формулировке,

любая квадратная комплексная матрица подобна квазидиагональной матрице, обладающей указанным выше свойством.

Замечание. Разложение (91.2), обладающее всеми свойствами из теоремы 91.1, для каждого оператора единственно в силу теоремы 90.3.

Корневые векторы. Пусть λ_j – собственное значение оператора $\mathcal{A}$. Вектор $x \in V$ называется ***корневым вектором*** оператора $\mathcal{A}$, ***отвечающим собственному значению*** λ_j, если $(\mathcal{A} - \lambda_j\mathcal{I})^k x = \theta$ при некотором $k \in \mathbb{Z}$, $k \geq 0$. *Высотой* корневого вектора называется наименьшее k, обладающее указанным свойством. Очевидно, высота нулевого корневого вектора (и только нулевого) равна нулю.

Отметим простейшие свойства корневых векторов, вытекающие из определения.

1°. *Корневые векторы высоты* 1 ***являются собственными векторами***, так как удовлетворяют условиям $(\mathcal{A} - \lambda_j\mathcal{I})x = \theta$, $x \neq \theta$.

2°. *Если x – корневой вектор высоты $k > 0$, то вектор $(\mathcal{A}-\lambda_j\mathcal{I})x$ является корневым вектором высоты $k-1$.*

3°. *Если x – корневой вектор высоты $k > 0$, то векторы*
$$x, (\mathcal{A} - \lambda_j\mathcal{I})x, \ldots, (\mathcal{A} - \lambda_j\mathcal{I})^{k-1}x$$
линейно независимы. Доказательство этого утверждения аналогично доказательству теоремы 90.1.

4°. *Корневые векторы различных высот линейно независимы.*

Корневые векторы высоты $k > 1$ называются ***присоединенными векторами*** $(k-1)$-*го порядка*. Итак, если x – присоединенный вектор $(k-1)$-го порядка, то $(\mathcal{A} - \lambda_j\mathcal{I})^k x = \theta$, $(\mathcal{A} - \lambda_j\mathcal{I})^{k-1}x \neq \theta$ или, другими словами, $(\mathcal{A} - \lambda_j\mathcal{I})^{k-1}x$ – собственный вектор оператора $\mathcal{A}$, отвечающий собственному значению λ_j. Таким образом, корневой вектор – это либо нулевой вектор, либо собственный вектор, либо присоединенный вектор.

Корневые подпространства. Множество всех корневых векторов оператора $\mathcal{A}$, отвечающих собственному значению λ_j, называется ***корневым подпространством*** оператора $\mathcal{A}$, ***отвечающим собственному значению*** λ_j. *Обозначение:* K_{λ_j}. Итак, $K_{\lambda_j} = \{x \in V \mid \exists k \in \mathbb{N} : (\mathcal{A} - \lambda_j\mathcal{I})^k x = \theta\}$. Опишем структуру корневого подпространства.

Пусть $N_k = \ker(\mathcal{A} - \lambda_j\mathcal{I})^k$. Тогда согласно (90.2)

$$N_1 \subset N_2 \subset \ldots \subset N_q = N_{q+1} = \ldots \tag{91.7}$$

1. Очевидно, что N_1 состоит из корневых векторов высоты, не превосходящей 1, т.е. из нулевого вектора θ и всех собственных векторов, отвечающих собственному значению λ_j, т.е. совпадает с собственным подпространством W_{λ_j}. Таким образом,

$$N_1 = W_{\lambda_j} \tag{91.8}$$

и, следовательно,

$$\dim N_1 = s_j\,, \tag{91.9}$$

где s_j – геометрическая кратность собственного значения λ_j.

2. Подпространство N_2 в (91.7) состоит из корневых векторов высоты, не превосходящей 2, а подпространство N_q состоит из корневых векторов всех высот, т.е. совпадает с корневым подпространством K_{λ_j}. Таким образом, q – максимальная высота корневого вектора, отвечающего собственному значению λ_j, $K_{\lambda_j} = \ker(\mathcal{A} - \lambda_j\mathcal{I})^q$ и

$$W_{\lambda_j} = N_1 \subset N_2 \subset \ldots \subset N_q = K_{\lambda_j}. \tag{91.10}$$

Это означает, что подпространства $K_{\lambda_1}, \ldots, K_{\lambda_p}$, построенные при доказательстве теоремы 91.1, совпадают с корневыми подпространствами, отвечающими собственным значениям $\lambda_1, \ldots, \lambda_p$ соответственно. Следовательно, корневые подпространства обладают свойствами (91.2)–(91.4).

3. Из (91.10) следует, что максимальная высота q корневых векторов, отвечающих собственному значению λ_j, совпадает с индексом нильпотентности оператора $\mathcal{A} - \lambda_j\mathcal{I}$ и согласно свойству 3° не превосходит размерности K_{λ_j}, т.е. алгебраической кратности собственного значения λ_j.

4. Из (91.10) следует также, что собственное подпространство W_{λ_j} является подпространством корневого подпространства K_{λ_j}, так что $s_j \leq m_j$. При этом

$$W_{\lambda_j} = K_{\lambda_j} \iff s_j = m_j\,. \tag{91.11}$$

Таким образом, цепочка (91.10) дает еще одно доказательство теорем 87.1 и 88.3.

§92. Жорданова форма

Перейдем к построению базиса, в котором матрица линейного оператора имеет наиболее простую форму. Этим базисом будет совокупность построенных специальным образом базисов корневых подпространств.

Канонический базис корневого подпространства. Пусть K_{λ_j} – корневое подпространство оператора $\mathcal{A}$, отвечающее собственному значению λ_j. Положим

$$\mathcal{B} = \mathcal{A} - \lambda_j\mathcal{I},\ N_k = \ker \mathcal{B}^k,\ n_k = \dim N_k,\ r_k = \operatorname{rg} \mathcal{B}^k.$$

Построим сначала само корневое подпространство K_{λ_j}. Для этого согласно (91.7) необходимо найти момент q, начиная с которого все ядра N_k будут совпадать с $N_q = K_{\lambda_j}$; при этом в силу (91.9), (91.3) $n_1 = s_j < n_2 < \ldots < n_q = m_j$, где s_j и m_j – геометрическая и алгебраическая кратности λ_j.

Теперь будем строить базис K_{λ_j}, последовательно просматривая подпространства $N_q, N_{q-1}, \dots, N_1$.

N_q) Пусть $f_1, \dots, f_{t_q}$ – векторы, дополняющие произвольный базис N_{q-1} до базиса N_q. Ясно, что:

1) они будут корневыми векторами высоты q;
2) их количество равно $n_q - n_{q-1}$;
3) $t_q = n_q - n_{q-1} = (n_q - n_{q-1}) - (n_{q+1} - n_q) = -n_{q+1} + 2n_q - n_{q-1}$, так как $n_{q+1} = n_q$;
4) никакая нетривиальная линейная комбинация этих векторов не принадлежит N_{q-1} (такие векторы будем называть *линейно независимыми над* N_{q-1}).

N_{q-1}) Построим векторы $\mathcal{B}f_1, \dots, \mathcal{B}f_{t_q}$. Эти векторы являются корневыми векторами высоты $q-1$, и они линейно независимы над N_{q-2}, так как в противном случае для нетривиального набора чисел $\alpha_1, \dots, \alpha_{t_q}$ будем иметь $\mathcal{B}^{q-2}\sum_{k=1}^{t_q}\alpha_k\mathcal{B}f_k = \theta$, т.е. $\mathcal{B}^{q-1}\sum_{k=1}^{t_q}\alpha_k f_k = \theta$, и $\sum_{k=1}^{t_q}\alpha_k f_k \in N_{q-1}$, что противоречит линейной независимости $f_1, \dots, f_{t_q}$ над N_{q-1}.

Дополним эти векторы векторами $g_1, \dots, g_{t_{q-1}} \in N_{q-1}$ так, чтобы векторы $\mathcal{B}f_1, \dots, \mathcal{B}f_{t_q}, g_1, \dots, g_{t_{q-1}}$ дополняли произвольный базис N_{q-2} до базиса N_{q-1}. Ясно, что:

1) они будут корневыми векторами высоты $q-1$;
2) их количество равно $n_{q-1} - n_{q-2}$;
3) $t_{q-1} = (n_{q-1} - n_{q-2}) - (n_q - n_{q-1}) = -n_q + 2n_{q-1} - n_{q-2}$;
4) они линейно независимы над N_{q-2}.

N_{q-2}) Аналогично строятся векторы

$$\mathcal{B}^2 f_1, \dots, \mathcal{B}^2 f_{t_q}, \mathcal{B}g_1, \dots, \mathcal{B}g_{t_{q-1}}, h_1, \dots, h_{t_{q-2}},$$

дополняющие произвольный базис N_{q-3} до базиса N_{q-2}. Для этой системы векторов справедливы те же факты 1–4, что и для векторов, построенных на предыдущих шагах .

Выполняя далее такие же построения в подпространствах N_{q-3}, N_{q-4}, ..., придем к подпространству N_1.

N_1) Здесь строятся векторы

$$\mathcal{B}^{q-1}f_1, \dots, \mathcal{B}^{q-1}f_{t_q}, \mathcal{B}^{q-2}g_1, \dots, \mathcal{B}^{q-2}g_{t_{q-1}}, \dots, \mathcal{B}v_1, \dots, \mathcal{B}v_{t_2},$$

которые дополняются векторами $u_1, \dots, u_{t_1}$ до базиса N_1. Таким образом, векторы

$$\mathcal{B}^{q-1}f_1, \dots, \mathcal{B}^{q-1}f_{t_q}, \mathcal{B}^{q-2}g_1, \dots, \mathcal{B}^{q-2}g_{t_{q-1}}, \dots, \mathcal{B}v_1, \dots, \mathcal{B}v_{t_2}, u_1, \dots, u_{t_1}$$

1) являются собственными векторами;
2) их количество равно $n_1 = n_1 - n_0$ (очевидно, $n_0 = \text{def}\, \mathcal{B}^0 = 0$);
3) $t_1 = (n_1 - n_0) - (n_2 - n_1) = -n_2 + 2n_1 - n_0$;
4) они линейно независимы.

Полученную за q шагов систему векторов удобно объединить в приводимую ниже таблицу, которую будем называть *жордановой лестницей*.

Теорема 92.1. *Построенная система векторов образует базис корневого подпространства* K_{λ_j}.

Доказательство. Количество векторов в построенной системе равно размерности пространства K_{λ_j}, так как $n_1 + (n_2 - n_1) + (n_3 - n_2) + \dots + (n_q - n_{q-1}) = n_q = \dim K_{\lambda_j}$. Эти векторы линейно независимы, так как если приравнять их линейную комбинацию нулевому вектору и последовательно применить операторы $\mathcal{B}^{q-1}, \mathcal{B}^{q-2}, \dots, \mathcal{B}$ к обеим частям полученного равенства, то все коэффициенты линейной комбинации окажутся равными нулю. ■

$t_q = -n_{q-1} + 2n_q - n_{q+1}$

$t_{q-1} = -n_{q-2} + 2n_{q-1} - n_q$

$t_1 = -n_0 + 2n_1 - n_2$

N_q	$f_1, \dots, f_{t_q}$			
N_{q-1}	$\mathcal{B}f_1, \dots, \mathcal{B}f_{t_q}$	$g_1, \dots, g_{t_{q-1}}$		
$\vdots$	$\vdots$	$\vdots$	$\ddots$	
N_1	$\mathcal{B}^{q-1}f_1, \dots, \mathcal{B}^{q-1}f_{t_q}$	$\mathcal{B}^{q-2}g_1, \dots, \mathcal{B}^{q-2}g_{t_{q-1}}$	$\dots$	$u_1, \dots, u_{t_1}$

$n_q - n_{q-1}$

$n_{q-1} - n_{q-2}$

$n_1 = n_1 - n_0$

Нумерация базиса. Будем нумеровать векторы построенного базиса по столбцам жордановой лестницы: внутри каждого столбца снизу вверх, а сами столбцы в произвольном порядке. Полученный таким образом базис называется *каноническим* (или *жордановым*) *базисом корневого подпространства* K_{λ_j}.

Матрица оператора $\mathcal{A}|K_{\lambda_j}$ в каноническом базисе. 1. Пусть $e_1, \dots, e_q$ – векторы первого столбца жордановой лестницы. Тогда

$$\begin{cases} e_1 = \mathcal{B}^{q-1}f_1, \\ e_2 = \mathcal{B}^{q-2}f_1, \\ \dots \\ e_q = f_1 \end{cases} \Rightarrow \begin{cases} \mathcal{B}e_1 = \theta, \\ \mathcal{B}e_2 = e_1, \\ \dots \\ \mathcal{B}e_q = e_{q-1} \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} (\mathcal{A} - \lambda_j\mathcal{I})e_1 = \theta, \\ (\mathcal{A} - \lambda_j\mathcal{I})e_2 = e_1, \\ \dots \\ (\mathcal{A} - \lambda_j\mathcal{I})e_q = e_{q-1} \end{cases} \Rightarrow \begin{cases} \mathcal{A}e_1 = \lambda_j e_1, \\ \mathcal{A}e_2 = \lambda_j e_2 + e_1, \\ \dots \\ \mathcal{A}e_q = \lambda_j e_q + e_{q-1}. \end{cases} \qquad \textbf{(92.1)}$$

Этой группе векторов канонического базиса соответствуют **первые** q столбцов матрицы $\mathcal{A}|K_{\lambda_j}$ в каноническом базисе, которые **согласно**

(92.1) имеют вид

$$\begin{bmatrix} \boxed{J_q(\lambda_j)} \\ O \end{bmatrix}, \qquad (92.2)$$

где $J_q(\lambda_j)$ – клетка Жордана (88.7) q-го порядка с λ_j на главной диагонали.

Точно так же устроены столбцы матрицы $\mathcal{A}|K_{\lambda_j}$, определяемые векторами второго столбца жордановой лестницы: диагональная клетка имеет тот же вид $J_q(\lambda_j)$, а все остальные элементы равны нулю. Таким образом, первая группа из t_q столбцов жордановой лестницы порождает клетки Жордана q-го порядка с λ_j на главной диагонали. Число этих клеток равно t_q.

2. Следующая группа из t_{q-1} столбцов жордановой лестницы определяет клетки $J_{q-1}(\lambda_j)$ на главной диагонали матрицы $\mathcal{A}|K_{\lambda_j}$. Число клеток $(q-1)$-го порядка равно t_{q-1}.

3. Рассмотрев все столбцы жордановой лестницы, получим матрицу A_j оператора $\mathcal{A}|K_{\lambda_j}$ в каноническом базисе. Из (92.2) и последующих рассуждений вытекает, что A_j – квазидиагональная матрица с клетками Жордана $J_k(\lambda_j)$ на главной диагонали. Всего этих клеток столько, сколько столбцов в жордановой лестнице, т.е. n_1 или, согласно (91.9), s_j (геометрическая кратность λ_j). Таким образом,

$$A_j = \begin{bmatrix} \boxed{J_{q_1}(\lambda_j)} & & & O \\ & \boxed{J_{q_2}(\lambda_j)} & & \\ & & \ddots & \\ O & & & \boxed{J_{q_{s_j}}(\lambda_j)} \end{bmatrix}, \qquad (92.3)$$

где $q_1 + \ldots + q_{s_j} = \dim K_{\lambda_j} = m_j$, а число клеток k-го порядка равно числу $t_k = -n_{k-1} + 2n_k - n_{k+1}$, $k = \overline{1,q}$.

4. Несмотря на некоторый произвол в выборе базиса, матрица A_j определена однозначно с точностью до порядка клеток, так как количество всех клеток равно геометрической кратности s_j собственного значения λ_j, а количество клеток k-го порядка равно числу $t_k = -n_{k-1} + 2n_k - n_{k+1}$, $k = \overline{1,q}$, или, согласно (80.2), $t_k = r_{k-1} - 2r_k + r_{k+1}$, $\quad k = \overline{1,q}$.

Таким образом, структура клеток Жордана в матрице A_j определяется только оператором $\mathcal{A}$. Порядок клеток определен порядком нумерации столбцов жордановой лестницы.

Жорданова форма матрицы линейного оператора в комплексном пространстве. Будем придерживаться прежних обозначений: m_j и s_j – алгебраическая и геометрическая кратности λ_j, $r_k = \operatorname{rg}(\mathcal{A} - \lambda_j\mathcal{I})^k$.

Теорема 92.2. *Пусть $\mathcal{A} \in \mathcal{L}(V,V)$ – линейный оператор, действующий в комплексном пространстве V, и его характеристический многочлен имеет вид*

$$f(\lambda) = a_n(\lambda_1 - \lambda)^{m_1} \ldots (\lambda_p - \lambda)^{m_p}, \quad \text{где} \quad \lambda_i \neq \lambda_j \text{ при } i \neq j.$$

Тогда в пространстве V существует базис e, в котором матрица оператора $\mathcal{A}$ имеет квазидиагональную форму

$$A_e = \begin{bmatrix} \boxed{A_1} & & & O \\ & \boxed{A_2} & & \\ & & \ddots & \\ O & & & \boxed{A_p} \end{bmatrix}, \qquad (92.4)$$

в которой матрицы A_j, $j = \overline{1,p}$, имеют вид (92.3).

Доказательство. В соответствии с теоремой 91.1, $V = K_{\lambda_1} \oplus \ldots \oplus K_{\lambda_p}$. В качестве искомого базиса e возьмем совокупность канонических базисов корневых подпространств $K_{\lambda_1}, \ldots, K_{\lambda_p}$. Согласно теореме 84.2 матрица A_e имеет вид (92.4), где A_j – матрица оператора $\mathcal{A}|K_{\lambda_j}$ в каноническом базисе K_{λ_j}. Следовательно, матрица A_j имеет вид (92.3). ■

Следствие. Для собственных значений оператора $\mathcal{A}$ имеют место соотношения $\lambda_1 + \ldots + \lambda_n = \operatorname{tr}\mathcal{A}$, $\lambda_1 \cdot \ldots \cdot \lambda_n = \det\mathcal{A}$.

Полученная форма матрицы линейного оператора называется *жордановой формой*, а построенный базис – *каноническим* (или *жордановым*) *базисом*.

Замечание 1. *Жорданова форма матрицы линейного оператора определена однозначно с точностью до порядка клеток Жордана.* Порядок клеток определен порядком нумерации корневых подпространств и тем порядком, в котором нумеруются столбцы жордановой лестницы в каноническом базисе K_{λ_j}.

Замечание 2. *Для операторов простой структуры, и только для них, жорданова форма совпадает с диагональной*, так как критерий (88.4) с учетом (91.11) равносилен условию

$$W_{\lambda_j} = K_{\lambda_j}, \quad j = \overline{1,p}. \qquad (92.5)$$

Замечание 3. В матричной формулировке теорема 92.2 означает, что *любая квадратная комплексная матрица подобна матрице, имеющей жорданову форму*. Это может быть доказано по известной схеме (88.6).

Матрица A_e, имеющая жорданову форму и подобная матрице A, называется *жордановой формой матрицы* A.

Теорема 92.3. *Две матрицы $A, B \in \mathbb{C}^{n\times n}$ подобны тогда и только тогда, когда их жордановы формы совпадают.*

Это утверждение следует из того, что квадратные матрицы одинакового порядка над общим полем подобны тогда и только тогда, когда они являются матрицами одного и того же линейного оператора (§82, п. 8°). ■

Теорема 92.4 (теорема Гамильтона–Кэли). *Линейный оператор, действующий в комплексном (или в вещественном) пространстве, является корнем своего характеристического многочлена.*

Доказательство. 1. Докажем сначала это утверждение для комплексного пространства V. Пусть $\mathcal{A} \in \mathcal{L}(V,V)$ и его характеристический многочлен имеет вид (91.1). Согласно теореме 91.1, $V = K_{\lambda_1} \oplus \ldots \oplus K_{\lambda_p}$ и, следовательно, для любого вектора $x \in V$ имеет место разложение $x = x_1 + \ldots + x_p$, где $x_j \in K_{\lambda_j}$, $j = \overline{1,p}$. Тогда

$$f(\mathcal{A})x = f(\mathcal{A})x_1 + \ldots + f(\mathcal{A})x_j + \ldots + f(\mathcal{A})x_p\,. \qquad (92.6)$$

Каждое слагаемое в этом разложении равно нулевому вектору, так как $f(\mathcal{A})x_j = (\lambda_1\mathcal{I}-\mathcal{A})^{m_1}\ldots(\lambda_j\mathcal{I}-\mathcal{A})^{m_j}\ldots(\lambda_p\mathcal{I}-\mathcal{A})^{m_p}x_j = \theta$, ибо операторы в этом произведении перестановочны, а $(\mathcal{A}-\lambda\mathcal{I})^{m_j}x_j = \theta$ в силу (91.10). Следовательно, $f(\mathcal{A})x = \theta$, $\forall x \in V$, т.е. $f(\mathcal{A}) = \mathcal{O}$.

2. Пусть V – вещественное линейное пространство. Возьмем какой-либо базис e пространства V, и пусть A_e – матрица оператора $\mathcal{A}$ в этом базисе. Для доказательства теоремы достаточно показать, что $f(A_e) = O$. Рассмотрим любое комплексное пространство V_1 той же размерности, что и V. Пусть f – произвольный базис V_1, тогда матрица A_e является матрицей оператора $\mathcal{B} \in \mathcal{L}(V_1,V_1)$ в базисе f, т.е. $A_e = B_f$. Значит, характеристические многочлены операторов $\mathcal{A}$ и $\mathcal{B}$ совпадают и, согласно п. 1 доказательства, $f(A_e) = O$. ■

§93. Инвариантные подпространства минимальной размерности

Теорема 93.1. *У всякого линейного оператора в комплексном пространстве существует одномерное инвариантное подпространство.*

Доказательство. Утверждение следует из существования собственного вектора для любого оператора, действующего в комплексном пространстве (§86): если e – собственный вектор оператора $\mathcal{A}$, то $\mathcal{L}(e)$ – одномерное подпространство, инвариантное относительно $\mathcal{A}$. ■

Теорема 93.2. *У всякого линейного оператора в вещественном пространстве существует одномерное или двумерное инвариантное подпространство.*

Доказательство. Пусть V – вещественное пространство, $\mathcal{A} \in \mathcal{L}(V,V)$, $e = (e_1,\ldots,e_n)$ – базис V и A – матрица оператора $\mathcal{A}$ в базисе e. Характеристический многочлен $f(\lambda)$ оператора $\mathcal{A}$ является многочленом с вещественными коэффициентами, так как $f(\lambda) = \det(A-\lambda I)$,

где $A \in \mathbb{R}^{n\times n}$. Пусть α_0 – корень характеристического многочлена $f(\lambda)$.

1. Если $\lambda_0 \in \mathbb{R}$, то λ_0 – собственное значение оператора $\mathcal{A}$. Тогда линейная оболочка $\mathcal{L}(e)$, натянутая на соответствующий собственный вектор e, образует одномерное подпространство, инвариантное относительно оператора $\mathcal{A}$.

2. Если $\lambda_0 = \alpha + i\beta$, $\beta \neq 0$, то $|A - \lambda_0 I| = 0$ и система уравнений

$$Az = \lambda_0 z \qquad (93.1)$$

над полем $\mathbb{C}$ имеет нетривиальное решение $z_0 = (x_1 + iy_1, \ldots, x_n + iy_n)^T$. В обозначениях $x = (x_1,\ldots,x_n)^T \in \mathbb{R}^n$, $y = (y_1,\ldots,y_n)^T \in \mathbb{R}^n$ система (93.1) может быть записана в виде $A(x+iy) = (\alpha+i\beta)(x+iy)$ или

$$\begin{cases} Ax = \alpha x - \beta y, \\ Ay = \beta x + \alpha y\,. \end{cases} \qquad (93.2)$$

Тогда, если $u = \sum_{i=1}^n x_i e_i$, $v = \sum_{i=1}^n y_i e_i$, то системе (93.2) соответствует система векторных уравнений

$$\begin{cases} \mathcal{A}u = \alpha u - \beta v, \\ \mathcal{A}v = \beta u + \alpha v\,, \end{cases} \qquad (93.3)$$

где u, v – векторы пространства V, одновременно не равные нулю. Из (93.3) следует, что $\mathcal{L}(u,v)$ – инвариантное подпространство. Нетрудно проверить, что $\dim \mathcal{L}(u,v) = 2$.

Итак, каждый вещественный корень характеристического многочлена порождает одномерное инвариантное пространство, а каждый комплексный (не вещественный) корень – двумерное инвариантное подпространство. ■

Глава XVI. Линейные операторы в унитарных (евклидовых) пространствах

§94. Сопряженный оператор

В этом параграфе V и W – два пространства, оба унитарных или оба евклидовых.

Теорема 94.1. *Если $\mathcal{A}, \mathcal{B}$ – линейные операторы из $\mathcal{L}(V, W)$ и $(\mathcal{A}x, y) = (\mathcal{B}x, y)$, $\forall x \in V$, $y \in W$, то $\mathcal{A} = \mathcal{B}$.*

Доказательство. Утверждение вытекает из следующей цепочки простейших импликаций:

$$(\mathcal{A}x, y) = (\mathcal{B}x, y),\ \forall x \in V,\ y \in W, \Rightarrow (\mathcal{A}x - \mathcal{B}x, y) = 0,\ \forall y \in W, \Rightarrow$$
$$\Rightarrow \mathcal{A}x - \mathcal{B}x = \theta \Rightarrow \mathcal{A}x = \mathcal{B}x, \quad \forall x \in V, \Rightarrow \mathcal{A} = \mathcal{B}. \blacksquare$$

Замечание. Очевидно, что из равенства $(x, \mathcal{A}y) = (x, \mathcal{B}y)$, $\forall x \in W$, $y \in V$, также следует, что $\mathcal{A} = \mathcal{B}$.

Пусть $\mathcal{A} \in \mathcal{L}(V, W)$. Отображение $\mathcal{A}^* : W \to V$ называется *сопряженным оператором к оператору* $\mathcal{A}$, если

$$(\mathcal{A}x, y) = (x, \mathcal{A}^*y), \quad \forall x \in V,\ y \in W. \tag{94.1}$$

Теорема 94.2. *Сопряженный оператор линеен.*

Доказательство. Пусть $y_1, y_2 \in W$, тогда согласно (94.1)

$$(\mathcal{A}x, y_1 + y_2) = (x, \mathcal{A}^*(y_1 + y_2)). \tag{94.2}$$

С другой стороны, $(\mathcal{A}x, y_1 + y_2) = (\mathcal{A}x, y_1) + (\mathcal{A}x, y_2) = (x, \mathcal{A}^*y_1) + (x, \mathcal{A}^*y_2) = (x, \mathcal{A}^*y_1 + \mathcal{A}^*y_2)$. Отсюда с учетом (94.2) получим, что $(x, \mathcal{A}^*(y_1 + y_2) - (\mathcal{A}^*y_1 + \mathcal{A}^*y_2)) = 0$, $\forall x \in V$, следовательно,

$$\mathcal{A}^*(y_1 + y_2) = \mathcal{A}^*y_1 + \mathcal{A}^*y_2, \quad \forall y_1, y_2 \in W. \tag{94.3}$$

Аналогично показывается, что

$$\mathcal{A}^*(\alpha y) = \alpha\mathcal{A}^*y, \quad \forall y \in W, \forall \alpha \in \mathbb{C}(\mathbb{R}). \tag{94.4}$$

Свойства (94.3), (94.4) доказывают линейность оператора $\mathcal{A}^*$. $\blacksquare$

Теорема 94.3. *Для любого оператора $\mathcal{A} \in \mathcal{L}(V, W)$ существует, и притом единственный, сопряженный оператор.*

Доказательство. Существование. Пусть $e_1, \dots, e_n$ – ортонормированный базис V. Тогда для любого вектора $x \in V$ согласно (70.6) имеет место разложение $x = \sum_{k=1}^n (x, e_k)e_k$. Следовательно, $\mathcal{A}x = \sum_{k=1}^n (x, e_k)\mathcal{A}e_k$ и в силу (70.7)

$$(\mathcal{A}x, y) = \sum_{k=1}^n (x, e_k)(\mathcal{A}e_k, y), \quad \forall y \in W. \tag{94.5}$$

Покажем, что оператор $\mathcal{B} \in \mathcal{L}(W, V)$, определенный равенством
$$\mathcal{B}y = \sum_{k=1}^n (y, \mathcal{A}e_k)e_k, \quad \forall y \in W,$$
является сопряженным к $\mathcal{A}$. Действительно, из того, что $(x, \mathcal{B}y) = (x, \sum_{k=1}^n (y, \mathcal{A}e_k)e_k) = \sum_{k=1}^n (\mathcal{A}e_k, y)(x, e_k)$, и равенства (94.5) следует, что $(\mathcal{A}x, y) = (x, \mathcal{B}y)$, $\forall x \in V$, $y \in W$, т.е. $\mathcal{B} = \mathcal{A}^*$.

Единственность вытекает из теоремы 94.1, так как для любых двух операторов $\mathcal{B}$ и $\mathcal{C}$, сопряженных к $\mathcal{A}$, имеет место соотношение $(x, \mathcal{B}y) = (\mathcal{A}x, y) = (x, \mathcal{C}y)$, $\forall x \in V$, $y \in W$. $\blacksquare$

Теорема 94.4. *Операция сопряжения линейного оператора обладает следующими свойствами:*

1) $(\mathcal{A} + \mathcal{B})^* = \mathcal{A}^* + \mathcal{B}^*$,
2) $(\alpha\mathcal{A})^* = \bar{\alpha}\mathcal{A}^*$,
3) $(\mathcal{A}\mathcal{B})^* = \mathcal{B}\mathcal{A}$,
4) $(\mathcal{A}^*)^{-1} = (\mathcal{A}^{-1})^*$,
5) $(\mathcal{A}^*)^* = \mathcal{A}$,

выполненными для любых операторов, для которых определены указанные операции.

Доказательство. Все свойства доказываются однотипно. Докажем свойства 3 и 4. Свойство 3 вытекает из (94.1) и теоремы 94.1, так как $(x, (\mathcal{A}\mathcal{B})^*y) = (\mathcal{A}\mathcal{B}x, y) = (\mathcal{B}x, \mathcal{A}^*y) = (x, \mathcal{B}^*\mathcal{A}^*y)$, $\forall x, y$. Свойство 4 (очевидно, для невырожденных $\mathcal{A} \in \mathcal{L}(V, V)$) вытекает из свойства 3, так как если $\mathcal{A}\mathcal{A}^{-1} = \mathcal{A}^{-1}\mathcal{A} = \mathcal{I}$, то $(\mathcal{A}^{-1})^*\mathcal{A}^* = \mathcal{A}^*(\mathcal{A}^{-1})^* = \mathcal{I}$. Это означает, что оператор $(\mathcal{A}^{-1})^*$ является обратным к $\mathcal{A}^*$. $\blacksquare$

§95. Биортогональные базисы

Здесь и далее будем рассматривать линейные операторы, действующие в одном пространстве V, унитарном или евклидовом.

Две системы векторов $x_1, \dots, x_k$ и $y_1, \dots, y_k$ в унитарном (евклидовом) пространстве называются *биортогональными системами*, если $(x_i, y_j) = \delta_{ij}$, где δ_{ij} – символ Кронекера (§70).

Каждая из двух биортогональных систем векторов линейно независима. В этом можно убедиться, если приравнять линейную комбинацию одной системы нулевому вектору и последовательно умножать обе части полученного равенства скалярно на векторы другой системы.

Биортогональные системы $e_1, \dots, e_n$ и $f_1, \dots, f_n$, образующие базисы пространства V, называют *биортогональной парой базисов*.

Итак,

$$(e_i, f_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \qquad (95.1)$$

Примером биортогональной пары базисов могут служить тройки некомпланарных векторов **a**, **b**, **c** и [**b**, **c**], [**c**, **a**], [**a**, **b**] геометрического пространства V_3, для которых (**a**, **b**, **c**) = 1.

Очевидно, что ортонормированный базис биортогонален самому себе.

Теорема 95.1. *Для любого базиса $e_1, \ldots, e_n$ унитарного (евклидова) пространства существует, и притом единственный, биортогональный базис $f_1, \ldots, f_n$.*

Доказательство. Согласно (95.1) вектор f_j, $j = \overline{1,n}$, ортогонален всем векторам $e_1, \ldots, e_n$, кроме e_j. Следовательно, $f_j \in L_j$, где $L_j = \mathcal{L}^{\perp}(e_1, \ldots, e_{j-1}, e_{j+1}, \ldots, e_n)$. Очевидно, $\dim L_j = 1$. Если g – базис L_j, то $f_j = \alpha g$. Из (95.1) следует, что $(f_j, e_j) = 1$, откуда получаем, что $\alpha = 1/(g, e_j)$. Тем самым доказаны существование и единственность векторов f_j, $j = \overline{1,n}$, биортогонального базиса. ■

Теорема 95.2. *В паре биортогональных базисов e и f унитарного (евклидова) пространства V матрицы операторов $\mathcal{A}$ и $\mathcal{A}^*$ связаны соотношением*

$$(\mathcal{A}^*)_f = (A_e)^H. \qquad (95.2)$$

Доказательство. Пусть $A_e = (a_{ij})$, $(\mathcal{A}^*)_f = (b_{ij})$. Тогда $\mathcal{A}e_j = \sum_{k=1}^n a_{kj}e_k$, $\mathcal{A}^* f_i = \sum_{k=1}^n b_{ki}f_k$. Умножив первое из этих равенств скалярно на f_i, получим, что $(\mathcal{A}e_j, f_i) = \sum_{k=1}^n a_{kj}(e_k, f_i) = a_{ij}$. С другой стороны, $(\mathcal{A}e_j, f_i) = (e_j, \mathcal{A}^* f_i) = (e_j, \sum_{k=1}^n b_{ki}f_k) = \sum_{k=1}^n \overline{b_{ki}}(e_j, f_k) = \overline{b_{ji}}$. Следовательно, $a_{ij} = \overline{b_{ji}}$, $i = \overline{1,n}$, $j = \overline{1,n}$. Это равносильно (95.2). ■

Следствие 1. *В ортонормированном базисе e*

$$(\mathcal{A}^*)_e = (A_e)^H.$$

Следствие 2. *Для любого оператора $\mathcal{A} \in \mathcal{L}(V, V)$*

$$\det \mathcal{A}^* = \overline{\det \mathcal{A}}, \quad \operatorname{rg} \mathcal{A}^* = \operatorname{rg} \mathcal{A}. \qquad (95.3)$$

Теорема 95.3. *Для любого оператора $\mathcal{A} \in \mathcal{L}(V, V)$*

$$\ker \mathcal{A} = \operatorname{im}^{\perp} \mathcal{A}^*, \quad \ker \mathcal{A}^* = \operatorname{im}^{\perp} \mathcal{A}. \qquad (95.4)$$

Доказательство. Для любых векторов $x \in \ker \mathcal{A}$, $y \in \operatorname{im} \mathcal{A}^*$ имеем $\mathcal{A}x = \theta$, $y = \mathcal{A}^* y_1$. Значит, $(x, y) = (x, \mathcal{A}^* y_1) = (\mathcal{A}x, y_1) = 0$. Следовательно, $\ker \mathcal{A} \subset \operatorname{im}^{\perp} \mathcal{A}^*$. С другой стороны (см. (80.2) и (95.3)), $\dim \ker \mathcal{A} = \dim V - \dim \operatorname{im} \mathcal{A} = \dim V - \dim \operatorname{im} \mathcal{A}^* = \dim \operatorname{im}^{\perp} \mathcal{A}^*$. Теорема 64.3 о монотонности размерности доказывает первое из равенств (95.4). Второе доказывается аналогично. ■

Теорема 95.4. *Если подпространство L инвариантно относительно оператора $\mathcal{A}$, то его ортогональное дополнение $L^{\perp}$ инвариантно относительно сопряженного оператора $\mathcal{A}^*$.*

Доказательство. Если $x \in L$, $y \in L^{\perp}$, то $(\mathcal{A}x, y) = 0$, так как по условию теоремы $\mathcal{A}x \in L$. С другой стороны, $(\mathcal{A}x, y) = (x, \mathcal{A}^* y)$. Следовательно, $(x, \mathcal{A}^* y) = 0$, $\forall x \in L$ и, значит, $\mathcal{A}^* y \in L^{\perp}$. ■

§96. Нормальный оператор

Пусть V – унитарное или евклидово пространство. Линейный оператор $\mathcal{A} \in \mathcal{L}(V, V)$ называется ***нормальным оператором***, если

$$\mathcal{A}\mathcal{A}^* = \mathcal{A}^*\mathcal{A}.$$

Квадратная матрица A (комплексная или вещественная) называется ***нормальной матрицей***, если

$$AA^H = A^H A.$$

Замечание 1. Из определения и теоремы 95.2 следует, что оператор нормален тогда и только тогда, когда в любом ортонормированном базисе его матрица нормальна.

Теорема 96.1. *Собственный вектор нормального оператора, отвечающий собственному значению λ, является собственным вектором сопряженного оператора, отвечающим собственному значению $\overline{\lambda}$.*

Доказательство. Легко проверить, что если $\mathcal{A}$ – нормальный оператор, то $\mathcal{A} - \lambda\mathcal{I}$ также нормален. Пусть теперь x – собственный вектор оператора $\mathcal{A}$, отвечающий собственному значению λ, тогда $(\mathcal{A} - \lambda\mathcal{I})x = \theta$ и $((\mathcal{A} - \lambda\mathcal{I})x, (\mathcal{A} - \lambda\mathcal{I})x) = 0$. Согласно (94.1) отсюда следует, что $(x, (\mathcal{A} - \lambda\mathcal{I})^*(\mathcal{A} - \lambda\mathcal{I})x) = 0$ или, с учетом нормальности оператора $\mathcal{A} - \lambda\mathcal{I}$, что $(x, (\mathcal{A} - \lambda\mathcal{I})(\mathcal{A} - \lambda\mathcal{I})^* x) = 0$, т.е. $((\mathcal{A} - \lambda\mathcal{I})^* x, (\mathcal{A} - \lambda\mathcal{I})^* x) = 0$ и $(\mathcal{A} - \lambda\mathcal{I})^* x = \theta$. Отсюда в силу теоремы 94.4 следует, что $(\mathcal{A}^* - \overline{\lambda}\mathcal{I})x = \theta$, т.е. $\mathcal{A}^* x = \overline{\lambda}x$. ■

Следствие 1. *Если $\mathcal{A}$ – нормальный оператор, то*

$$\ker \mathcal{A} = \ker \mathcal{A}^*, \qquad (96.1)$$

так как нетривиальные векторы ядра являются собственными векторами, отвечающими нулевому собственному значению.

Следствие 2. *Если $\mathcal{A}$ – нормальный оператор, то*

$$\ker \mathcal{A} = \operatorname{im}^{\perp} \mathcal{A}, \quad \ker \mathcal{A}^* = \operatorname{im}^{\perp} \mathcal{A}^*.$$

Это следует из (95.4) и (96.1).

Теорема 96.2. *Собственные векторы нормального оператора, отвечающие различным собственным значениям, попарно ортогональны.*

Доказательство. Пусть $\mathcal{A}x = \lambda x$, $\mathcal{A}y = \mu y$, $\lambda \neq \mu$. Тогда $(\mathcal{A}x, y) = (\lambda x, y) = \lambda(x, y)$. С другой стороны, $(\mathcal{A}x, y) = (x, \mathcal{A}^* y) =$ { в силу теоремы 96.1 } $= (x, \overline{\mu} y) = \mu(x, y)$. Следовательно, $\lambda(x, y) = \mu(x, y)$ и так как $\lambda \neq \mu$, то $(x, y) = 0$. ■

Теорема 96.3 (теорема Шура). *Для любого оператора, действующего в унитарном пространстве, существует ортонормированный базис, в котором он имеет треугольную матрицу.*

Доказательство этой теоремы отличается от доказательства теоремы 89.2 только тем, что на каждом шаге строится ортонормированный базис инвариантного подпространства. ■

Ортонормированный базис унитарного (евклидова) пространства, в котором матрица линейного оператора имеет треугольную форму, называется *базисом Шура* для этого оператора.

Теорема 96.4 (критерий нормальности). *Оператор, действующий в унитарном пространстве, нормален тогда и только тогда, когда существует ортонормированный базис из собственных векторов этого оператора.*

Доказательство. Необходимость. Пусть $\mathcal{A}$ – нормальный оператор и e – его базис Шура. Тогда

$$A_e = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ 0 & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & a_{nn} \end{bmatrix}, \quad (A_e)^H = \begin{bmatrix} \overline{a_{11}} & 0 & \ldots & 0 \\ \overline{a_{12}} & \overline{a_{22}} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ \overline{a_{1n}} & \overline{a_{2n}} & \ldots & \overline{a_{nn}} \end{bmatrix}$$

и, согласно (95.2) и замечанию 1, $A_e(A_e)^H = (A_e)^H A_e$. Сравнив диагональные элементы матриц, стоящих в левой и правой частях последнего соотношения, получим равенства

$$|a_{11}|^2 + |a_{12}|^2 + \ldots + |a_{1n}|^2 = |a_{11}|^2, \ |a_{22}|^2 + \ldots + |a_{2n}|^2 =$$
$$= |a_{22}|^2, \ldots, |a_{n-1,n-1}|^2 + |a_{n-1,n}|^2 = |a_{n-1,n-1}|^2,$$

из которых следует, что

$$a_{12} = a_{13} = \ldots = a_{1n} = 0, \ a_{23} = a_{24} = \ldots = a_{2n} = 0, \ldots, a_{n-1,n} = 0.$$

Следовательно, матрица A_e имеет диагональную форму. Таким образом, базис Шура является ортонормированным базисом из собственных векторов оператора $\mathcal{A}$.

Достаточность. Пусть e – ортонормированный базис из собственных векторов оператора $\mathcal{A}$, тогда

$$A_e = \begin{bmatrix} \lambda_1 & & O \\ & \ddots & \\ O & & \lambda_n \end{bmatrix}, \quad (A_e)^H = \begin{bmatrix} \overline{\lambda_1} & & O \\ & \ddots & \\ O & & \overline{\lambda_n} \end{bmatrix}.$$

Из перестановочности диагональных матриц следует, что A_e – нормальная матрица. Это означает (замечание 1), что $\mathcal{A}$ – нормальный оператор. ■

Следствие 3. *В унитарном пространстве нормальный оператор $\mathcal{A}$ и его сопряженный $\mathcal{A}^*$ имеют общий ортонормированный базис из собственных векторов.*

Теорема 96.5. *Если любой собственный вектор оператора $\mathcal{A}$, действующего в унитарном пространстве V, является собственным вектором сопряженного оператора $\mathcal{A}^*$, то $\mathcal{A}$ – нормальный оператор.*

Доказательство. Любой оператор, действующий в комплексном пространстве, обладает хотя бы одним собственным вектором (теорема 86.6). Пусть $\dim V = n$ и e_1 – собственный вектор оператора $\mathcal{A}$, тогда e_1 – собственный вектор оператора $\mathcal{A}^*$ и согласно теореме 95.4 подпространство $L_{n-1} = \mathcal{L}^{\perp}(e_1)$ инвариантно относительно оператора $\mathcal{A}$. В этом подпространстве (в силу теоремы 86.6) существует собственный вектор $e_2 \in L_{n-1}$ оператора $\mathcal{A}$, при этом $e_2 \perp e_1$. Вектор e_2 также является собственным вектором $\mathcal{A}^*$, и по-прежнему $\mathcal{L}(e_2)$ инвариантно относительно $\mathcal{A}^*$ и, следовательно, $L_{n-2} = \mathcal{L}^{\perp}(e_2)$ (имеется в виду ортогональное дополнение $\mathcal{L}(e_2)$ до L_{n-1}) инвариантно относительно оператора $\mathcal{A}$. Поступая аналогично, построим ортогональную систему собственных векторов $e_1, \ldots, e_n$ оператора $\mathcal{A}$. В таком случае векторы $e_1/|e_1|, \ldots, e_n/|e_n|$ образуют ортонормированный базис пространства V, состоящий из собственных векторов оператора $\mathcal{A}$, и в силу теоремы 96.4 $\mathcal{A}$ – нормальный оператор. ■

Замечание 2. Отметим, что доказанная теорема является обратной к теореме 96.1 для случая унитарного пространства.

Унитарно подобные матрицы. Подобные комплексные (вещественные) матрицы A и $B = Q^{-1}AQ$ называются *унитарно* (соответственно *ортогонально*) *подобными*, если матрица преобразования подобия Q унитарна (соответственно ортогональна), т.е. если $Q^H Q = QQ^H = I$ (соответственно $Q^T Q = QQ^T = I$). Очевидно, что соотношение подобия для унитарно подобных матриц A и B имеет вид $B = Q^H AQ$, а для ортогонально подобных – $B = Q^T AQ$.

Замечание 3. Из теоремы 70.5 и свойства 8° §82 следует, что две комплексные (вещественные) квадратные матрицы одинакового порядка унитарно (соответственно ортогонально) подобны тогда и только тогда, когда они являются матрицами одного и того же оператора в унитарном (соответственно евклидовом) пространстве в ортонормированных базисах.

Теорема 96.6. *Квадратная комплексная матрица является нормальной тогда и только тогда, когда она унитарно подобна диагональной матрице.*

Доказательство может быть проведено по известной схеме (88.6), если использовать ортонормированные базисы e, f и учесть замечание 3. ■

Доказанная теорема представляет собой матричную формулировку теоремы 96.4.

§97. Унитарный (ортогональный) оператор

Линейный оператор $\mathcal{U}$, действующий в унитарном (евклидовом) пространстве, называется *унитарным* (соответственно *ортогональным*) оператором, если $\mathcal{U}^*\mathcal{U} = \mathcal{U}\mathcal{U}^* = \mathcal{I}$.

Из определения вытекают следующие факты.

1°. Оператор $\mathcal{U}$ унитарен (ортогонален) только тогда, когда в любом ортонормированном базисе он имеет унитарную (соответственно ортогональную) матрицу.

2°. Для унитарного (ортогонального) оператора $\mathcal{U}$

$$|\det\mathcal{U}| = 1. \tag{97.1}$$

3°. Унитарный (ортогональный) оператор нормален.

Теорема 97.1 (критерии унитарности). *В унитарном (евклидовом) пространстве V следующие утверждения равносильны:*

1) оператор $\mathcal{U}$ унитарен (ортогонален);

2) $\mathcal{U}^*\mathcal{U} = \mathcal{I}$;

3) $\mathcal{U}\mathcal{U}^* = \mathcal{I}$;

4) оператор $\mathcal{U}$ сохраняет скалярное произведение, т.е.

$$(\mathcal{U}x,\mathcal{U}y) = (x,y), \quad \forall x,y \in V;$$

5) оператор $\mathcal{U}$ сохраняет длину, т.е.

$$|\mathcal{U}x| = |x|, \quad \forall x \in V;$$

6) оператор $\mathcal{U}$ переводит любой ортонормированный базис V в ортонормированный базис;

7) оператор $\mathcal{U}$ переводит хотя бы один ортонормированный базис V в ортонормированный базис.

Доказательство. $1 \Leftrightarrow 2 \Leftrightarrow 3$. Любое из равенств $\mathcal{U}^*\mathcal{U} = \mathcal{I}$ или $\mathcal{U}\mathcal{U}^* = \mathcal{I}$ означает невырожденность $\mathcal{U}$ и существование $\mathcal{U}^{-1}$. Умножение обеих частей этих равенств на $\mathcal{U}^{-1}$ (справа или слева) приводит к тому, что $\mathcal{U}^* = \mathcal{U}^{-1}$. Следовательно, имеют место импликации $2 \Rightarrow 1$, $3 \Rightarrow 1$. Переход $1 \Rightarrow 2$, $1 \Rightarrow 3$ очевиден.

$1 \Rightarrow 4$. Так как $\mathcal{U}^*\mathcal{U} = \mathcal{I}$, то

$$(\mathcal{U}x,\mathcal{U}y) = (x,\mathcal{U}^*\mathcal{U}y) = (x,y), \quad \forall x,y \in V.$$

$4 \Rightarrow 1$. Так как $(x,\mathcal{U}^*\mathcal{U}y) = (\mathcal{U}x,\mathcal{U}y) = (x,y)$, $\forall x,y \in V$, то на основании теоремы 94.1 $\mathcal{U}^*\mathcal{U} = \mathcal{I}$.

$4 \Rightarrow 5$. Очевидно, так как $|\mathcal{U}x| = \sqrt{(\mathcal{U}x,\mathcal{U}x)} = \sqrt{(x,x)} = |x|$, $\forall x \in V$.

$5 \Rightarrow 4$. Это следует из легко проверяемых соотношений:

$$(x,y) = (|x+y|^2 - |x|^2 - |y|^2)/2$$

в евклидовом пространстве и

$$(x,y) = (|x+y|^2 - |x-y|^2 + i|x+iy|^2 - i|x-iy|^2)/4$$

в унитарном пространстве.

$4 \Rightarrow 6$. Очевидно, так как $(\mathcal{U}e_i,\mathcal{U}e_j) = (e_i,e_j) = \delta_{ij}$.

$6 \Rightarrow 4$. Если $e_1,\dots,e_n$ – ортонормированный базис V и $x = \sum_{i=1}^n x_ie_i$, $y = \sum_{i=1}^n y_ie_i$, то $\mathcal{U}x = \sum_{i=1}^n x_i\mathcal{U}e_i$, $\mathcal{U}y = \sum_{i=1}^n y_i\mathcal{U}e_i$, и так как $\mathcal{U}e_1,\dots,\mathcal{U}e_n$ – ортонормированный базис, то, согласно (70.7), $(\mathcal{U}x,\mathcal{U}y) = \sum_{i=1}^n x_i\overline{y_i} = (x,y)$, $\forall x,y \in V$.

$6 \Rightarrow 7$. Очевидно.

$7 \Rightarrow 6$. Этот переход фактически доказан в п. $6 \Rightarrow 4$. ∎

Следствие. *Унитарный (ортогональный) оператор на любом инвариантном подпространстве индуцирует унитарный (соответственно ортогональный) оператор,* так как сохраняет скалярное произведение любой пары векторов этого подпространства.

Теорема 97.2 (спектральная характеристика унитарного оператора). *Нормальный оператор в унитарном пространстве унитарен тогда и только тогда, когда все его собственные значения по модулю равны единице.*

Доказательство. Необходимость справедлива в унитарном и евклидовом пространстве и означает, что все собственные значения унитарного (и ортогонального) оператора $\mathcal{U}$ по модулю равны единице. Докажем это. Пусть x – собственный вектор оператора $\mathcal{U}$ и λ – отвечающее ему собственное значение. Тогда $(\mathcal{U}x,\mathcal{U}x) = (\lambda x,\lambda x) = |\lambda|^2(x,x)$. С другой стороны, $(\mathcal{U}x,\mathcal{U}x) = (x,\mathcal{U}^*\mathcal{U}x) = (x,x)$. Из этих равенств следует, что $|\lambda| = 1$.

Достаточность. Если $\mathcal{U}$ – нормальный оператор, то согласно теореме 96.4 в пространстве V существует ортонормированный базис $e_1,\dots,e_n$ из собственных векторов оператора $\mathcal{U}$. Тогда для любого вектора $x = \sum_{i=1}^n x_ie_i \in V$ имеем $\mathcal{U}x = \sum_{i=1}^n x_i\lambda_ie_i$, где $|\lambda_i| = 1$, $i = \overline{1,n}$. В силу ортонормированности базиса e согласно (70.7) получаем отсюда, что $(x,x) = \sum_{i=1}^n |x_i|^2$ и $(\mathcal{U}x,\mathcal{U}x) = \sum_{i=1}^n |\lambda_i|^2|x_i|^2 = \sum_{i=1}^n |x_i|^2$. Следовательно, $|\mathcal{U}x| = |x|$, $\forall x \in V$, и $\mathcal{U}$ – унитарный оператор (теорема 97.1). ∎

Теорема 97.3. *Если подпространство L инвариантно относительно унитарного (ортогонального) оператора $\mathcal{U}$, то его ортогональное дополнение $L^\perp$ также инвариантно относительно $\mathcal{U}$.*

Доказательство. Пусть $y \in L^\perp$. Покажем, что $\mathcal{U}y \in L^\perp$, т.е. $(x,\mathcal{U}y) = 0$, $\forall x \in L$. Оператор $\mathcal{U}$ индуцирует на подпространстве L унитарный (ортогональный) оператор $\mathcal{U}|L$. Значит, оператор $\mathcal{U}|L$ обратим и его образ совпадает со всем подпространством L, т.е. $\mathrm{im}\,\mathcal{U}|L = L$ (теорема 83.5). Это означает, что для любого вектора $x \in L$ существует вектор $x_1 \in L$ такой, что $x = \mathcal{U}x_1$. Тогда $(x,\mathcal{U}y) = (\mathcal{U}x_1,\mathcal{U}y) = (x_1,y) = 0$, так как $x_1 \in L$, $y \in L^\perp$. ∎

Каноническая форма матрицы унитарного оператора. Из того что унитарный оператор нормален и все его собственные значения по модулю равны единице, следует, что *в пространстве V существует ортонормированный базис e, в котором матрица унитарного оператора $\mathcal{U}$ имеет диагональную форму*

$$U_e = \begin{bmatrix} \lambda_1 & & O \\ & \ddots & \\ O & & \lambda_n \end{bmatrix}, \quad \text{где } |\lambda_i| = 1,\ i = \overline{1,n},$$

или, в матричной формулировке: ***квадратная комплексная матрица унитарна тогда и только тогда, когда она унитарно подобна диагональной матрице, у которой все диагональные элементы по модулю равны единице.***

Каноническая форма матрицы ортогонального оператора. Пусть Q – ортогональный оператор, действующий в евклидовом пространстве E. Согласно теореме 97.2 (п. "необходимость") и равенству (97.1) для его собственных значений λ и определителя (так как $\lambda \in \mathbb{R}$, $\det Q \in \mathbb{R}$) имеем

$$\lambda = \pm 1, \quad \det Q = \pm 1 . \tag{97.2}$$

В любом ортонормированном базисе e оператор Q имеет ортогональную матрицу Q_e, следовательно,

$$Q_e^{-1} = Q_e^T . \tag{97.3}$$

Примеры. 1. В одномерном пространстве матрица Q_e ортогонального оператора Q имеет вид

$$Q_e = [\pm 1] . \tag{97.4}$$

Иными словами, либо $Q_e = I$, либо $Q_e = -I$.

2. В двумерном пространстве матрица Q_e ортогонального оператора Q имеет вид $Q_e = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$, причем $\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}^{-1} = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$ в силу (97.3), и, с учетом (97.2), (5.3),

$$\pm \begin{bmatrix} \delta & -\beta \\ -\gamma & \alpha \end{bmatrix} = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} . \tag{97.5}$$

а) Если $\det Q = 1$, то (97.5) означает, что $\begin{bmatrix} \delta & -\beta \\ -\gamma & \alpha \end{bmatrix} = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$, откуда следует, что $Q_e = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$, $\alpha^2+\beta^2 = 1$. Положив $\alpha = \cos\varphi$, $\beta = -\sin\varphi$, получим

$$Q_e = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} . \tag{97.6}$$

б) Если $\det Q = -1$, то (97.5) означает, что $\begin{bmatrix} -\delta & \beta \\ \gamma & -\alpha \end{bmatrix} = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$, откуда следует, что $Q_e = \begin{bmatrix} \alpha & \beta \\ \beta & -\alpha \end{bmatrix}$, $\alpha^2 + \beta^2 = 1$; при этом $f(\lambda) = \det(Q_e - \lambda I) = \lambda^2 - 1$. Следовательно, оператор Q в двумерном пространстве имеет два различных собственных значения: $\lambda_1 = 1$,

$\lambda_2 = -1$, и для него существует ортонормированный базис f, в котором его матрица Q_f имеет вид

$$Q_f = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} . \tag{97.7}$$

Теорема 97.4. *Для любого ортогонального оператора Q в евклидовом пространстве существует ортонормированный базис e, в котором его матрица имеет квазидиагональную форму с клетками вида (97.4) и (97.6) на главной диагонали.*

Доказательство. Воспользуемся индукцией по размерности n пространства V. Для $n = 1, 2$ утверждение вытекает из (97.4), (97.6), (97.7). Пусть $n \geq 3$ и теорема верна для ортогональных операторов в пространствах размерности $k \leq (n-1)$. Докажем теорему для размерности n. Согласно теореме 93.2 для оператора Q существует одномерное либо двумерное инвариантное подпространство L. Если $\dim L = 1$, то $Q|L$ – ортогональный оператор (согласно следствию) в одномерном пространстве и в ортонормированном базисе e_1 пространства L его матрица имеет вид (97.4). Если же $\dim L = 2$, то $Q|L$ – ортогональный оператор в двумерном пространстве и в ортонормированном базисе e_1, e_2 пространства L его матрица имеет вид (97.6) или (97.7). Согласно теореме 97.3 ортогональное дополнение $L^\perp$ инвариантно относительно оператора Q. По индуктивному предположению в пространстве $L^\perp$ существует ортонормированный базис $e_2, \ldots, e_n$ (или $e_3, \ldots, e_n$), в котором матрица оператора $Q|L^\perp$ имеет требуемый вид. Тогда в базисе $e_1, \ldots, e_n$ матрица оператора Q также будет иметь требуемый вид. ■

После очевидной перестановки векторов построенного базиса матрица ортогонального оператора Q будет иметь вид

$$Q_e = \begin{bmatrix} 1 & & & & & & & & O \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & & & & & \boxed{\begin{matrix} \cos\varphi_1 & -\sin\varphi_1 \\ \sin\varphi_1 & \cos\varphi_1 \end{matrix}} & & \\ & & & & & & & \ddots & \\ O & & & & & & & & \boxed{\begin{matrix} \cos\varphi_k & -\sin\varphi_k \\ \sin\varphi_k & \cos\varphi_k \end{matrix}} \end{bmatrix} . \tag{97.8}$$

Матрица (97.8) называется ***канонической формой матрицы ортогонального оператора.***

Простым вращением называется оператор в евклидовом пространстве, который в некотором ортонормированном базисе имеет матрицу вида

$$\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & O & \\ & & 1 & & & & \\ & & & \boxed{\begin{matrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{matrix}} & & & \\ & & & & 1 & & \\ & O & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}. \qquad (97.9)$$

Простым отражением называется оператор в евклидовом пространстве, который в некотором ортонормированном базисе имеет матрицу вида

$$\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & O & & \\ & & 1 & & & & \\ & & & -1 & & & \\ & & & & 1 & & \\ & O & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}. \qquad (97.10)$$

Из определения следует, что простое вращение и простое отражение – ортогональные операторы, так как их матрицы (97.9) и (97.10) в ортонормированном базисе ортогональны. Простое вращение представляет собой поворот в некоторой двумерной плоскости и оставляет неизменным $(n-2)$-мерное подпространство, ортогональное этой плоскости. Простое отражение меняет направление всех векторов некоторого одномерного подпространства и оставляет неизменным его $(n-1)$-мерное ортогональное дополнение.

Теорема 97.5. *Всякий ортогональный оператор может быть представлен как произведение некоторого числа простых вращений и простых отражений.*

Это утверждение следует из того, что матрица вида (97.8) может быть представлена в виде произведения некоторого числа матриц вида (97.9), (97.10). ■

§98. Самосопряженный оператор

Линейный оператор $\mathcal{A}$, действующий в унитарном (евклидовом) пространстве, называется *самосопряженным*, если $\mathcal{A} = \mathcal{A}^*$. Самосопряженный оператор в унитарном пространстве называют *эрмитовым*, а в евклидовом пространстве – *симметрическим*.

Квадратная матрица A (комплексная или вещественная) называется *самосопряженной*, если $A = A^H$. Комплексную самосопряженную матрицу называют ***эрмитовой***, а вещественную – ***симметрической*** или ***вещественно-эрмитовой*** (очевидно, для симметрической матрицы: $A = A^T$).

Из определения вытекают следующие факты.

1°. Самосопряженный оператор нормален.

2°. Оператор самосопряжен тогда и только тогда, когда в любом ортонормированном базисе он имеет самосопряженную матрицу.

3°. Определитель самосопряженного оператора вещественен.

4°. Если подпространство L инвариантно относительно самосопряженного оператора $\mathcal{A}$, то $L^\perp$ также инвариантно относительно $\mathcal{A}$ (теорема 95.4).

5°. Самосопряженный оператор на любом инвариантном подпространстве индуцирует самосопряженный оператор.

Теорема 98.1 (спектральная характеристика самосопряженного оператора). *Нормальный оператор в унитарном (евклидовом) пространстве самосопряжен тогда и только тогда, когда все корни его характеристического многочлена вещественны.*.

Доказательство. Необходимость. В унитарном пространстве это утверждение означает, что все собственные значения эрмитова оператора вещественны, и вытекает из равенств $\mathcal{A}x = \lambda x$ и, с учетом теоремы 96.1, $\mathcal{A}x = \overline{\lambda}x$. Докажем утверждение для евклидова пространства E. Пусть e – ортонормированный базис E, тогда A_e – самосопряженная (вещественная) матрица. Рассмотрим произвольное унитарное пространство U той же размерности, что и пространство E, и в нем произвольный ортонормированный базис f. Тогда матрице A_e отвечает самосопряженный оператор $\mathcal{B} \in \mathcal{L}(U,U)$, для которого матрица A_e является матрицей в базисе f: $A_e = B_f$. Следовательно, характеристические многочлены операторов $\mathcal{A}$ и $\mathcal{B}$ совпадают и по доказанному выше (применительно к оператору $\mathcal{B}$) все корни характеристического многочлена оператора $\mathcal{A}$ вещественны.

Достаточность. Пусть $\mathcal{A}$ – нормальный оператор и все корни его характеристического многочлена вещественны. Тогда как в евклидовом, так и в унитарном пространстве существует ортонормированный базис $e_1, \dots, e_n$ из собственных векторов оператора $\mathcal{A}$. Если $x = \sum_{i=1}^n x_i e_i$ – любой вектор пространства, то $\mathcal{A}x = \sum_{i=1}^n x_i \lambda_i e_i$ и $\mathcal{A}^*x = \sum_{i=1}^n x_i \overline{\lambda_i} e_i = \sum_{i=1}^n x_i \lambda_i e_i$, так как $\lambda_i \in \mathbb{R}$. Следовательно, $\mathcal{A}x = \mathcal{A}^*x$, $\forall x \in V$, откуда следует, что $\mathcal{A} = \mathcal{A}^*$. ■

Итак, для самосопряженных операторов основное поле ($\mathbb{R}$ или $\mathbb{C}$) не играет такой "роковой" роли, как это было с другими операторами; все факты, касающиеся комплексных пространств, справедливы и в вещественном пространстве.

В частности, как следует из теоремы 98.1, *оператор, действующий в унитарном (или евклидовом) пространстве, самосопряжен тогда и только тогда, когда существует ортонормированный базис, в котором его матрица имеет вещественную диагональную* *форму*, или, в матричной формулировке: ***квадратная матрица (комплексная или вещественная) является самосопряженной тогда и только тогда, когда она унитарно подобна вещественной диагональной матрице.***

§99. Знакоопределенные операторы

Теорема 99.1. *Если в унитарном пространстве V для любого вектора x имеет место равенство $(x,\mathcal{B}x)=0$, то $\mathcal{B}=\mathcal{O}$.*

Доказательство. Для любых векторов $y,z\in V$ имеем $(\mathcal{B}(y+z),y+z)=0$ и $(\mathcal{B}(iy+z),iy+z)=0$, т.е. $(\mathcal{B}y,y)+(\mathcal{B}z,y)+(\mathcal{B}y,z)+(\mathcal{B}z,z)=0$, $(\mathcal{B}y,y)-i(\mathcal{B}z,y)+i(\mathcal{B}y,z)+(\mathcal{B}z,z)=0$ или, с учетом условия леммы, $(\mathcal{B}z,y)+(\mathcal{B}y,z)=0$, $-i(\mathcal{B}z,y)+i(\mathcal{B}y,z)=0$. Прибавив к первому равенству второе, умноженное на i, получим, что $(\mathcal{B}y,z)=0$, $\forall y,z\in V$. Следовательно, $\mathcal{B}=\mathcal{O}$ (теорема 94.1). ■

Теорема 99.2. *Линейный оператор $\mathcal{A}$ в унитарном пространстве V эрмитов тогда и только тогда, когда*

$$(\mathcal{A}x,x)\in\mathbb{R},\quad \forall x\in V. \tag{99.1}$$

Доказательство. Необходимость. Если $\mathcal{A}$ – эрмитов оператор, то $(\mathcal{A}x,x)=(x,\mathcal{A}^*x)=(x,\mathcal{A}x)$, $\forall x\in V$; следовательно,

$$(\mathcal{A}x,x)=\overline{(\mathcal{A}x,x)} \text{ и } (\mathcal{A}x,x)\in\mathbb{R},\ \forall x\in V.$$

Достаточность. Пусть $(\mathcal{A}x,x)\in\mathbb{R}$. Тогда $(\mathcal{A}x,x)=(x,\mathcal{A}x)$ и $(\mathcal{A}x,x)=(x,\mathcal{A}^*x)$. Отсюда следует, что

$$(x,(\mathcal{A}-\mathcal{A}^*)x)=0,\quad \forall x\in V. \tag{99.2}$$

Отсюда на основании теоремы 99.1 следует, что $\mathcal{A}=\mathcal{A}^*$. ■

Замечание. Условие (99.1) формально верно и в евклидовом пространстве, однако лишено смысла, так как справедливо для любых линейных операторов $\mathcal{A}$.

Теорема 99.2 позволяет для самосопряженных операторов $\mathcal{A}$ говорить о знаке числа $(\mathcal{A}x,x)$. Самосопряженный оператор в унитарном (евклидовом) пространстве называется *положительно определенным*, если

$$(\mathcal{A}x,x)>0,\quad \forall x\neq\theta,$$

неотрицательно определенным (*отрицательно определенным* или *неположительно определенным*), если $(\mathcal{A}x,x)\geq 0$, $\forall x\neq\theta$ (соответственно $(\mathcal{A}x,x)<0$ или $(\mathcal{A}x,x)\leq 0$). Обозначение: $\mathcal{A}>\mathcal{O}$, $\mathcal{A}\geq\mathcal{O}$, $\mathcal{A}<\mathcal{O}$, $\mathcal{A}\leq\mathcal{O}$ соответственно.

Теорема 99.3. *Самосопряженный оператор $\mathcal{A}$ в унитарном (и евклидовом) пространстве положительно определен (соответственно $\mathcal{A}\geq\mathcal{O}$, $\mathcal{A}<\mathcal{O}$, $\mathcal{A}\leq\mathcal{O}$) тогда и только тогда, когда все его собственные значения $\lambda>0$ ($\lambda\geq 0$, $\lambda<0$, $\lambda\leq 0$).*

Доказательство. Докажем вариант теоремы, относящийся к $\mathcal{A}>\mathcal{O}$. Необходимость. Если $\mathcal{A}>\mathcal{O}$, то $(\mathcal{A}x,x)>0$ для любого вектора $x\neq\theta$. В частности, это верно и для собственного вектора x, поэтому $(\mathcal{A}x,x)=(\lambda x,x)=\lambda(x,x)>0$, где $(x,x)>0$. Значит, $\lambda>0$.

Достаточность. Если $\mathcal{A}$ – самосопряженный оператор, то существует ортонормированный базис из собственных векторов $e_1,\ldots,e_n$ оператора $\mathcal{A}$. При этом соответствующие собственные значения $\lambda_i>0$, $i=\overline{1,n}$. Тогда для любого вектора $x=\sum_{i=1}^n x_ie_i\neq\theta$ имеем

$$(\mathcal{A}x,x)=\left(\sum_{i=1}^n x_i\lambda_ie_i,\sum_{i=1}^n x_ie_i\right)=\sum_{i=1}^n\lambda_i|x_i|^2>0.\ ■$$

Следствие. Если $\mathcal{A}>\mathcal{O}$ (или $\mathcal{A}<\mathcal{O}$), то $\mathcal{A}$ обратим, так как $\det\mathcal{A}=\lambda_1\cdot\ldots\cdot\lambda_n$.

Теорема 99.4. *Оператор, обратный к положительно (отрицательно) определенному оператору, положительно (соответственно отрицательно) определен.*

Доказательство. Если $\mathcal{A}$ – самосопряженный оператор, то $\mathcal{A}^{-1}$ – тоже самосопряженный оператор, так как согласно теореме 94.4 $(\mathcal{A}^{-1})^*=(\mathcal{A}^*)^{-1}=\mathcal{A}^{-1}$. Если $\mathcal{A}>\mathcal{O}$ ($\mathcal{A}<\mathcal{O}$), то все собственные значения оператора $\mathcal{A}^{-1}$ положительны (соответственно отрицательны), так как они обратны собственным значениям оператора $\mathcal{A}$. Из теоремы 99.3 следует, что $\mathcal{A}^{-1}>\mathcal{O}$ ($\mathcal{A}^{-1}<\mathcal{O}$). ■

Теорема 99.5. *Для любого неотрицательно (положительно) определенного оператора $\mathcal{A}$ существует, и притом единственный, неотрицательно (соответственно положительно) определенный оператор $\mathcal{B}$ такой, что $\mathcal{B}^2=\mathcal{A}$.*

Доказательство. Существование. Построим оператор $\mathcal{B}$. Пусть $e_1,\ldots,e_n$ – ортонормированный базис пространства из собственных векторов оператора $\mathcal{A}$ и $\mathcal{A}e_i=\lambda_ie_i$, $i=\overline{1,n}$. По условию теоремы $\lambda_i\geq 0$ ($\lambda_i>0$), $i=\overline{1,n}$. Положим

$$\mathcal{B}e_i=\sqrt{\lambda_i}e_i,\quad i=\overline{1,n}. \tag{99.3}$$

Из (99.3) следует, что оператор $\mathcal{B}\geq\mathcal{O}$ ($\mathcal{B}>\mathcal{O}$), так как $\mathcal{B}$ – нормальный оператор (ибо существует ортонормированный базис $e_1,\ldots,e_n$ из собственных векторов $\mathcal{B}$), при этом он самосопряжен (ибо его собственные значения $\sqrt{\lambda_i}\in\mathbb{R}$) и, кроме того, $\sqrt{\lambda_i}\geq 0$ ($\sqrt{\lambda_i}>0$). Оператор $\mathcal{B}$ – искомый, так как в силу (99.3) $\mathcal{B}^2e_i=\lambda_ie_i=\mathcal{A}e_i$, $i=\overline{1,n}$.

Единственность. Пусть существует другой оператор $\mathcal{C}\geq\mathcal{O}$ ($\mathcal{C}>\mathcal{O}$) такой, что $\mathcal{C}^2=\mathcal{A}$. Тогда существует ортонормированный базис $f_1,\ldots,f_n$ пространства из собственных векторов оператора $\mathcal{C}$. Если $\mathcal{C}f_i=\mu_if_i$, $i=\overline{1,n}$, то $\mathcal{A}f_i=\mathcal{C}^2f_i=\mu_i^2f_i$, $i=\overline{1,n}$. Значит, числа $\mu_1^2,\ldots,\mu_n^2$ являются собственными значениями оператора $\mathcal{A}$ и, следовательно, совпадают с числами $\lambda_1,\ldots,\lambda_n$. Покажем, что

$$\mathcal{C}e_i=\sqrt{\lambda_i}e_i,\quad i=\overline{1,n}, \tag{99.4}$$

тем самым в силу (99.3) будет доказано, что $\mathcal{C} = \mathcal{B}$.

Разложим вектор e_i по базису f:

$$e_i = \sum_{k=1}^{n} \alpha_k f_k . \tag{99.5}$$

Отметим, что в этом равенстве участвуют собственные векторы e_i, $f_1, \dots, f_n$ оператора $\mathcal{A}$, отвечающие собственным значениям λ_i, $\mu_1^2, \dots, \mu_n^2$. Из линейной независимости собственных векторов, отвечающих различным собственным значениям, следует, что в разложении (99.5) отличными от нуля будут коэффициенты α_k лишь при тех f_k, которые отвечают собственному значению $\mu_k^2 = \lambda_i$. Поэтому $\mathcal{C}e_i = \sum_{k=1}^{n} \alpha_k \mathcal{C} f_k = \sum_{k=1}^{n} \alpha_k \sqrt{\lambda_i} f_k = \sqrt{\lambda_i}\sum_{k=1}^{n} \alpha_k f_k = \sqrt{\lambda_i} e_i$, т.е. (99.4). ■

Оператор $\mathcal{B}$ называется *квадратным корнем из оператора* $\mathcal{A}$.

§100. Разложения линейного оператора

Линейный оператор $\mathcal{A} \in \mathcal{L}(V, V)$ в унитарном (евклидовом) пространстве V называется *косоэрмитовым* (соответственно *кососимметрическим*), если $\mathcal{A}^* = -\mathcal{A}$.

Квадратная комплексная матрица называется *косоэрмитовой*, если $A^H = -A$. Квадратная вещественная матрица называется *кососимметрической*, если $A^T = -A$.

Из определения следует, что оператор $\mathcal{A}$ косоэрмитов (кососимметричен) тогда и только тогда, когда его матрица в любом ортонормированном базисе пространства косоэрмитова (соответственно кососимметрична).

Теорема 100.1. *Линейный оператор $\mathcal{A}$ в унитарном пространстве эрмитов тогда и только тогда, когда оператор $i\mathcal{A}$ косоэрмитов.*

Доказательство. Заметим, что $(i\mathcal{A})^* = -i\mathcal{A}^*$ (теорема 94.4). Следовательно, $(i\mathcal{A})^* = -i\mathcal{A}$ тогда и только тогда, когда $\mathcal{A} = \mathcal{A}^*$. ■

Из теоремы следует, что все свойства самосопряженных операторов переносятся на косоэрмитовы (кососимметрические) операторы с той лишь разницей, что в этом случае все корни характеристического многочлена чисто мнимые.

Теорема 100.2. *Линейный оператор $\mathcal{A}$ в унитарном (евклидовом) пространстве может быть представлен, и притом единственным образом, в виде суммы*

$$\mathcal{A} = \mathcal{B} + \mathcal{C} \tag{100.1}$$

эрмитова (симметрического) оператора $\mathcal{B}$ и косоэрмитова (кососимметрического) оператора $\mathcal{C}$.

Доказательство. Положим

$$\mathcal{B} = \frac{1}{2}(\mathcal{A} + \mathcal{A}^*), \quad \mathcal{C} = \frac{1}{2}(\mathcal{A} - \mathcal{A}^*). \tag{100.2}$$

Тогда, как легко проверить, $\mathcal{B}^* = \mathcal{B}$, $\mathcal{C}^* = -\mathcal{C}$ и $\mathcal{A} = \mathcal{B} + \mathcal{C}$. Единственность такой пары операторов следует из того, что для любой другой пары операторов $\mathcal{B}_1$ и $\mathcal{C}_1$ таких, что $\mathcal{B}_1^* = \mathcal{B}_1$, $\mathcal{C}_1^* = -\mathcal{C}_1$ и $\mathcal{A} = \mathcal{B}_1 + \mathcal{C}_1$, имеем $\mathcal{A}^* = \mathcal{B}_1 - \mathcal{C}_1$ или $\frac{1}{2}(\mathcal{A} + \mathcal{A}^*) = \mathcal{B}_1$, $\frac{1}{2}(\mathcal{A} - \mathcal{A}^*) = \mathcal{C}_1$, т.е. в силу (100.2) $\mathcal{B}_1 = \mathcal{B}$, $\mathcal{C}_1 = \mathcal{C}$. ■

Разложение (100.1) называется *эрмитовым разложением* оператора $\mathcal{A}$.

Теорема 100.3. *Линейный оператор $\mathcal{A} \in \mathcal{L}(V, V)$ в унитарном (евклидовом) пространстве нормален тогда и только тогда, когда операторы $\mathcal{B}$ и $\mathcal{C}$ в эрмитовом разложении (100.1) этого оператора перестановочны.*

Доказательство. Если $\mathcal{A} = \mathcal{B} + \mathcal{C}$, то $\mathcal{A}^* = \mathcal{B} - \mathcal{C}$ и $\mathcal{A}\mathcal{A}^* = \mathcal{B}^2 - \mathcal{B}\mathcal{C} + \mathcal{C}\mathcal{B} - \mathcal{C}^2$, $\mathcal{A}^*\mathcal{A} = \mathcal{B}^2 - \mathcal{C}\mathcal{B} + \mathcal{B}\mathcal{C} - \mathcal{C}^2$, т.е. $\mathcal{A}\mathcal{A}^* - \mathcal{A}^*\mathcal{A} = \frac{1}{2}(\mathcal{C}\mathcal{B} - \mathcal{B}\mathcal{C})$. Значит, $\mathcal{A}\mathcal{A}^* = \mathcal{A}^*\mathcal{A}$ тогда и только тогда, когда $\mathcal{C}\mathcal{B} = \mathcal{B}\mathcal{C}$. ■

Теорема 100.4. *Любой оператор $\mathcal{A}$ в унитарном (евклидовом) пространстве может быть представлен в виде произведения*

$$\mathcal{A} = \mathcal{B}\mathcal{U} \tag{100.3}$$

неотрицательного оператора $\mathcal{B}$ и унитарного оператора $\mathcal{U}$. При этом оператор $\mathcal{B}$ определен однозначно, а если $\mathcal{A}$ обратим, то и оператор $\mathcal{U}$ определен однозначно.

Доказательство. 1. Рассмотрим оператор $\mathcal{A}^*\mathcal{A}$.

а) $\mathcal{A}^*\mathcal{A}$ – самосопряженный оператор, так как $(\mathcal{A}^*\mathcal{A})^* = \mathcal{A}^*\mathcal{A}$.

б) $\mathcal{A}^*\mathcal{A} \geq \mathcal{O}$, так как $(\mathcal{A}^*\mathcal{A}x, x) = (\mathcal{A}x, \mathcal{A}x) \geq 0$, $\forall x \neq \theta$.

в) Существует ортонормированный базис пространства из собственных векторов $e_1, \dots, e_n$ оператора $\mathcal{A}^*\mathcal{A}$. При этом можно считать, что векторы $e_1, \dots, e_r$ отвечают ненулевым собственным значениям $\rho_1^2, \dots, \rho_r^2$, а остальные – нулевым, так что

$$\rho_k > 0 \text{ при } k \leq r \text{ и } \rho_k = 0 \text{ при } k > r . \tag{100.4}$$

Отметим, что

$$r = \operatorname{rg} \mathcal{A}^*\mathcal{A} . \tag{100.5}$$

2. Согласно (80.1), $\operatorname{im} \mathcal{A} = \mathcal{L}(\mathcal{A}e_1, \dots, \mathcal{A}e_n)$. Для векторов $\mathcal{A}e_1, \dots, \mathcal{A}e_n$ имеем

$$(\mathcal{A}e_i, \mathcal{A}e_j) = (\mathcal{A}^*\mathcal{A}e_i, e_j) = (\rho_i^2 e_i, e_j) = \begin{cases} \rho_i^2, & i = j, \\ 0, & i \neq j. \end{cases}$$

Следовательно, $\mathcal{A}e_{r+1} = \ldots = \mathcal{A}e_n = \theta$ (в силу (100.5)), а $\mathcal{A}e_1, \dots, \mathcal{A}e_r$ – ненулевые попарно ортогональные векторы, причем $|\mathcal{A}e_k| = \rho_k$, $k = \overline{1, r}$. Следовательно, векторы

$$g_k = \frac{1}{\rho_k}\mathcal{A}e_k , \quad k = \overline{1, r}, \tag{100.6}$$

образуют ортонормированный базис $\operatorname{im}\mathcal{A}$. Отметим, что тем самым получено (в силу (64.3)) равенство

$$r = \operatorname{rg}\mathcal{A}. \tag{100.7}$$

Дополним векторы $g_1, \ldots, g_r$ до ортонормированного базиса $g_1, \ldots, g_r, g_{r+1}, \ldots, g_n$ всего пространства. Тогда из (100.6) с учетом (100.4) получим, что

$$\mathcal{A}e_k = \rho_k g_k, \quad k = \overline{1,n}. \tag{100.8}$$

3. Построим операторы $\mathcal{B}$ и $\mathcal{U}$ из (100.3). Положим

$$\mathcal{U}e_k = g_k, \quad \mathcal{B}g_k = \rho_k g_k, \; k = \overline{1,n}. \tag{100.9}$$

Тогда $\mathcal{U}$ – унитарный оператор, так как переводит ортонормированный базис e в ортонормированный базис g; $\mathcal{B} \geq \mathcal{O}$ как нормальный оператор (ибо ортонормированный базис g состоит из его собственных векторов), собственные значения которого $\rho_k \geq 0$, $k = \overline{1,n}$. При этом $\mathcal{A} = \mathcal{B}\mathcal{U}$, так как, согласно (100.8) и (100.9), $(\mathcal{B}\mathcal{U})e_k = \mathcal{B}(\mathcal{U}e_k) = \mathcal{B}g_k = \rho_k g_k = \mathcal{A}e_k$, $k = \overline{1,n}$.

4. Докажем единственность. Пусть $\mathcal{A} = \mathcal{B}\mathcal{U}$ – разложение (100.3) для оператора $\mathcal{A}$. Тогда $\mathcal{A}^* = \mathcal{U}^*\mathcal{B}$ и $\mathcal{A}\mathcal{A}^* = \mathcal{B}^2$. Таким образом, $\mathcal{B}$ – квадратный корень из оператора $\mathcal{A}\mathcal{A}^*$, который на основании теоремы 99.4 существует и определен однозначно. Если оператор $\mathcal{A}$ обратим, то из (100.5) и (100.7) следует, что обратим и оператор $\mathcal{A}^*\mathcal{A}$, при этом $\rho_k \neq 0$, $k = \overline{1,n}$, поэтому согласно (100.9) обратим оператор $\mathcal{B}$. В силу (100.3) отсюда следует, что $\mathcal{U} = \mathcal{B}^{-1}\mathcal{A}$. ■

Отметим, что попутно доказаны следующие факты:

а) $\operatorname{rg}\mathcal{A} = \operatorname{rg}\mathcal{A}^*\mathcal{A} = \operatorname{rg}\mathcal{A}\mathcal{A}^*$ (это следует из (100.5), (100.7) и равенства $\mathcal{A}\mathcal{A}^* = \mathcal{B}^2$);

б) оператор $\mathcal{U}$ переводит ортонормированный базис из собственных векторов оператора $\mathcal{A}^*\mathcal{A}$ в ортонормированный базис из собственных векторов оператора $\mathcal{A}\mathcal{A}^*$ (это следует из (100.9));

в) операторы $\mathcal{A}^*\mathcal{A}$ и $\mathcal{A}\mathcal{A}^*$ имеют одинаковые собственные значения (это следует из того, что $\mathcal{A}\mathcal{A}^* = \mathcal{B}^2$).

Арифметические значения квадратных корней из собственных значений оператора $\mathcal{A}^*\mathcal{A}$ называются *сингулярными числами* оператора $\mathcal{A}$.

Разложение (100.3) называется *полярным разложением оператора* $\mathcal{A}$.

Замечание. Для любого оператора $\mathcal{A}$ имеет место разложение вида $\mathcal{A} = \mathcal{V}\mathcal{C}$, где $\mathcal{V}$ – унитарный оператор, $\mathcal{C} \geq \mathcal{O}$. Оно может быть получено из полярного разложения сопряженного оператора $\mathcal{A}^* = \mathcal{B}\mathcal{U}$ переходом к оператору $\mathcal{A}$: $\mathcal{A} = \mathcal{U}^*\mathcal{B}$, где $\mathcal{U}^*$, очевидно, унитарный оператор.

Теорема 100.5. *Оператор $\mathcal{A}$ нормален тогда и только тогда, когда в любом его полярном разложении операторы $\mathcal{B}$ и $\mathcal{U}$ перестановочны.*

Доказательство. Пусть $\mathcal{A} = \mathcal{B}\mathcal{U}$ – полярное разложение оператора $\mathcal{A}$. Тогда $\mathcal{A}\mathcal{A}^* = \mathcal{B}^2$, $\mathcal{A}^*\mathcal{A} = \mathcal{U}^*\mathcal{B}^2\mathcal{U}$. Достаточность очевидна, так как если $\mathcal{B}\mathcal{U} = \mathcal{U}\mathcal{B}$, то $\mathcal{A}^*\mathcal{A} = \mathcal{U}^*\mathcal{B}^2\mathcal{U} = \mathcal{U}^*\mathcal{B}(\mathcal{B}\mathcal{U}) = \mathcal{U}^*\mathcal{B}(\mathcal{U}\mathcal{B}) = \mathcal{U}^*(\mathcal{B}\mathcal{U})\mathcal{B} = \mathcal{U}^*(\mathcal{U}\mathcal{B})\mathcal{B} = \mathcal{U}^*\mathcal{U}\mathcal{B}^2 = \mathcal{B}^2 = \mathcal{A}\mathcal{A}^*$.

Необходимость. Если $e_1, \ldots, e_n$ – ортонормированный базис из собственных векторов $\mathcal{A}^*\mathcal{A}$, то

$$\mathcal{A}^*\mathcal{A}e_k = \rho_k^2 e_k, \quad k = \overline{1,n}. \tag{100.10}$$

С другой стороны, $\mathcal{A}^*\mathcal{A} = \mathcal{U}^*\mathcal{B}^2\mathcal{U}$ и, значит, $\mathcal{U}^*\mathcal{B}^2\mathcal{U}e_k = \rho_k^2 e_k$, т.е.

$$\mathcal{B}^2\mathcal{U}e_k = \rho_k^2\mathcal{U}e_k, \quad k = \overline{1,n}. \tag{100.11}$$

Так как $\mathcal{U}$ унитарен, то векторы $\mathcal{U}e_1, \ldots, \mathcal{U}e_n$ образуют ортонормированный базис, поэтому в соответствии с (99.3) из (100.11) получим

$$\mathcal{B}\mathcal{U}e_k = \rho_k\mathcal{U}e_k, \quad k = \overline{1,n}. \tag{100.12}$$

С другой стороны, $\mathcal{A}^*\mathcal{A} = \mathcal{A}\mathcal{A}^* = \mathcal{B}^2$ и согласно (100.10) $\mathcal{B}^2 e_k = \rho_k^2 e_k$ или, в соответствии с (99.3), $\mathcal{B}e_k = \rho_k e_k$, откуда умножением на $\mathcal{U}$ слева получим, что

$$\mathcal{U}\mathcal{B}e_k = \rho_k\mathcal{U}e_k, \quad k = \overline{1,n}. \tag{100.13}$$

Сравнение равенств (100.12) и (100.13) доказывает, что $\mathcal{B}\mathcal{U} = \mathcal{U}\mathcal{B}$. ■

Глава XVII. Квадратичные формы

В §58 рассматривалась одна из основных задач аналитической геометрии на плоскости – приведение общего уравнения линии второго порядка к каноническому виду. Теория квадратичных форм, излагаемая здесь, основной своей целью ставит решение этой задачи и задач, связанных с нею, в пространстве любой размерности.

§101. Квадратичные формы в линейном пространстве

Билинейные формы. Пусть V – линейное пространство над полем P[1]. Отображение $\mathcal{A} : V \times V \to P$ называется *билинейной формой* в пространстве V, если для любых $x, y, z \in V$, $\alpha \in P$:

$$\begin{aligned} &1)\ \mathcal{A}(x+y,z) = \mathcal{A}(x,z) + \mathcal{A}(y,z);\\ &2)\ \mathcal{A}(\alpha x,y) = \alpha\mathcal{A}(x,y);\\ &3)\ \mathcal{A}(x,y+z) = \mathcal{A}(x,y) + \mathcal{A}(x,z);\\ &4)\ \mathcal{A}(x,\alpha y) = \alpha\mathcal{A}(x,y). \end{aligned} \tag{101.1}$$

Билинейная форма называется *симметричной*, если $\mathcal{A}(y,x) = \mathcal{A}(x,y)$, $\forall x, y \in V$.

Примеры. 1. Скалярное произведение (x,y) в евклидовом пространстве является симметричной билинейной формой.

2. В любом линейном пространстве V, если $f(x)$, $g(x)$ – линейные формы, то функция $\mathcal{A}(x,y) = f(x)g(y)$ является симметричной билинейной формой.

3. В n-мерном пространстве V с базисом $e_1, \dots, e_n$ отображение $\mathcal{A} : V \times V \to P$, определенное правилом

$$\mathcal{A}(x,y) = \sum_{i,j=1}^{n} a_{ij} x_i y_j, \tag{101.2}$$

$\forall x = \sum_{i=1}^n x_i e_i$, $y = \sum_{i=1}^n y_i e_i$, где a_{ij} $(i,j = \overline{1,n})$ – фиксированные числа, является билинейной формой (в силу линейности координат).

Теорема 101.1. *Пусть V – линейное пространство над полем P и $e_1, \dots, e_n$ – базис V. Для любых чисел $a_{ij} \in P$, $i,j = \overline{1,n}$, существует, и притом единственная, билинейная форма $\mathcal{A}(x,y)$ в пространстве V, для которой $\mathcal{A}(e_i, e_j) = a_{ij}$, $i,j = \overline{1,n}$.*

Доказательство. Пусть $e_1, \dots, e_n$ – базис пространства V и a_{ij}, $i,j = \overline{1,n}$, – заданные числа. Отображение (101.2) является билинейной формой в пространстве V (пример 3), причем $\mathcal{A}(e_i, e_j) = a_{ij}$, $i,j = \overline{1,n}$ (так как векторы e_i и e_j имеют очевидные наборы координат в базисе $e_1, \dots, e_n$). Покажем, что любая билинейная форма $\mathcal{B}(x,y)$ в пространстве V, для которой $\mathcal{B}(e_i, e_j) = a_{ij}$, $i,j = \overline{1,n}$, совпадает с $\mathcal{A}(x,y)$. В самом деле, для любых векторов $x = \sum_{i=1}^n x_i e_i$, $y = \sum_{i=1}^n y_i e_i$ в силу (101.1) имеем $\mathcal{B}(x,y) = \mathcal{B}(\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j) = \sum_{i=1}^n \sum_{j=1}^n x_i y_j \mathcal{B}(e_i, e_j) = \sum_{i,j=1}^n a_{ij} x_i y_j$, т.е. $\mathcal{B}(x,y) = \mathcal{A}(x,y)$, $\forall x, y \in V$. ■

Таким образом, в примере 3 приведен самый общий вид билинейной формы в n-мерном пространстве.

Представление билинейной формы в виде (101.2) называется *общим видом билинейной формы в базисе e*. Матрица $A_e = (a_{ij}) \in P^{n\times n}$, элементы которой определены равенством $a_{ij} = \mathcal{A}(e_i, e_j)$, $i,j = \overline{1,n}$, называется *матрицей билинейной формы $\mathcal{A}(x,y)$ в базисе e*.

Как следует из теоремы 101.1, билинейная форма однозначно определяется своим заданием на базисных векторах, при этом матрица билинейной формы представляет собой таблицу ее значений на парах базисных векторов и в силу (101.2) однозначно определяет значения этой формы для любой пары векторов.

Общий вид (101.2) билинейной формы $\mathcal{A}(x,y)$ может быть записан в компактной форме: если x_e и y_e – координатные столбцы векторов x и y в базисе e, то

$$\mathcal{A}(x,y) = x_e^T A_e y_e, \qquad \mathcal{A}(x,y) = y_e^T A_e^T x_e. \tag{101.3}$$

Первое из равенств (101.3) проверяется непосредственно, второе равенство можно получить транспонированием обеих частей первого.

Выражение, стоящее в правой части (101.2), называется *билинейной формой от переменных $x_1, \dots, x_n$ и $y_1, \dots, y_n$*. Правые части равенств (101.3) представляют собой компактные формы записи этой билинейной формы.

Теорема 101.2. *Произвольная матрица $A = (a_{ij}) \in P^{n\times n}$ является матрицей единственной билинейной формы в заданном базисе пространства.*

Доказательство. Пусть $e_1, \dots, e_n$ – заданный базис пространства V. Матрица $A = (a_{ij}) \in P^{n\times n}$ является матрицей билинейной формы (101.2). Единственность этой билинейной формы вытекает из теоремы 101.1. ■

Таким образом, соответствие между билинейными формами и их матрицами является биективным отображением множества всех билинейных форм в n-мерном пространстве V над полем P на множестве всех матриц $P^{n\times n}$.

Теорема 101.3. *Матрицы билинейной формы $\mathcal{A}(x,y)$ в базисах e и $f = eQ$ связаны соотношением $A_f = Q^T A_e Q$.*

Доказательство. Согласно (101.3), с одной стороны, $\mathcal{A}(x,y) = x_e^T A_e y_e = \{x_e = Q x_f, y_e = Q y_f\} = x_f^T Q^T A_e Q y_f$. С другой стороны, $\mathcal{A}(x,y) = x_f^T A_f y_f$. Отсюда с учетом произвольности x, y следует утверждение теоремы. ■

Следствие. $\operatorname{rg} A_e = \operatorname{rg} A_f$.

[1] В этой главе предполагается, что характеристика основного поля равна 0.

Теорема 101.4. *Билинейная форма симметрична тогда и только тогда, когда ее матрица в любом базисе симметрична.*

Доказательство. Необходимость проверяется непосредственно. Достаточность. Если $A_e^T = A_e$, то, согласно (101.3), $\mathcal{A}(x,y) = y_e^T A_e^T x_e = y_e^T A_e x_e = \mathcal{A}(y,x)$. ■

Все матрицы одной билинейной формы имеют одинаковый ранг (см. следствие из теоремы 101.3). *Рангом билинейной формы* называется ранг ее матрицы в произвольном базисе. Билинейная форма $\mathcal{A}(x,y)$ называется *вырожденной*, если $\operatorname{rg}\mathcal{A}(x,y) < \dim V$, и *невырожденной*, если $\operatorname{rg}\mathcal{A}(x,y) = \dim V$.

Теорема 101.5. *Билинейная форма $\mathcal{A}(x,y)$ вырождена тогда и только тогда, когда существует вектор $x \neq \theta$ такой, что*

$$\mathcal{A}(x,y) = 0, \quad \forall y \in V. \tag{101.4}$$

Доказательство. Пусть $e_1, \ldots, e_n$ – базис V и $A_e = (a_{ij})$ – матрица билинейной формы $\mathcal{A}(x,y)$ в этом базисе. Соотношение (101.4) равносильно системе равенств $\mathcal{A}(x,e_j) = 0$, $j = \overline{1,n}$, или, если учесть разложение $x = \sum_{i=1}^n x_i e_i$, системе равенств $\sum_{i=1}^n x_i \mathcal{A}(e_i,e_j) = 0$, $j = \overline{1,n}$, т.е.

$$\left\{\begin{array}{l} a_{11}x_1 + a_{21}x_2 + \ldots + a_{n1}x_n = 0, \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ a_{1n}x_1 + a_{2n}x_2 + \ldots + a_{nn}x_n = 0. \end{array}\right. \tag{101.5}$$

Значит, билинейная форма $\mathcal{A}(x,y)$ вырождена тогда и только тогда, когда однородная система линейных уравнений (101.5) имеет нетривиальное решение, т.е. (теорема 28.5) когда $\operatorname{rg} A_e < n$. ■

Квадратичные формы. Пусть $\mathcal{A}(x,y)$ – симметричная билинейная форма в пространстве V над полем P. *Квадратичной формой* называется отображение $\mathcal{A}: V \to P$, которое каждому вектору $x \in V$ ставит в соответствие число $\mathcal{A}(x,x)$. Обозначение: $\mathcal{A}(x,x)$ или $\mathcal{A}(x)$. Билинейная форма $\mathcal{A}(x,y)$ при этом называется *полярной билинейной формой* к квадратичной форме $\mathcal{A}(x,x)$.

Теорема 101.6. *Полярная билинейная форма для любой квадратичной формы определена однозначно.*

Это утверждение вытекает из того, что если $\mathcal{A}(x,y)$ – полярная билинейная форма для квадратичной формы $\mathcal{A}(x,x)$, то для любых $x, y \in V$

$$\mathcal{A}(x,y) = \tfrac{1}{2}(\mathcal{A}(x+y,x+y) - \mathcal{A}(x,x) - \mathcal{A}(y,y)). \quad ■$$

Таким образом, полярная билинейная форма $\mathcal{A}(x,y)$ однозначно восстанавливается по квадратичной форме $\mathcal{A}(x,x)$. Единственность полярной билинейной формы $\mathcal{A}(x,y)$ позволяет переносить ее характеристики на квадратичную форму $\mathcal{A}(x,x)$.

Матрицей квадратичной формы $\mathcal{A}(x,x)$ в базисе e называется матрица полярной к ней билинейной формы $\mathcal{A}(x,y)$ в этом базисе.

Из свойств билинейной формы вытекают следующие свойства квадратичных форм.

1°. Матрица квадратичной формы симметрична.

2°. Любая симметрическая матрица является матрицей единственной квадратичной формы в заданном базисе.

3°. Матрицы квадратичной формы в базисах e и $f = eQ$ связаны соотношением

$$A_f = Q^T A_e Q. \tag{101.6}$$

4°. В базисе e квадратичная форма $\mathcal{A}(x,x)$ с матрицей $A_e = (a_{ij})$ может быть записана в следующем виде: $\forall x = \sum_{i=1}^n x_i e_i$

$$\mathcal{A}(x,x) = \sum_{i,j=1}^n a_{ij}x_i x_j, \quad a_{ij} = a_{ji}, \tag{101.7}$$

или, в компактной форме,

$$\mathcal{A}(x,x) = x_e^T A_e x_e, \quad A_e^T = A_e. \tag{101.8}$$

Представление квадратичной формы в виде (101.7) или (101.8) называется *общим видом квадратичной формы $\mathcal{A}(x,x)$ в базисе e*. Выражение $f(x_1,\ldots,x_n) = \sum_{i,j=1}^n a_{ij}x_i x_j$, где $a_{ij} = a_{ji}$, называется *квадратичной формой от переменных $x_1,\ldots,x_n$*.

5°. *Рангом квадратичной формы* называется ранг ее матрицы в произвольном базисе. Очевидно, $\operatorname{rg}\mathcal{A}(x,x) = \operatorname{rg}\mathcal{A}(x,y)$. Квадратичная форма $\mathcal{A}(x,x)$ называется *вырожденной*, если $\operatorname{rg}\mathcal{A}(x,x) < \dim V$, и *невырожденной*, если $\operatorname{rg}\mathcal{A}(x,x) = \dim V$.

Канонический вид квадратичной формы. Выбирая подходящим образом базис V, можно менять вид матрицы квадратичной формы и, следовательно, ее общий вид (101.7).

Базис $e = (e_1,\ldots,e_n)$ называется *каноническим базисом квадратичной формы $\mathcal{A}(x,x)$*, если матрица квадратичной формы в этом базисе диагональна: $A_e = \operatorname{diag}(\lambda_1,\ldots,\lambda_n)$.

В каноническом базисе квадратичная форма $\mathcal{A}(x,x)$ согласно (101.7) имеет вид $\mathcal{A}(x,x) = \lambda_1 x_1^2 + \ldots + \lambda_n x_n^2$, который называется *каноническим видом* квадратичной формы, а числа $\lambda_1,\ldots,\lambda_n$ – ее *каноническими коэффициентами*. Канонический вид называют также *суммой квадратов*. Очевидно, что число ненулевых квадратов совпадает с рангом $\mathcal{A}(x,x)$. Итак, если e – канонический базис и $r = \operatorname{rg}\mathcal{A}(x,x)$, то

$$\mathcal{A}(x,x) = \lambda_1 x_1^2 + \ldots + \lambda_r x_r^2, \quad \forall x = \sum_{i=1}^n x_i e_i. \tag{101.9}$$

Теорема 101.7. *Для любой квадратичной формы существует канонический базис.*

Доказательство. Пусть $e = (e_1, \dots, e_n)$ – базис V. Квадратичная форма $\mathcal{A}(x, x)$ с матрицей $A_e = (a_{ij})$ имеет в этом базисе вид (101.7). Обозначим $g(x_1, \dots, x_n) = \sum_{i,j=1}^n a_{ij} x_i x_j$.

Переход к другому базису $f = eQ$ равносилен преобразованию координат по известному закону $x_e = Qx_f$, $|Q| \neq 0$. В терминах координат утверждение теоремы означает, что *любая квадратичная форма $g(x_1, \dots, x_n)$ от переменных $x_1, \dots, x_n$ невырожденным преобразованием координат приводится к сумме квадратов*. Это мы и будем доказывать. Будем считать, что $A_e \neq O$ (если $A_e = O$, то e – искомый базис). Обозначим через Δ_k *угловые миноры* k-го порядка матрицы A_e, т.е. $\Delta_k = M_{1,\dots,k}^{1,\dots,k}$, $k = \overline{1,n}$, и положим $\Delta_0 = 1$.

I. Рассмотрим сначала случай, когда $\Delta_k \neq 0$, $k = \overline{1, n-1}$. В этом случае процесс приведения к каноническому виду состоит из $n-1$ однотипных шагов.

Первый шаг основан на том, что $\Delta_1 \neq 0$, т.е. $a_{11} \neq 0$. Соберем все члены квадратичной формы $g(x_1, \dots, x_n)$, содержащие x_1, и выделим из них полный квадрат:

$$\mathcal{A}(x,x) = g(x_1, \dots, x_n) = a_{11} x_1^2 + 2\sum_{k=2}^n a_{1k} x_1 x_k + \sum_{i,k=2}^n a_{ik} x_i x_k =$$
$$= a_{11}\left(x_1 + \sum_{k=2}^n \frac{a_{1k}}{a_{11}} x_k\right)^2 - a_{11}\left(\sum_{k=2}^n \frac{a_{1k}}{a_{11}} x_k\right)^2 + \sum_{i,k=2}^n a_{ik} x_i x_k.$$

Перейдем к новым координатам:

$$x_1' = x_1 + \sum_{k=2}^n \frac{a_{1k}}{a_{11}} x_k \quad \text{и} \quad x_j' = x_j \text{ при } j \neq 1,$$

очевидно выполнив при этом невырожденное преобразование координат с матрицей

$$Q_1 = \begin{bmatrix} 1 & -a_{12}/a_{11} & \dots & -a_{1n}/a_{11} \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Тогда квадратичная форма $\mathcal{A}(x,x)$ в новых координатах примет вид $\mathcal{A}(x,x) = a_{11} x_1'^2 + h(x_2', \dots, x_n')$, где $h(x_2', \dots, x_n')$ – квадратичная форма от переменных $x_2', \dots, x_n'$. При этом матрица $A_1 = Q_1^T A_e Q_1$ квадратичной формы $\mathcal{A}(x,x)$ в новом базисе будет иметь вид

$$A_1 = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22}' & \dots & a_{2n}' \\ \dots & \dots & \dots & \dots \\ 0 & a_{n2}' & \dots & a_{nn}' \end{bmatrix}.$$

Каждая строка (столбец) матрицы A_1, начиная со второй, полу чена из соответствующей строки (соответственно столбца) матриц A_e вычитанием из нее первой строки (соответственно столбца) A_e умноженной на некоторое число, поэтому угловые миноры матриц

A_1 совпадают с $\Delta_1, \dots, \Delta_n$. Следовательно, $\Delta_1 = a_{11}$, $\Delta_2 = a_{11} a_{22}'$ и

$$a_{22}' = \Delta_2/\Delta_1 \neq 0. \qquad (101.10)$$

Второй шаг основан на том, что $\Delta_2 \neq 0$, т.е. $a_{22}' \neq 0$, и состоит в применении действий первого шага к квадратичной форме $h(x_2', \dots, x_n')$: выделяется полный квадрат среди всех членов, содержащих x_2', выполняется невырожденное преобразование координат

$$x_2'' = x_2' + \sum_{k=3}^n \frac{a_{2k}'}{a_{22}'} x_k' \quad \text{и} \quad x_j'' = x_j' \text{ при } j \neq 2$$

и квадратичная форма $\mathcal{A}(x,x)$ приводится к виду $\mathcal{A}(x,x) = a_{11} x_1''^2 + a_{22}' x_2''^2 + v(x_3'', \dots, x_n'')$, где $v(x_3'', \dots, x_n'')$ – квадратичная форма от переменных $x_3'', \dots, x_n''$, а ее матрица – к виду

$$A_2 = \begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22}' & 0 & \dots & 0 \\ 0 & 0 & a_{33}'' & \dots & a_{3n}'' \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & a_{n3}'' & \dots & a_{nn}'' \end{bmatrix}.$$

По-прежнему угловые миноры A_2 совпадают с $\Delta_1, \Delta_2, \dots, \Delta_n$, поэтому

$$a_{33}'' = \Delta_3/\Delta_2 \neq 0. \qquad (101.11)$$

Повторяя этот процесс, через $(n-1)$ шагов мы придем к базису, в котором матрица квадратичной формы $\mathcal{A}(x,x)$ имеет диагональную форму: $A_{n-1} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, где с учетом (101.10), (101.11) и обозначения $\Delta_0 = 1$

$$\lambda_i = \Delta_i/\Delta_{i-1}, \quad i = \overline{1,n}. \qquad (101.12)$$

Отметим, что каждый шаг (для определенности i-й шаг) процесса начинается с проверки условия $\Delta_i \neq 0$, так как только при выполнении этого условия можно провести i-й шаг.

Метод приведения к каноническому виду, описанный здесь, называется *методом Лагранжа*.

II. Пусть теперь среди угловых миноров $\Delta_1, \dots, \Delta_{n-1}$ могут встретиться нулевые. Модифицируем метод Лагранжа.

Опишем i-й шаг. Пусть после $(i-1)$-го шага матрица квадратичной формы $\mathcal{A}(x,x)$ имеет вид

$$A_{i-1} = \begin{bmatrix} a_{11} & & & \\ & \ddots & & O \\ & & a_{i-1,i-1} & \\ O & & & \boxed{C} \end{bmatrix}, \text{ где } C = \begin{bmatrix} a_{ii} & \dots & a_{in} \\ \dots & \dots & \dots \\ a_{ni} & \dots & a_{nn} \end{bmatrix}$$

(здесь штрихи опущены для упрощения записи). Будем считать, что

$C \neq O$ (равенство $C = O$ означает, что канонический базис уже построен).

1. Если $a_{ii} \neq 0$, то выполним i-й шаг метода Лагранжа.
2. Пусть $a_{ii} = 0$.

а) Если среди диагональных элементов матрицы C существует элемент $a_{jj} \neq 0$, $j > i$, то перенумеруем переменные (т.е. векторы базиса): $x_i' = x_j$, $x_j' = x_i$ и $x_k' = x_k$ при $k \neq i, j$. Тогда в матрице A_{i-1} поменяются местами строки (столбцы) с номерами i и j, поэтому в позиции (i,i) окажется ненулевой элемент $a_{ii}' = a_{jj}$, с помощью которого выполним i-й шаг метода Лагранжа.

б) Пусть все диагональные элементы матрицы C равны нулю, тогда в ней существует внедиагональный элемент $a_{kj} \neq 0$, $k, j \geq i$, $k \neq i$. Это означает, что в квадратичной форме от переменных $x_i, \ldots, x_n$ отсутствуют квадраты $x_i^2, \ldots, x_n^2$, но содержится член вида $2a_{kj}x_kx_j$. Перейдем к новым координатам, положив $x_k = x_k' + x_j'$, $x_j = x_k' - x_j'$ и $x_s' = x_s$ при $s \neq k, j$. Тогда квадратичная форма будет иметь квадраты $x_k'^2$, $x_j'^2$ и мы окажемся в ситуации "а". Отметим, что все преобразования координат были невырожденными. ■

Теорема 101.8. *Если в матрице квадратичной формы $\mathcal{A}(x, x)$ ранга r первые r угловых миноров отличны от нуля: $\Delta_k \neq 0$, $k = \overline{1,r}$, то существует базис e, в котором матрица квадратичной формы имеет диагональный вид $A_e = \mathrm{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)$, где*

$$\lambda_k = \Delta_k/\Delta_{k-1}, \; k = \overline{1,r}. \tag{101.13}$$

Доказательство. Для квадратичной формы $\mathcal{A}(x,x)$ в условиях теоремы выполнимы первые r шагов метода Лагранжа. После r-го шага матрица A_r квадратичной формы имеет вид

$$A_r = \begin{bmatrix} \lambda_1 & & & O \\ & \ddots & & \\ & & \lambda_r & \\ O & & & \boxed{C} \end{bmatrix},$$

где, согласно (101.12), $\lambda_k = \Delta_k/\Delta_{k-1}$, $k = \overline{1,r}$, а C – некоторая матрица. Так как $\mathrm{rg}\, A_r = r$ и $\lambda_k \neq 0$, $k = \overline{1,r}$, то $\mathrm{rg}\, C = 0$ и $C = O$. Следовательно, матрица A_r имеет искомый вид. ■

Соотношения (101.13) называются *формулами Якоби.*

§102. Квадратичные формы в вещественном пространстве

Ни канонический базис, ни канонический вид квадратичной формы не определены однозначно. Например, любая перестановка векторов канонического базиса приводит вновь к каноническому базису.

Что общего у различных канонических видов одной и той же квадратичной формы? В (101.9) мы отмечали, что число ненулевых канонических коэффициентов равно рангу квадратичной формы и не зависит от выбора канонического базиса. В вещественном пространстве можно говорить и о знаках канонических коэффициентов.

Пусть квадратичная форма $\mathcal{A}(x,x)$ ранга r в вещественном пространстве V приведена к каноническому виду (101.9). Число π положительных квадратов в (101.9) и число $\nu = r - \pi$ называются *положительным* и *отрицательным индексами инерции* квадратичной формы $\mathcal{A}(x,x)$, их разность $\sigma = \pi - \nu$ называется *сигнатурой* $\mathcal{A}(x,x)$. Корректность определения вытекает из следующей теоремы.

Теорема 102.1 (закон инерции). *Положительный и отрицательный индексы инерции вещественной квадратичной формы не зависят от выбора канонического базиса.*

Доказательство. Пусть e и f – канонические базисы квадратичной формы $\mathcal{A}(x,x)$ ранга r и для $x = \sum_{i=1}^n x_ie_i = \sum_{i=1}^n y_if_i$

$$\begin{aligned} \mathcal{A}(x,x) &= a_1x_1^2 + \ldots + a_px_p^2 - a_{p+1}x_{p+1}^2 - \ldots - a_rx_r^2, \\ \mathcal{A}(x,x) &= b_1y_1^2 + \ldots + b_qy_q^2 - b_{q+1}y_{q+1}^2 - \ldots - b_ry_r^2, \end{aligned} \tag{102.1}$$

где $a_i > 0$, $b_i > 0$, $i = \overline{1,r}$. Докажем, что $p \leq q$. Пусть $p > q$. Рассмотрим подпространства $L_1 = \mathcal{L}(e_1, \ldots, e_p)$, $L_2 = \mathcal{L}(f_{q+1}, \ldots, f_n)$. Согласно (65.2), $\dim L_1 \cap L_2 = p + (n-q) - \dim(L_1 + L_2)$. Так как $\dim(L_1 + L_2) \leq n$, $p > q$, то $\dim L_1 \cap L_2 > 0$. Следовательно, существует ненулевой вектор $x_0 \in L_1 \cap L_2$. Пусть $x_0 = \alpha_1e_1 + \ldots + \alpha_pe_p = \beta_{q+1}f_{q+1} + \ldots + \beta_nf_n$. Тогда согласно (102.1)

$$\mathcal{A}(x_0, x_0) = a_1\alpha_1^2 + \ldots + a_p\alpha_p^2 = -b_{q+1}\beta_{q+1}^2 - \ldots - b_r\beta_r^2. \tag{102.2}$$

Так как $x_0 \neq \theta$, то $a_1\alpha_1^2 + \ldots + a_p\alpha_p^2 > 0$, $-b_{q+1}\beta_{q+1}^2 - \ldots - b_r\beta_r^2 < 0$. Это противоречит (102.2), и, значит, $p \leq q$. Аналогично показывается, что $q \leq p$. Следовательно, $p = q$. ■

Обозначим символами $P(\Delta_0, \Delta_1, \ldots, \Delta_k)$ и $V(\Delta_0, \Delta_1, \ldots, \Delta_k)$ число совпадений и перемен знаков в последовательности $\Delta_0, \Delta_1, \ldots, \Delta_k$.

Теорема 102.2 (сигнатурное правило Якоби). *Пусть Δ_k – угловой минор k-го порядка матрицы квадратичной формы $\mathcal{A}(x,x)$ ранга r и $\Delta_k \neq 0$, $k = \overline{1,r}$. Тогда*

$$\pi = P(\Delta_0, \Delta_1, \ldots, \Delta_r), \quad \nu = V(\Delta_0, \Delta_1, \ldots, \Delta_r).$$

Утверждение теоремы вытекает из формул Якоби (101.13), так как $\lambda_i > 0$, если $\mathrm{sgn}\,\Delta_i = \mathrm{sgn}\,\Delta_{i-1}$, и $\lambda_i < 0$, если $\mathrm{sgn}\,\Delta_i \neq \mathrm{sgn}\,\Delta_{i-1}$. ■

Знакоопределенные квадратичные формы. Квадратичная форма $\mathcal{A}(x,x)$ называется *положительно* (*отрицательно*) *определенной*, если $\mathcal{A}(x,x) > 0$ (соответственно $\mathcal{A}(x,x) < 0$), $\forall x \neq \theta$. Такие формы называют *знакоопределенными* (или *знакопостоянными*).

Квадратичная форма, для которой существуют векторы x и y такие, что $\mathcal{A}(x,x) > 0$, $\mathcal{A}(y,y) < 0$, называются ***знакопеременными***.

Пример. Примером положительно определенной квадратичной формы в вещественном пространстве может служить скалярный квадрат в евклидовом пространстве, т.е. отображение $\mathcal{A} : E \to \mathbb{R}$, определенное равенством $\mathcal{A}(x) = (x,x)$, $\forall x \in E$. Это следует из того, что скалярное произведение (x,y) является симметричной билинейной формой (в силу аксиом 1–3 из (68.1)), а скалярный квадрат – положительно определенной квадратичной формой (в силу аксиомы 4).

Теорема 102.3. *Квадратичная форма $\mathcal{A}(x,x)$ положительно (отрицательно) определена тогда и только тогда, когда ее положительный (соответственно отрицательный) индекс инерции совпадает с размерностью пространства.*

Доказательство. Докажем сначала вариант теоремы, относящийся к положительно определенной форме. Пусть e – канонический базис квадратичной формы $\mathcal{A}(x,x)$ и $\lambda_1,\ldots,\lambda_n$ – ее канонические коэффициенты. Необходимость. Если $\mathcal{A}(x,x) > 0$, $\forall x \neq \theta$, то, в частности, $\mathcal{A}(e_i,e_i) > 0$, $i = \overline{1,n}$. Но $\mathcal{A}(e_i,e_i) = \lambda_i$, следовательно, все канонические коэффициенты положительны и $\pi = n$.

Достаточность. Если $\lambda_i > 0$, $i = \overline{1,n}$, то согласно (101.9)

$$\mathcal{A}(x,x) = \sum_{i=1}^{n} \lambda_i x_i^2 > 0,\ \forall x = \sum_{i=1}^{n} x_i e_i \neq \theta.$$

Для доказательства второго утверждения достаточно рассмотреть квадратичную форму $-\mathcal{A}(x,x)$:

$$\mathcal{A}(x,x) < 0 \iff -\mathcal{A}(x,x) > 0 \iff \nu = \dim V. \ ■$$

Следствие. *Определитель матрицы положительно определенной квадратичной формы положителен,* так как если e – канонический базис, то в произвольном базисе f, согласно (101.6), $|A_f| = \lambda_1 \cdot \ldots \cdot \lambda_n \cdot |Q|^2 > 0$.

Теорема 102.4 (критерий Сильвестра). *Квадратичная форма $\mathcal{A}(x,x)$ положительно (отрицательно) определена тогда и только тогда, когда угловые миноры Δ_k, $k = \overline{1,n}$, ее матрицы в произвольном базисе положительны (соответственно чередуют знаки, начиная с отрицательного):*

$$\Delta_k > 0,\ k = \overline{1,n} \quad (\Delta_k \Delta_{k-1} < 0,\ k = \overline{0,n}).$$

Доказательство. Докажем первое утверждение.

Необходимость. Пусть $\mathcal{A}(x,x)$ – положительно определенная квадратичная форма, A_e – ее матрица в произвольном базисе $e = (e_1,\ldots,e_n)$. Рассмотрим подпространства $L_k = \mathcal{L}(e_1,\ldots,e_k)$, $k = \overline{1,n}$, и квадратичные формы $\mathcal{A}(x,x)$ на L_k. Очевидно, $\mathcal{A}(x,x) > 0$, $\forall x \in L_k$, $x \neq \theta$, и, значит, ее матрица A_k в базисе $e_1,\ldots,e_k$ пространства L_k имеет положительный определитель $|A_k|$ (следствие 1). Но $|A_k| = \Delta_k$, $k = \overline{1,n}$, следовательно, $\Delta_k > 0$, $k = \overline{1,n}$.

Достаточность вытекает из теорем 102.2, 102.3.

Для доказательства второго утверждения рассмотрим квадратичную форму $-\mathcal{A}(x,x)$ с угловыми минорами δ_k. Отрицательная определенность $\mathcal{A}(x,x)$ равносильна положительности формы $-\mathcal{A}(x,x)$, т.е. условию $\delta_k > 0$, $k = \overline{1,n}$, или $(-1)^k \Delta_k > 0$, $k = \overline{1,n}$. Так как $\Delta_0 = 1 > 0$, то это условие означает, что $\Delta_k \Delta_{k-1} < 0$, $k = \overline{1,n}$. ■

Как следует из примера положительно определенной квадратичной формы, скалярное произведение в евклидовом пространстве является билинейной формой, полярной к положительно определенной квадратичной форме. Оказывается, что такими билинейными формами исчерпываются все скалярные произведения в вещественном пространстве.

Теорема 102.5. *Пусть V – вещественное линейное пространство. Отображение $\mathcal{A} : V \times V \to \mathbb{R}$ есть скалярное произведение в пространстве V тогда и только тогда, когда оно является билинейной формой, полярной к положительно определенной квадратичной форме.*

Доказательство. Необходимость рассмотрена в примере. Достаточность. Пусть $\mathcal{A}(x,y)$ – билинейная форма, полярная к положительно определенной квадратичной форме. Тогда отображение $\mathcal{A} : V \times V \to \mathbb{R}$, определенное правилом

$$(x,y) = \mathcal{A}(x,y), \quad \forall x,y \in V, \tag{102.3}$$

является скалярным произведением, так как оно удовлетворяет всем аксиомам (68.1) скалярного произведения. ■

Замечание. Матрица билинейной формы, задающей скалярное произведение, совпадает с матрицей Грама базисных векторов: $A_e = G(e_1,\ldots,e_k)$.

§103. Квадратичные формы в комплексном пространстве

В комплексном пространстве полным аналогом вещественных квадратичных форм являются так называемые эрмитовы квадратичные формы.

Полуторалинейные формы. Пусть V – комплексное линейное пространство. Отображение $\mathcal{A} : V \times V \to \mathbb{C}$ называется ***полуторалинейной формой***, если для любых $x, y, z \in V$, $\alpha \in \mathbb{C}$:

1) $\mathcal{A}(x+y,z) = \mathcal{A}(x,z) + \mathcal{A}(y,z)$;
2) $\mathcal{A}(\alpha x,y) = \alpha\mathcal{A}(x,y)$;
3) $\mathcal{A}(x,y+z) = \mathcal{A}(x,y) + \mathcal{A}(x,z)$;
4) $\mathcal{A}(x,\alpha y) = \overline{\alpha}\mathcal{A}(x,y)$.

Полуторалинейную форму называют ***эрмитовой***, **если**

$$\mathcal{A}(y,x) = \overline{\mathcal{A}(x,y)},\ \forall x,y \in V.$$

Примеры. 1. Скалярное произведение (x, y) в унитарном пространстве является эрмитовой полуторалинейной формой.

2. Примером полуторалинейной формы в n-мерном комплексном пространстве V с базисом $e_1, \dots, e_n$ может служить функция, определенная равенством: $\forall x = \sum_{i=1}^n x_i e_i$, $y = \sum_{i=1}^n y_i e_i$

$$\mathcal{A}(x, y) = \sum_{i,j=1}^{n} a_{ij} x_i \overline{y_j}, \tag{103.1}$$

где a_{ij} $(i, j = \overline{1, n})$ – фиксированные комплексные числа.

В действительности выражение (103.1) дает *общий вид полуторалинейной формы в базисе* e. Это доказывается так же, как и аналогичное соотношение (101.2) для билинейной формы. Следует отметить, что для полуторалинейных форм остаются в силе и все другие утверждения (а также их доказательства), касающиеся билинейных форм. Незначительные изменения в них носят очевидный характер. В частности:

– компактная форма (103.1) имеет вид $\mathcal{A}(x, y) = x_e^T A_e \overline{y_e}$ или $\mathcal{A}(x, y) = y_e^H A_e^T x_e$;

– если e и $f = eQ$ – два базиса пространства, то $A_f = Q^T A_e \overline{Q}$;

– полуторалинейная форма эрмитова тогда и только тогда, когда ее матрица в любом базисе эрмитова.

Теорема 103.1. *Полуторалинейная форма эрмитова тогда и только тогда, когда* $\mathcal{A}(x, x) \in \mathbb{R}$, $\forall x \in V$.

Доказательство. Необходимость очевидна, так как

$$\mathcal{A}(x, x) = \overline{\mathcal{A}(x, x)}, \ \forall x \in V.$$

Достаточность вытекает из равенства: $\forall x, y \in V$

$$\mathcal{A}(x, y) = \tfrac{1}{4}\big(\mathcal{A}(x + y, x + y) - \mathcal{A}(x - y, x - y) + \\ + i\mathcal{A}(x + iy, x + iy) - i\mathcal{A}(x - iy, x - iy)\big), \tag{103.2}$$

которое справедливо для любой полуторалинейной формы $\mathcal{A}(x, y)$. ■

Эрмитовы формы. Пусть $\mathcal{A}(x, y)$ – эрмитова полуторалинейная форма в комплексном пространстве V. *Эрмитовой квадратичной формой* (или, короче, *эрмитовой формой*) называется отображение $\mathcal{A} : V \to \mathbb{C}$, которое каждому вектору $x \in V$ ставит в соответствие число $\mathcal{A}(x, x)$. Полуторалинейная форма $\mathcal{A}(x, y)$ при этом называется *полярной полуторалинейной формой* к эрмитовой форме $\mathcal{A}(x, x)$.

Эрмитова квадратичная форма обладает всеми свойствами (с очевидными изменениями) обычных квадратичных форм (§101). Отметим некоторые из них.

1. Существует взаимно однозначное соответствие между эрмитовыми полуторалинейными и эрмитовыми квадратичными формами пространства. Это следует из (103.2).

2. Матрица эрмитовой квадратичной формы в любом базисе эрмитова.

3. Общий вид эрмитовой квадратичной формы: $\forall x = \sum_{i=1}^n x_i e_i$

$$\mathcal{A}(x, x) = \sum_{i,j=1}^{n} a_{ij} x_i \overline{x_j}, \quad a_{ij} = \overline{a_{ji}} \ (i, j = \overline{1, n}),$$

или, в компактной форме,

$$\mathcal{A}(x, x) = x_e^T A_e \overline{x_e} = x_e^H A_e^T x_e, \quad A_e^H = A_e.$$

4. Канонический вид эрмитовой квадратичной формы:

$$\mathcal{A}(x, x) = \sum_{k=1}^{r} \lambda_k |x_k|^2, \quad \forall x = \sum_{i=1}^{n} x_i e_i,$$

где $r = \operatorname{rg} \mathcal{A}(x, x)$, $\lambda_k \in \mathbb{R}$ $(k = \overline{1, r})$.

5. Метод Лагранжа приведения к каноническому виду применим и для эрмитовой формы. Изменение алгоритма, описанного в доказательстве теоремы 101.7, состоит лишь в том, что на каждом шаге выделяется полный квадрат модуля (например, на первом шаге: $f(x_1, \dots, x_n) = a_{11}\left|x_1 + \sum_{k=1}^n \frac{a_{1k}}{a_{11}} x_k\right|^2 + g(x_2, \dots, x_n)$).

6. Остаются справедливыми и формулы Якоби (101.13).

7. Как следует из теоремы 103.1, эрмитова квадратичная форма принимает только вещественные значения. Это свойство сближает ее с вещественной квадратичной формой (§102). В частности, для эрмитовых квадратичных форм справедливы закон инерции, сигнатурное правило Якоби, критерий Сильвестра. Практически без изменения остаются и доказательства этих фактов.

8. И, наконец, отображение $\mathcal{A} : V \times V \to \mathbb{C}$ есть скалярное произведение в комплексном пространстве V тогда и только тогда, когда оно является полуторалинейной формой, полярной к положительно определенной эрмитовой форме.

§104. Квадратичные формы в евклидовом (унитарном) пространстве

Мы отмечали (§102), что ни канонический базис, ни канонический вид квадратичной формы не определены однозначно – однозначно лишь число всех ненулевых квадратов (в произвольном линейном пространстве) и число положительных и отрицательных квадратов (в вещественном и комплексном пространстве). В евклидовом пространстве (в унитарном пространстве для эрмитовых квадратичных форм) положение иное, если рассматривать ортонормированные базисы.

Теорема 104.1. *Для любой квадратичной формы* $\mathcal{A}(x, x)$ *в евклидовом пространстве* E *существует, и притом единственный, симметрический оператор* $\mathcal{H} \in \mathcal{L}(E, E)$ *такой, что*

$$\mathcal{A}(x, x) = (\mathcal{H}x, x), \quad \forall x \in E. \tag{104.1}$$

Доказательство. Существование. Пусть $e = (e_1, \ldots, e_n)$ – ортонормированный базис пространства E и A_e – матрица квадратичной формы $\mathcal{A}(x, x)$ в этом базисе. В силу симметричности матрицы A_e и ортонормированности базиса e существует симметрический оператор $\mathcal{H} \in \mathcal{L}(E, E)$, который в базисе e имеет матрицу A_e, так что $H_e = A_e$. Тогда для любого вектора $x = \sum_{i=1}^n x_i e_i$ координаты вектора $\mathcal{H}x$ находятся согласно (77.3): $(\mathcal{H}x)_e = H_e x_e = A_e x_e$. Отсюда в силу (70.7) следует равенство $(\mathcal{H}x, x) = x_e^T A_e x_e$, которое с учетом (101.8) доказывает (104.1).

Единственность. Если $\mathcal{H}_1, \mathcal{H}_2$ – симметрические операторы, удовлетворяющие (104.1), то $(\mathcal{H}_1 x, x) = (\mathcal{H}_2 x, x)$, $\forall x \in E$, или

$$((\mathcal{H}_1 - \mathcal{H}_2)x, x) = 0, \ \forall x \in E. \tag{104.2}$$

Отсюда следует [2], что $\mathcal{H}_1 = \mathcal{H}_2$. ■

Теорема 104.2. *Для любой квадратичной формы в евклидовом пространстве E существует ортонормированный базис, в котором она имеет канонический вид.*

Доказательство. Пусть $\mathcal{H}$ – симметрический оператор, для которого $(\mathcal{H}x, x) = \mathcal{A}(x, x)$, $\forall x \in E$. Если $e_1, \ldots, e_n$ – ортонормированный базис из собственных векторов $\mathcal{H}$, то для любого вектора $x = \sum_{i=1}^n x_i e_i$, согласно (104.1) и (70.7), имеем

$$\mathcal{A}(x, x) = \left(\sum_{i=1}^n x_i \lambda_i e_i, \sum_{i=1}^n x_i e_i\right) = \sum_{i=1}^n \lambda_i x_i^2, \tag{104.3}$$

где $\lambda_1, \ldots, \lambda_n$ – собственные значения оператора $\mathcal{H}$, отвечающие собственным векторам $e_1, \ldots, e_n$. ■

Операция построения ортонормированного базиса, в котором квадратичная форма имеет канонический вид, называется *приведением квадратичной формы к главным осям*. Из теоремы 104.2 следует, что *любая квадратичная форма приводится к главным осям.*

Замечание 1. Доказывая теорему 104.2, мы попутно показали, что канонический базис квадратичной формы $\mathcal{A}(x, x)$ совпадает с ортонормированным базисом из собственных векторов соответствующего симметрического оператора $\mathcal{H}$, а канонические коэффициенты – с отвечающими им собственными значениями. Собственные значения оператора $\mathcal{H}$ являются корнями уравнения $|A_e - \lambda I| = 0$, которые, вообще говоря, уже не зависят от оператора $\mathcal{H}$ и инвариантно связаны только с самой квадратичной формой.

Таким образом, при приведении квадратичной формы к главным осям *канонические коэффициенты определены однозначно.* Это позволяет находить канонический вид квадратичной формы, минуя вычисление канонического базиса.

Что же касается канонического базиса, то он определен с той же степенью произвола, с какой определена полная ортонормированная система из собственных векторов оператора.

[2] Оператор $\mathcal{H}_1 - \mathcal{H}_2$ симметричен, для него существует ортонормированный базис из собственных векторов, а все собственные значения в силу (104.2) равны нулю.

Теорема 104.3 (о паре квадратичных форм). *Для любой пары квадратичных форм $\mathcal{A}(x, x)$ и $\mathcal{B}(x, x)$ в вещественном пространстве V, одна из которых положительно определена, существует общий базис, в котором обе квадратичные формы имеют канонический вид.*

Доказательство. Пусть $\mathcal{B}(x, x) > 0$ и $\mathcal{B}(x, y)$ – билинейная форма, полярная к квадратичной форме $\mathcal{B}(x, x)$. В соответствии с теоремой 102.5 введем скалярное произведение (102.3). Тогда пространство V становится евклидовым и в нем согласно теореме 104.2 существует ортонормированный базис $e_1, \ldots, e_n$, в котором квадратичная форма $\mathcal{A}(x, x)$ имеет канонический вид (104.3). При этом в силу (70.7) для любого вектора $x = \sum_{i=1}^n x_i e_i$

$$\mathcal{B}(x, x) = (x, x) = \sum_{i=1}^n x_i^2. \ ■ \tag{104.4}$$

Замечание 2. Один из способов поиска общего базиса $e_1, \ldots, e_n$, в котором формы $\mathcal{A}(x, x)$ и $\mathcal{B}(x, x)$ имеют канонический вид (104.3) и (104.4), состоит в следующем.

Пусть A и B – матрицы квадратичных форм $\mathcal{A}(x, x)$ и $\mathcal{B}(x, x)$ в некотором базисе f и пусть $e = fQ$. Тогда $Q^T AQ = \Lambda$, где $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, и $Q^T BQ = I$. Следовательно, $A = (Q^T)^{-1}\Lambda Q^{-1}$, $B = (Q^T)^{-1}Q^{-1}$, $B^{-1} = QQ^T$, $B^{-1}A = Q\Lambda Q^{-1}$ и $B^{-1}AQ = Q\Lambda$. Последнее равенство означает, что столбцы матрицы Q (т.е. координаты векторов искомого канонического базиса e в исходном базисе f) являются собственными векторами матрицы $B^{-1}A$, отвечающими собственным значениям $\lambda_1, \ldots, \lambda_n$. Таким образом, канонические коэффициенты $\lambda_1, \ldots, \lambda_n$ являются корнями уравнения $|B^{-1}A - \lambda I| = 0$ или, в равносильной записи, уравнения

$$|A - \lambda B| = 0, \tag{104.5}$$

а векторы канонического базиса – нетривиальными решениями однородной системы уравнений

$$Ax = \lambda Bx. \tag{104.6}$$

Теорема 104.3 обеспечивает существование полного набора вещественных корней уравнения (104.5) и для каждого кратного корня – наличие соответствующего числа линейно независимых решений системы (104.6). Вообще говоря, описанный алгоритм (как и доказательство теоремы 104.3) решает поставленную задачу в усиленной форме: положительно определенная форма $\mathcal{B}(x, x)$ приводится к каноническому виду с коэффициентами, равными 1. В действительности это не требуется: без существенных изменений приведенный здесь алгоритм применим и для случая, когда $Q^T BQ = \mathrm{diag}(\mu_1, \ldots, \mu_n)$. Таким образом, коэффициенты преобразуемых форм не определяются однозначно. Однако, как легко показать, *отношения λ_i/μ_i канонических коэффициентов не зависят от способа одновременного приведения форм $\mathcal{A}(x, x)$ и $\mathcal{B}(x, x)$ к каноническому виду.*

В вычислительной алгебре задачу (104.6) называют *обобщенной проблемой собственных значений.*

Замечание 3. Теоремы 104.1–104.3 с очевидными изменениями (см. §103) переносятся и на эрмитовы квадратичные формы в комплексном пространстве.

Глава XVIII. Поверхности второго порядка

§105. Гиперповерхности второго порядка в евклидовом пространстве

Общее уравнение. Пусть $\mathcal{A}(x,x)$ – ненулевая квадратичная форма, $g(x)$ – линейная форма, заданные в евклидовом пространстве E, и c – вещественная константа. Множество всех векторов $x \in E$, удовлетворяющих условию

$$\mathcal{A}(x,x)+2g(x)+c=0, \tag{105.1}$$

называется *гиперповерхностью второго порядка в евклидовом пространстве* E[1]. Уравнение (105.1) называется *общим уравнением гиперповерхности второго порядка.*

Пусть $e=(e_1,\dots,e_n)$ – базис пространства E, $A=(a_{ij})\in\mathbb{R}^{n\times n}$ – матрица квадратичной формы $\mathcal{A}(x,x)$ в этом базисе, $b_i=g(e_i)$, $i=\overline{1,n}$, – коэффициенты линейной формы и $x=\sum_{i=1}^n x_ie_i$. Тогда в соответствии с (101.7), (81.1) общее уравнение (105.1) может быть записано в виде

$$\sum_{i,j=1}^n a_{ij}x_ix_j+2\sum_{i=1}^n b_ix_i+c=0,\quad a_{ij}=a_{ji}, \tag{105.2}$$

или, в компактной форме,

$$X^TAX+2b^TX+c=0,\quad A=A^T, \tag{105.3}$$

где $X=(x_1,\dots,x_n)^T$, $b=(b_1,\dots,b_n)^T$.

Приведенные уравнения. Исследование гиперповерхности второго порядка в n-мерном пространстве проводится по той же схеме, по которой исследуются линии второго порядка на плоскости (§58, 59). Отметим основные моменты этого исследования с общих позиций.

Пусть e – ортонормированный базис E и пусть в этом базисе уравнение гиперповерхности имеет вид (105.2).

1. Приведем квадратичную форму $\mathcal{A}(x,x)$ к главным осям, т.е. найдем ортонормированный базис f, в котором квадратичная форма $\mathcal{A}(x,x)$ имеет канонический вид. Отметим, что канонические коэффициенты $\lambda_1,\dots,\lambda_n$ будут при этом определены однозначно (§104,

[1]В случае, когда $\dim E=2$, геометрический образ, определяемый уравнением (105.1), называется линией второго порядка. Однако всюду в дальнейшем мы употребляем общий термин "поверхность", не оговаривая особо этот случай.

замечание 1). Не нарушая общности рассуждений, можно считать, что $\lambda_k\neq0$ при $k\le r$ (где $r=\operatorname{rg}\mathcal{A}(x,x)$) и $\lambda_k=0$ при $k>r$.

При переходе от базиса e к базису $f=eQ$ уравнение (105.2) ортогональным преобразованием координат (теорема 70.5) преобразуется в уравнение

$$\sum_{k=1}^r\lambda_kx_k'^2+2\sum_{k=1}^n b_k'x_k'+c=0. \tag{105.4}$$

2. Пусть $a\in E$. Отображение $\varphi: E\to E$, определенное равенством $\varphi(x)=x+a$, $\forall x\in E$, называется *параллельным переносом пространства E на вектор a*. Очевидно, $\operatorname{im}\varphi=E$.

Если в уравнении (105.4) $\lambda_k\neq0$ для некоторого k, то $\lambda_kx_k'^2+2b_k'x_k'=\lambda_k(x_k'+b_k'/\lambda_k)^2-b_k'^2/\lambda_k$. Положив $x_k''=x_k'+b_k'/\lambda_k$, можно добиться исчезновения переменной x_k'' в первой степени. Выполнение таких преобразований для всех $k=\overline{1,r}$ равносильно параллельному переносу на вектор с координатами $(b_1'/\lambda_1,\dots,b_r'/\lambda_r,0,\dots,0)$. В результате этого переноса уравнение (105.4) преобразуется в уравнение

$$\lambda_1x_1''^2+\ldots+\lambda_rx_r''^2+2\sum_{k=r+1}^n b_k'x_k''+c'=0, \tag{105.5}$$

где $c'=c-\sum_{k=1}^r b_k'^2/\lambda_k$.

3. Исследуем уравнение (105.5). Возможны два случая:

а) $b_{r+1}'=\ldots=b_n'=0$;

б) $\exists b_k'\neq0$, $k>r$.

В случае "а" уравнение (105.5) имеет вид

$$\lambda_1x_1''^2+\ldots+\lambda_rx_r''^2+c'=0.$$

В случае "б" выполним еще одно ортогональное преобразование координат. Матрица S этого преобразования строится следующим образом. Рассмотрим арифметический вектор $(b_{r+1}',\dots,b_n')\in\mathbb{R}^{n-r}$. По условию он ненулевой и его можно нормировать: если $\alpha=\left(\sum_{k=r+1}^n b_k'^2\right)^{1/2}$, то вектор $s_1=(b_{r+1}'/\alpha,\dots,b_n'/\alpha)$ представляет собой ортонормированную систему из одного вектора. Дополним этот вектор до ортонормированного базиса $s_1,s_2,\dots,s_{n-r}$ пространства $\mathbb{R}^{n-r}$. В обозначениях $s_i=(s_{i,r+1},\dots,s_{in})$, $i=\overline{1,n-r}$, искомая матрица S имеет вид

$$S=\begin{bmatrix} 1 & & & & & \\ & \ddots & & & O & \\ & & 1 & & & \\ & & & s_{1,r+1} & \dots & s_{1n} \\ & O & & \dots & \dots & \dots \\ & & & s_{n-r,r+1} & \dots & s_{n-r,n} \end{bmatrix}.$$

Матрица S ортогональна (так как ее строки ортонормированы), поэтому преобразование базиса с помощью матрицы перехода S приводит к новому ортонормированному базису. При этом преобразование координат осуществляется по закону $x_k''' = x_k''$ при $k = \overline{1,r}$, $x_r''' = \alpha^{-1}\sum_{k=r+1}^{n} b_k' x_k''$, $x_k''' = \sum_{i=r+1}^{n} s_{ki} x_i''$ при $k = \overline{r+2,n}$, и приводит уравнение (105.5) к виду $\lambda_1 x_1'''^2 + \ldots + \lambda_r x_r'''^2 + 2\alpha x_{r+1}''' + c' = 0$. В последнем уравнении можно освободиться от свободного члена c', если выполнить параллельный перенос на вектор, у которого все координаты, кроме $(r+1)$-й (равной $c'/(2\alpha)$), равны нулю.

Итак, с помощью параллельного переноса и перехода к новому ортонормированному базису общее уравнение (105.2) приводится к одному из двух типов уравнений:

$$\lambda_1 x_1^2 + \ldots + \lambda_r x_r^2 + a_0 = 0, \quad \lambda_1 \ldots \lambda_r \neq 0, \tag{105.6}$$

$$\lambda_1 x_1^2 + \ldots + \lambda_r x_r^2 + b_0 x_{r+1} = 0, \quad \lambda_1 \ldots \lambda_r b_0 \neq 0, \tag{105.7}$$

которые принято называть ***приведенными уравнениями гиперповерхности.***

Инварианты гиперповерхности. Величины и функции, определяемые коэффициентами общего уравнения (105.2), которые не изменяются при преобразованиях координат, называются ***инвариантами гиперповерхности относительно этих преобразований.***

Аппарат инвариантов (как и в случае линий второго порядка на плоскости) позволяет по общему уравнению однозначно найти тип приведенного уравнения и все его коэффициенты.

Рассмотрим матрицу

$$B = \left[\begin{array}{c|c} A & b \\ \hline b^T & c \end{array}\right], \tag{105.8}$$

полученную окаймлением матрицы A квадратичной формы столбцом и строкой из коэффициентов линейной формы и свободным членом c.

Теорема 105.1. *Характеристические многочлены матриц A и B являются инвариантами гиперповерхности относительно ортогонального преобразования координат.*

Доказательство. Утверждение теоремы следует из того, что соотношения $A_f = Q^T A_e Q$ и $B_f = \tilde{Q}^T B_e \tilde{Q}$ (см. доказательство теоремы 58.2) в случае ортогональности Q означают подобие матриц A_e и A_f, B_e и B_f, а следовательно, и совпадение характеристических многочленов (теорема 86.2). ■

Следствие. Инвариантами гиперповерхности относительно ортогонального преобразования координат являются:

а) канонические коэффициенты $\lambda_1, \ldots, \lambda_n$ (как корни характеристического многочлена A);

б) величины $\operatorname{tr} A$, $\det A$ (как коэффициенты характеристического многочлена);

в) числа $\operatorname{rg} A$, $\operatorname{rg} B$.

Теорема 105.2. *Величины* $\operatorname{tr} A$, $\det A$, $\det B$, $\operatorname{rg} A$, $\operatorname{rg} B$ *являются инвариантами гиперповерхности относительно параллельного переноса.*

Доказательство. Инвариантность величин $\operatorname{tr} A$, $\det A$, $\operatorname{rg} A$ следует из того, что при параллельном переносе матрица A не изменяется, а инвариантность $\det B$, $\operatorname{rg} B$ – из того, что матрица B при параллельном переносе подвергается элементарным преобразованиям (см. доказательство теоремы 58.2). ■

Классификация гиперповерхностей. Классификация осуществляется на основе приведенных уравнений (105.6) и (105.7).

1. Пусть $r = n$. Тогда уравнение (105.2) приводится только к уравнению типа (105.6), которое в этом случае имеет вид

$$\lambda_1 x_1^2 + \ldots + \lambda_n x_n^2 + a_0 = 0, \tag{105.9}$$

при этом $\det A = \lambda_1 \ldots \lambda_n$, $\det B = \lambda_1 \ldots \lambda_n a_0$, $a_0 = \det B / \det A$. Все коэффициенты в (105.9) определены однозначно вследствие инвариантности чисел $\lambda_1, \ldots, \lambda_n$, $\det A$, $\det B$. В силу закона инерции (теорема 102.1) число положительных и число отрицательных коэффициентов $\lambda_1, \ldots, \lambda_n$ определены однозначно, так что уравнение (105.9) однозначно преобразуется в уравнение

$$\frac{x_1^2}{a_1^2} + \ldots + \frac{x_k^2}{a_k^2} - \frac{x_{k+1}^2}{a_{k+1}^2} - \ldots - \frac{x_n^2}{a_n^2} = 1, \tag{105.10}$$

если $\det B \neq 0$; а в случае, когда $\det B = 0$, – в уравнение

$$\frac{x_1^2}{a_1^2} + \ldots + \frac{x_k^2}{a_k^2} - \frac{x_{k+1}^2}{a_{k+1}^2} - \ldots - \frac{x_n^2}{a_n^2} = 0. \tag{105.11}$$

Гиперповерхность, определяемая уравнением (105.10), при $k = n$ называется *эллипсоидом*, при $k = 0$ – ***мнимым эллипсоидом***, при $0 < k < n$ – *гиперболоидом*. Гиперповерхность, определяемая уравнением (105.11), называется *конусом*.

2. Пусть $r = n - 1$. Тогда уравнение (105.2) приводится к уравнению типа (105.6), если $\det B = 0$, и к уравнению (105.7), если $\det B \neq 0$. Эти уравнения однозначно преобразуются в первом случае в уравнение

$$\frac{x_1^2}{a_1^2} + \ldots + \frac{x_k^2}{a_k^2} - \frac{x_{k+1}^2}{a_{k+1}^2} - \ldots - \frac{x_{n-1}^2}{a_{n-1}^2} = c, \tag{105.12}$$

где $c = 1$, если $\operatorname{rg} B = n$, и $c = 0$, если $\operatorname{rg} B = n - 1$; а во втором случае в уравнение

$$\frac{x_1^2}{a_1^2} + \ldots + \frac{x_k^2}{a_k^2} - \frac{x_{k+1}^2}{a_{k+1}^2} - \ldots - \frac{x_{n-1}^2}{a_{n-1}^2} = 2px_n, \quad p > 0. \tag{105.13}$$

Гиперповерхность, определяемая уравнением (105.13), называется ***параболоидом***.

3. Пусть $0 < r < n-1$. Тогда уравнение (105.2) приводится к уравнению типа (105.6), если $\operatorname{rg} B \le r+1$, и к уравнению типа (105.7), если $\operatorname{rg} B = r+2$. Эти уравнения однозначно преобразуются в первом случае в уравнение

$$\frac{x_1^2}{a_1^2} + \ldots + \frac{x_k^2}{a_k^2} - \frac{x_{k+1}^2}{a_{k+1}^2} - \ldots - \frac{x_r^2}{a_r^2} = c, \qquad (105.14)$$

где $c = 1$, если $\operatorname{rg} B = r+1$, и $c = 0$, если $\operatorname{rg} B = r$; во втором случае в уравнение

$$\frac{x_1^2}{a_1^2} + \ldots + \frac{x_k^2}{a_k^2} - \frac{x_{k+1}^2}{a_{k+1}^2} - \ldots - \frac{x_r^2}{a_r^2} = 2px_{r+1}, \quad p > 0. \qquad (105.15)$$

Гиперповерхности, определяемые уравнениями (105.12), (105.14), (105.15), называются ***цилиндрами***. Так как условия, определяющие эти уравнения, взаимно исключают друг друга, то ***общее уравнение (105.2) определяет одну и только одну из перечисленных поверхностей***.

§106. Алгебраические поверхности второго порядка

Общее уравнение. Под общим уравнением алгебраической поверхности второго порядка в системе координат $Oxyz$ пространства понимают уравнение вида

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + \\ +2a_{23}yz + 2b_1x + 2b_2y + 2b_3z + c = 0, \qquad (106.1)$$

где не все коэффициенты a_{ij} $(i,j=\overline{1,3})$ равны нулю, $a_{ij} = a_{ji}$ $(i,j=\overline{1,3})$. Группа членов второго порядка образует здесь квадратичную форму от переменных x, y, z (или, в векторной терминологии, от радиус-вектора точки (x,y,z)), а группа линейных членов – линейную форму, так что уравнение (106.1) является уравнением гиперповерхности второго порядка в евклидовом пространстве V_3. Обозначения, соответствующие общей теории гиперповерхностей, в данном случае имеют вид: $X = (x,y,z)^T$, $b = (b_1, b_2, b_3)^T$,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}, \; B = \begin{bmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{12} & a_{22} & a_{23} & b_2 \\ a_{13} & a_{23} & a_{33} & b_3 \\ b_1 & b_2 & b_3 & c \end{bmatrix}.$$

В этих обозначениях общее уравнение (106.1) может быть записано в компактной форме:

$$X^TAX + 2b^TX + c = 0, \quad A = A^T. \qquad (106.2)$$

Приведенные уравнения. Пусть общее уравнение (106.1) задано в прямоугольной декартовой системе координат. Согласно общей теории гиперповерхностей (§105) уравнение (106.1) с помощью ортогонального преобразования координат (т.е. простым вращением и простым отражением) и параллельного переноса (т.е. переноса начала) приводится к одному и только одному из двух типов приведенных уравнений (105.6) и (105.7). В соответствии с общей схемой классификации гиперповерхностей эти уравнения можно разбить на следующие пять простейших уравнений в зависимости от значения $r = \operatorname{rg} A$.

1. При $r = 3$:

$$\lambda_1x^2 + \lambda_2y^2 + \lambda_3z^2 + a_0 = 0, \quad \lambda_1\lambda_2\lambda_3 \ne 0. \qquad (106.3)$$

2. При $r = 2$:

$$\lambda_1x^2 + \lambda_2y^2 + 2b_0z = 0, \quad \lambda_1\lambda_2b_0 \ne 0 \qquad (106.4)$$

или

$$\lambda_1x^2 + \lambda_2y^2 + c_0 = 0, \quad \lambda_1\lambda_2 \ne 0. \qquad (106.5)$$

3. **При $r = 1$:**

$$\lambda_1x^2 + 2p_0y = 0, \quad \lambda_1p_0 \ne 0 \qquad (106.6)$$

или

$$\lambda_1x^2 + q_0 = 0, \quad \lambda_1 \ne 0. \qquad (106.7)$$

Отметим, что все коэффициенты уравнений (106.3)–(106.7) определены однозначно общими инвариантами гиперповерхностей: $I_1 = \operatorname{tr} A$, $I_3 = \det A$, $K_4 = \det B$ и специальными инвариантами алгебраической поверхности второго порядка: I_2 – сумма главных миноров второго порядка матрицы A и K_2, K_3 – суммы главных миноров второго и третьего порядков матрицы B.

Величины I_2, K_2, K_3 – инварианты ортогонального преобразования координат, так как являются коэффициентами характеристических многочленов матриц A, B (докажите!). Однако при параллельном переносе величина K_3 остается неизменной, только если $I_2 = I_3 = K_4 = 0$, а число K_2 – если $I_2 = I_3 = K_3 = K_4 = 0$. В этом несложно убедиться непосредственно аналогично тому, как это было сделано для линии второго порядка (теорема 58.2).

Для уравнений (106.3)–(106.7) легко вычислить все инварианты, а именно:

(106.3): $I_3 \ne 0$;

(106.4): $I_3 = 0,\ K_4 \neq 0$;
(106.5): $I_3 = 0,\ K_4 = 0,\ I_2 \neq 0$;
(106.6): $I_3 = 0,\ K_4 = 0,\ I_2 = 0,\ K_3 \neq 0$;
(106.7): $I_3 = 0,\ K_4 = 0,\ I_2 = 0,\ K_3 = 0,\ I_1 \neq 0$.

Эти условия необходимы и достаточны для каждого из перечисленных уравнений, так как взаимно исключают друг друга. При этом все коэффициенты уравнений определены однозначно, ибо $\lambda_1, \lambda_2, \lambda_3$ – корни характеристического уравнения: $-\lambda^3 + I_1\lambda^2 - I_2\lambda + I_3 = 0$; $a_0 = \dfrac{K_4}{I_3}$; $b_0 = \left(-\dfrac{K_4}{I_2}\right)^{1/2}$; $c_0 = \dfrac{K_3}{I_2}$; $p_0 = \left(-\dfrac{K_3}{I_1}\right)^{1/2}$; $q_0 = \dfrac{K_2}{I_1}$.

Канонические уравнения.

Теорема 106.1. *Для любой алгебраической поверхности второго порядка существует прямоугольная декартова система координат $Oxyz$, в которой уравнение этой поверхности имеет один из следующих видов:*

1) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$, *эллипсоид;*

2) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = -1$, *мнимый эллипсоид;*

3) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 0$, *вырожденный эллипсоид;*

4) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 1$, *однополостный гиперболоид;*

5) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = -1$, *двуполостный гиперболоид;*

6) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 0$, *конус;*

7) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 2z$, *эллиптический параболоид;*

8) $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 2z$, *гиперболический параболоид;*

9) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$, *эллиптический цилиндр;*

10) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = -1$, *мнимый эллиптический цилиндр;*

11) $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$, *гиперболический цилиндр;*

12) $y^2 = 2px\ (p > 0)$, *параболический цилиндр;*

13) $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 0$, *пара пересекающихся плоскостей;*

14) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 0$, *пара мнимых пересекающихся плоскостей;*

15) $y^2 = a^2\ (a > 0)$, *пара параллельных плоскостей;*

16) $y^2 = -a^2\ (a > 0)$, *пара мнимых параллельных плоскостей;*

17) $y^2 = 0$, *пара совпадающих плоскостей.*

Доказательство. Будем исходить из уравнений (106.3)–(106.7). Рассмотрим уравнение (106.3). В зависимости от сочетаний знаков $\lambda_1, \lambda_2, \lambda_3, a_0$ оно может быть записано по-разному.

1. Пусть знаки чисел $\lambda_1, \lambda_2, \lambda_3$ совпадают. При этом:

а) если $a_0\lambda_1 < 0$, то уравнение (106.3) может быть записано как уравнение 1, где $a^2 = -a_0/\lambda_1$, $b^2 = -a_0/\lambda_2$, $c^2 = -a_0/\lambda_3$;

б) если $a_0\lambda_1 > 0$ – как уравнение 2;

в) если $a_0 = 0$ – как уравнение 3.

2. Пусть знаки чисел $\lambda_1, \lambda_2, \lambda_3$ не совпадают. Не ограничивая общности рассуждений, можно считать, что $\lambda_1 > 0$, $\lambda_2 > 0$, $\lambda_3 < 0$. При этом уравнение (106.3) может быть записано как уравнение 4 (если $a_0\lambda_3 > 0$), как уравнение 5 (если $a_0\lambda_3 < 0$) и как уравнение 6 (если $a_0 = 0$).

Аналогично уравнение (106.4) в зависимости от знаков λ_1, λ_2 может быть записано (после возможных перенумераций неизвестных и изменения направления одной из осей координат) как уравнение 7 (если $\lambda_1\lambda_2 > 0$) и как уравнение 8 (если $\lambda_1\lambda_2 < 0$).

Уравнение (106.5) может быть записано как уравнение 9 (если $\lambda_1\lambda_2 > 0$, $c_0\lambda_1 < 0$), как уравнение 10 (если $\lambda_1\lambda_2 > 0$, $c_0\lambda_1 > 0$), как уравнение 14 (если $\lambda_1\lambda_2 > 0$, $c_0 = 0$), как уравнение 11 (если $\lambda_1\lambda_2 < 0$, $c_0 \neq 0$) и как уравнение 13 (если $\lambda_1\lambda_2 < 0$, $c_0 = 0$).

Уравнение (106.6) после стандартных преобразований может быть записано как уравнение 12.

Уравнение (106.7) может быть записано как уравнение 15 (если $q_0\lambda_1 < 0$), как уравнение 16 (если $q_0\lambda_1 > 0$) и как уравнение 17 (если $q_0 = 0$).

Таким образом, уравнениями 1–17 исчерпываются все поверхности второго порядка. ■

Уравнения 1–17 называются ***каноническими уравнениями алгебраических поверхностей второго порядка.***

Геометрические свойства. Канонические уравнения позволяют исследовать геометрические свойства поверхностей. Чтобы представить форму поверхности, проще всего изучить ее сечения плоскостями, параллельными координатным плоскостям.

Эллипсоид (рис. 1). Поверхность, определяемая уравнением 1, называется ***эллипсоидом.*** Числа a, b, c в каноническом уравнении называются ***полуосями*** эллипсоида. Как следует из уравнения, координатные плоскости являются плоскостями симметрии, а начало координат – центром симметрии эллипсоида. Эллипсоид – ограниченная поверхность, заключенная в параллелепипеде $|x| \le a$, $|y| \le b$, $|z| \le c$. Линии пересечения эллипсоида с любой плоскостью являются поэтому ограниченными линиями второго порядка, т.е. эллипсами.

Мнимый эллипсоид. Нет ни одной точки пространства, координаты которой удовлетворяют уравнению 2. Принято говорить об этом уравнении как об уравнении ***мнимого эллипсоида.***

Вырожденный эллипсоид. Уравнению 3 удовлетворяет только начало координат. Это уравнение принято называть уравнением ***вырожденного эллипсоида*** (или ***мнимого конуса***).

Однополостный гиперболоид (рис. 2). Поверхность, определяемая уравнением 4, называется ***однополостным гиперболоидом.***

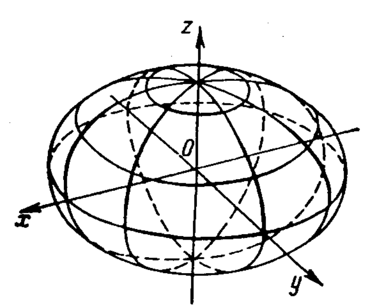


Рис. 1

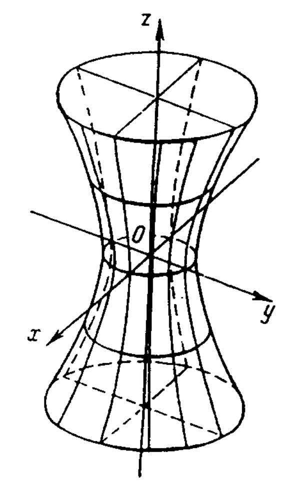


Рис. 2

Как следует из канонического уравнения, координатные плоскости являются плоскостями симметрии, а начало координат – центром симметрии однополостного гиперболоида. Сечения гиперболоида плоскостями $z = h$ представляют собой эллипсы с полуосями $a(1+h^2/c^2)^{1/2}$, $b(1+h^2/c^2)^{1/2}$, которые неограниченно возрастают при $h \to \infty$. Эллипс, получающийся при $h = 0$, называется ***горловым эллипсом*** гиперболоида.

Плоскость $y = h$ пересекает гиперболоид:

а) при $|h| < b$ по гиперболе с полуосями $a(1-h^2/b^2)^{1/2}$, $c(1+h^2/b^2)^{1/2}$, которые убывают от a и c до нуля при $|h| \to b$;

б) при $|h| = b$ по паре пересекающихся прямых;

в) при $|h| > b$ по гиперболе с полуосями $c(-1+h^2/b^2)^{1/2}$, $a(-1+h^2/b^2)^{1/2}$, которые неограниченно возрастают при $|h| \to \infty$.

Сечения гиперболоида плоскостями $x = h$ обладают аналогичными свойствами.

Важной особенностью однополостного гиперболоида является наличие прямых, целиком лежащих на этой поверхности.

Теорема 106.2. *Через каждую точку однополостного гиперболоида проходят две различные прямые, целиком расположенные на этой поверхности.*

Доказательство. Рассмотрим прямые L и L^*, заданные как пересечение плоскостей

$$L: \begin{cases} \alpha\left(\frac{x}{a} - \frac{z}{c}\right) = \beta\left(1 - \frac{y}{b}\right), \\ \beta\left(\frac{x}{a} + \frac{z}{c}\right) = \alpha\left(1 + \frac{y}{b}\right), \\ \alpha^2 + \beta^2 \neq 0 \end{cases} \quad \text{и} \quad L^*: \begin{cases} \gamma\left(\frac{x}{a} - \frac{z}{c}\right) = \delta\left(1 + \frac{y}{b}\right), \\ \delta\left(\frac{x}{a} + \frac{z}{c}\right) = \gamma\left(1 - \frac{y}{b}\right), \\ \gamma^2 + \delta^2 \neq 0. \end{cases}$$

Прямые L и L^* целиком лежат на гиперболоиде (чтобы убедиться в этом, достаточно почленно перемножить уравнения плоскостей). При этом через каждую точку $M_0(x_0, y_0, z_0)$ гиперболоида проходит единственная прямая из семейства L и единственная прямая из семейства L^*: эти прямые (т.е. пары чисел α, β и γ, δ) находятся из однородных систем уравнений

$$\begin{cases} \alpha\left(\frac{x_0}{a} - \frac{z_0}{c}\right) = \beta\left(1 - \frac{y_0}{b}\right), \\ \beta\left(\frac{x_0}{a} + \frac{z_0}{c}\right) = \alpha\left(1 + \frac{y_0}{b}\right) \end{cases} \quad \text{и} \quad \begin{cases} \gamma\left(\frac{x_0}{a} - \frac{z_0}{c}\right) = \delta\left(1 + \frac{y_0}{b}\right), \\ \delta\left(\frac{x_0}{a} + \frac{z_0}{c}\right) = \gamma\left(1 - \frac{y_0}{b}\right), \end{cases}$$

матрицы которых вырождены (т.е. системы имеют нетривиальные решения) и имеют ранг, равный 1 (т.е. все решения каждой из систем пропорциональны и определяют единственную прямую). Остается добавить, что прямые L и L^* не совпадают (достаточно проверить неколлинеарность их направляющих векторов). ■

Прямые, все точки которых лежат на поверхности, называются ***прямолинейными образующими*** этой поверхности. Итак, однополостный гиперболоид покрыт двумя различными семействами прямолинейных образующих (рис. 3).

Двуполостный гиперболоид (рис. 4). Поверхность, определяемая уравнением 5, называется ***двуполостным гиперболоидом.*** Как следует из канонического уравнения, координатные плоскости являются плоскостями симметрии, а начало координат – центром симметрии двуполостного гиперболоида.

Плоскости $z = h$ пересекают этот гиперболоид:

а) при $|h| > c$ по эллипсам, размеры которых неограниченно возрастают при $|h| \to \infty$;

б) при $|h| = c$ в единственной точке.

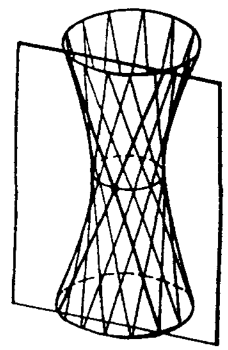


Рис. 3

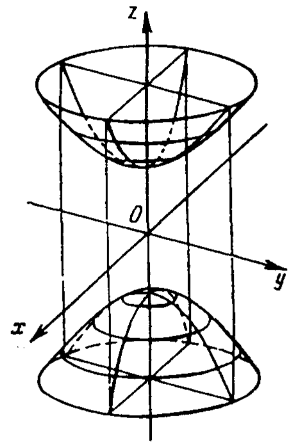


Рис. 4

В слое между плоскостями $z = c$ и $z = -c$ нет ни одной точки гиперболоида. Таким образом, двуполостный гиперболоид состоит из двух симметричных полостей, расположенных в полупространствах $z \geq c$ и $z \leq -c$.

Плоскости $x = h$ и $y = h$ пересекают гиперболоид по гиперболам.

К о н у с (рис. 5). Поверхность, определяемая уравнением 6, называется *конусом*. Как следует из канонического уравнения, координатные плоскости являются плоскостями симметрии, а начало координат – центром симметрии конуса.

Плоскости $z = h$, $h \neq 0$, пересекают конус по эллипсам, размеры которых неограниченно возрастают при $|h| \to \infty$. Прямая, проходящая через любую точку $M_0(x_0, y_0, z_0)$ конуса и начало координат, является прямолинейной образующей конуса, так как любая точка этой прямой имеет координаты (x_0t, y_0t, z_0t).

Плоскости $y = h$ и $x = h$, где $h \neq 0$, пересекают конус по гиперболам (рис. 6) с полуосями $c|h|/b$, $a|h|/b$ и $c|h|/a$, $b|h|/a$ соответственно; плоскости $x = 0$ и $y = 0$ – по парам пересекающихся прямых.

Плоскими сечениями конуса являются и параболы (рис. 7). Так, параболой будет сечение конуса плоскостью $z = h + cx/a$, где $h \neq 0$, ибо числа x, y являются на этой плоскости аффинными (не прямоугольными) координатами, а уравнение линии, высекаемой на ней

конусом, имеет в этих координатах вид

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{(h + cx/a)^2}{c^2} = 0.$$

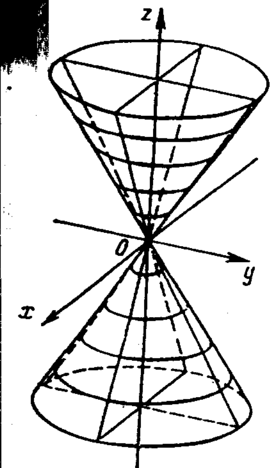


Рис. 5

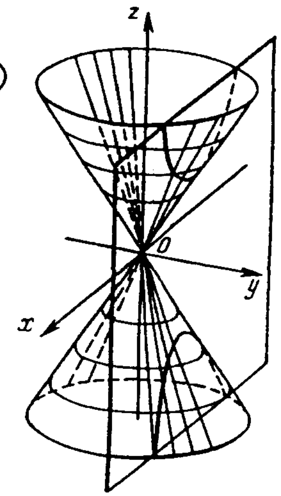


Рис. 6

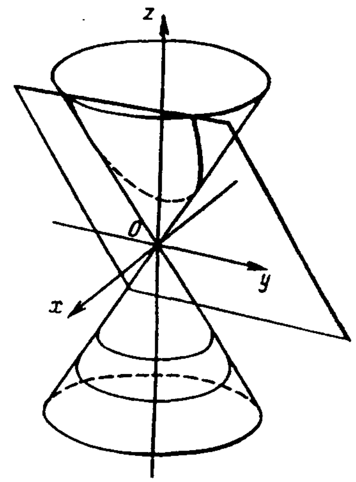


Рис. 7

Традиционными преобразованиями это уравнение приводится к уравнению параболы

$$y = 2\frac{hb^2}{ac}\left(x + \frac{ha}{2c}\right).$$

Таким образом, и эллипс, и гипербола, и парабола являются плоскими сечениями конуса. На этом основании эти линии обычно называют *коническими сечениями*.

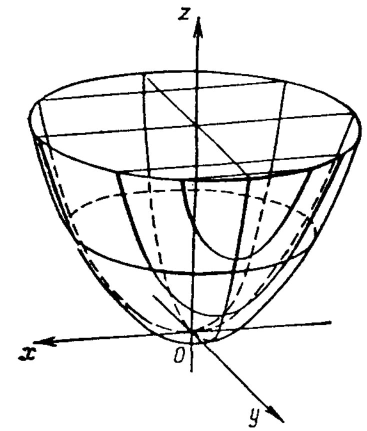


Рис. 8

Эллиптический параболоид (рис. 8). Поверхность, определяемая уравнением 7, называется *эллиптическим параболоидом*. Как следует из канонического уравнения, плоскости $x = 0$ и $y = 0$ являются плоскостями симметрии. Центра симметрии нет. Плоскости $z = h$, $h > 0$, пересекают параболоид по эллипсам, размеры которых неограниченно возрастают при $h \to \infty$. Плоскости $y = h$ и $x = h$ пересекают параболоид по параболам, ветви которых направлены вверх.

Гиперболический параболоид (рис. 9). Поверхность, определяемая уравнением 8, называется *гиперболическим параболоидом*. Плоскости $x = 0$ и $y = 0$ являются плоскостями симметрии. Центра симметрии нет. Плоскости $y = h$ и $x = h$ пересекают параболоид по параболам, ветви первой параболы направлены вверх, а второй - вниз.

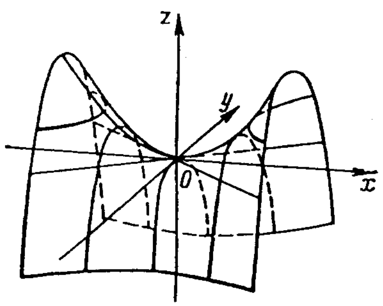


Рис. 9

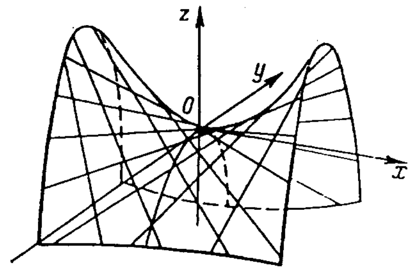


Рис. 10

Плоскости $z = h$ пересекают параболоид:

а) при $h < 0$ по гиперболам $-\frac{x^2}{a_1^2} + \frac{y^2}{b_1^2} = 1$, где $a_1^2 = -2a^2h$, $b_1^2 = -2b^2h$;

б) при $h > 0$ по гиперболам $\frac{x^2}{a_2^2} - \frac{y^2}{b_2^2} = 1$, где $a_2^2 = 2a^2h$, $b_2^2 = 2b^2h$;

в) при $h = 0$ по паре пересекающихся прямых.

Как и однополостный гиперболоид, гиперболический параболоид покрыт двумя различными семействами прямолинейных образующих (рис. 10). Те же рассуждения, что и в случае однополостного гиперболоида, показывают, что *через любую точку $M_0(x_0, y_0, z_0)$ гиперболического параболоида проходят две различные прямые, целиком лежащие на этой поверхности*. Эти прямые определяются плоскостями

$$L : \begin{cases} \alpha z = \beta\left(\frac{x}{a} + \frac{y}{b}\right), \\ 2\beta = \alpha\left(\frac{x}{a} - \frac{y}{b}\right), \\ \alpha^2 + \beta^2 \neq 0 \end{cases} \quad \text{и} \quad L^* : \begin{cases} \gamma z = \delta\left(\frac{x}{a} - \frac{y}{b}\right), \\ 2\delta = \gamma\left(\frac{x}{a} + \frac{y}{b}\right), \\ \gamma^2 + \delta^2 \neq 0, \end{cases}$$

где коэффициенты α, β, γ, δ находятся из условия, что эти прямые проходят через точку $M_0(x_0, y_0, z_0)$.

Цилиндрические поверхности. Все остальные поверхности называются цилиндрами. Поверхности, определяемые уравнениями 9, 11 и 12, называются соответственно *эллиптическим цилиндром* (рис. 11), *гиперболическим цилиндром* (рис. 12) и *параболическим цилиндром* (рис. 13). Уравнения этих поверхностей не зависят от z, поэтому все сечения плоскостями $z = h$ совпадают. Для цилиндрических поверхностей достаточно найти сечение плоскостью $z = 0$, чтобы выяснить форму поверхности. Отметим, наконец, что прямая, проходящая через точку $M_0(x_0, y_0, z_0)$ поверхности и параллельная оси Oz, является прямолинейной образующей, так как любая точка этой прямой имеет координаты $(x_0, y_0, z_0 + t)$.

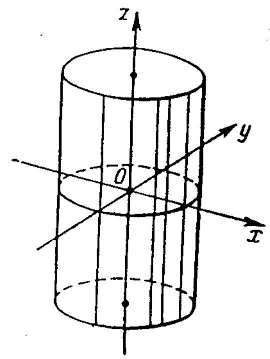


Рис. 11

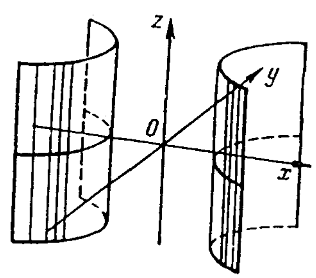


Рис. 12

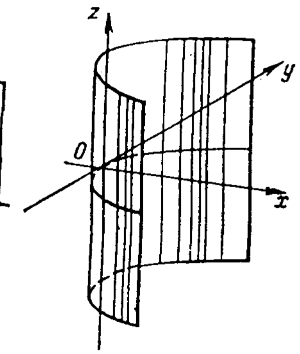


Рис. 13

Нет ни одной точки пространства, координаты которой удовлетворяют уравнению 10. Принято называть его уравнением *мнимого эллиптического цилиндра*. Уравнение 13 определяет *пару пересекающихся плоскостей* (рис. 14).

Точки поверхности, определяемой уравнением 14, составляют прямую. Это уравнение называют уравнением *пары мнимых пересекающихся плоскостей* (рис. 15).

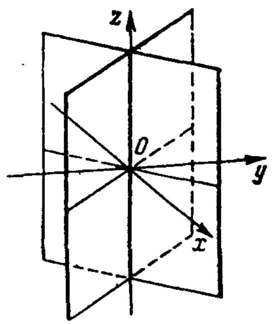


Рис. 14

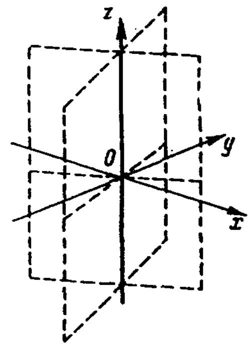


Рис. 15

Поверхности, определяемые уравнениями 15–17, называют соответственно ***парой параллельных плоскостей*** (рис. 16), ***парой мнимых параллельных плоскостей*** и ***парой совпадающих плоскостей*** (рис. 17).

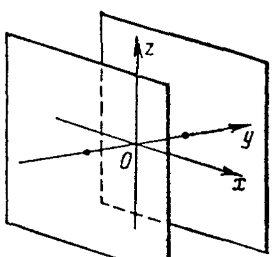


Рис. 16

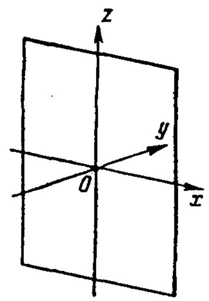


Рис. 17

Глава XIX. Линейные нормированные пространства

§107. Норма вектора

Пусть V – линейное пространство, вещественное или комплексное. ***Нормой*** в линейном пространстве V называется отображение $\|\cdot\| : V \to \mathbb{R}$, ставящее в соответствие каждому вектору $x \in V$ число $\|x\| \in \mathbb{R}$ и удовлетворяющее аксиомам: $\forall x, y \in V$, $\alpha \in \mathbb{R}(\mathbb{C})$

1) $\|x\| \geq 0$,
   $\|x\| = 0$ тогда и только тогда, когда $x = \theta$;
2) $\|\alpha x\| = |\alpha| \cdot \|x\|$;
3) $\|x + y\| \leq \|x\| + \|y\|$ (неравенство треугольника).

Линейное пространство V с заданной на нем нормой $\|\cdot\|$ называется ***линейным нормированным пространством***. Число $\|x\|$ называется ***нормой вектора*** x.

Примеры. 1. В евклидовом (унитарном) пространстве норма может быть введена как длина, т.е. $\|x\| = |x|$. Эта норма называется ***евклидовой нормой*** и обозначается символом $\|x\|_E$. Итак,

$$\|x\|_E = \sqrt{(x, x)}. \qquad (107.1)$$

Справедливость аксиом нормы вытекает из свойств длины (§ 69).

2. В арифметических пространствах $\mathbb{R}^n$, $\mathbb{C}^n$ обычно рассматривают следующие нормы вектора $x = (x_1, \ldots, x_n)$:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|; \quad \|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}; \quad \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|,$$

которые можно записать единообразно в виде

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \text{ где } p = 1, 2, \infty$$

(здесь $\|x\|_\infty = \lim\limits_{p \to \infty} \|x\|_p$). Отметим, что $\|\cdot\|_2$ совпадает с $\|\cdot\|_E$ для естественного скалярного произведения (70.7).

3. Аналогичные нормы рассматриваются в произвольном конечномерном линейном пространстве V (вещественном или комплексном): если $e_1, \ldots, e_n$ – базис V и $x = \sum_{i=1}^{n} x_i e_i$, то

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \quad \text{где } p = 1, 2, \infty. \qquad (107.2)$$

4. В пространствах матриц $\mathbb{R}^{m\times n}$ и $\mathbb{C}^{m\times n}$ рассматривают следующие нормы матрицы $A=(a_{ij})$:

$$\|A\|_1 = \max_{1\le j\le n}\sum_{i=1}^{m}|a_{ij}|; \quad \|A\|_\infty = \max_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}|;$$
$$\|A\|_E = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}.$$

Сходимость по норме.

Утверждение 1. *В нормированном пространстве V отображение $\rho: V\times V\to\mathbb{R}$, определенное равенством*
$$\rho(x,y)=\|x-y\|, \quad \forall x,y\in V,$$
является метрикой.

Аксиомы метрики непосредственно вытекают из аксиом нормы. ■

Таким образом, в нормированном пространстве можно ввести расстояние между векторами и, значит, пользоваться предельным переходом. Последовательность векторов $\{x^{(k)}\}$ в нормированном пространстве V называется *сходящейся по норме* к вектору $a\in V$, если $\lim\limits_{k\to\infty}\|x^{(k)}-a\|=0$, при этом вектор a называется *пределом последовательности* $\{x^{(k)}\}$ *по норме* $\|\cdot\|$.

Обозначение: $\lim\limits_{k\to\infty}x^{(k)}=a$ или $x^{(k)}\to a$.

Утверждение 2. *Сходящаяся по норме последовательность имеет единственный предел.*

Доказательство повторяет доказательство аналогичной теоремы для числовой последовательности [7] и основано на аксиоме треугольника: $\|a-b\|=\|a-x^{(k)}+x^{(k)}-b\|\le\|x^{(k)}-a\|+\|x^{(k)}-b\|$, где a и b – два предела последовательности $x^{(k)}$. ■

Пусть $x_0\in V$ и $r>0$. Множество $S(x_0,r)=\{x\in V\,|\,\|x-x_0\|=r\}$ называется *сферой радиуса r с центром x_0 по норме* $\|\cdot\|$, а множество $B(x_0,r)=\{x\in V\,|\,\|x-x_0\|\le r\}$ – замкнутым *шаром радиуса r с центром x_0 по норме* $\|\cdot\|$.

На рис. 1 точками плоскости изображены сферы радиуса единица с центром $x_0=(0,0)$ по нормам $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$ в арифметическом пространстве $\mathbb{R}^2$.

В дальнейшем сферы и шары по евклидовой норме будем обозначать символами $B_E(x_0,r)$ и $S_E(x_0,r)$.

Утверждение 3. *Из любой последовательности векторов $x^{(k)}\in B_E(x_0,r)$ (или $S_E(x_0,r)$) можно выделить подпоследовательность, сходящуюся по норме $\|\cdot\|_E$ к вектору $a\in B_E(x_0,r)$ ($S_E(x_0,r)$ соответственно).*

Доказательство. Без ограничения общности можно считать, что $x_0=\theta$. Доказательство проведем для сферы $S_E(r)=\{x\in$

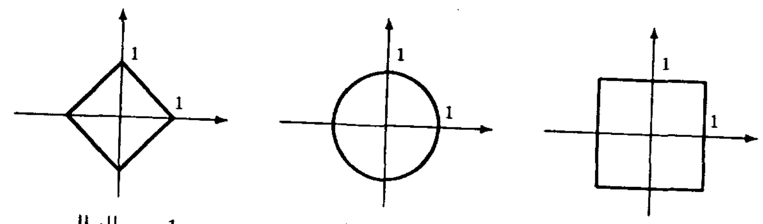


Рис.1

$V\,|\,\|x\|_E=r\}$. Пусть $e_1,\dots,e_n$ – ортонормированный базис пространства V и $x^{(k)}=\sum_{i=1}^n x_i^{(k)}e_i\in S_E(r)$. Тогда
$$\|x^{(k)}\|_E=\left(\sum_{i=1}^n|x_i^{(k)}|^2\right)^{1/2}=r.$$

Это означает ограниченность координат векторов $x^{(k)}$ рассматриваемой последовательности. Согласно теореме Больцано–Вейерштрасса [7] из этой последовательности можно выделить сходящуюся (покоординатно) подпоследовательность $\{x^{(k_m)}\}$. Пусть $x^{(k_m)}$ имеет координаты $x_1^{(k_m)},\dots,x_n^{(k_m)}$, сходящиеся соответственно к $a_1,\dots,a_n$. Положим $a=\sum_{i=1}^n a_ie_i$. Тогда
$$\|x^{(k_m)}-a\|_E=\left(\sum_{i=1}^n\left|x_i^{(k_m)}-a_i\right|^2\right)^{1/2}\to 0,$$
следовательно, подпоследовательность $\{x^{(k_m)}\}$ сходится к вектору a по евклидовой норме.

Покажем, что $a\in S_E(r)$. Действительно, в очевидном неравенстве $\big|\|x\|-\|y\|\big|\le\|x-y\|$ положим $x=x^{(k_m)}$, $y=a$. Тогда $\big|\|x^{(k_m)}\|_E-\|a\|_E\big|\le\|x^{(k_m)}-a\|_E$, откуда следует, что
$$\|x^{(k_m)}\|_E-\|x^{(k_m)}-a\|_E\le\|a\|_E\le\|x^{(k_m)}\|_E+\|x^{(k_m)}-a\|_E$$
или, с учетом того, что $\|x^{(k_m)}-a\|_E\to 0$,
$$\|x^{(k_m)}\|_E-\varepsilon\le\|a\|_E\le\|x^{(k_m)}\|_E+\varepsilon, \quad \forall\varepsilon>0,$$
если m достаточно велико. Следовательно, $\|a\|_E=r$ и $a\in S_E(r)$. ■

Возникает естественный вопрос: может ли одна и та же последовательность быть сходящейся по одной норме и не сходящейся по другой? Ответим на этот вопрос для конечномерных пространств.

Эквивалентность норм в конечномерном пространстве. Две нормы $\|\cdot\|_1$ и $\|\cdot\|_2$ в линейном пространстве V называются *эквивалентными*, если существуют такие числа $c_1>0$, $c_2>0$, что для любого вектора $x\in V$ выполняются неравенства
$$\|x\|_1\le c_1\|x\|_2 \quad\text{и}\quad \|x\|_2\le c_2\|x\|_1.$$

Теорема 107.1. *В конечномерном пространстве любые две* *нормы эквивалентны.*

Доказательство. Выберем в пространстве V произвольный базис $e = (e_1, \ldots, e_n)$ и введем скалярное произведение согласно (70.7). Тогда V – евклидово (унитарное) пространство и в нем можно рассматривать евклидову норму (107.1). Так как в этом скалярном произведении e – ортонормированный базис, то $\|x\|_E = |x| = \sqrt{|x_1|^2 + \ldots + |x_n|^2}$. Покажем, что любая норма $\|\cdot\|$ в линейном пространстве V эквивалентна евклидовой норме $\|\cdot\|_E$. Тем самым теорема будет доказана.

1. Докажем сначала существование числа $c_1 > 0$ такого, что для любого вектора $x \in V$

$$\|x\| \le c_1 \|x\|_E. \tag{107.3}$$

Пусть $x \in V$ и $x = \sum_{i=1}^n x_i e_i$. Тогда $\|x\| \le \sum_{i=1}^n |x_i|\,\|e_i\|$ и в силу неравенства Коши–Буняковского

$$\|x\| \le \left(\sum_{i=1}^n |x_i|^2 \sum_{i=1}^n \|e_i\|^2\right)^{1/2} = c_1\|x\|_E, \text{ где } c_1 = \left(\sum_{i=1}^n \|e_i\|^2\right)^{1/2} > 0.$$

2. Докажем теперь существование числа $c_2 > 0$ такого, что для любого вектора $x \in V$

$$\|x\|_E \le c_2\|x\|. \tag{107.4}$$

а) Рассмотрим единичную сферу по норме $\|\cdot\|$: $S = \{x \in V \mid \|x\| = 1\}$ и покажем, что множество координат всех векторов сферы S ограничено. Пусть это не так. Тогда для любого $m \in \mathbb{N}$ существует вектор $x^{(m)}$ такой, что $\|x^{(m)}\| = 1$ и $\left|x_k^{(m)}\right| > m$ для некоторого $k \in \{1, \ldots, n\}$. Следовательно,

$$\|x^{(m)}\|_E = \left(\sum_{i=1}^n \left|x_i^{(m)}\right|^2\right)^{1/2} \ge \left|x_k^{(m)}\right| > m. \tag{107.5}$$

Рассмотрим последовательность векторов

$$y^{(m)} = x^{(m)} \Big/ \|x^{(m)}\|_E\,. \tag{107.6}$$

Очевидно, что $\|y^{(m)}\|_E = 1$ и, значит, $y^{(m)} \in S_E(1)$. В силу утверждения 3 из последовательности $\{y^{(m)}\}$ можно выделить подпоследовательность $\{y^{(m_j)}\}$, сходящуюся по $\|\cdot\|_E$ к вектору $a \in S_E(1)$, т.е.

$$\|y^{(m_j)} - a\|_E \to 0, \tag{107.7}$$

$$\|a\|_E = 1. \tag{107.8}$$

Из (107.7) с учетом (107.3) следует, что

$$\|y^{(m_j)} - a\| \to 0. \tag{107.9}$$

Таким образом, последовательность $\{y^{(m_j)}\}$ сходится к вектору a по норме $\|\cdot\|$.

С другой стороны, в силу (107.6), (107.5) с учетом $\|x^{(m_j)}\| = 1$

$$\|y^{(m_j)}\| = \|x^{(m_j)}\| / \|x^{(m_j)}\|_E < 1/m_j.$$

Следовательно, $\|y^{(m_j)}\| \to 0$. Отсюда и из (107.9) в силу единственности предела вытекает, что $a = \theta$, что противоречит (107.8).

Таким образом, существует число M такое, что для всех $x = \sum_{i=1}^n x_i e_i \in S$ выполняется неравенство

$$|x_i| \le M, \quad i = \overline{1,n}. \tag{107.10}$$

б) Пусть $x = \sum_{i=1}^n x_i e_i$ – любой вектор пространства V, тогда вектор $x/\|x\| \in S$ и для его координат имеет место неравенство (107.10). Значит, $|x_i|/\|x\| \le M$, т.е. $|x_i| \le M\|x\|$, $i = \overline{1,n}$, и $\|x\|_E = \sqrt{|x_1|^2 + \ldots + |x_n|^2} \le M\sqrt{n}\,\|x\| = c_2\|x\|$, где $c_2 = M\sqrt{n} > 0$. ■

Следствие. *В конечномерном пространстве из сходимости по одной норме следует сходимость по любой другой норме*, так как

$$\|x^{(k)} - a\|_1 \le c_1\|x^{(k)} - a\|_2.$$

§108. Линейные операторы в нормированных пространствах

Пусть V и W – линейные нормированные пространства (оба вещественные или оба комплексные). Линейное пространство $\mathcal{L}(V, W)$, как и любое линейное пространство, можно сделать нормированным, введя на нем норму оператора. Однако для нормы оператора нет той свободы выбора, которая имеет место для векторных норм, так как оператор тесно связан с векторами и, следовательно, его норма тоже должна быть связана с векторными нормами пространств V и W.

Согласованные нормы. Норма оператора $\mathcal{A} \in \mathcal{L}(V, W)$ называется *согласованной* с векторными нормами $\|\cdot\|_V$, $\|\cdot\|_W$ пространств V и W, если для любого вектора $x \in V$

$$\|\mathcal{A}x\|_W \le \|\mathcal{A}\| \cdot \|x\|_V.$$

Замечание 1. В дальнейшем в символе $\|\cdot\|_V$ индекс V будет опускаться, если из контекста ясно, о каком пространстве идет речь.

Теорема 108.1. *Собственное значение линейного оператора $\mathcal{A} \in \mathcal{L}(V, V)$ не превосходит по абсолютной величине любую его согласованную норму.*

Доказательство. Если $\mathcal{A}x = \lambda x$, то для любой согласованной нормы оператора имеем $\|\mathcal{A}x\| = |\lambda| \cdot \|x\|$ и $\|\mathcal{A}x\| \le \|\mathcal{A}\| \cdot \|x\|$, откуда следует, что $|\lambda| \le \|\mathcal{A}\|$. ■

Ограниченный оператор. Линейный оператор $\mathcal{A} \in \mathcal{L}(V, W)$ называется *ограниченным*, если единичный шар в V он переводит в ограниченное по норме пространства W множество, т.е. если существует число $c > 0$ такое, что для всех векторов $x \in V$, для которых $\|x\| = 1$, выполняется неравенство $\|\mathcal{A}x\| \le c$, или, в равносильной формулировке, если существует число $c > 0$ такое, что $\|\mathcal{A}x\| \le c\|x\|$ для всех векторов $x \in V$.

Теорема 108.2. *В конечномерных нормированных пространствах V и W любой линейный оператор $\mathcal{A} \in \mathcal{L}(V, W)$ ограничен.*

Доказательство. Пусть $e_1, \ldots, e_n$ – базис пространства V и $x = \sum_{i=1}^n x_i e_i$. Тогда согласно аксиомам нормы и неравенству Коши–Буняковского

$$\|\mathcal{A}x\| \le \sum_{i=1}^n |x_i| \cdot \|\mathcal{A}e_i\| \le \left(\sum_{i=1}^n \|\mathcal{A}e_i\|^2\right)^{1/2} \left(\sum_{i=1}^n |x_i|^2\right)^{1/2} = M\|x\|_E,$$

где $M = \left(\sum_{i=1}^n \|\mathcal{A}e_i\|^2\right)^{1/2}$. Отсюда с учетом (107.4) получим, что $\|\mathcal{A}x\| \le c\|x\|$, где $c = Mc_2 > 0$. ■

Подчиненная норма. Пусть V, W – конечномерные пространства и $\mathcal{A} \in \mathcal{L}(V, W)$. Из теоремы 108.2 следует ограниченность оператора $\mathcal{A}$, т.е. существование числа $c > 0$ такого, что $\|\mathcal{A}x\| \le c\|x\|$, $\forall x \in V$ или, в равносильной формулировке, $\|\mathcal{A}x\| / \|x\| \le c$, $\forall x \ne \theta$.

Это означает, что числовое множество

$$\left\{ \frac{\|\mathcal{A}x\|}{\|x\|} \;\middle|\; x \ne \theta \right\}$$

ограничено сверху и, следовательно, для него существует точная верхняя грань. Положим

$$\mu(\mathcal{A}) = \sup_{x \ne \theta} \frac{\|\mathcal{A}x\|}{\|x\|}. \tag{108.1}$$

Теорема 108.3. *Отображение $\mu(\mathcal{A})$ является нормой в пространстве $\mathcal{L}(V, W)$.*

Доказательство. Очевидно, что $\mu(\mathcal{A}) \ge 0$ для любого оператора $\mathcal{A} \in \mathcal{L}(V, W)$, при этом равенство $\mu(\mathcal{A}) = 0$ означает, что $\|\mathcal{A}x\| = 0$, $\forall x \in V$, т.е. $\mathcal{A}x = \theta$, $\forall x \in V$ или $\mathcal{A} = \mathcal{O}$. Аксиомы 2 и 3 вытекают из свойств точной верхней грани. ■

Норма $\mu(\mathcal{A})$ называется *нормой* оператора $\mathcal{A}$, *подчиненной (порожденной)* векторным нормам пространств V и W.

Обозначение: $\|\mathcal{A}\|$. Итак,

$$\|\mathcal{A}\| = \sup_{x \ne \theta} \frac{\|\mathcal{A}x\|}{\|x\|} = \sup_{\|x\|=1} \|\mathcal{A}x\|.$$

Свойства подчиненной нормы. 1. Подчиненная норма обладает свойством согласованности: $\|\mathcal{A}x\| \le \|\mathcal{A}\| \cdot \|x\|$, $\forall x \in V$, так как, согласно (108.1), $\|\mathcal{A}x\|/\|x\| \le \mu(\mathcal{A}) = \|\mathcal{A}\|$.

2. Подчиненная норма – наименьшая из всех согласованных норм, так как является точной верхней гранью множества, тогда как любая согласованная норма – верхней гранью этого множества.

3. Подчиненная норма обладает *свойством мультипликативности*, т.е. $\|\mathcal{AB}\| \le \|\mathcal{A}\| \cdot \|\mathcal{B}\|$, так как $\|\mathcal{AB}\| = \sup_{\|x\|=1} \|\mathcal{AB}x\| \le \sup_{\|x\|=1} \|\mathcal{A}\| \cdot \|\mathcal{B}x\| = \|\mathcal{A}\| \sup_{\|x\|=1} \|\mathcal{B}x\| = \|\mathcal{A}\| \cdot \|\mathcal{B}\|$.

Замечание 2. Подчиненная норма определяется не только самим оператором, но и векторными нормами пространств V и W. Меняя векторные нормы, можно изменить и норму оператора. Рассмотрим наиболее часто употребляемые нормы.

Спектральная норма. Пусть V, W – евклидовы (унитарные) пространства. Норма линейного оператора $\mathcal{A} \in \mathcal{L}(V, W)$, порожденная евклидовыми нормами вектора, называется *спектральной нормой*. Обозначение: $\|\mathcal{A}\|_2$. Итак,

$$\|\mathcal{A}\|_2 = \sup_{\|x\|_E=1} \|\mathcal{A}x\|_E = \sup_{(x,x)=1} \sqrt{(\mathcal{A}x, \mathcal{A}x)}.$$

Теорема 108.4. *Спектральная норма оператора равна максимальному сингулярному числу этого оператора.*

Доказательство. Пусть $e_1, \ldots, e_n$ – ортонормированный базис из собственных векторов оператора $\mathcal{A}^*\mathcal{A}$, а $\rho_1, \ldots, \rho_n$ – сингулярные числа оператора $\mathcal{A}$ (§100), $x = \sum_{i=1}^n x_i e_i$, $\rho_k = \max_{1 \le i \le n} \rho_i$. Тогда $\|\mathcal{A}x\|_E^2 = (\mathcal{A}x, \mathcal{A}x) = (\mathcal{A}^*\mathcal{A}x, x) = \left(\sum_{i=1}^n \rho_i^2 x_i e_i, \sum_{j=1}^n x_j e_j\right) = \sum_{i=1}^n \rho_i^2 |x_i|^2 \le \rho_k^2 \sum_{i=1}^n |x_i|^2$. Отсюда следует, что $\|\mathcal{A}x\|_E \le \rho_k$, если $\|x\|_E = 1$, и $\|\mathcal{A}x\|_E = \rho_k$, если $x = e_k$ (очевидно, $\|e_k\|_E = 1$). Значит, $\rho_k = \max_{\|x\|_E=1} \|\mathcal{A}x\|_E = \sup_{\|x\|_E=1} \|\mathcal{A}x\|_E = \|\mathcal{A}\|_2$. ■

Следствие 1. *Спектральная норма нормального оператора равна абсолютному значению максимального по модулю собственного значения этого оператора.*

Теорема 108.5. *Сингулярные числа линейного оператора в евклидовом (унитарном) пространстве не изменяются при умножении оператора на ортогональный (унитарный) оператор.*

Доказательство. Пусть $\mathcal{B} = \mathcal{UAV}$, где $\mathcal{U}^*\mathcal{U} = \mathcal{I}$, $\mathcal{V}^*\mathcal{V} = \mathcal{I}$. Тогда $\mathcal{B}^*\mathcal{B} = \mathcal{V}^*\mathcal{A}^*\mathcal{AV}$, следовательно, матрицы операторов $\mathcal{B}^*\mathcal{B}$ и $\mathcal{A}^*\mathcal{A}$ подобны и их собственные значения совпадают. ■

Следствие 2. *Спектральная норма линейного оператора не изменяется при умножении оператора на ортогональный (унитарный) оператор.*

§109. Матричные нормы оператора

Спектральная норма линейного оператора является, по существу, единственной подчиненной нормой, не связанной явно с базисами пространств V и W. Если же в пространствах V и W зафиксированы базисы, то возможность введения операторных норм существенно расширяется.

Пусть $e = (e_1, \ldots, e_n)$ и $f = (f_1, \ldots, f_m)$ – базисы пространств V и W. Введем в V и W векторную норму $\|\cdot\|_p$ согласно (107.2). Будем считать, что в пространствах V и W введены нормы одинакового типа. Обозначим через $\|\mathcal{A}\|_p$ норму оператора, подчиненную векторным нормам $\|\cdot\|_p$, через $A = (a_{ij})$ – матрицу оператора $\mathcal{A}$ в базисах e и f.

Теорема 109.1. *Для любого оператора $\mathcal{A} \in \mathcal{L}(V, W)$*

$$\|\mathcal{A}\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|.$$

Доказательство. Пусть $x = \sum_{j=1}^{n} x_j e_j$, тогда

$$\mathcal{A}x = \sum_{j=1}^{n} x_j \mathcal{A} e_j = \sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} f_i = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij} x_j \Big) f_i.$$

Согласно (107.2)

$$\|\mathcal{A}x\|_1 = \sum_{i=1}^{m} \Big| \sum_{j=1}^{n} a_{ij} x_j \Big| \le \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}||x_j| = \sum_{j=1}^{n} |x_j| \sum_{i=1}^{m} |a_{ij}|. \qquad (109.1)$$

Пусть k-й столбец матрицы A имеет максимальную столбцовую сумму, т.е. $\sum_{i=1}^{m} |a_{ik}| = \max\limits_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|$. Тогда из (109.1) следует, что $\|\mathcal{A}x\|_1 \le \sum_{j=1}^{n} |x_j| \sum_{i=1}^{m} |a_{ik}| = \|x\|_1 \sum_{i=1}^{m} |a_{ik}|$. Это означает, что $\|\mathcal{A}x\|_1 \le \sum_{i=1}^{m} |a_{ik}|$ для любого вектора x, у которого $\|x\|_1 = 1$, и $\|\mathcal{A}x\|_1 = \sum_{i=1}^{m} |a_{ik}|$ для вектора $x = e_k$ (очевидно, $\|e_k\|_1 = 1$). Следовательно, $\sum_{i=1}^{m} |a_{ik}| = \sup\limits_{\|x\|_1 = 1} \|\mathcal{A}x\|_1$. ■

Теорема 109.2. *Для любого оператора $\mathcal{A} \in \mathcal{L}(V, W)$*

$$\|\mathcal{A}\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|.$$

Доказательство аналогично доказательству теоремы 109.1. Оно предоставляется читателю. ■

Замечание. Векторные нормы $\|\cdot\|_2$ пространств V и W порождают спектральную норму оператора $\|\mathcal{A}\|_2$, так как норма $\|\cdot\|_2$ совпадает с евклидовой нормой $\|\cdot\|_E$, если в пространствах V и W ввести скалярное произведение так, чтобы базисы e и f стали ортонормированными.

Евклидова норма оператора. Можно строить норму линейного оператора и как норму вектора в линейном пространстве $\mathcal{L}(V, W)$. Одна из таких норм представляет интерес.

Евклидовой нормой оператора $\mathcal{A}$ называется число

$$\|\mathcal{A}\|_E = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}. \qquad (109.2)$$

Корректность этого определения следует из легко проверяемых для $\|\mathcal{A}\|_E$ аксиом нормы.

Как видно из (109.2), евклидова норма оператора легко вычисляема (по сравнению, например, с $\|\mathcal{A}\|_2$). При этом она обладает многими свойствами подчиненных норм. Перечислим их.

1. *Свойство согласованности:* $\|\mathcal{A}\|_E \le \|\mathcal{A}\|_E \|x\|_E$, так как в силу неравенства Коши–Буняковского $\|\mathcal{A}x\|_E^2 = \sum_{i=1}^{m} \big|\sum_{j=1}^{n} a_{ij} x_j\big|^2 \le \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} |a_{ij}|^2 \sum_{j=1}^{n} |x_j|^2 \Big) = \sum_{j=1}^{n} |x_j|^2 \cdot \|\mathcal{A}\|_E^2 = \|\mathcal{A}\|_E^2 \|x\|_E^2$.

2. *Свойство мультипликативности:* $\|\mathcal{AB}\|_E \le \|\mathcal{A}\|_E \|\mathcal{B}\|_E$, ибо $\|\mathcal{AB}\|_E^2 = \sum_{i,j} \big|\sum_k a_{ik} b_{kj}\big|^2 \le \sum_{i,j} \big(\sum_k |a_{ik}||b_{kj}|\big)^2 \le \sum_{i,j} \big(\sum_k |a_{ik}|^2 \cdot \sum_k |b_{kj}|^2\big) = \sum_{i,k} |a_{ik}|^2 \sum_{j,k} |b_{kj}|^2 = \|\mathcal{A}\|_E^2 \|\mathcal{B}\|_E^2$.

3. $\|\mathcal{A}\|_E^2 = \operatorname{tr} \mathcal{A}^*\mathcal{A} = \operatorname{tr} \mathcal{A}\mathcal{A}^*$. Эти равенства легко проверяются непосредственно.

4. $\|\mathcal{A}\|_E^2 = \rho_1^2 + \ldots + \rho_n^2$, где $\rho_1, \ldots, \rho_n$ – сингулярные числа оператора $\mathcal{A}$. Это следует из свойства 3 и теоремы 92.2 (следствие).

5. $\|\mathcal{A}\|_E \ge \|\mathcal{A}\|_2$ в силу теоремы 108.4.

6. $\|\mathcal{A}\|_E$ не изменяется при умножении оператора $\mathcal{A}$ на ортогональные (унитарные) операторы. Это следует из свойства 4 и теоремы 108.5.

§110. Экстремальные свойства собственных значений самосопряженных операторов

Пусть $\mathcal{A}$ – самосопряженный оператор в евклидовом (унитарном) пространстве V. Построим в пространстве V ортонормированный базис

$$e_1, \ldots, e_n \qquad (110.1)$$

из собственных векторов оператора $\mathcal{A}$, отвечающих собственным значениям

$$\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_n. \qquad (110.2)$$

Такой нумерации собственных векторов и собственных значений будем придерживаться всюду в этом параграфе. Под **нормой** $\|\cdot\|$ будем понимать евклидову норму $\|\cdot\|_E$, так что если $x = \sum_{i=1}^n x_i e_i$, то $\|x\| = \sqrt{(x,x)} = \left(\sum_{i=1}^n x_i^2\right)^{1/2}$.

Теорема 110.1. *Для самосопряженного оператора $\mathcal{A}$*

$$\lambda_1 = \max_{\|x\|=1}(\mathcal{A}x, x), \qquad \lambda_n = \min_{\|x\|=1}(\mathcal{A}x, x).$$

Доказательство. Для любого вектора $x = \sum_{i=1}^n x_i e_i \in V$ имеем $\mathcal{A}x = \sum_{i=1}^n x_i \lambda_i e_i \in V$ и, с учетом ортонормированности базиса, $(\mathcal{A}x, x) = \sum_{i=1}^n \lambda_i |x_i|^2$. В силу (110.2) отсюда следует, что $\lambda_1 \ge (\mathcal{A}x, x) \ge \lambda_n$, если $\|x\| = 1$; причем $(\mathcal{A}e_1, e_1) = \lambda_1$, $(\mathcal{A}e_n, e_n) = \lambda_n$ и $\|e_1\| = 1$, $\|e_n\| = 1$. Следовательно, λ_1 и λ_n – наибольшее и наименьшее значения $(\mathcal{A}x, x)$ на единичной евклидовой сфере. ∎

Замечание. Эта теорема дает экстремальные свойства и квадратичной формы в евклидовом (унитарном) пространстве: *на единичной сфере квадратичная форма $\mathcal{A}(x,x)$ принимает экстремальные значения на тех векторах, которые являются собственными векторами самосопряженного оператора $\mathcal{H}$* (теорема 104.1).

Кстати, в терминах квадратичной формы $f(x_1, \ldots, x_n)$ от переменных $x_1, \ldots, x_n$ задача поиска экстремальных значений квадратичной формы представляет собой классическую задачу на условный экстремум. Одним из методов ее решения является метод Лагранжа [7].

Теорема 110.2. *Если L – линейная оболочка собственных векторов $e_{i_1}, \ldots, e_{i_k}$ $(i_1 < \ldots < i_k)$ из (110.1) самосопряженного оператора $\mathcal{A}$, то*

$$\lambda_{i_1} = \max_{\|x\|=1, x\in L}(\mathcal{A}x, x), \qquad \lambda_{i_n} = \min_{\|x\|=1, x\in L}(\mathcal{A}x, x). \tag{110.3}$$

Доказательство аналогично доказательству теоремы 110.1. ∎

Теорема 110.3 (теорема Куранта–Фишера). *Для собственных значений самосопряженного оператора $\mathcal{A}$ справедливо представление*

$$\lambda_k = \max_{L_k} \min_{\|x\|=1, x\in L_k}(\mathcal{A}x, x), \tag{110.4}$$

где максимум берется по всевозможным k-мерным подпространствам L_k пространства V.

Доказательство. Пусть L_k – произвольное k-мерное подпространство и W_{n-k+1} – линейная оболочка собственных векторов $e_k, \ldots, e_n$ из (110.1) оператора $\mathcal{A}$. Так как $\dim L_k + \dim W_{n-k+1} = n+1$, то $L_k \cap W_{n-k+1} \ne \{\theta\}$. Пусть $x_0 \in L_k \cap W_{n-k+1}$ и $\|x_0\| = 1$. Тогда, согласно (110.3), $(\mathcal{A}x_0, x_0) \le \lambda_k$, поэтому $\min_{\|x\|=1, x\in L_k}(\mathcal{A}x_0, x_0) \le \lambda_k$, так что и $\max_{L_k} \min_{\|x\|=1, x\in L_k}(\mathcal{A}x_0, x_0) \le \lambda_k$. Равенство в (110.4) достигается для $L_k = \mathcal{L}(e_1, \ldots, e_k)$. ∎

§111. Линейные операторные уравнения

Рассмотрим проблему решения систем линейных алгебраических уравнений с точки зрения свойств линейного оператора.

Пусть V, W – евклидовы (унитарные) пространства, $\mathcal{A} \in \mathcal{L}(V, W)$, $u \in W$. Уравнение

$$\mathcal{A}z = u \tag{111.1}$$

называется ***линейным операторным уравнением***, вектор u – ***правой частью***, вектор z – ***решением***. Очевидно, в матричной записи операторное уравнение превращается в систему линейных алгебраических уравнений и, следовательно, все свойства систем уравнений можно переносить на операторные уравнения и наоборот. Отметим некоторые из этих свойств в терминах операторных уравнений.

Однородное уравнение

$$\mathcal{A}^* w = \theta \tag{111.2}$$

называется *сопряженным* к уравнению (111.1).

Теорема 111.1 (альтернатива Фредгольма). *Либо основное уравнение* (111.1) *имеет решение при любой правой части $u \in W$, либо сопряженное к нему уравнение имеет нетривиальное решение.*

Доказательство. Пусть $r = \operatorname{rg}\mathcal{A}$, $n = \dim W$. Возможны два случая: либо $r = n$, либо $r < n$. Условие $r = n$ равносильно условию $\operatorname{im}\mathcal{A} = W$, которое означает, что уравнение (111.1) имеет решение при любом $u \in W$. При этом так как $\operatorname{rg}\mathcal{A} = \operatorname{rg}\mathcal{A}^*$, то $\ker\mathcal{A}^* = \{\theta\}$ и уравнение (111.2) не имеет ненулевого решения. Условие $r < n$ равносильно условию $\operatorname{def}\mathcal{A}^* > 0$, которое означает существование ненулевого вектора $w \in \ker\mathcal{A}^*$, т.е. ненулевого решения (111.2). При этом $\operatorname{im}\mathcal{A} \ne W$ и уравнение (111.1) имеет решение не для любого $u \in W$. ∎

Замечание 1. Альтернатива Фредгольма для оператора $\mathcal{A}$, действующего в одном пространстве V, означает, что либо основное уравнение имеет единственное решение при любом $u \in V$, либо сопряженное к нему уравнение имеет нетривиальное решение.

Теорема 111.2 (теорема Фредгольма). *Операторное уравнение* (111.1) *имеет решение тогда и только тогда, когда его правая часть ортогональна всем решениям сопряженного уравнения.*

Доказательство. Уравнение (111.1) имеет решение тогда и только тогда, когда $u \in \operatorname{im}\mathcal{A}$ или, с учетом (95.4), когда $u \in \ker^\perp \mathcal{A}^*$. Это равносильно ортогональности вектора u всем векторам $\ker\mathcal{A}^*$, т.е. решениям уравнения (111.2). ∎

Нормальное решение. Пусть уравнение (111.1) разрешимо, т.е. имеет хотя бы одно решение. Обозначим через H множество всех его решений. ***Нормальным решением*** уравнения (111.1) называется такое его решение z_0, что

$$\|z_0\|_E = \inf_{z \in H} \|z\|_E.$$

Другими словами, нормальное решение – это решение наименьшей длины. Корректность определения вытекает из следующей теоремы.

Теорема 111.3. *Для любого разрешимого уравнения* (111.1) *нормальное решение существует и единственно.*

Доказательство. Существование. Пусть z – решение уравнения (111.1). Совокупность H всех решений этого уравнения является линейным многообразием $H = z + \ker\mathcal{A}$, так как для множеств H и $z + \ker\mathcal{A}$, как легко показать, имеет место двустороннее вложение. Из свойств линейного многообразия в евклидовом (унитарном) пространстве (теорема 73.1) следует существование и единственность нормального вектора сдвига z_0, ортогонального направляющему подпространству $\ker\mathcal{A}$. Причем вектор z_0 имеет наименьшую длину среди всех решений уравнения (111.1). Таким образом, z_0 – нормальное решение уравнения (111.1).

Единственность. Пусть z_1 – нормальное решение (111.1). Тогда $z_1 \in H$, т.е. $z_1 = z_0 + w$, где $w \in \ker\mathcal{A}$, $(w, z_0) = 0$, и $\|z_1\|_E^2 = \|z_0\|_E^2 + \|w\|_E^2$. Так как $\|z_0\|_E = \|z_1\|_E$, то $w = \theta$ и $z_1 = z_0$. ■

Замечание 2. Доказательство теоремы дает правило для отыскания нормального решения (теорема 73.2): z_0 – перпендикуляр, опущенный из любого решения z уравнения (111.1) на $\ker\mathcal{A}$.

Псевдорешение. Рассматривается уравнение (111.1), не обязательно разрешимое. Вектор $r = \mathcal{A}z - u$ называется *невязкой* вектора z, функция $F(z) = \|\mathcal{A}u - z\|_E^2$ – *функционалом невязки.*

Очевидно, вектор z является решением (111.1) тогда и только тогда, когда его невязка $r = \theta$, т.е. когда $\|r\|_E = 0$. Поскольку нулевое значение нормы является наименьшим, то решение z можно рассматривать как вектор, невязка которого имеет наименьшую норму, или, как принято говорить, *минимизирующий функционал невязки.*

Задача отыскания векторов, минимизирующих функционал невязки, имеет смысл и тогда, когда уравнение (111.1) неразрешимо (например, вследствие погрешностей измерения $\mathcal{A}$ и u). В этом случае, если z минимизирует функционал невязки, то расстояние $\rho(\mathcal{A}z, u) = \|\mathcal{A}z - u\|_E$ минимально и, следовательно, при таком z левая часть уравнения $\mathcal{A}z$ "ближе" всего к правой части u. Для многих задач вычислительной математики это свойство служит оценкой качества приближенного решения z.

Вектор $z^+ \in V$ называется *псевдорешением* уравнения (111.1), если

$$\|\mathcal{A}z^+ - u\|_E^2 = \inf_{z \in V} \|\mathcal{A}z - u\|_E^2. \tag{111.3}$$

Другими словами, псевдорешение – это вектор пространства V, минимизирующий функционал невязки.

Очевидно, если уравнение (111.1) разрешимо, то псевдорешение совпадает с решением в обычном смысле.

Теорема 111.4. *Псевдорешение существует для любого операторного уравнения* (111.1).

Доказательство. Согласно определению (111.3), $\|\mathcal{A}z^+ - u\|_E = \inf_{z \in V} \|\mathcal{A}z - u\|_E = \inf_{z \in V} |\mathcal{A}z - u| = \inf_{z \in V} \rho(\mathcal{A}z, u) = \inf_{y \in \operatorname{im}\mathcal{A}} \rho(y, u)$.

Это означает, что вектор $\mathcal{A}z^+$ – это вектор из $\operatorname{im}\mathcal{A}$, который ближе всего расположен к вектору u. В силу теоремы 74.2 о кратчайшем расстоянии отсюда следует, что $\mathcal{A}z^+$ – ортогональная проекция вектора u на $\operatorname{im}\mathcal{A}$. Пусть $u = g + h$, где $g \in \operatorname{im}\mathcal{A}$, $h \perp \operatorname{im}\mathcal{A}$. Тогда z^+ является решением в обычном смысле уравнения

$$\mathcal{A}z = g \tag{111.4}$$

Очевидно, оно имеет решение, так как $g \in \operatorname{im}\mathcal{A}$). ■

Уравнение

$$\mathcal{A}^*\mathcal{A}z = \mathcal{A}^*u \tag{111.5}$$

называется *нормальным уравнением* для уравнения (111.1).

Теорема 111.5. *Вектор z^+ пространства V является псевдорешением уравнения* (111.1) *тогда и только тогда, когда z^+ – решение нормального уравнения* (111.5).

Доказательство. Выше было показано, что z^+ – псевдорешение (111.1) тогда и только тогда, когда z^+ – решение в обычном смысле уравнения (111.4). Покажем, что уравнения (111.4) и (111.5) равносильны. Действительно, $\mathcal{A}^*u = \mathcal{A}^*g$, так как $u = g + h$, где $h \in \operatorname{im}^\perp \mathcal{A} = \ker\mathcal{A}^*$. При этом если z – решение (111.4), то $\mathcal{A}z = g$, поэтому $\mathcal{A}^*\mathcal{A}z = \mathcal{A}^*u$ и z является решением (111.5). Если же z – решение (111.5), то $\mathcal{A}^*\mathcal{A}z = \mathcal{A}^*g$ или $\mathcal{A}^*(\mathcal{A}z - g) = \theta$. Значит, $\mathcal{A}z - g \in \ker\mathcal{A}^* = \operatorname{im}^\perp\mathcal{A}$. Но $\mathcal{A}z \in \operatorname{im}\mathcal{A}$, $g \in \operatorname{im}\mathcal{A}$, следовательно, $\mathcal{A}z - g \in \operatorname{im}\mathcal{A}$. Отсюда с учетом соотношения $\mathcal{A}z - g \in \operatorname{im}^\perp\mathcal{A}$, полученного выше, следует, что $\mathcal{A}z - g = \theta$, т.е. z является решением (111.4). ■

Псевдорешение наименьшей длины называется *нормальным псевдорешением.* Из теорем 111.3, 111.4 следует, что нормальное псевдорешение существует и единственно для любого уравнения (111.1).

Предметный указатель

Учебное издание

ИЛЬИН Владимир Александрович, КИМ Галина Динховна

**ЛИНЕЙНАЯ АЛГЕБРА
И АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ**

Зав. редакцией *Л.Н. Николова*
Редактор *Т.В. Властовская*
Технический редактор *Н.И. Матюшина*
Компьютерная верстка *Л.В. Крицков, А.В. Ильин*
Корректор *Н.И. Коновалова*

Изд. лиц. № 040414 от 18.04.97
Подписано в печать 31.03.98.
Формат 60×90 1/16. Бумага кн.-журн.
Гарнитура литературная. Офсетная печать.
Усл. печ. л. 18,14 Уч.-изд. л. 22,28.
Тираж 3000 экз.   Заказ № 5508 Изд. № 6245.

Оригинал-макет и иллюстрации подготовлены Л. В. Крицковым с использованием издательской системы TEX на факультете вычислительной математики и кибернетики МГУ.

Ордена "Знак Почета" издательство Московского университета.
103009, Москва, ул. Б. Никитская, 5/7.

Отпечатано в Производственно-издательском комбинате ВИНИТИ
140010, г. Люберцы, Московской обл., Октябрьский пр-т, 403.
Тел. 554-21-86