Академик АН СССР

Л. С. Понтрягин

АЛГЕБРА

ЗНАКОМСТВО С ВЫСШЕЙ МАТЕМАТИКОЙ (в 4 книгах)

«...они будут учитывать мои юношеские воспоминания о возможностях восприятия молодого человека, с тем, чтобы нынешнее поколение молодых людей, начиная со школьников старших классов, могло знакомиться по ним с высшей математикой и приобретать правильный здоровый вкус к ней».

Теория определителей •

Корни многочленов и комплексные числа •

Приведение матриц к каноническому виду •

УРСС

Лев Семенович
ПОНТРЯГИН (1908–1988)

Выдающийся советский математик, академик АН СССР, Герой Социалистического Труда (1969). Родился 3 сентября 1908 г. в Москве. В 14 лет потерял зрение в результате несчастного случая. Окончил Московский государственный университет им. М.В.Ломоносова (1929). С 1930 г. работал в Московском университете, где в 1935 г. получил ученое звание профессора, и одновременно с 1939 г. занимал должность заведующего отделом Математического института им. В.А.Стеклова АН СССР.

Основные работы Л.С.ПОНТРЯГИНА

относятся к теории дифференциальных уравнений, топологии, теории колебаний, теории управления, вариационному исчислению, алгебре.

В топологии он открыл общий закон двойственности и в связи с этим построил теорию характеров непрерывных групп; получил ряд результатов в теории гомотопий (классы Понтрягина).

В теории колебаний главные результаты работ Л.С.Понтрягина относятся к асимптотике релаксационных колебаний.

В теории управления он выступил как создатель математической теории оптимальных процессов, в основе которой лежит так называемый принцип максимума Понтрягина.

Ему принадлежат также существенные результаты в области вариационного исчисления, дифференциальных игр, теории размерности, теории регулирования.

Работы школы Л.С.Понтрягина оказали большое влияние на развитие теории управления и вариационного исчисления во всем мире.

ЗНАКОМСТВО С ВЫСШЕЙ МАТЕМАТИКОЙ

Л. С. Понтрягин

АЛГЕБРА

Издание второе, стереотипное

Москва • 2004

УРСС

ББК 22.1о, 22.144

Понтрягин Лев Семенович

Алгебра. Изд. 2-е, стереотипное. — М.: Едиториал УРСС, 2004. — 136 с. (Знакомство с высшей математикой.)

ISBN 5-354-00614-7

«Знакомство с высшей математикой» — серия небольших научно-популярных книг. Настоящая книга является третьей в этой серии. В ней приведены основные результаты алгебры, включая теорию определителей. Этот раздел составляет основную часть книги. Кроме того, в книге содержится раздел, посвященный корням многочленов и комплексным числам.

Для школьников старших классов, интересующихся математикой. Может быть полезна также преподавателям средней и высшей школы.

Рецензент: член-корреспондент АН СССР *И. Р. Шафаревич*

Издательство «Едиториал УРСС». 117312, г. Москва, пр-т 60-летия Октября, 9.
Лицензия ИД № 05175 от 25.06.2001 г. Подписано к печати 03.12.2003 г.
Формат 60×84/16. Тираж 1000 экз. Печ. л. 8,5. Зак. № 2-1182/384.

Отпечатано в типографии ООО «РОХОС». 117312, г. Москва, пр-т 60-летия Октября, 9.

ISBN 5-354-00614-7

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Несколько лет тому назад я решил написать серию небольших книг под названием «Знакомство с высшей математикой», предназначенную для школьников, интересующихся математикой. Две книги из этой серии «Метод координат» и «Анализ бесконечно малых» уже опубликованы. «Алгебра» является третьей книгой в этой серии. В ней приведены основные результаты алгебры, включая теорию определителей. Этот раздел составляет главную часть книги. Кроме того, книга содержит раздел, посвященный корням многочленов и комплексным числам. При этом правила перемножения комплексных чисел, взятых в тригонометрической форме, доказываются без использования формул тригонометрии для косинуса суммы и синуса суммы.

Большую работу при редактировании этой книги проделал С. М. Асеев, за что я выражаю ему благодарность.

Глава 1

ТЕОРИЯ ОПРЕДЕЛИТЕЛЕЙ

§ 1. Векторные пространства

В этом параграфе будет рассмотрено важнейшее для линейной алгебры понятие векторного пространства.

Определение 1. Последовательность

$$x=(x^1, x^2, \ldots, x^n) \tag{1}$$

из n действительных чисел, расположенных в определенном порядке, указанном в формуле (1), называется *n-мерным вектором*. Совокупность A^n всех n-мерных векторов называется *n мерным векторным пространством*. Числа

$$x^1, x^2, \ldots, x^n$$

называются *координатами вектора* x (см. (1)).

Вектор считается равным нулю и обозначается через $\mathbf{0}$, если все его координаты равны нулю.

Произвольный вектор x из A^n может быть умножен на действительное число c. Для получения этого произведения каждая координата вектора x умножается на число c. Таким образом,

$$cx=(cx^1, cx^2, \ldots, cx^n). \tag{2}$$

Каждые два вектора x (см. (1)) и

$$y=(y^1, y^2, \ldots, y^n)$$

из A^n могут быть сложены. При этом координаты их складываются, так что

$$x+y=(x^1+y^1, x^2+y^2, \ldots, x^n+y^n) \tag{3}$$

(см. гл. 4, пример 1).

А) Если

$$x_1, x_2, \ldots, x_k \tag{4}$$

— совокупность из k векторов пространства A^n, то, пользуясь операциями (2) и (3), можно составить их линейную форму:

$$z = c^1x_1 + c^2x_2 + \ldots + c^kx_k, \tag{5}$$

где $c^1, c^2, \ldots, c^k$ — действительные числа.

В дальнейшем мы будем пользоваться сокращенными обозначениями для суммирования. Именно, если в одночлене один и тот же обозначенный греческой буквой индекс встречается один раз вверху и один раз внизу, то это означает сумму по всем значениям этого индекса. Таким образом, формула (5) переписывается в виде

$$z = c^\alpha x_\alpha.$$

Совокупность из k векторов (4) считается *линейно независимой*, если вектор, определенный формулой (5), обращается в нуль лишь при условии, когда

$$c^1 = c^2 = \ldots = c^k = 0.$$

В противном случае, т. е. если существуют такие числа

$$c^1, c^2, \ldots, c^k, \tag{6}$$

не все равные нулю, что вектор z, определяемый формулой (5), обращается в нуль, совокупность (4) называется *линейно зависимой*. Если к совокупности линейно зависимых векторов (4) присоединить еще несколько векторов $x_{k+1}, x_{k+2}, \ldots, x_l$, то расширенная совокупность

$$x_1, x_2, \ldots, x_l$$

будет линейно зависимой. Именно, если совокупность коэффициентов (6) осуществляет линейную зависимость векторов (4), то мы имеем соотношение

$$c^\alpha x_\alpha + 0x_{k+1} + \ldots + 0x_l = 0, \quad \alpha = 1, \ldots, k,$$

причем не все числа последовательности

$$c^1, c^2, \ldots, c^k, c^{k+1} = 0, \ldots, c^l = 0$$

равны нулю.

Если в векторном пространстве A^n, состоящем из всех векторов вида (1), рассмотреть те векторы, для которых i-я координата обращается в нуль, то мы получим также векторное пространство A^{n-1}, но размерности

$n-1$, которое называется *координатным подпространством* пространства A^n.

Докажем теперь следующее важное предложение.

В) *Совокупность* (4) *векторов n-мерного пространства A^n при $k > n$ всегда линейно зависима.*

Для доказательства достаточно рассмотреть случай $k = n+1$. Доказательство будем вести индуктивно. Отметим прежде всего, что для $n = 1$ утверждение В) справедливо. Пусть x, y — два вектора пространства A^1, причем

$$x = (x^1), \quad y = (y^1).$$

Если числа x^1, y^1 оба равны нулю, то мы имеем соотношение

$$cx + dy = 0$$

при произвольных c и d, так что в этом случае векторы x и y линейно зависимы. Если хотя бы одно из чисел x^1, y^1 не равно нулю, то мы имеем следующее соотношение:

$$y^1 x - x^1 y = 0.$$

Таким образом, при $n = 1$ утверждение В) верно.

Пусть теперь

$$x_1, \; x_2, \; \ldots, \; x_{n+1} \tag{7}$$

— совокупность из $n+1$ векторов пространства A^n. Сосредоточим внимание на последних координатах всех векторов совокупности (7). Если все эти координаты равны нулю, то совокупность векторов (7) принадлежит к $(n-1)$-мерному пространству и, согласно предположению индукции, линейно зависима. Допустим теперь, что хотя бы один вектор, например x_{n+1}, имеет последнюю координату, отличную от нуля. Тогда для каждого вектора x_j совокупности (7) при $j \leqslant n$ можно подобрать такое число d_j, что вектор y_j, определяемый формулой

$$y_j = x_j - d_j x_{n+1}, \qquad j = 1, \ldots, n, \tag{8}$$

имеет последнюю координату, равную нулю, и потому лежит в $(n-1)$-мерном векторном пространстве. Следовательно, по предположению индукции векторы $y_1, y_2, \ldots$ $\ldots, y_n$ линейно зависимы. Таким образом, существует совокупность чисел

$$c^1, \; c^2, \; \ldots, \; c^n,$$

(причем не все эти числа равны нулю) такая, что имеется соотношение

$$c^\alpha y_\alpha = 0, \quad \alpha = 1, \ldots, n.$$

Из этого и из формулы (8) вытекает

$$c^\alpha x_\alpha - (c^\alpha d_\alpha) x_{n+1} = 0, \quad \alpha = 1, \ldots, n.$$

Таким образом, совокупность векторов (7) линейно зависима. Итак, предложение B) доказано.

Базис векторного пространства

C) Обозначим через e_i вектор из A^n, все координаты которого равны нулю, за исключением i-й координаты, которая равна 1, т. е. положим

$$e_1 = (1, 0, \ldots, 0),$$
$$e_2 = (0, 1, \ldots, 0),$$
$$\cdots\cdots\cdots$$
$$e_n = (0, 0, \ldots, 1).$$

Ясно, что вектор x (см. (1)) записывается в виде

$$x = x^\alpha e_\alpha, \quad \alpha = 1, \ldots, n,$$

причем коэффициенты $x^1, x^2, \ldots, x^n$, входящие в эту формулу, однозначно определены вектором x. Таким образом, каждый вектор x в векторном пространстве A^n выражается в виде линейной формы относительно векторов

$$e_1, e_2, \ldots, e_n, \tag{9}$$

причем коэффициенты этой линейной формы определены однозначно. Система векторов, обладающая этим свойством, называется *базисом векторного пространства* A^n. Построенный нами базис (9) не является единственным. Оказывается, что любая линейно независимая система векторов

$$e'_1, e'_2, \ldots, e'_n, \tag{10}$$

содержащая n векторов пространства A^n, является базисом этого пространства.

Действительно, пусть x — произвольный вектор пространства A^n. Тогда векторы

$$x, e'_1, e'_2, \ldots, e'_n$$

линейно зависимы, поскольку содержат $n+1$ вектор (см. B)). Таким образом мы имеем соотношение

$$cx + c^\alpha e'_\alpha = 0, \quad \alpha = 1, \ldots, n, \tag{11}$$

причем в этом соотношении не все коэффициенты равны нулю. В частности, коэффициент c не может быть равен нулю, так как если бы он был равен нулю, то оказалось бы, что векторы системы (10) линейно зависимы. Пользуясь этим, разделим соотношение (11) на величину $-c$. Для того чтобы не вводить новые обозначения, будем считать просто, что $c=-1$, а остальные коэффициенты сохраним. Таким образом, получаем

$$x = c^{\alpha} e'_{\alpha}, \quad \alpha = 1, \ldots, n.$$

Если бы вектор x мог быть записан аналогично через систему (10), но с другими коэффициентами, т. е. если бы имело место соотношение

$$x = d^{\alpha} e'_{\alpha}, \quad \alpha = 1, \ldots, n,$$

то, вычитая последнее соотношение из предыдущего, мы получили бы соотношение

$$(c^{\alpha} - d^{\alpha}) e'_{\alpha} = 0, \quad \alpha = 1, \ldots, n,$$

что ввиду линейной независимости векторов системы (10) приводит к равенству

$$c^{i} = d^{i}, \quad i = 1, \ldots, n.$$

D) *Пусть*

$$x_1, x_2, \ldots, x_p \tag{12}$$

— некоторая линейно независимая система векторов из A^n. Если $p < n$, то систему (12) можно дополнить векторами $x_{p+1}, \ldots, x_n$ так, что система

$$x_1, x_2, \ldots, x_p, x_{p+1}, \ldots, x_n$$

линейно независима.

Докажем утверждение D). Если предложение D) неверно, то, присоединяя к системе (12) вектор e_1 (см. C)), мы получим расширенную систему

$$e_1, x_1, \ldots, x_p$$

обязательно линейно зависимую. Вектор e_1 ненулевой. Поэтому среди векторов системы (12) обязательно найдется вектор x_q, выражающийся в виде линейной комбинации через остальные векторы расширенной совокупности. Значит, любой вектор пространства A^n можно выразить в виде линейной комбинации векторов

$$e_1, x_1, \ldots, x_{q-1}, x_{q+1}, \ldots, x_p.$$

Исключим вектор x_q из расширенной совокупности векторов и добавим к ней вектор e_2. Из линейной независимости векторов e_1, e_2 следует, что среди векторов $x_1, \ldots, x_{q-1}, x_{q+1}, \ldots, x_p$ обязательно найдется вектор, который можно исключить. Продолжая этот процесс, мы придем к выводу, что любой вектор пространства A^n можно представить в виде линейной комбинации векторов $e_1, \ldots, e_p$, что невозможно, так как векторы (9) линейно независимы. Следовательно, предложение D) верно.

Из предложения C) видно, что данное в начале параграфа определение n-мерного векторного пространства A^n при помощи координат вектора (см. (1)) не является инвариантным, так как оно зависит от случайно выбранного базиса (9), а базисов в пространстве A^n существует бесчисленное множество. Поэтому мы дадим другое инвариантное определение n-мерного пространства A^n.

Определение 2. Множество A элементов называется *векторным пространством*, если в нем определены две операции: сложение и умножение на действительные числа, причем выполнено условие: если c и d — два числа, а x, y, z — элементы пространства A, т. е. три вектора, то имеют место соотношения

$$
\begin{gathered}
(x+y)+z=x+(y+z),\\
x+y=y+x,\\
(c+d)x=cx+dx,\\
c(x+y)=cx+cy,\\
c(dx)=(cd)x,\\
1\cdot x=x.
\end{gathered}
$$

Кроме того, предполагается, что имеется нулевой вектор 0, удовлетворяющий условиям

$$
\begin{gathered}
x+0=x,\\
0x=0.
\end{gathered}
$$

При помощи операций, имеющихся в векторном пространстве A, можно составить линейную комбинацию для произвольной системы из векторов

$$x_1, \ldots, x_n,$$

т. е. построить вектор

$$z=c^\alpha x_\alpha, \quad \alpha=1, \ldots, n,$$

где $c^1, \ldots, c^n$ — действительные числа.

Таким образом, мы можем ввести понятие линейной зависимости и линейной независимости векторов, как это было сделано раньше (см. С)

E) Если в векторном пространстве A имеется n линейно независимых векторов, но нет $n+1$ линейно независимых векторов, то считается, что векторное пространство A имеет *размерность* n и оно обозначается через A^n. Выбирая в пространстве A^n n линейно независимых векторов, мы получим базис и при помощи него координатную запись любого вектора.

Векторное подпространство

F) Пусть A — векторное пространство и B — такое подмножество его векторов, что наряду с двумя векторами x и y из B в B входит также $x+y$, а наряду с вектором x в B входит и его произведение на произвольное действительное число, т. е. вектор cx. Тогда B, очевидно, представляет собой векторное пространство в силу тех операций, которые имеются в A. Множество B называется *векторным подпространством* пространства A. Если векторное пространство A имеет конечную размерность n (см. E)), то его подпространство B имеет размерность, не превосходящую n.

G) Пусть A^n — n-мерное векторное пространство, а B^p и C^q — два его векторных подпространства размерности p и q соответственно. Если для векторных подпространств B^p и C^q нулевой вектор является единственным общим вектором, а $p+q=n$, то векторное пространство A^n распадается в прямую сумму своих векторных подпространств B^p и C^q. Именно, каждый вектор x из A^n представляется в виде суммы

$$x = y + z,$$

где y — вектор из B^p, а z — вектор из C^q, причем слагаемые y и z однозначно определены вектором x.

Для доказательства этого утверждения выберем в подпространстве B^p базис

$$e_1,\ e_2,\ \ldots,\ e_p,$$

состоящий из p элементов, а в подпространстве C^q базис

$$e_{p+1},\ e_{p+2},\ \ldots,\ e_n,$$

состоящий из q элементов. Объединяя эти две последовательности векторов, мы получим последовательность

$$e_1, e_2, \ldots, e_n. \tag{13}$$

Докажем, что эта совокупность составляет базис пространства A^n. Так как число векторов системы (13) равно размерности пространства A^n, то достаточно доказать, что векторы системы (13) линейно независимы. Допустим, что имеет место противоположное, именно, имеет место соотношение

$$c^\alpha e_\alpha = 0, \quad \alpha = 1, \ldots, n, \tag{14}$$

причем не все коэффициенты c^i, входящие в это соотношение, равны нулю. Обозначим через y сумму первых p членов суммы (14) и через z сумму остальных членов. Тогда получим два вектора y и z, причем оба они одновременно не могут обращаться в нуль, так что хотя бы один отличен от нуля. При этом соотношение (14) переписывается в виде

$$y + z = 0,$$

или, иначе,

$$y = -z,$$

где $y \in B^p$, $-z \in C^q$.

Таким образом, мы приходим к выводу, что векторные подпространства B^p и C^q имеют общий вектор, отличный от нуля, что противоречит предположению.

Таким образом, система (13) является базисом пространства A^n и каждый вектор x из A^n может быть записан в виде

$$x = c^\alpha e_\alpha, \quad \alpha = 1, \ldots, n.$$

Обозначая через y сумму первых p слагаемых последней суммы, а через z — сумму остальных слагаемых, приходим к выводу, что

$$x = y + z,$$

где $y \in B^p$, $z \in C^q$. Таким образом, утверждение G) доказано.

§ 2. Линейные отображения векторных пространств и матрицы

В этом параграфе будет показано, как линейное отображение одного векторного пространства в другое задается матрицей.

Здесь будем пользоваться сокращенными обозначениями для суммирования (см. § 1, А)).

Определение 3. Пусть A^p и B^q — два векторных пространства размерности p и q соответственно. Отображение φ пространства B^q в пространство A^p, т. е. такое отображение, которое каждому вектору $y \in B^q$ ставит в соответствие вектор

$$x = \varphi y$$

из пространства A^p, называется *линейным*, если выполнено следующее условие: каковы бы ни были векторы y_1, y_2 из B^q и два действительных числа c^1, c^2, имеем

$$\varphi(c^1 y_1 + c^2 y_2) = c^1 \varphi y_1 + c^2 \varphi y_2.$$

Множество всех $y \in B^q$, переходящих при отображении φ в нулевой вектор пространства A^p, называется *ядром* линейного отображения φ. Согласно общепринятым обозначениям оно записывается в виде $\varphi^{-1}0$. Непосредственно проверяется (я здесь этого делать не буду), что ядро отображения φ есть векторное подпространство пространства B^q. Точно так же легко проверяется, что совокупность всех векторов $x \in A^p$ вида φy, где $y \in B^q$, есть векторное подпространство пространства A^p. Оно обозначается через φB^q. Если $\varphi B^q = A^p$, то про отображение φ говорят, что оно есть отображение *на* пространство A^p.

Заметим, что множество D действительных чисел является одномерным векторным пространством. В случае если A^p есть множество D, отображение φ называется *линейной формой*, заданной на пространстве B^q.

Если ψ и χ — два линейных отображения пространства B^q в пространство A^p, а c, d — два действительных числа, то определяется линейное отображение

$$\varphi = c\psi + d\chi,$$

которое задается формулой

$$\varphi y = c\psi y + d\chi y. \qquad (15)$$

Легко проверяется, что отображение φ, заданное этой формулой, является линейным отображением пространства B^q в пространство A^p. Таким образом, множество всех линейных отображений пространства B^q в пространство A^n является векторным пространством. Позже мы покажем, что это векторное пространство имеет конечную размерность pq.

Допустим, что наряду с упомянутыми векторными пространствами A^p и B^q имеется еще третье векторное пространство C^r размерности r и дано линейное отображение η векторного пространства C^r в векторное пространство B^q. Тогда мы можем определить линейное отображение векторного пространства C^r в векторное пространство A^p, ставя в соответствие каждому его вектору z вектор

$$x = \varphi\eta z. \tag{16}$$

Тот факт, что отображение, заданное формулой (16), является линейным, непосредственно проверяется. Оно по определению считается *произведением отображений* η и φ и записывается как $\varphi\eta$. Если имеется еще и четвертое векторное пространство D^s размерности s и $\varkappa$ есть линейное отображение пространства D^s в пространство C^r, то, кроме произведения $\varphi\eta$, определено произведение $\eta\varkappa$, а также два произведения

$$(\varphi\eta)\varkappa, \quad \varphi(\eta\varkappa).$$

Легко проверяется, что оба отображения, выписанные в предыдущей строке, совпадают между собой. Именно, имеет место равенство

$$(\varphi\eta)\varkappa = \varphi(\eta\varkappa), \tag{17}$$

т. е. произведение отображений ассоциативно. Действительно, оба выражения, стоящие в равенстве (17), можно описать следующим образом. Если u — произвольный вектор из пространства D^s, то к вектору u применяем отображение $\varkappa$. К полученному вектору $\varkappa u$ применяем отображение η. К полученному вектору $\eta\varkappa u$ применяем отображение φ, так что оба выражения, стоящие в равенстве (17), получаются последовательным применением к вектору u отображений $\varkappa$, η, φ. Отсюда и вытекает равенство (17).

Расскажем теперь, как отображение φ можно записать в координатной форме при помощи таблицы, называемой матрицей.

А) Пусть

$$e_1, e_2, \ldots, e_p, \tag{18}$$

$$f_1, f_2, \ldots, f_q \tag{19}$$

— базисы пространств A^p и B^q соответственно. Вектор φf_j имеет в базисе (18) некоторые координаты, зависящие от номера j, так что этот вектор можно записать в виде

$$\varphi f_j = x_j^\alpha e_\alpha, \quad \alpha = 1, \ldots, p. \tag{20}$$

Если

$$y = y^\beta f_\beta, \quad \beta = 1, \ldots, q,$$

есть некоторый вектор из B^q, то

$$\varphi y = x = x^\alpha e_\alpha, \quad \alpha = 1, \ldots, p. \tag{21}$$

Умножая соотношение (20) на y^j и суммируя его по j, получим

$$x^\alpha e_\alpha = x_\beta^\alpha y^\beta e_\alpha, \quad \alpha = 1, \ldots, p, \quad \beta = 1, \ldots, q. \tag{22}$$

При этом мы использовали линейность отображения φ. Из соотношения (22) следует соотношение

$$x^i = x_\beta^i y^\beta, \quad \beta = 1, \ldots, q. \tag{23}$$

Последняя формула дает нам выражение координат вектора $x = \varphi y$ в базисе (18) через координаты вектора y в базисе (19). Таким образом, это соотношение представляет собой координатную запись отображения φ. Числа x_j^i, $i = 1, \ldots, p$; $j = 1, \ldots, q$, естественно записать в виде прямоугольной таблицы

$$\begin{Vmatrix} x_1^1 & x_2^1 & \ldots & x_q^1 \\ x_1^2 & x_2^2 & \ldots & x_q^2 \\ \cdot & \cdot & \cdot & \cdot \\ x_1^p & x_2^p & \ldots & x_q^p \end{Vmatrix}. \tag{24}$$

Эта таблица называется *матрицей*. Число ее строк равно p, а число столбцов равно q. Кратко мы будем говорить, что (24) есть матрица размеров (p, q). Таким образом, p есть высота матрицы, q — ее ширина. Вся матрица (24) обозначается обычно одной буквой. В данном случае обозначим ее через X. Кратко можно записать

$$X = \|x_j^i\|, \quad i = 1, \ldots, p; \quad j = 1, \ldots, q.$$

Таким образом, линейное отображение φ определяет матрицу X и само определяется этой матрицей, если заданы базисы (18), (19) обоих пространств A^p и B^q (см. гл. 4, примеры 2—4).

Если φ — линейная форма, заданная на B^q, т. е. отображает векторное пространство B^q в векторное пространство D действительных чисел, то число φy

записывается в форме

$$\varphi y = d_\beta y^\beta, \quad \beta = 1, \ldots, q. \tag{25}$$

Если ψ и χ — два линейных отображения пространства B^q в пространство A^p, c и d — два действительных числа, то определено отображение φ (см. (15)). Пусть отображениям φ, ψ, χ соответствуют матрицы X, Y, Z. Положим

$$X = \| x^i_j \|, \quad Y = \| y^i_j \|, \quad Z = \| z^i_j \|.$$

В силу (23) мы имеем

$$(\varphi y)^\alpha = x^\alpha_\beta y^\beta, \quad \beta = 1, \ldots, p;$$

$$(\psi y)^\alpha = y^\alpha_\beta y^\beta, \quad \beta = 1, \ldots, p;$$

$$(\chi y)^\alpha = z^\alpha_\beta y^\beta, \quad \beta = 1, \ldots, p.$$

В силу (15) получаем

$$x^i_j = c y^i_j + d z^i_j.$$

Если положить

$$X = cY + dZ,$$

то получим

$$X = \| c y^i_j + d z^i_j \|.$$

Последняя формула определяет умножение матрицы размеров (p, q) на число и суммирование двух матриц этого размера. Из нее видно, что пространство всех матриц размеров (p, q) есть векторное пространство размерности pq.

В) Пусть A^p, B^q, C^r — три векторных пространства указанных размерностей, φ — линейное отображение пространства B^q в пространство A^p, а ψ — линейное отображение пространства C^r в пространство B^q. Пусть

$$e_1, e_2, \ldots, e_p; \; f_1, f_2, \ldots, f_q; \; g_1, g_2, \ldots, g_r$$

— базисы рассматриваемых трех векторных пространств A^p, B^q и C^r соответственно. В этих базисах отображению ψ соответствует некоторая матрица $Y = \| y^j_k \|$ размеров (q, r). Именно, если

$$z = z^\gamma g_\gamma, \quad y = \psi z = y^\beta f_\beta,$$

то в силу предложения А) имеем

$$y^j = y^j_\gamma z^\gamma \tag{26}$$

и матрица $\| y^j_k \|$, $j = 1, \ldots, q$; $k = 1, \ldots, r$, соответствует отображению ψ. Если теперь $X = \| x^i_j \|$ — матрица, соответствующая отображению φ, то в силу А)

$$x^i = x^i_\beta y^\beta. \tag{27}$$

Подставляя координаты вектора y из формулы (26) в последнюю формулу, получаем

$$x^i = x^i_\beta y^\beta_\gamma z^\gamma.$$

Таким образом, отображению $\varphi\psi$ соответствует матрица, задаваемая формулой

$$\| x^i_\beta y^\beta_k \|, \quad i = 1, \ldots, p; \quad \beta = 1, \ldots, q; \quad k = 1, \ldots, r. \tag{28}$$

Матрица (28) считается *произведением матриц* X и Y и записывается в виде XY, так что мы имеем

$$XY = \| x^i_\beta y^\beta_k \|. \tag{29}$$

Здесь отображению φ соответствует матрица X размеров (p, q), а отображению ψ соответствует матрица Y размеров (q, r). Отображению $\varphi\psi$ соответствует матрица XY размеров (p, r). Здесь мы установили операцию умножения матриц X и Y в случае, если матрица X имеет ширину q, а матрица Y — высоту q. Говорят, что матрица XY образуется путем перемножения строк матрицы X на столбцы матрицы Y.

Если наряду с пространствами A^p, B^q, C^r рассматривается еще четвертое пространство D^s и отображение χ пространства D^s в пространство C^r, причем отображению χ соответствует матрица Z размеров (r, s), то определены произведения матриц

$$XY, \; YZ, \; (XY)Z, \; X(YZ).$$

Поскольку имеет место ассоциативность при перемножении отображений (см. (17)), то для соответствующих матриц она тоже имеет место, т. е. мы имеем

$$(XY)Z = X(YZ).$$

Правую часть равенства (27), определяющую координаты вектора x через координаты вектора y, можно трактовать как произведение матрицы X на одностолбцовую матрицу $\hat{y}$, где $\hat{y}$ представляет собой выписанные в виде столбца координаты вектора y, т. е. матрицу размеров $(q, 1)$. Само равенство (27) теперь может быть записано в виде

$$\hat{x} = X\hat{y},$$

где матрица $\hat{x}$ представляет собой столбец, в котором выписаны координаты вектора x. Она имеет размеры $(p, 1)$.

C) Если в предложении A) число q равно числу p, то матрица X, соответствующая отображению φ, имеет размеры (p, p) и является квадратной. Особый интерес представляет случай, когда пространство B^q совпадает с пространством A^p. Тогда матрица X отображения φ пространства A^p самого в себя является квадратной размеров (p, p), но при этом выделяется играющая особую роль единичная матрица. Именно, если отображение φ есть тождественное отображение, т. е. имеет место соотношение

$$\varphi x = x,$$

то мы имеем

$$e_i = \varphi e_i,$$

и потому в формуле (20) надо считать $f_i = e_i$ и все числа x_j^i, при которых $i \neq j$, равны нулю, а числа x_i^i равны единице. Такая матрица называется *единичной* и обозначается через E. Для ее записи применяется так называемый *символ Кронекера*. Это

$$\delta_j^i = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Таким образом, мы имеем

$$E = \| \delta_j^i \|.$$

Если отображение ψ в предложении B) является тождественным отображением пространства $B^q = A^p$ на себя, то ему соответствует матрица $Y = E$, и из (29) имеем

$$XE = X.$$

Точно так же, если отображение φ в предложении B) является тождественным отображением пространства $B^q = A^p$ на себя, то соответствующая ему матрица является единичной матрицей E, и мы имеем

$$EY = Y.$$

Особо большую роль единичная матрица играет при рассмотрении отображений пространства A^p самого на себя. Если φ — такое отображение, а X — соответствующая ему матрица, то мы имеем соотношение

$$XE = EX = X,$$

так что в множестве всех матриц, описывающих отображение пространства A^p в себя, единичная матрица E играет роль единицы в отношении умножения.

Билинейные формы

D) Пусть A^n — n-мерное векторное пространство и

$$e_1, e_2, \ldots, e_n \tag{30}$$

— некоторый его базис. Функция $f(x, y)$ от двух переменных векторов x и y пространства A^n называется *билинейной формой*, если она является линейной формой относительно каждого из векторов x и y (см. определение 3). Положим

$$a_{ij} = f(e_i, e_j). \tag{31}$$

Пусть

$$x = x^\alpha e_\alpha, \quad y = y^\beta e_\beta$$

— запись векторов x и y в базисе (30). Оказывается, что

$$f(x, y) = a_{\alpha\beta} x^\alpha y^\beta. \tag{32}$$

Числа a_{ij} называются *коэффициентами* билинейной формы $f(x, y)$ в базисе (30). Они естественным образом составляют матрицу, если i принять за номер строки, а j за номер столбца:

$$A = \| a_{ij} \|, \quad i, j = 1, \ldots, n.$$

Таким образом, билинейной форме $f(x, y)$ соответствует в базисе (30) квадратная матрица A порядка n. Если функция $f(x, y)$ не меняется при перестановке ее аргументов, т. е. имеет место соотношение

$$f(x, y) = f(y, x),$$

то билинейная форма $f(x, y)$ называется *симметричной*. Симметричной билинейной форме $f(x, y)$ соответствует симметричная матрица A, т. е. матрица A, удовлетворяющая условию $a_{ij} = a_{ji}$. Если в билинейной форме $f(x, y)$ заменить вектор y вектором x, т. е. положить $y = x$, то мы получим квадратичную форму $f(x, x)$, соответствующую билинейной форме $f(x, y)$. Квадратичная форма $f(x, x)$ является функцией одного вектора x пространства A^n. Таким образом, каждой билинейной форме соответствует квадратичная форма. Ясно, что для получения любой квадратичной формы достаточно использовать лишь симметричные билинейные формы. Если исходная билинейная форма несимметрична, то квадратичная форма, соответствующая ей, получается из симметричной билинейной формы $\tilde{f}(x, y)$, которая задается формулой

$$\tilde{f}(x, y) = \frac{1}{2}(f(x, y) + f(y, x)).$$

Если коэффициенты билинейной формы $f(x, y)$ суть a_{ij}, то коэффициенты $\tilde{a}_{ij}$ билинейной формы $\tilde{f}(x, y)$ задаются по формуле

$$\tilde{a}_{ij} = \frac{1}{2}(a_{ij} + a_{ji}).$$

Квадратичная форма $f(x, x)$ называется *положительно определенной*, если она положительна при любом $x \neq 0$:

$$f(x, x) > 0 \text{ при } x \neq 0.$$

Докажем соотношение (32). Так как $f(x, y)$ есть линейная форма относительно x, то мы имеем

$$f(x, y) = f(x^\alpha e_\alpha, y) = x^\alpha f(e_\alpha, y). \tag{33}$$

Далее, так как $f(e_\alpha, y)$ есть линейная форма относительно y, то мы имеем

$$f(e_\alpha, y) = f(e_\alpha, y^\beta e_\beta) = y^\beta f(e_\alpha, e_\beta) = y^\beta a_{\alpha\beta}.$$

Подставляя последнюю формулу в формулу (33), получаем формулу (32).

Каждой билинейной форме $f(x, y)$ соответствует в базисе (30) матрица A ее коэффициентов. В связи с этим возникает вопрос, нельзя ли выбрать базис (30), т. е. систему координат в пространстве A^n, таким образом, чтобы матрица A получила наиболее простой вид. Эта задача будет решаться позже для случая симметричных форм или, что то же самое, для случая квадратичных форм.

Е) Симметричная билинейная форма (x, y) двух переменных x и y однозначно определяется соответствующей квадратичной формой как функцией одной переменной.

В самом деле, мы имеем

$$f(x+y, x+y) = f(x, x) + f(x, y) + f(y, x) + f(y, y).$$

Но так как $f(x, y) = f(y, x)$, то из этого получаем

$$f(x, y) = \frac{1}{2}[f(x+y, x+y) - f(x, x) - f(y, y)].$$

Здесь слева стоит билинейная форма $f(x, y)$, а справа — квадратичная форма от переменных x, y и $x+y$.

§ 3. Определители

В линейной алгебре играют важную роль так называемые определители или детерминанты.

Каждой квадратной матрице

$$X = \| x^i_j \|, \qquad i = 1, \ldots, n; \quad j = 1, \ldots, n,$$

определенным образом ставится в соответствие некоторое число $D(X)$, являющееся вполне определенной функцией элементов матрицы X. Эта функция $D(X)$ называется *определителем* или *детерминантом* матрицы X.

Определители обладают рядом замечательных свойств, которыми и объясняется их роль в алгебре. Сформулируем заранее свойства a) и b), являющиеся основными.

a) Пусть

$$x^i = (x^i_1, x^i_2, \ldots, x^i_n)$$

— i-я строка матрицы X. Оказывается, что если рассматривать x^i как n-мерный вектор, то определитель $D(X)$ является его линейной формой, т. е. записывается в виде

$$D(X) = d^\alpha x^i_\alpha, \qquad \alpha = 1, \ldots, r,$$

где коэффициенты этой линейной формы, т. е. числа

$$d^1, d^2, \ldots, d^n,$$

уже не зависят от элементов i-й строки матрицы X.

b) Пусть X' — матрица, получающаяся из матрицы X перестановкой местами двух ее строк. Тогда оказывается, что

$$D(X') = -D(X),$$

т. е. при перестановке двух строк матрицы X определитель ее меняет знак.

Если присоединить к этим двум свойствам a) и b) еще нижеследующее свойство c), то этими тремя свойствами, как мы докажем позже, функция $D(X)$ определяется однозначно.

c) Если $X = E$ — единичная матрица, то мы имеем

$$D(E) = 1,$$

т. е. определитель единичной матрицы равен единице.

Конструктивное определение функции $D(X)$ довольно сложно. Изложению полной конструкции предпошлю некоторое ее описание.

Функция $D(X)$ определяется как сумма взятых с определенными знаками произведений вида

$$x^1_{j_1} x^2_{j_2} \ldots x^n_{j_n}, \tag{34}$$

где последовательность

$$b = (j_1, j_2, \ldots, j_n)$$

состоит из расположенных в некотором порядке натуральных чисел

$$a = (1, 2, \ldots, n).$$

Таким образом, произведение (34) содержит один множитель, взятый из каждой строки, и один множитель, взятый из каждого столбца. Знак, с которым произведение (34) входит в сумму, составляющую $D(X)$, определяется довольно сложно. Он зависит от того, в каком отношении порядок чисел последовательности b находится к порядку чисел последовательности a. Перед тем как перейти к описанию выбора знака в произведении (34), дадим формулу для определителя $D(X)$ в случае $n = 2$.

При $n = 2$ имеем

$$D(X) = x^1_1 x^2_2 - x^1_2 x^2_1$$

(см. гл. 4, пример 5).

А) Перестановки. Пусть

$$a = (1, 2, \ldots, n)$$

— последовательность натуральных чисел от 1 до n и

$$b = (j_1, j_2, \ldots, j_n)$$

— последовательность тех же чисел, расположенных в каком-то порядке, быть может, и прежнем. Последовательность b называется *перестановкой* последовательности a. В дальнейших рассмотрениях не играет существенной роли, что последовательность a состоит из первых n натуральных чисел, расположенных в возрастающем порядке. Важно, что a есть последовательность из n различных символов, расположенных в определенном порядке, а b есть последовательность тех же символов, расположенных также в определенном порядке, быть может, совпадающем с порядком последовательности a. Последовательность символов b называется перестановкой последовательности a.

Если (u, v) — произвольная пара символов из последовательности a (подразумевается, что u и v — два различных символа), то в последовательности b они могут стоять в том же порядке, что и в a, или в противоположном. В первом случае говорят, что порядок в последовательности b на паре (u, v) сохранен, а во втором нарушен. Если число таких пар (u, v), на которых порядок при переходе от a к b нарушен, четно, то последовательность b называется *четной перестановкой* последовательности a, в противном случае — *нечетной*. Если перестановка b четная, то ей ставится в соответствие число $+1$, если перестановка b нечетная, то ей ставится в соответствие число -1. Это записывается в следующем виде:

$$\left[\frac{a}{b}\right] = +1, \quad \left[\frac{a}{b}\right] = -1.$$

Согласно определению последовательность a может рассматриваться как перестановка последовательности b, и потому определено число $\left[\frac{b}{a}\right] = \pm 1$. Из самого способа подсчета четности или нечетности следует равенство

$$\left[\frac{a}{b}\right] = \left[\frac{b}{a}\right].$$

Это следует из того, что пару (u, v) можно брать как в последовательности a, так и в последовательности b, и факт нарушения порядка имеет место в обоих случаях. Если последовательность b' получается из последовательности b путем перестановки в b местами двух символов, то мы будем говорить, что b' получена из b путем парной перестановки. Среди парных перестановок следует выделить парные перестановки двух соседних символов. Оказывается, что имеют место четыре следующих свойства перестановок:

1) При парной перестановке двух соседних символов в b четность перестановки b меняется на противоположную. Таким образом, если b' получена из b путем парной перестановки соседних символов, то мы имеем

$$\left[\frac{a}{b}\right] = -\left[\frac{a}{b'}\right]. \tag{35}$$

2) Всякая парная перестановка символов последовательности b может быть получена путем последовательного применения нечетного числа парных перестановок

соседних символов. Отсюда следует, что при парной перестановке справедлива та же формула (35), что и при парной перестановке соседних символов, т. е. если b' получена из b путем парной перестановки, то

$$\left[\frac{a}{b}\right] = -\left[\frac{a}{b'}\right].$$

3) Всякая перестановка b может быть получена из a путем конечного числа парных перестановок.

4) Если a, b и c — три перестановки из n символов, то имеет место формула

$$\left[\frac{a}{c}\right] = \left[\frac{a}{b}\right]\left[\frac{b}{c}\right].$$

Докажем свойства 1)—4).

Докажем 1). В последовательности b возьмем два соседних символа j_k, j_{k+1}. Если на паре (j_k, j_{k+1}) есть нарушение порядка, то на паре (j_{k+1}, j_k) его уже не будет, и наоборот, если на паре (j_k, j_{k+1}) нет нарушения порядка, то на паре (j_{k+1}, j_k) оно уже будет. Если последовательность b' получена из последовательности b путем перестановки символов j_k и j_{k+1}, то на подсчет числа нарушений порядка может влиять только пара (j_k, j_{k+1}), других изменений при переходе от b к b' не произойдет. Таким образом, свойство 1) доказано.

Докажем свойство 2). Пусть j_k, j_l — два символа из последовательности b, подвергающейся парной перестановке, причем $k < l$. Отрезок символов в b, начинающийся с j_k и кончающийся j_l, заменим символами

$$1,\ 2,\ \ldots,\ p-2,\ p-1,\ p.$$

В этой последовательности произведем парную перестановку двух последних соседних между собой символов, т. е. чисел $p-1$ и p. Тогда получим последовательность

$$1,\ 2,\ \ldots,\ p-2,\ p,\ p-1.$$

В этой последовательности произведем парную перестановку соседних символов $p-2$ и p. Тогда получим последовательность

$$1,\ 2,\ \ldots,\ p,\ p-2,\ p-1.$$

Передвигая таким образом число p налево, мы поставим его в конце концов на первое место, т. е. получим последовательность

$$p,\ 1,\ 2,\ \ldots,\ p-2,\ p-1,$$

произведя при этом $p-1$ парную перестановку соседних символов. Теперь точно таким же способом в последней последовательности перегоним символ 1 в конец последовательности, производя парные перестановки соседних символов. При этом мы должны будем произвести $p-2$ парных перестановки соседних символов. После этого произойдет парная перестановка символа 1 и символа p в результате $2p-3$ парных перестановок соседних символов. Итак, утверждение 2) доказано.

Утверждение 3) будем доказывать индуктивно по числу n. Пусть b — та перестановка, которую мы хотим получить из последовательности a путем конечного числа парных перестановок. В последовательности b имеется число n. Путем одной парной перестановки его можно поставить на последнее место. Это очевидно. Если n уже стоит на последнем месте, нам никакой перестановки делать не нужно. Пусть теперь в последовательности b символ n уже стоит на последнем месте. Тогда последовательность

$$\hat{b}=(j_1, j_2, \ldots, j_{n-1})$$

является перестановкой последовательности

$$\hat{a}=(1, 2, \ldots, n-1).$$

В силу предположения индукции последовательность $\hat{b}$ можно получить из последовательности $\hat{a}$ путем конечного числа парных перестановок. Таким образом, мы перевели последовательность a в последовательность b путем конечного числа парных перестановок. Утверждение 3) доказано.

Докажем 4). Пусть μ — число парных перестановок, при помощи которых можно перевести последовательность a в последовательность b, а ν — число тех парных перестановок, при помощи которых последовательность b можно перевести в последовательность c. Тогда последовательность a можно перевести в последовательность c при помощи $\mu+\nu$ парных перестановок. Таким образом.

$$\left[\frac{a}{b}\right]=(-1)^{\mu}, \quad \left[\frac{b}{c}\right]=(-1)^{\nu}, \quad \left[\frac{a}{c}\right]=(-1)^{\mu+\nu}.$$

Итак, свойство 4) доказано.

Следовательно, предложение А) полностью доказано.

Для иллюстрации предложения А) рассмотрим все перестановки последовательности

$$a = (1, 2, 3),$$

состоящей из трех символов, и выясним четность и нечетность этих перестановок.

Пусть

$$b = (1, 2, 3),$$

так что b совпадает с a. В последовательности b ни на одной паре символов не нарушается порядок, имеющийся в a, поэтому перестановка b четна.

Пусть

$$b = (2, 1, 3).$$

В этой последовательности b имеется нарушение порядка на паре $(1, 2)$. В последовательности a она стоит в порядке $(1, 2)$, а в последовательности b — в порядке $(2, 1)$. На других парах нарушения порядка нет, поэтому последняя перестановка b в этом случае нечетна.

Пусть

$$b = (1, 3, 2).$$

В этой перестановке имеется нарушение порядка на паре $(2, 3)$. В последовательности a она стоит в порядке $(2, 3)$, а в последовательности b — в порядке $(3, 2)$. В других парах нарушения порядка нет, поэтому перестановка b нечетна.

Пусть

$$b = (2, 3, 1).$$

В этой перестановке b имеется нарушение порядка на парах $(1, 2)$ и $(1, 3)$. Других нарушений порядка нет, поэтому перестановка b в этом случае четна.

Пусть

$$b = (3, 1, 2).$$

В этой перестановке b имеется нарушение порядка на парах $(1, 3)$ и $(2, 3)$. Других нарушений порядка нет, поэтому перестановка в этом случае четна.

Пусть

$$b = (3, 2, 1).$$

В этой перестановке b имеется нарушение порядка на парах $(1, 2)$, $(1, 3)$, $(2, 3)$. Таким образом, перестановка b нечетна.

Перейдем теперь к конструктивному описанию определителя.

B) Согласно определению, данному в предложении A), последовательность a' из n различных символов называется перестановкой последовательности a из тех же n символов, если в a' переписаны те же символы, которые входят в a, но в некотором порядке, быть может, совпадающем с a, а может быть, отличном от a. Оказывается, что переход от последовательности a к последовательности a' однозначно определяет переход от любой перестановки b последовательности a к некоторой определенной перестановке b' последовательности a'. Таким образом, переход от последовательности a к последовательности a' определяет некоторую операцию перехода от любой перестановки b к некоторой перестановке b'. Эту операцию мы обозначим через φ и будем называть *операцией перестановки*. Операция перестановки φ строится следующим образом. Выпишем под последовательностью a последовательность b так, чтобы под каждым элементом x последовательности a оказался некоторый элемент y последовательности b. Эти два элемента составят столбик $\begin{smallmatrix} x \\ y \end{smallmatrix}$. Так мы получили последовательность из n столбиков. Если мы подвергнем эту последовательность из n столбиков некоторой перестановке, то она определит некоторую перестановку последовательности a и некоторую перестановку последовательности b, причем перестановки эти тесно связаны между собой. Поэтому обозначим через φa перестановку, полученную из a, и через φb перестановку, полученную из b. Ясно, что перестановку столбиков можно выбрать так, чтобы $\varphi(a) = a'$. Тогда положим $b' = \varphi b$. Так мы получили некоторую перестановку b' последовательности b, однозначно определенную перестановкой a' последовательности a. Оказывается, что переход от последовательности b к ее перестановке b' определяет ту же операцию перестановки φ, которую определил переход от a к a'.

Докажем это. Пусть c — некоторая перестановка последовательности a и c' — та перестановка последовательности c, которая индуцируется переходом от a к a'. Докажем, что переход от b к b' индуцирует тот же самый переход от c к c'. Для того чтобы убедиться в этом, достаточно выписать под последовательностью a сперва последовательность b, а под b последовательность c так, чтобы образовались трехэтажные столбики вида $\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}$.

Оказывается, что имеют место нижеследующие утверждения 1), 2), 3):

1) Если a и b — две различные перестановки, то φa и φb также различны. Для доказательства этого выпишем последовательности a и b друг под другом. Так как последовательности a и b различны, то существует столбик $\begin{smallmatrix} x \\ y \end{smallmatrix}$, в котором нижний элемент не совпадает с верхним. Ясно, что при любой перестановке полученных так столбиков из столбика получится столбик из тех же двух элементов, различных между собой, так что $\varphi a \neq \varphi b$.

2) Если φ_1 и φ_2 — две операции перестановки, совпадающие на a, т. е. если

$$\varphi_1 a = \varphi_2 a,$$

то операции перестановки φ_1 и φ_2 одинаковы на любой другой перестановке b, т. е. из $\varphi_1 a = \varphi_2 a$ следует $\varphi_1 b = \varphi_2 b$, иначе говоря, $\varphi_1 = \varphi_2$. Для доказательства этого напишем под последовательностью a последовательность b. Каждой из перестановок φ_1 и φ_2 последовательности a соответствует перестановка столбиков. Но так как на a φ_1 и φ_2 совпадают, то перестановки столбиков, соответствующие φ_1 и φ_2, одинаковы. Поэтому $\varphi_1 b = \varphi_2 b$.

3) Имеет место равенство

$$\left[\frac{a}{b}\right] = \left[\frac{\varphi a}{\varphi b}\right]. \tag{36}$$

Для доказательства этой формулы выпишем друг под другом последовательности a и b. Операцию перестановки φ можно рассматривать как перестановку полученных столбиков. Согласно предложению А) она может быть осуществлена путем некоторого числа парных перестановок. То же число парных перестановок достаточно для перехода от a к φa и от b к φb. Таким образом, равенство (36) доказано.

Определение 4. Пусть

$$X = \| x_j^i \|, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n, \tag{37}$$

— квадратная матрица размеров (n, n). Ее мы будем называть *квадратной матрицей порядка* n. Пусть

$$a = (i_1, i_2, \ldots, i_n)$$

— некоторая перестановка последовательности натуральных чисел $1, 2, \ldots, n$,

$$b = (j_1, j_2, \ldots, j_n)$$

— перестановка последовательности a. Тогда определитель $D(X)$ задается формулой

$$D(X) = \sum_b \left[\frac{a}{b}\right] x_{j_1}^{i_1} x_{j_2}^{i_2} \ldots x_{j_n}^{i_n}. \tag{38}$$

Здесь суммирование ведется по всем перестановкам b последовательности a.

Докажем, что если вместо последовательности a взять любую ее перестановку

$$a' = (i'_1, i'_2, \ldots, i'_n) = \varphi a,$$

а вместо последовательности b взять последовательность

$$b' = (j'_1, j'_2, \ldots, j'_n) = \varphi b,$$

то определитель $D(X)$ можно задать формулой

$$D(X) = \sum_{b'} \left[\frac{a'}{b'}\right] x_{j'_1}^{i'_1} \ldots x_{j'_n}^{i'_n}. \tag{39}$$

Докажем формулу (39). Заметим прежде всего, что если выписать друг под другом последовательности a и b так, как это сделано в предложении В), то операция φ перестановки столбиков равносильна перемене порядка множителей в одночлене $x_{j_1}^{i_1} \ldots x_{j_n}^{i_n}$, причем одночлен этот не меняется. Таким образом, мы имеем равенство

$$x_{j_1}^{i_1} \ldots x_{j_n}^{i_n} = x_{j'_1}^{i'_1} \ldots x_{j'_n}^{i'_n}. \tag{40}$$

В силу утверждения 3) предложения В)

$$\left[\frac{a'}{b'}\right] = \left[\frac{a}{b}\right]. \tag{41}$$

Из формул (40) и (41) следует равенство (39).

Таким образом, в формуле (38) последовательность a может быть заменена любой ее перестановкой a', а b — соответствующей перестановкой b' (см. В)).

Определение 5. Каждой матрице

$$X = \| x_j^i \|, \quad i = 1, \ldots, p, \quad j = 1, \ldots, q,$$

размеров (p, q) взаимно однозначно ставится в соответствие *транспонированная к ней матрица* размеров (q, p), которая обозначается через X^{T} Для того чтобы записать матрицу X^{T} в виде формулы, положим

$$X^{\mathsf{T}} = U = \| u_j^i \|, \quad i = 1, \ldots, q, \quad j = 1, \ldots, p.$$

Тогда элементы матрицы U задаются формулой

$$u_j^i = x_i^j.$$

Для того чтобы наглядно описать процесс получения матрицы X^{T}, отметим, что матрица X^{T} получается из матрицы X путем поворота последней вокруг главной диагонали. Главная диагональ матрицы X — это линия, идущая из верхнего левого угла x_1^1 направо вниз под углом 45°. Тогда при повороте матрицы X ее строки превращаются в столбцы матрицы X^{T}, а столбцы X в строки X^{T}.

C) Пусть X — матрица размеров (p, q) и Y — матрица размеров (q, r). Тогда можно составить произведение

$$Z = XY,$$

причем матрица Z имеет размеры (p, r). Оказывается, что имеет место соотношение

$$Z^{\mathsf{T}} = Y^{\mathsf{T}} X^{\mathsf{T}}. \tag{42}$$

Для доказательства этой формулы положим

$$X^{\mathsf{T}} = U, \quad Y^{\mathsf{T}} = V, \quad Z^{\mathsf{T}} = W.$$

Пусть

$$X = \| x_j^i \|, \quad Y = \| y_k^j \|, \quad Z = \| z_k^i \|.$$

Для транспонированных матриц U, V, W имеем

$$u_j^i = x_i^j, \quad v_k^j = y_j^k, \quad w_k^i = z_i^k.$$

В силу этих формул получаем

$$w_k^i = z_i^k = x_\beta^k y_i^\beta = y_i^\beta x_\beta^k = v_\beta^i u_k^\beta.$$

Таким образом, формула (42) доказана.

Теорема 1. *Пусть X — произвольная квадратная матрица порядка n* (см. (37)), *а X^{T} — транспонированная к ней матрица. Тогда X^{T} также есть квадратная матрица порядка n. Оказывается, что имеет место равенство*

$$D(X^{\mathsf{T}}) = D(X). \tag{43}$$

Таким образом, матрица X и транспонированная к ней матрица X^{T} имеют равные определители.

Доказательство. Для доказательства формулы (43) положим

$$X^{\mathrm{T}} = U = \| u_j^i \|.$$

При этом

$$u_j^i = x_i^j. \tag{44}$$

В силу определения 4 имеем

$$D(U) = \sum_b \left[\frac{a}{b}\right] u_{j_1}^{i_1} u_{j_2}^{i_2} \ldots u_{j_n}^{i_n}, \tag{45}$$

где

$$a = (i_1, i_2, \ldots, i_n)$$

— некоторая фиксированная перестановка из n первых натуральных чисел, а

$$b = (j_1, j_2, \ldots, j_n)$$

— произвольная перестановка из n первых натуральных чисел. Суммирование ведется по всем перестановкам b. Заменяя элементы матрицы U элементами матрицы X по формуле (44), получаем

$$D(U) = \sum_b \left[\frac{a}{b}\right] x_{i_1}^{j_1} x_{i_2}^{j_2} \ldots x_{i_n}^{j_n}. \tag{46}$$

Здесь в отличие от формулы (38) последовательность нижних индексов (см. (46)) считается фиксированной, а последовательность верхних индексов произвольной и суммирование ведется по этой произвольной перестановке верхних индексов. В формуле (38) фиксированной является последовательность верхних индексов, а последовательность нижних индексов является произвольной, и по ней ведется суммирование. Для того чтобы преобразовать сумму (46) в сумму (38), мы в каждом слагаемом суммы (46) переставим множители таким образом, чтобы верхние индексы составили некоторую фиксированную последовательность

$$\hat{a} = (\hat{i}_1, \hat{i}_2, \ldots, \hat{i}_n).$$

При этом окажется, что нижние индексы составят произвольную перестановку

$$\hat{b} = (\hat{j}_1, \hat{j}_2, \ldots, \hat{j}_n)$$

из первых n натуральных чисел. Выпишем теперь последовательность b под последовательностью a так, чтобы образовалась двухэтажная пара строк. Заметим, что перестановке множителей в слагаемом суммы соответствует перестановка столбцов в этой полученной двухэтажной последовательности. Вместо того чтобы говорить о перестановке множителей, мы можем теперь говорить о перестановке столбцов.

Обозначим через φ_b такую операцию перестановки (см. В)), что

$$\varphi_b b = \hat{a}.$$

Из утверждения 1) в В) следует, что двум различным перестановкам b_1, b_2 соответствуют различные операции перестановки φ_{b_1}, φ_{b_2}. Действительно, если бы имело место равенство $\varphi_{b_1} = \varphi_{b_2}$ при $b_1 \neq b_2$, то две различные операции перестановки φ_{b_1} и φ_{b_2} переводили бы b_1 и b_2 в одну и ту же перестановку $\hat{a}$, что, согласно утверждению 1) п. В), невозможно. Таким образом, операции перестановок φ_b все различны и число их равно числу перестановок из n элементов. Положим теперь

$$\hat{b} = \varphi_b a. \tag{47}$$

Как было установлено в утверждении 2) в В), двум различным перестановкам b_1 и b_2 соответствуют две различные операции φ_{b_1} и φ_{b_2}. Поэтому число так полученных перестановок (см. (47)) первых n натуральных чисел равно числу перестановок из n элементов. Таким образом, $\varphi_b a = \hat{b}$ пробегает совокупность всех различных перестановок из n первых натуральных чисел. Так как в силу утверждения 3) в В) $\left[\frac{a}{b}\right] = \left[\frac{\varphi_b a}{\varphi_b b}\right]$, то имеем

$$\left[\frac{\varphi_b a}{\varphi_b b}\right] = \left[\frac{\varphi_b b}{\varphi_b a}\right] = \left[\frac{\hat{a}}{\hat{b}}\right].$$

Теперь формула (45) приобретает вид

$$D(U) = \sum_{\hat{b}} \left[\frac{\hat{a}}{\hat{b}}\right] x_{\hat{j}_1}^{\hat{i}_1} x_{\hat{j}_2}^{\hat{i}_2} \ldots x_{\hat{j}_n}^{\hat{i}_n},$$

где

$$\hat{a} = (\hat{i}_1, \hat{i}_2, \ldots, \hat{i}_n)$$

— некоторая фиксированная последовательность из n первых натуральных чисел, а

$$\hat{b} = (\hat{j}_1, \hat{j}_2, \ldots, \hat{j}_n)$$

— произвольная перестановка из первых n натуральных чисел. Так как $\varphi_b(a) = \hat{b}$ пробегает совокупность всех различных перестановок из первых n натуральных чисел, то суммирование здесь можно вести по $\hat{b}$. Таким образом, формула (46) переписывается в виде:

$$D(U) = \sum_{\hat{b}} \left[\frac{a}{\hat{b}}\right] x_{\hat{i}_1}^{\hat{i}_1} x_{\hat{i}_2}^{\hat{i}_2} \ldots x_{\hat{i}_n}^{\hat{i}_n},$$

а эта формула совпадает с формулой (38). Таким образом, теорема 1 доказана.

Так как при переходе к транспонированной матрице строки и столбцы меняются местами, то в силу теоремы 1 из всякого результата, относящегося к определителю, сформулированному в терминах строк, следует аналогичный результат для определителей, сформулированный в терминах столбцов. Из свойств а) и b), сформулированных в начале этого параграфа в отношении строк, будут следовать следующие свойства а′) и b′), относящиеся к столбцам:

а′) Определитель $D(X)$ матрицы X является линейной формой относительно элементов каждого столбца матрицы X.

b′) При перестановке местами столбцов матрицы X определитель $D(X)$ меняет знак.

§ 4. Решение системы линейных уравнений

Здесь будут описаны некоторые свойства определителей, дающие, в частности, возможность их вычисления не при помощи громоздкой формулы (38), а несколько проще.

А) Если в матрице

$$X = \|x_j^i\|, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n,$$

переставить местами две соседние строки — или, как мы будем говорить, пользуясь терминологией, введенной в § 3, «произвести парную перестановку соседних строк», — то вновь полученная матрица U будет иметь определитель, отличающийся от определителя матрицы X знаком, т. е. будет иметь место формула

$$D(U) = -D(X). \tag{48}$$

Так как каждая парная перестановка символов некоторой последовательности может быть получена в результате нечетного числа парных перестановок соседних символов, то из сказанного следует, что если матрица U получена из матрицы X путем перестановки местами двух ее различных строк, то имеет место формула

$$D(U) = -D(X). \tag{49}$$

Точно так же, если матрица V получена из матрицы X парной перестановкой двух столбцов, то имеет место равенство

$$D(V) = -D(X). \tag{50}$$

Из формулы (49), в частности, следует, что если в матрице X есть две равные строки с различными номерами, то определитель $D(X)$ матрицы X равен нулю:

$$D(X) = 0.$$

То же верно и для столбцов (см. (50)).

При доказательстве формулы (48) будем считать, что матрица U получается из матрицы X перестановкой k-й и $(k+1)$-й строк матрицы X. Поэтому элементы матрицы $U = \|u^i_j\|$ определяются формулами

$$u^i_j = x^i_j, \quad i \neq k, \quad i \neq k+1,$$

$$u^k_j = x^{k+1}_j, \quad u^{k+1}_j = x^k_j.$$

Согласно определению 4 (см. (38)) имеем

$$D(U) = \sum_b \left[\frac{a}{b}\right] u^1_{j_1} u^2_{j_2} \ldots u^n_{j_n},$$

где

$$a = (1, 2, \ldots, n), \quad b = (j_1, j_2, \ldots, j_n).$$

Производя в правой части последней формулы замену элементов матрицы U на элементы матрицы X, получим

$$D(U) = \sum_b \left[\frac{a}{b}\right] x^1_{j_1} x^2_{j_2} \ldots x^{k+1}_{j_k} x^k_{j_{k+1}} \ldots x^n_{j_n},$$

что можно иначе записать формулой

$$D(U) = \sum_b \left[\frac{a'}{b}\right] x^{i_1}_{j_1} x^{i_2}_{j_2} \ldots x^{i_n}_{j_n},$$

где

$$a' = (1, 2, \ldots, k-1, k+1, k, k+2, \ldots, n).$$

Так как в силу утверждения 4) п. А) § 3

$$\left[\frac{a'}{b}\right] = \left[\frac{a'}{a}\right]\left[\frac{a}{b}\right]$$

и очевидно, что $\left[\frac{a'}{a}\right] = -1$, то $\left[\frac{a'}{b}\right] = -\left[\frac{a}{b}\right]$. Пользуясь этим соотношением, получаем

$$D(U) = \sum_b \left[\frac{a'}{b}\right] x^{i_1}_{j_1} x^{i_2}_{j_2} \dots x^{i_n}_{j_n} = -D(X).$$

Соответствующее утверждение для столбцов, т. е. формула (50), следует из равноправия столбцов и строк. Таким образом, предложение А) доказано.

В) Пусть

$$X = \| x^i_j \|, \quad i = 1, \dots, p, \quad j = 1, \dots, q,$$

— произвольная, вообще говоря, неквадратная матрица размеров (p, q). Выделим две возрастающие последовательности натуральных чисел:

$$i_1 < i_2 < \dots < i_r, \quad r \leqslant p; \quad j_1 < j_2 < \dots < j_s, \quad s \leqslant q. \tag{51}$$

Рассмотрим теперь матрицу $\hat{X}$, состоящую из элементов матрицы X, входящих в строки с номерами $i_1, i_2, \dots, i_r$ и столбцы с номерами $j_1, j_2, \dots, j_s$. Именно, положим

$$\hat{X} = \left\| \begin{matrix} x^{i_1}_{j_1} & x^{i_1}_{j_2} & \dots & x^{i_1}_{j_s} \\ x^{i_2}_{j_1} & x^{i_2}_{j_2} & \dots & x^{i_2}_{j_s} \\ \hline x^{i_r}_{j_1} & x^{i_r}_{j_2} & \dots & x^{i_r}_{j_s} \end{matrix} \right\|.$$

Матрица $\hat{X}$ имеет размеры (r, s). Она некоторым определенным образом составлена из элементов матрицы X. Такого рода матрицы широко употребляются в алгебре, но для них нет специального названия. Я их буду называть подматрицами матрицы X. Для того чтобы задать матрицу $\hat{X}$, нужно указать номера строк и столбцов, отмеченных формулой (51). Будем употреблять несколько другую, но эквивалентную формулировку. Именно, говорят, что матрица $\hat{X}$ получена из матрицы X вычеркиванием всех строк, номера которых не совпадают с числами $i_1, i_2, \dots, i_r$, и всех столбцов, номера которых не совпадают с $j_1, j_2, \dots, j_s$.

С) Пусть

$$X = \| x^i_j \|, \quad i = 1, \dots, n, \quad j = 1, \dots, n, \tag{52}$$

— произвольная квадратная матрица порядка n. Пусть далее k — номер некоторой строки матрицы X, l — номер некоторого столбца матрицы X. Таким образом, в матрице X есть элемент x_l^k. В формуле (38), описывающей определитель $D(X)$, имеются члены, содержащие множители x_l^k. Их сумму обозначим через σ_l^k. Опишем эту сумму.

Вычеркнем из матрицы X k-ю строку и l-й столбец. Так полученную подматрицу матрицы X размеров $(n-1, n-1)$ обозначим через X_k^l. Определитель ее называется *минором* элемента x_l^k матрицы X. Оказывается, что сумма σ_l^k задается формулой

$$\sigma_l^k = (-1)^{k+l} x_l^k D(X_k^l). \qquad (53)$$

Докажем предложение С), именно формулу (53). Рассмотрим сперва случай, когда $k=1$, $l=1$, т. е. когда элемент x_1^1 стоит в верхнем левом углу матрицы X. Будем считать, что при построении суммы (38)

$$a = (1, 2, \dots, n), \quad b = (j_1, j_2, \dots, j_n),$$

где b — произвольная перестановка последовательности a. Слагаемое в сумме (38), содержащее множитель x_1^1 имеет вид

$$x_1^1 x_{j_2}^2 \dots x_{j_n}^n.$$

Все эти слагаемые соответствуют перестановке b, задаваемой формулой

$$b = (1, j_2, \dots, j_n).$$

Положим

$$\hat{a} = (2, 3, \dots, n), \quad \hat{b} = (j_2, j_3, \dots, j_n).$$

Теперь мы можем написать

$$a = (1, \hat{a}), \quad b = (1, \hat{b}).$$

Ясно, что для получения перестановки b из последовательности a путем парных перестановок нужно переставлять только символы, входящие в $\hat{a}$, таким образом, чтобы получить перестановку $\hat{b}$. Поэтому

$$\left[\frac{a}{b}\right] = \left[\frac{\hat{a}}{\hat{b}}\right].$$

Из этого следует, что

$$\sigma_1^1 = x_1^1 \sum_{\hat{b}} \left[\frac{\hat{a}}{\hat{b}}\right] x_{j_2}^2 x_{j_3}^3 \dots x_{j_n}^n.$$

Сумма, входящая в правую часть последнего равенства, очевидно, равна определителю матрицы X_1^1, так что мы получаем

$$\sigma_1^1 = x_1^1 D(X_1^1).$$

Для получения формулы (53) при произвольных k и l произведем парную перестановку k-й строки матрицы X с предшествующей строкой, т. е. со строкой с номером $k-1$. В полученной матрице произведем парную перестановку строки, стоящей на $(k-1)$-м месте, с предшествующей строкой. Так будем продолжать, пока не переставим k-ю строку матрицы X на первое место. При этом мы произведем $k-1$ парных перестановок соседних строк. Точно так же передвинем l-й столбец на первое место. Для этого будем передвигать его постепенно налево. Тогда мы произведем $l-1$ парных перестановок соседних столбцов. Полученную в результате этих операций матрицу обозначим через U. Так как при перестановках строк и столбцов матрицы X было произведено всего $(k-1)+(l-1)$ парных перестановок соседних строк и столбцов, то мы имеем (см. A))

$$D(U) = (-1)^{k+l} D(X). \tag{54}$$

В матрице U в левом верхнем углу стоит элемент x_l^k, а при вычеркивании из матрицы U первой строки и первого столбца мы получим подматрицу, которая совпадает с подматрицей матрицы X, получающейся вычеркиванием k-й строки и l-го столбца. Из формулы (54) следует формула (53). Итак, предложение C) доказано.

Дадим теперь определение разложения определителя матрицы по элементам k-й строки и элементам l-го столбца.

D) Имеют место две следующие важные формулы:

$$D(X) = (-1)^{k+\alpha} x_\alpha^k D(X_k^\alpha), \tag{55}$$

$$D(X) = (-1)^{l+\beta} x_l^\beta D(X_\beta^l). \tag{56}$$

Формула (55) дает разложение определителя по элементам k-й строки, а формула (56) дает разложение того же определителя по элементам l-го столбца.

Для доказательства формулы (55) соберем в сумме (38) сначала все члены, содержащие элемент x_1^k. При этом их сумма дается формулой (53). Затем соберем все члены суммы (38), содержащие множитель x_2^k, и, продолжая так далее, наконец, соберем все слагаемые, содержащие множитель x_n^k. Этим будет исчерпана без перекрытий вся сумма (38), так как в каждый ее член входит ровно один множитель из k-й строки. Таким образом, формула (55) верна. Точно так же доказывается формула (56).

E) Пусть X — произвольная квадратная матрица порядка n (см. (52)). Положим

$$x^k = (x_1^k, x_2^k, \ldots, x_n^k).$$

Здесь в правой части выписаны подряд все элементы k-й строки матрицы X. Мы хотим рассматривать их как координаты вектора x^k. Таким образом, мы будем понимать строку как вектор. Точно так же положим

$$x_l = (x_l^1, x_l^2, \ldots, x_l^n).$$

Здесь в правой части равенства выписаны элементы l-го столбца. Мы будем рассматривать их как координаты вектора x_l. Таким образом, мы будем понимать столбец матрицы X как некоторый вектор. Обратим теперь внимание на зависимость определителя $D(X)$ от элементов k-й строки и от элементов l-го столбца. Для этого положим $D(X) = D(x^k)$ и $D(X) = D(x_l)$. Из предложения C) следует, что функции $D(x^k)$ и $D(x_l)$ являются линейными формами относительно векторов x^k и x_l. Пусть

$$u = (u_1, u_2, \ldots, u_n)$$

— некоторый n-мерный вектор. Составим теперь сумму

$$D(X_k^u) = (-1)^{k+\alpha} u_\alpha D(X_k^\alpha). \tag{57}$$

Правая часть последнего равенства отличается от правой части равенства (55) только тем, что вместо вектора x^k, составленного из элементов k-й строки матрицы X, поставлены элементы вектора u и полученная так матрица обозначена через X_k^u. Таким образом, правая часть равенства (57) представляет собой определитель матрицы X_k^u, в которой k-я строка заменена строкой u. Проделаем то же самое для столбцов. Пусть

$$\theta = (\theta^1, \theta^2, \ldots, \theta^n)$$

— некоторый вектор размерности n. Составим сумму

$$D(X_\theta^l) = (-1)^{l+\beta} \theta^\beta D(X_\beta^l). \tag{58}$$

Правая часть последнего равенства отличается от правой части равенства (56) только тем, что вместо l-го столбца матрицы X поставлен столбец θ и полученная так матрица обозначена через X^l_θ. Так что правая часть равенства (58) представляет собой определитель матрицы X^l_θ, в которой l-й столбец заменен столбцом θ. Пусть теперь

$$u = x^m,$$

т. е. вектор u состоит из элементов строки с номером m матрицы X. Таким образом, сумма (57) дает нам определитель матрицы $X^{x^m}_k$, полученной из матрицы X заменой k-й строки матрицы X на m-ю ее строку. Таким образом, при $m = k$ формула (57) дает нам определитель $D(X)$, а при $m \neq k$ — нуль, так как это есть определитель матрицы, которая имеет две совпадающие строки. Итак, мы получаем

$$(-1)^{m+\alpha} x^m_\alpha D\left(X^\alpha_k\right) = \delta^m_k D(X). \tag{59}$$

Здесь в правой части стоит символ Кронекера δ^m_k, который равен единице, когда его индексы нижний и верхний равны между собой, и нулю, если индексы не равны между собой. Проделывая те же выкладки для l-го столбца, мы получим следующую формулу:

$$(-1)^{m+\beta} x^\beta_m D\left(X^l_\beta\right) = \delta^l_m D(X). \tag{60}$$

В случае, если $D(X) \neq 0$, мы можем соотношения (59) и (60) разделить на $D(X)$. Для того чтобы переписать эти соотношения в удобной для нас форме, положим

$$y^l_k = (-1)^{k+l} \frac{D\left(X^l_k\right)}{D(X)}. \tag{61}$$

Таким образом, формулы (59) и (60) переписываются теперь в виде

$$x^m_\alpha y^\alpha_k = \delta^m_k, \tag{62}$$

$$y^l_\beta x^\beta_m = \delta^l_m. \tag{63}$$

Положим теперь

$$Y = \|y^i_l\|.$$

Пользуясь этим обозначением, мы можем переписать формулу (62) в виде

$$XY = E, \tag{64}$$

где E — единичная матрица, а формулу (63) в виде

$$YX = E. \qquad (65)$$

Формулы (64) и (65) показывают, что матрица Y является *обратной* для матрицы X как справа, так и слева и потому обозначается через X^{-1}. Таким образом, мы доказали, что если определитель $D(X)$ не равен нулю, то у матрицы X существует обратная матрица $X^{-1} = Y$.

До сих пор рассмотрение определителей ничем не было оправдано. Теперь, пользуясь результатами, изложенными в предложении E), мы можем показать, что определители дают нам аппарат для решения важной и естественной алгебраической задачи, — именно, для решения системы уравнений первой степени с несколькими неизвестными. Мы будем рассматривать системы из n уравнений с n неизвестными. Перенеся все члены, содержащие неизвестные величины в левую сторону уравнения, а известные величины в правую сторону уравнения, мы можем записать систему таких уравнений в виде

$$a^i_\alpha x^\alpha = c^i, \quad i = 1, 2, \ldots, n, \quad \alpha = 1, 2, \ldots, n. \qquad (66)$$

Здесь координаты вектора

$$x = (x^1, x^2, \ldots, x^n)$$

являются неизвестными величинами, а сам вектор x — неизвестным вектором. Коэффициенты при этих неизвестных составляют матрицу

$$A = \|a^i_j\|; \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n.$$

Здесь координаты вектора

$$c = (c^1, c^2, \ldots, c^n)$$

являются правыми известными частями системы уравнений (66).

F) Будем трактовать векторы x и c как матрицы размеров $(n, 1)$, т. е. как одностолбцовые матрицы высоты n. Тогда систему (66) можно переписать в следующем виде:

$$Ax = c. \qquad (67)$$

Здесь слева стоит матрица A размеров (n, n), умноженная на одностолбцовую матрицу x размеров $(n, 1)$, а справа — одностолбцовая матрица c размеров $(n, 1)$. Если определитель $D(A)$ матрицы A отличен от нуля:

$$D(A) \neq 0,$$

то уравнение (67) имеет решение

$$x = A^{-1}c \tag{68}$$

и притом единственное.

Следует отметить, что если $D(A) = 0$, то решение системы (67) все же может существовать.

Для получения решения системы (67) в случае $D(A) \neq 0$ достаточно умножить соотношение (67) на матрицу A^{-1} слева. Решение это единственно. Допустим, что существуют два решения x_1 и x_2. Тогда мы имеем два соотношения

$$Ax_1 = c, \quad Ax_2 = c.$$

Вычитая второе из этих равенств из первого, получим

$$A(x_1 - x_2) = 0.$$

Умножив это соотношение на A^{-1}, получим

$$x_1 - x_2 = 0.$$

Таким образом, $x_1 = x_2$, и наши решения совпадают между собой.

Дадим теперь более традиционную запись решения системы уравнений (66).

Теорема 2. *Решение системы* (66) *дается формулой*

$$x^l = \frac{D(A_c^l)}{D(A)}, \tag{69}$$

где матрица A_c^l получается из матрицы A заменой ее l-го столбца столбцом c. Таким образом, для получения неизвестной величины x^l нужно в матрице A заменить ее l-й столбец столбцом c и определитель этой матрицы разделить на определитель $D(A)$.

Доказательство. Для доказательства формулы (69) мы воспользуемся формулой (68), написав ее в развернутом виде.

Пусть

$$A^{-1} = B = \| b_j^i \|.$$

Тогда в силу (61) имеем

$$b_j^i = (-1)^{i+j} \frac{D(A_i^j)}{D(A)}.$$

Формула (68) запишется в виде

$$x^i = b^i_\beta c^\beta = (-1)^{i+\beta} \frac{D(A^i_\beta)}{D(A)} c^\beta =$$

$$= \frac{(-1)^{i+\beta} c^\beta D(A^i_\beta)}{D(A)} = \frac{D(A^i_c)}{D(A)}$$

(см. (58)). Таким образом, теорема доказана.

§ 5. Элементарные преобразования матриц

В этом параграфе будут описаны так называемые элементарные преобразования матриц и их определителей. Будет показано, что элементарным преобразованием матрицу можно привести к более простому виду, удобному для вычисления ее определителя. Такими удобными для вычисления определителей видами матриц являются так называемые треугольные и диагональные матрицы. Их определение будет дано в предложении А) этого параграфа. После описания элементарных операций мы применим полученные конструкции для доказательства важнейшей теоремы, — именно, теоремы о том, что определитель произведения двух квадратных матриц равен произведению их определителей. Это будет заключительная теорема настоящего параграфа.

А) Пусть

$$X = \| x^i_j \|$$

— квадратная матрица порядка n. *Главной диагональю* матрицы X называется прямолинейный отрезок, ведущий из левого верхнего угла x^1_1 в правый нижний, т. е. к элементу x^n_n. Элементы, лежащие на этом прямолинейном отрезке, т. е. элементы

$$x^1_1, \; x^1_2, \; \ldots, \; x^n_n, \tag{70}$$

называются *диагональными элементами* матрицы X, они составляют ее главную диагональ. Матрица X называется *треугольной*, если все ее элементы, лежащие справа от диагонали, равны нулю, т. е. если выполнено соотношение

$$x^j_i = 0 \quad \text{при} \quad j > i.$$

Частным случаем треугольной матрицы является так называемая *диагональная* матрица, в которой от нуля отличны только элементы матрицы, стоящие на ее главной диагонали, т. е. элементы (70). Оказывается, что определитель треугольной матрицы равен произведению ее диагональных элементов, т. е. если X — треугольная матрица, то

$$D(X) = x_1^1 x_2^2 \ldots x_n^n. \tag{71}$$

Докажем последнюю формулу. Определитель $D(X)$ треугольной матрицы X будем вычислять по формуле (38). Пусть

$$\left[\frac{a}{b}\right] x_{j_1}^1 x_{j_2}^2 \ldots x_{j_n}^n \tag{72}$$

— отличное от нуля слагаемое в сумме (38). Здесь мы предположим, что

$$a = (1, 2, \ldots, n), \quad b = (j_1, j_2, \ldots, j_n).$$

Исходя из слагаемого (72), определим натуральное число k, обладающее тем свойством, что при $i \leqslant k$ $j_i = i$. Если натурального числа, обладающего этим свойством, нет, то считаем, что $k = 0$. Докажем, что $k = n$. Допустим противоположное, т. е. что $k < n$. Тогда в произведении (72) существует множитель $x_{j_{k+1}}^{k+1}$. Так как $j_{k+1} \neq k + 1$, то имеются две возможности: $j_{k+1} > k + 1$ и $j_{k+1} < k + 1$. Если имеет место первое из этих неравенств, то в силу треугольности матрицы X множитель $x_{j_{k+1}}^{k+1} = 0$ и одночлен (71) вообще не рассматривается. Если имеет место второе неравенство, т. е. $j_{k+1} < k + 1$, то при $i = j_{k+1}$ мы имеем $j_i = i = j_{k+1}$. Таким образом, в ряду нижних индексов одночлена (71) имеется два равных: j_i и j_{k+1}, что невозможно, так как b есть перестановка натуральных чисел a. Так, мы имеем

$$j_1 = 1, \; j_2 = 2, \; \ldots, \; j_n = n,$$

и потому $\left[\frac{a}{b}\right] = 1$. Таким образом, единственное слагаемое, отличное от нуля, входящее в сумму (38), является произведением диагональных членов матрицы X, и мы получаем формулу (71).

В § 4 были доказаны важнейшие свойства определителя $D(X)$ матрицы X, сформулированные еще в начале § 3 (см. a), b) § 3). Эти свойства определителя могут быть сформулированы как в терминах строк, как это сделано в начале § 3, так и в терминах столбцов, что сделано в конце § 3 (см. a'), b') § 3). Однако определитель $D(X)$ является не единственной функцией матрицы, удовлетворяющей этим условиям. Если умножить функцию $D(X)$ на произвольную константу, то мы получим функцию, удовлетворяющую тем же условиям. Поэтому мы рассмотрим произвольную функцию матрицы X, удовлетворяющую указанным условиям, причем условия эти будем рассматривать в терминах столбцов, т. е. в виде a′), b′) § 3. Для того чтобы сформулировать высказанные условия, будем записывать матрицу X в следующем виде:

$$X = \| x_1 x_2 \ldots x_n \|.$$

Здесь $x_1, x_2, \ldots, x_n$ обозначены столбцы матрицы X. Их мы можем трактовать как одностолбцовые матрицы или как векторы.

B) Обозначим через $F(X)$ функцию квадратной матрицы X порядка n, удовлетворяющую двум следующим условиям:

1) При перестановке двух столбцов матрицы X функция $F(X)$ меняет знак (см. b′) § 3). В виде формулы это запишем так:

$$F(X) = -F(\| x_1 x_2 \ldots x_q \ldots x_p \ldots x_n \|), \qquad p < q.$$

Из этого следует, что если два столбца матрицы X совпадают, то $F(X) = 0$.

2) $F(X)$ является линейной формой относительно столбца x_k матрицы X. В виде формулы запишем это так. Пусть

$$x_k = cy + dz.$$

Тогда

$$F(\| x_1 x_2 \ldots x_k \ldots x_n \|) =$$
$$= cF(\| x_1 x_2 \ldots y \ldots x_n \|) + dF(\| x_1 x_2 \ldots z \ldots x_n \|).$$

В дальнейшем мы всегда будем обозначать через $F(X)$ функцию матрицы X, удовлетворяющую условиям 1) и 2).

C) Введем так называемые элементарные операции над матрицей X, при которых функция $F(X)$ либо меняет знак, либо остается неизменной.

1) Операция перестановки столбцов. При ней функция матрицы меняет знак.

2) Прибавление к одному из столбцов линейной комбинации нескольких других столбцов с действительными коэффициентами. При этой операции функция $F(X)$ не меняет своего значения. Напишем это в виде формулы,

для определенности, для n-го столбца следующим образом. Пусть

$$c^\alpha x_\alpha, \quad \alpha = 1, 2, \ldots, n-1,$$

— линейная форма относительно первых $n-1$ столбцов матрицы X. Тогда

$$F(X) = F(\|x_1 x_2 \ldots x_n + c^\alpha x_\alpha\|), \quad \alpha = 1, 2, \ldots, n-1. \tag{73}$$

Докажем последнюю формулу. В силу свойства 2) предложения B), которым обладает функция $F(X)$, мы имеем

$$F(X) = c^\alpha F(\|x_1 x_2 \ldots x_\alpha\|) + F(\|x_1 x_2 \ldots x_n\|). \tag{74}$$

Так как матрица $\|x_1 x_2 \ldots x_i\|$, $i \leqslant n-1$, имеет два равных столбца, — именно, i-й и последний, — то в правой части формулы (74) все члены, кроме последнего, равны нулю, и мы получаем из нее формулу (73).

Операция 2), прибавление к некоторому столбцу линейной комбинации нескольких других столбцов, часто будет состоять в прибавлении только одного столбца, умноженного на действительный коэффициент.

Теперь мы займемся доказательством важной теоремы 3. Перед тем как сформулировать ее, напомню, что линейной зависимостью между столбцами матрицы X называется зависимость между ее столбцами как векторами. Именно, столбцы $x_1, x_2, \ldots, x_n$ матрицы X считаются линейно зависимыми, если существуют действительные числа c^i, $i = 1, \ldots, n$, не все равные нулю, такие, что имеет место соотношение

$$c^\alpha x_\alpha = 0. \tag{75}$$

Теперь мы можем сформулировать и доказать теорему.

Теорема 3. *Определитель $D(X)$ квадратной матрицы $X = \|x_j^i\|$ тогда и только тогда равен нулю, когда столбцы матрицы X линейно зависимы.*

Доказательство. Теорема 3 состоит из двух утверждений: 1) Если столбцы матрицы линейно зависимы, то определитель ее равен нулю. 2) Если определитель $D(X)$ матрицы X равен нулю, то столбцы матрицы X линейно зависимы.

Первое утверждение мы докажем для функции $F(X)$. Так как определитель $D(X)$ матрицы X есть частный случай функции $F(X)$, то первая часть утверждения теоремы будет доказана.

Допустим, что имеет место линейная зависимость (75) между столбцами матрицы X, при этом хотя бы один коэффициент c^i, входящий в линейную зависимость (75), не равен нулю. Допустим, что это коэффициент c^n. Будем считать, что он равен единице. Тогда в силу формулы (73) мы получим

$$F(X)=F(\|x_1x_2\ldots x_{n-1}0\|)=0.$$

Допустим, что определитель $D(X)$ матрицы X равен нулю, и докажем, что столбцы ее линейно зависимы. Доказательство будем вести индуктивно, по порядку n матрицы X. При $n=1$ утверждение очевидно.

Если все элементы последней строки матрицы X равны нулю, то так как столбцов она имеет n, а размерность каждого столбца как вектора равна $n-1$, то в силу предложения В) § 1 эти столбцы как векторы линейно зависимы. Таким образом, мы должны рассмотреть случай, когда не все элементы последней строки матрицы X равны нулю. Допустим для определенности, что отличен от нуля последний элемент последней строки. Путем перестановки двух столбцов матрицы X, от чего может измениться только знак определителя $D(X)$, мы можем всегда добиться этого. Таким образом, нам следует рассмотреть лишь случай, когда $x_n^n\neq 0$. Рассмотрим теперь столбец матрицы X с номером i, т. е. вектор x_i. Можно подобрать такие действительные числа c_i, что вектор

$$x_i+c_ix_n$$

имеет своей последней координатой нуль. Произведем $n-1$ элементарных операций. Прибавляя к каждому столбцу с номером i, $i\leqslant n-1$, последний столбец с коэффициентом c_i, мы получим матрицу $\hat{X}$ со столбцами

$$x_1+c_1x_n,\ x_2+c_2x_n,\ \ldots,\ x_{n-1}+c_{n-1}x_n,\ x_n. \qquad (76)$$

В последней строке этой матрицы будут идти нули до последнего элемента x_n^n. Вычеркнем теперь в матрице $\hat{X}$ последний столбец и последнюю строку. Тогда мы получим матрицу $\hat{X}_n^n$. Разлагая определитель матрицы $\hat{X}$ по элементам последней строки, мы получим

$$D(\hat{X})=x_n^nD(\hat{X}_n^n).$$

Так как $D(\hat{X})=0$, $x_n^n\neq 0$, то

$$D(\hat{X}_n^n)=0.$$

Матрица $\hat{X}_n^n$ имеет порядок $n-1$, и по предположению индукции столбцы ее линейно зависимы. Столбцы эти совпадают с первыми $n-1$ столбцами, выписанными в строке (76), за исключением того, что в матрице $\hat{X}$ последние элементы в столбцах (76) все равны нулю. Именно поэтому можно выписать вместо зависимости столбцов матрицы $\hat{X}_n^n$ зависимость между первыми $n-1$ столбцами в формуле (76). Коэффициенты этой линейной зависимости обозначим через b^i. Тогда мы получим

$$b^{\alpha}(x_{\alpha}+c_{\alpha}x_n)=b^{\alpha}x_{\alpha}+b^{\alpha}c_{\alpha}x_n=0.$$

Это соотношение представляет собой линейную зависимость между столбцами матрицы X. Таким образом, теорема доказана.

Следствие теоремы 3. *Система линейных уравнений*

$$x_{\alpha}^{i}\xi^{\alpha}=0, \quad i=1, \ldots, n, \tag{77}$$

где $\xi^1, \xi^2, \ldots, \xi^n$ — неизвестные, тогда и только тогда имеет нетривиальное решение, т. е. такое решение, у которого не все ξ^j, $j=1, \ldots, n$, равны нулю, когда определитель матрицы X равен нулю:

$$D(X)=0.$$

Доказательство. Если определитель $D(X)$ матрицы X равен нулю, то в силу теоремы 3 столбцы матрицы X линейно зависимы, т. е. система (77) имеет нетривиальное решение. Если, наоборот, система (77) имеет нетривиальное решение, то столбцы матрицы X линейно зависимы и, следовательно, в силу теоремы 3 определитель $D(X)$ равен нулю.

Замечание. Если в матрице X столбцы линейно зависимы, то функция $F(X)$ обращается в нуль.

Это утверждение было доказано при доказательстве первой части теоремы.

Докажем теперь еще одну теорему. Она у нас будет играть вспомогательную роль, поэтому мы назовем ее леммой.

Лемма 1. *Если определитель $D(X)$ матрицы X отличен от нуля, то матрицу X элементарными преобразованиями можно привести к диагональному виду, причем все ее диагональные элементы отличны от нуля.*

Доказательство. Доказательство будем вести индуктивно по порядку n матрицы X. При $n=1$ утверждение очевидно.

Разложим определитель матрицы X по элементам последней строки (см. D) § 4). Запишем это разложение (см. (55)) в виде

$$D(X)=(-1)^{n+\alpha}x^n_\alpha D(X^\alpha_n).$$

Так как определитель $D(X)\neq 0$, то хотя бы одно из слагаемых, стоящих в правой части последнего равенства, отлично от нуля. Перестановкой столбцов матрицы X можно добиться того, чтобы отличным от нуля было последнее слагаемое, т. е. чтобы имели место неравенства

$$x^n_n\neq 0,\quad D(X^n_n)\neq 0.$$

Применяя к матрице X элементарную операцию прибавления последнего столбца, умноженного на некоторое число, к каждому из предыдущих столбцов матрицы X, можно достичь того, чтобы в полученной матрице вся последняя строка состояла из нулей, кроме последнего ее элемента x^n_n. Полученную матрицу вновь обозначим через X. В этой матрице X все элементы последней строки, кроме последнего элемента x^n_n, равны нулю. Добьемся теперь того же для последнего столбца.

Для этого составим $n-1$ уравнений с $n-1$ неизвестными

$$u^1,\ u^2,\ \ldots,\ u^{n-1}.$$

Система уравнений будет следующая:

$$x^i_\alpha u^\alpha=-x^i_n,\qquad i=1,\ 2,\ \ldots,\ n-1.$$

Эта система из $n-1$ уравнений с $n-1$ неизвестными уже рассмотрена нами в предыдущем параграфе. Так как определитель матрицы коэффициентов X^n_n отличен от нуля, система имеет решение. Следовательно, числа $u^1, u^2, \ldots, u^{n-1}$ существуют. Произведем теперь над матрицей X элементарную операцию,— именно, прибавим к последнему ее столбцу все предыдущие, т. е. столбцы

$$x_1,\ x_2,\ \ldots,\ x_{n-1},$$

умножив каждый столбец x_i на число u^i. Тогда в последнем столбце так полученной матрицы обратятся в нуль все элементы, кроме последнего x^n_n. Полученную матрицу вновь обозначим через X. В ней в последнем столбце и последней строке есть единственный элемент x^n_n, отличный от нуля. Определитель матрицы X^n_n для нее отличен от нуля. Каждая элементарная операция над матрицей X, меняющая только ее первые $n-1$ столбцов, т. е. столбцы $x_1, x_2, \ldots, x_{n-1}$, порождает элементарную операцию над столбцами матрицы X_n^n и наоборот. Это происходит потому, что в последней строке матрицы X все элементы, стоящие на местах $1, 2, \ldots, n-1$, равны нулю. По предположению индукции элементарными операциями матрицу X_n^n можно привести к диагональному виду, а все эти элементарные операции можно рассматривать как полученные в результате элементарных операций над столбцами $x_1, x_2, \ldots, x_{n-1}$ матрицы X. Так как матрица X_n^n приведена теперь по предположению индукции к диагональному виду, то полученная матрица X оказалась диагональной.

Итак, лемма 1 доказана.

Докажем теперь еще одну лемму.

Лемма 2. *Оказывается, что для функции $F(X)$ матрицы X, удовлетворяющей условиям* 1), 2) *предложения* В), *всегда выполнено соотношение*

$$F(X) = D(X) F(E), \tag{78}$$

где E — единичная матрица.

Доказательство. В случае, если $D(X)=0$, формула (78) верна, так как функция $F(X)=0$ (см. замечание к теореме 3).

Теперь рассмотрим случай, когда $D(X)\neq 0$. Для доказательства приведем матрицу X к диагональному виду, пользуясь элементарными операциями, что возможно в силу леммы 1. Каждая элементарная операция, заключающаяся в перестановке столбцов матрицы X, одновременно меняет знак функций $F(X)$ и $D(X)$. Таким образом, нам осталось доказать лемму 2 лишь в предположении, что матрица X диагональная и на диагонали стоят отличные от нуля числа.

Итак, мы будем предполагать, что матрица X диагональная с отличными от нуля диагональными членами, и будем рассматривать функцию $F(X)$. Функция эта является линейной формой относительно каждого столбца матрицы X. Поэтому из каждого столбца этой матрицы X мы можем вынести за знак функции $F(X)$ числовой множитель, — именно, из столбца с номером i числовой множитель x_i^i. В результате получим соотношение

$$F(X) = x_1^1 x_2^2 \ldots x_n^n F(E).$$

Оно получается в результате вынесения из каждого столбца множителя, стоящего на диагонали. В конце концов мы получим диагональную матрицу, где на диагонали стоят единицы, т. е. матрицу E. В силу предложения A) произведение $x_1^1 x_2^2 \ldots x_n^n$, стоящее перед $F(E)$, равно $D(X)$. Таким образом, формула (78) доказана.

Перейдем теперь к главной теореме настоящего параграфа, для которой были доказаны предыдущие две леммы.

Теорема 4. *Пусть X и Y — две квадратные матрицы одного и того же порядка n. Тогда имеет место следующая формула:*

$$D(XY) = D(X)D(Y). \tag{79}$$

Таким образом, определитель произведения двух квадратных матриц одного порядка равен произведению определителей этих матриц.

Доказательство. Положим

$$XY = Z = \| z_k^i \|, \quad X = \| x_j^i \|, \quad Y = \| y_k^j \|.$$

Покажем теперь, что элементарным операциям над матрицей Y соответствуют элементарные операции над матрицей Z. Начнем с перестановки столбцов. Пусть $V = \| v_k^j \|$ — матрица, полученная из матрицы Y перестановкой столбцов с номерами p и q. Тогда мы имеем

$$v_k^j = y_k^j \quad \text{при} \quad k \neq p, \quad k \neq q, \quad v_q^j = y_p^j, \quad v_p^j = y_q^j.$$

Положим

$$W = XV.$$

Тогда

$$w_k^i = x_\alpha^i y_k^\alpha = z_k^i \quad \text{при} \quad k \neq p, \quad k \neq q.$$

Далее,

$$\begin{aligned} w_p^i &= x_\alpha^i v_p^\alpha = x_\alpha^i y_q^\alpha = z_q^i, \\ w_q^i &= x_\alpha^i v_q^\alpha = x_\alpha^i y_p^\alpha = z_p^i. \end{aligned} \tag{80}$$

Таким образом, перестановке двух столбцов матрицы Y соответствует перестановка столбцов матрицы Z с теми же номерами.

Докажем теперь, что

$$F(Y) = D(XY)$$

является линейной формой относительно k-го столбца y_k матрицы Y.

Пусть h_1 и h_2 — два n-мерных вектора. Подставим вместо k-го столбца матрицы Y линейную комбинацию этих двух векторов, т. е. положим

$$y_k = c_1 h_1 + c_2 h_2.$$

Полученную таким образом матрицу обозначим через Z.

Подставляя на место k-го столбца y_k матрицы Y столбцы h_1 и h_2, мы получим матрицы, которые обозначим через H_1 и H_2. Положим, далее,

$$P = XH_1, \quad S = XH_2;$$

k-й столбец матрицы P обозначим через p_k, а k-й столбец матрицы S обозначим через s_k. Мы имеем теперь

$$p_k^i = x_\alpha^i h_1^\alpha, \quad s_k^i = x_\alpha^i h_2^\alpha.$$

Отсюда видно, что при замене столбца y_k матрицы Y линейной комбинацией векторов h_1 и h_2 с коэффициентами c_1 и c_2 в матрице Z k-й столбец заменяется линейной комбинацией $c_1 p_k + c_2 s_k$. Так как определитель $D(Z)$ является линейной формой k-го ее столбца, то мы в конечном счете получаем

$$D(Z) = c_1 D(XH_1) + c_2 D(XH_2) = c_1 F(H_1) + c_2 F(H_2).$$

Таким образом, доказано, что функция $F(Y)$ является линейной формой k-го столбца матрицы Y и функция $F(Y)$ удовлетворяет обоим условиям 1) и 2) предложения В). Таким образом, в силу леммы 2 имеем

$$F(Y) = D(Y) F(E).$$

Но так как $D(XE) = D(X)$, то последняя формула переписывается в виде

$$D(XY) = F(Y) = D(X) D(Y).$$

Итак, формула (79) доказана и доказательство теоремы завершено.

Замечание к теореме 4. Квадратная матрица X имеет тогда и только тогда обратную матрицу, когда определитель ее отличен от нуля.

Уже было доказано (см. Е) § 4), что если $D(X) \neq 0$, то матрица X имеет обратную матрицу. Но если определитель матрицы равен нулю, то уравнение

$$XY = E$$

неразрешимо относительно Y. В силу теоремы 4 мы имеем

$$D(X)D(Y)=D(E)=1,$$

что невозможно, если $D(X)=0$ (см. гл. 4, пример 6).

§ 6. Ранг матрицы

Здесь будет описано понятие ранга матрицы. Затем на этой основе будет дан критерий разрешимости системы линейных уравнений, число неизвестных в которой может не равняться числу уравнений.

А) Пусть

$$X=\|x_j^i\|, \quad i=1, \ldots, p, \quad j=1, \ldots, q,$$

— произвольная матрица размеров (p, q). Как было рассказано в § 2 (см. предложение А)), этой матрице соответствует линейное отображение φ векторного пространства B^q размерности q в векторное пространство A^p размерности p. Как было сказано там же, образ φB^q пространства B^q в пространстве A^p является векторным подпространством пространства A^p. *Рангом* матрицы X называется размерность векторного пространства φB^q. Пусть

$$e_1, e_2, \ldots, e_p, \tag{81}$$

$$f_1, f_2, \ldots, f_q \tag{82}$$

— базисы пространств A^p и B^q, при помощи которых матрица X определяет отображение φ. φf_l есть вектор в пространстве A^p, который, будучи записан в базисе (81), является l-м столбцом матрицы X. В виде формулы это запишем так:

$$\varphi f_l = x_l.$$

Таким образом, среди столбцов матрицы X имеется столько линейно независимых столбцов, каков ранг матрицы X. Обозначим это число линейно независимых столбцов через r. Оказывается, что матрица X имеет квадратную подматрицу порядка r с определителем, не равным нулю, и не имеет квадратной подматрицы порядка $s > r$ с определителем, отличным от нуля.

Оказывается, далее, что матрица X имеет r линейно независимых строк. Таким образом, число линейно независимых строк матрицы X равно числу линейно независимых столбцов матрицы: и то и другое равны рангу r матрицы X (см. гл. 4, пример 7).

Докажем предложение А). Изменим прежде всего нумерацию базисных векторов (82) так, чтобы линейно независимыми столбцами матрицы X стали столбцы

$$x_1, x_2, \ldots, x_r.$$

Составленную из этих столбцов матрицу обозначим через U. Тогда мы будем иметь

$$U = \|u_j^i\|, \quad u_j^i = x_j^i, \quad i = 1, \ldots, p, \quad j = 1, \ldots r.$$

Заметим прежде всего, что

$$p \geqslant r,$$

так как, если бы было $p < r$, то мы имели бы r линейно независимых векторов в пространстве A^p размерности $p < r$, что невозможно (см. предложение В) § 1).

Если $p = r$, то матрица U квадратная и столбцы ее линейно независимы, поэтому определитель ее отличен от нуля (см. теорему 3). Таким образом, в случае $p = r$ для матрицы X найдена квадратная подматрица U порядка r с определителем, отличным от нуля.

Рассмотрим случай, когда $p > r$. Вычеркнув из матрицы U строку с номером k, получим некоторую матрицу U^k размеров $(p-1, r)$. Допустим, что столбцы матрицы U^k линейно зависимы. Тогда существуют числа

$$c^1, c^2, \ldots, c^r,$$

не все равные нулю, такие, что

$$u_\alpha^i c^\alpha = 0 \quad \text{при} \quad i \neq k.$$

Но при $i = k$ это равенство невозможно, так как тогда оказалось бы, что столбцы матрицы U линейно зависимы. Таким образом,

$$u_\alpha^k c_\alpha = \gamma \neq 0,$$

и мы получаем

$$c^\alpha x_\alpha^i = \gamma e_k^i, \quad \alpha = 1, 2, \ldots, r.$$

Итак, базисный вектор e_k пространства A^p является линейной комбинацией векторов $x_1, x_2 \ldots, x_r$, составляющих базис подпространства φB^q, так что вектор e_k оказывается вектором, принадлежащим векторному подпространству φB^q пространства A^p. Следовательно, мы пришли к выводу, что если у матрицы U вычеркнуть k-ю строку и полученная матрица будет иметь линейно зависимые столбцы, то базисный вектор e_k пространства A^p принадлежит векторному подпространству φB^q. Здесь число k может принимать значения $1, 2, \ldots, p > r$. Если бы оказалось, что при вычеркивании из матрицы U любой строки с номером $k \leqslant p$ мы получаем матрицу U^k, столбцы которой линейно зависимы, то оказалось бы, что в пространстве φB^q имеется $p > r$ линейно независимых векторов, что невозможно, так как $p > r$. Таким образом, найдется такой номер k строки матрицы U, что при вычеркивании k-й строки в этой матрице мы получим матрицу U^k, столбцы которой линейно независимы. Этот процесс можно продолжать до тех пор, пока высота матрицы U, получаемой путем последовательного вычеркивания строк, не станет равной r. Тогда оставшаяся матрица будет квадратной и в ней столбцы будут линейно независимы, так что определитель этой матрицы отличен от нуля (см. теорему 3). Итак, мы нашли в матрице X квадратную матрицу порядка r, определитель которой отличен от нуля.

Допустим теперь, что у матрицы X есть квадратная подматрица Θ порядка s, определитель которой отличен от нуля. Изменим нумерацию строк и столбцов матрицы X таким образом, чтобы

$$\Theta = \|\theta_j^i\|, \quad \theta_j^i = x_j^i, \quad i = 1, 2, \ldots, s, \quad j = 1, 2, \ldots, s.$$

Покажем теперь, что у матрицы X, полученной изменением порядка строк и столбцов, первые s строк линейно независимы. Допустим противоположное. Тогда имеет место соотношение

$$c_\alpha x_\gamma^\alpha = 0, \quad \alpha = 1, 2, \ldots, s, \quad \gamma = 1, 2, \ldots, q, \tag{83}$$

причем не все числа

$$c_1, c_2, \ldots, c_s$$

равны нулю. Из отношения (83), в частности, получаем

$$c_\alpha x_j^\alpha = 0, \quad \alpha = 1, 2, \ldots, s, \quad j = 1, 2, \ldots, s.$$

Это соотношение показывает, что во взятой нами квадратной подматрице Θ строки линейно зависимы, что невозможно, так как определитель ее отличен от нуля.

Транспонируя матрицу X, т. е. беря матрицу X^{T}, мы из сказанного можем сделать вывод, что если у матрицы X имеется квадратная подматрица U порядка s с определителем, отличным от нуля, то в матрице X^{T} имеется

s линейно независимых строк. Это значит, что в исходной матрице X имеется s линейно независимых столбцов. Но так как число линейно независимых столбцов матрицы X равно ее рангу r, то справедливо неравенство $s \leqslant r$.

Таким образом, мы показали, что в исходной матрице X ранга r имеется квадратная подматрица порядка r с определителем, отличным от нуля, но не имеется квадратной подматрицы большего порядка с определителем, отличным от нуля. Одновременно мы показали, что в матрице X имеется ровно r линейно независимых строк и ровно r линейно независимых столбцов.

Итак, предложение А) доказано.

Применим теперь вышесказанное к решению общей системы линейных уравнений. Именно, мы рассмотрим систему уравнений вида

$$a_\alpha^i y^\alpha = b^i, \quad i = 1, 2, \ldots, p, \quad \alpha = 1, 2, \ldots, q. \qquad (84)$$

Здесь

$$y = (y^1, y^2, \ldots, y^q)$$

— неизвестные величины. Числа a_j^i составляют коэффициенты системы уравнений (84) и образуют матрицу

$$A = \|a_j^i\|, \quad i = 1, 2, \ldots p, \quad j = 1, 2, \ldots, q, \qquad (85)$$
$$b = (b^1, b^2, \ldots, b^p)$$

— известные правые части. Таким образом, матрица A, составленная из коэффициентов системы уравнений (84), имеет p строк, т. е. высоту p, и q столбцов, т. е. ширину q.

Нижеследующая теорема отвечает на вопрос: при каких условиях система (84) имеет решение.

Теорема 5. *Присоединим к матрице A, имеющей ширину q, еще один столбец, — именно, столбец b* (см. (85)). *Полученную так матрицу обозначим через B. Матрица B имеет размеры* $(p, q+1)$. *Оказывается, что система* (84) *имеет решение тогда и только тогда, когда матрицы A и B имеют одинаковые ранги.*

Доказательство. Пусть ранг матриц A и B равен r. Это значит, что в матрице A существует r линейно независимых столбцов (см. А)). Присоединяя к этим столбцам столбец b (см. (85)), мы получим $r+1$ линейно зависимых столбцов, так как в противном случае матрица B имела бы как минимум ранг $r+1$. Выписывая эту линейную зависимость, мы видим, что система (84) разрешима. Если система (84) разрешима, то столбец b линейно зависит от столбцов матрицы A и, следовательно, ранг матрицы B равен рангу матрицы A (см. A)).

Итак, теорема 5 доказана.

В случае, если система уравнений (84) имеет решение, опишем совокупность всех решений, так как решение может быть не единственно.

Обозначим ранг матриц A и B через r, считая уже, что ранги этих матриц равны. Меняя нумерацию неизвестных системы (84) и нумерацию ее уравнений, мы можем добиться того, чтобы квадратная подматрица Θ порядка r матрицы A, определитель которой отличен от нуля, стояла в верхнем левом углу матрицы A, так что мы имеем

$$\Theta = \|a_j^i\|, \quad i = 1, 2, \ldots, r, \quad j = 1, 2, \ldots, r.$$

Как это видно из доказательства предложения A), первые r строк матрицы B линейно независимы, остальные линейно зависимы от них. Для того чтобы удобно записать этот факт, обозначим компоненты вектора b через a_{q+1}^i, $i = 1, 2, \ldots, p$. Запишем теперь в виде формулы тот факт, что последние строки матрицы B линейно зависят от первых r ее строк:

$$a_j^i = u_\alpha a_j^\alpha, \quad i = r+1, r+2, \ldots, p,$$
$$\alpha = 1, 2, \ldots, r, \quad j = 1, 2, \ldots, q+1,$$

где $u_{r+1}, u_{r+2}, \ldots, u_p$ суть числа. Это показывает, что в системе уравнений (84) последние $p-r$ уравнений линейно зависят от первых r уравнений и потому являются их следствием, так что их можно отбросить, не меняя решения. Таким образом, нам достаточно решить систему уравнений

$$a_\alpha^i y^\alpha = b^i, \quad i = 1, 2, \ldots, r, \quad \alpha = 1, 2, \ldots, q,$$

или, что то же самое, систему уравнений

$$a_\alpha^i y^\alpha = b^i_\beta y^\beta, \quad i = 1, 2, \ldots, r, \quad \alpha = 1, 2, \ldots, r,$$
$$\beta = r+1, \ldots, q.$$

Здесь неизвестным $y_{r+1}, \ldots, y_q$ можно придать произвольные значения и решать полученную систему по способу, указанному в теореме 2. Таким образом, совокупность решений системы уравнений (84) зависит от $q-r$ параметров $y_{r+1}, \ldots, y_q$.

§ 7. Евклидовы векторные пространства

Определение 6. Векторное пространство E^n называется *евклидовым*, если в нем определено скалярное произведение. Именно, в E^n задана некоторая симметричная билинейная форма $f(x, y)$ с положительно определенной квадратичной формой $f(x, x)$ и скалярное произведение (x, y) задается формулой

$$(x, y) = f(x, y).$$

Длина вектора $|x| > 0$ определяется формулой

$$|x|^2 = (x, x).$$

Пусть

$$e_1, e_2, \ldots, e_n$$

— некоторый базис пространства E^n, так что

$$x = x^\alpha e_\alpha, \quad y = y^\beta e_\beta.$$

Тогда в силу предложения D) § 2 мы имеем

$$(x, y) = f(e_\alpha, e_\beta) \quad x^\alpha y^\beta. \tag{86}$$

Из предложения E) § 2 следует, что, зная длину каждого вектора, можно вычислить скалярное произведение двух любых векторов этого пространства.

Ортонормальная система

A) Система векторов

$$e_1, e_2, \ldots, e_p \tag{87}$$

называется *ортонормальной*, если имеет место соотношение

$$(e_i, e_j) = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \quad i = 1, \ldots, p, \quad j = 1, \ldots, p. \tag{88}$$

Таким образом, каждые два различных вектора системы (87) ортогональны между собой, а длина каждого вектора (87) равна единице.

Ортонормальная система (87) всегда линейно независима.

Действительно, если имеет место соотношение

$$c^\alpha e_\alpha = 0,$$

то, умножая это соотношение скалярно на e_i, получаем $c_i = 0$. Таким образом, если $p = n$, то система

$$e_1, e_2, \ldots, e_n \tag{89}$$

является ортонормальным базисом пространства E^n. В случае ортонормального базиса (89) скалярное произведение (86) в координатной форме записывается в виде

$$(x, y) = f(x, y) = \sum_\alpha x^\alpha y^\alpha = \delta_{\alpha\beta} x^\alpha y^\beta.$$

Базис (89), удовлетворяющий условию (88), называется ортогональным, потому что каждые два различных его вектора ортогональны между собой; он нормирован потому, что длина каждого вектора равна единице.

В дальнейшем будет установлено (см. предложение С)), что в евклидовом векторном пространстве E^n всегда имеется ортонормальный базис.

В) В евклидовом векторном пространстве E^n каждая линейная форма φx (см. определение 3) задается в виде скалярного произведения

$$\varphi x = (u, x) \tag{90}$$

некоторого вектора u на вектор x, причем вектор u однозначно определяется линейной формой φ.

Выберем в пространстве E^n некоторый ортонормальный базис. Тогда линейная форма φ определяется равенством (см. (25) § 2)

$$\varphi x = d_\beta x^\beta.$$

Тогда вектор $u = (u^1, u^2, \ldots, u^n)$ задается формулой

$$u^i = d_i.$$

При таком выборе u равенство (90) выполнено. Докажем, что вектор u однозначно определяется линейной формой φ. В самом деле, допустим, что одновременно имеют место два равенства

$$\varphi x = (u, x), \quad \varphi x = (u^1, x).$$

Тогда мы имеем

$$(u - u^1, x) = 0$$

при произвольном x. Следовательно, при $x = u - u^1$ имеем

$$(u - u^1, u - u^1) = 0.$$

Это возможно лишь тогда, когда

$$u - u^1 = 0.$$

C) Каждая линейно независимая система векторов

$$x_1, x_2, \ldots, x_p \tag{91}$$

однозначно определяет ортонормальную систему векторов

$$e_1, e_2, \ldots, e_p. \tag{92}$$

Процесс перехода от системы векторов (91) к системе векторов (92) называется *ортонормированием*. Опишем его.

Так как система векторов (91) линейно независима, то вектор $x_1 \neq 0$ и, следовательно, его длина $|x_1| \neq 0$. Мы полагаем

$$e_1 = \frac{x_1}{|x_1|}.$$

Далее полагаем

$$e_2' = x_2 - (x_2, e_1)\, e_1.$$

Мы имеем

$$(e_2', e_1) = (x_2, e_1) - (x_2, e_1) = 0.$$

Таким образом, векторы e_1 и e_2' ортогональны между собой. При этом вектор $e_2' \neq 0$, так как векторы x_1 и x_2 линейно независимы, а следовательно, линейно независимы векторы e_1 и e_2'. Теперь нормируем вектор e_2'. Именно, положим

$$e_2 = e_2'/|e_2'|.$$

Таким образом, получаем

$$(e_2, e_2) = 1.$$

Пусть $e_1, e_2, \ldots, e_l$ уже построены так, что они составляют ортонормальную систему. Построим вектор e_{l+1}. Положим

$$e_{l+1}' = x_{l+1} - (x_{l+1}, e_1)\, e_1 - \\ - (x_{l+1}, e_2)\, e_2 - \ldots - (x_{l+1}, e_l)\, e_l$$

Прежде всего, вектор $e_{l+1}' \neq 0$, так как из линейной независимости векторов $x_1, x_2, \ldots, x_{l+1}$ следует линейная независимость векторов

$$e_1, e_2, \ldots, e_l, x_{l+1}.$$

Умножая вектор e'_{i+1} на произвольный вектор e_j, $j \leqslant i$, получаем

$$(e'_{i+1}, e_j) = 0.$$

Таким образом, построенный вектор e'_{i+1} ортогонален ко всем векторам $e_1, e_2, \ldots, e_i$ и отличен от нуля. Нормируя вектор e'_{i+1}, т. е. полагая

$$e_{i+1} = \frac{e'_{i+1}}{|e'_{i+1}|},$$

мы получаем вектор e_{i+1}.

D) Каждое векторное подпространство B^p евклидова векторного пространства E^n само естественным образом является евклидовым векторным пространством. Действительно, каждая пара векторов из B^p является парой векторов из E^n, и потому для нее определено скалярное произведение. Вектор z пространства E^n, ортогональный каждому вектору из подпространства B^p, считается ортогональным всему пространству B^p. Обозначим через C совокупность всех векторов пространства E^n, ортогональных векторному подпространству B^p. Оказывается, что C является векторным подпространством пространства E^n размерности $q = n - p$, причем векторное пространство E^n распадается в прямую сумму своих подпространств B^p и $C = C^q$. Подпространство C^q называется *ортогональным дополнением* подпространства B^p. Оказывается, что подпространство B^p является ортогональным дополнением подпространства C^q.

Для доказательства данного утверждения выберем в евклидовом векторном пространстве B^p ортонормальный базис

$$e_1, e_2, \ldots, e_p.$$

Данную систему векторов дополним до максимальной линейно независимой системой векторами $x_{p+1}, \ldots, x_n$ (см. D) § 1). Тогда мы получим максимальную линейно независимую систему

$$e_1, e_2, \ldots, e_p, x_{p+1}, \ldots, x_n.$$

Теперь подвергнем эту систему процессу ортонормирования (см. C)). При этом первые p векторов $e_1, e_2, \ldots, e_p$ не изменятся, а векторы $x_{p+1}, x_{p+2}, \ldots, x_n$ заменятся другими векторами. Вновь полученную систему запишем в виде

$$e_1, e_2, \ldots, e_p, e_{p+1}, \ldots, e_n.$$

Последняя система является ортонормальным базисом пространства E^n, так что каждый вектор x пространства E^n записывается в виде

$$x = x^\alpha e_\alpha.$$

Для того чтобы вектор x был нормальным ко всему пространству B^p, необходимо и достаточно, чтобы он был нормальным к каждому вектору e_i, $i = 1, 2, \ldots, p$. Таким образом, должно быть выполнено условие

$$(x, e_i) = 0, \quad i = 1, 2, \ldots, p.$$

Из последнего условия вытекает

$$x^1 = 0, \ldots, x^p = 0.$$

Таким образом, вектор z, нормальный ко всему пространству B^p, записывается в виде

$$z = x^\beta e_\beta; \quad \beta = p + 1, \ldots, n.$$

Совокупность всех векторов такого вида составляет векторное подпространство C^q пространства E^n размерности $q = n - p$, которое является ортогональным дополнением подпространства B^p. Ясно, что подпространство B^p в свою очередь является ортогональным дополнением подпространства C^q.

Дадим геометрическое описание скалярного произведения двух векторов пространства E^n.

Е) Пусть

$$x, y$$

— два линейно независимых вектора пространства E^n. Совокупность всех векторов

$$u = cx + dy,$$

где c и d — произвольные действительные числа, составляет двумерное векторное подпространство E^2 векторного пространства E^n. Множество E^2 является векторным двумерным евклидовым пространством, иначе говоря, евклидовой плоскостью Таким образом, векторы x и y являются линейно независимыми векторами плоскости E^2, и ясно, что такое угол между ними. Обозначим его через φ. Оказывается, что скалярное произведение векторов x и y выражается в следующем виде:

$$(x, y) = |x|\,|y| \cos\varphi. \tag{93}$$

Для доказательства формулы (93) ортонормируем пару векторов x, y. Мы получим тогда два вектора e_1, e_2, составляющих базис двумерного евклидова пространства E^2. При этом

$$x = |x| e_1, \quad y = |y| (e_1 \cos\varphi + e_2 \sin\varphi).$$

Таким образом,

$$(x, y) = (|x| e_1, |y| (e_1 \cos\varphi + e_2 \sin\varphi)) = |x| |y| \cos\varphi.$$

Таким образом, формула (93) доказана. В случае, если x и y линейно зависимы, следует считать, что угол φ между ними равен нулю или π. При этом оба они выражаются через один и тот же вектор

$$x = |x| e_1, \quad y = \pm |y| e_1.$$

Тогда мы имеем

$$(x, y) = |x| |y| \cos 0$$

или

$$(x, y) = |x| |y| \cos \pi.$$

Таким образом, формула (93) имеет место для любых двух векторов x и y из E^n.

F) Отображение φ евклидова векторного пространства E^n на евклидово векторное пространство $\bar{E}^n$ называется изоморфным, если оно линейно (см. определение 3) и сохраняет скалярное произведение, т. е. если для двух любых векторов x и y из E^n имеет место формула

$$(\varphi x, \varphi y) = (x, y).$$

Как это следует из E) § 2, сохранение скалярного произведения вытекает из сохранения длины векторов при отображении φ. Именно, если для любого вектора $x \in E^n$ имеет место соотношение

$$|\varphi x| = |x|,$$

то при отображении φ сохраняется и скалярное произведение.

В евклидовом векторном пространстве определено расстояние $\rho(x, y)$ между каждыми двумя его векторами x и y. Оно задается формулой

$$\rho(x, y) = |x - y|.$$

Ортогональные матрицы

G) Пусть E^n — евклидово векторное пространство размерности n и

$$e_1, e_2, \ldots, e_n \tag{94}$$

— некоторый его ортонормальный базис (см. предложение A)). Пусть далее, φ — некоторое изоморфное отображение пространства E^n на себя (см. F)).

Матрица

$$X = \|x_j^i\|; \quad i = 1, \ldots, n, \quad j = 1, \ldots, n,$$

соответствующая отображению φ в ортонормальном базисе (94), называется *ортогональной*. Если φ_1 и φ_2 — два изоморфных отображения пространства E^n на себя, то отображение $\varphi_1\varphi_2$ есть тоже изоморфное отображение, поэтому произведение двух ортогональных матриц есть также ортогональная матрица. Матрица X, соответствующая отображению φ, определяется соотношениями

$$\varphi e_j = x_j^\alpha e_\alpha.$$

Дадим теперь четыре характеристики ортогональных матриц.

Мы имеем

$$\varphi e_j = x_j = (x_j^1, x_j^2, \ldots, x_j^n).$$

Это значит, что элементы j-го столбца матрицы X являются координатами вектора φe_j в базисе (94). Так как отображение φ сохраняет скалярное произведение, то

$$(\varphi e_j, \varphi e_k) = \delta_{jk}.$$

Переписывая это выражение в координатной форме, получим

$$\sum_{\alpha=1}^{n} x_j^\alpha x_k^\alpha = \delta_{jk}. \tag{95}$$

Ясно, что если для матрицы X выполнено это соотношение, то она является ортогональной. Можно сказать, что столбцы матрицы X составляют ортонормальную систему векторов. Оказывается, что соотношение (95) эквивалентно соотношению

$$X^{\mathrm{T}} X = E. \tag{96}$$

Условия (95) и (96), наложенные на матрицу X, сформулированы в терминах столбцов. Оказывается, что их можно сформулировать в терминах строк. Именно, имеет место соотношение

$$\sum_{\beta=1}^{n} x_\beta^j x_\beta^k = \delta^{jk}. \tag{97}$$

Соотношение (97) можно переписать в виде формулы

$$XX^{\mathrm{T}} = E. \tag{98}$$

Оказывается, что если матрица X удовлетворяет одному из соотношений (95)—(98), то она ортогональна. Относительно (95) это утверждение уже было высказано как очевидное. Нам остается показать, что каждое из соотношений (96)—(98) эквивалентно соотношению (95).

Оказывается далее, что определитель ортогональной матрицы $D(X)$ удовлетворяет условию

$$D(X) = \pm 1. \tag{99}$$

Докажем, прежде всего, эквивалентность соотношений (95) и (96). Для этого положим

$$X^{\mathrm{T}} = U = \| u_j^i \|,$$

так что

$$u_j^i = x_i^j.$$

Соотношение (96) записывается в виде

$$u_\alpha^i x_k^\alpha = \delta_k^i.$$

Заменяя здесь u_α^i через x_i^α, мы получим формулу (95). Таким образом, соотношения (95) и (96) эквивалентны.

Эквивалентность соотношений (97) и (98) доказывается аналогично.

Докажем соотношение (99). В силу теоремы 1

$$D(X^{\mathrm{T}}) = D(X).$$

В силу этого, теоремы 4 и соотношения (96) мы имеем

$$D(X)\,D(X) = 1,$$

откуда и следует формула (99). Таким образом, в частности, определитель ортогональной матрицы отличен от нуля, и поэтому ортогональная матрица имеет обратную (см. замечание к теореме 4).

Теперь докажем, что соотношения (96) и (98) эквивалентны. Для этого соотношение (96) умножим справа на X^{-1}. Тогда мы получим

$$X^{\mathsf{T}} = X^{-1}.$$

Умножая это соотношение на X слева, получим

$$XX^{\mathsf{T}} = E.$$

Таким образом, из соотношения (96) вытекает соотношение (98). Аналогично из соотношения (98) следует соотношение (96). Таким образом, соотношения (96) и (98) эквивалентны.

Итак, все соотношения (95)—(98) эквивалентны друг другу и означают ортогональность матрицы X (см. гл. 4, пример 8).

Глава 2
КОРНИ МНОГОЧЛЕНОВ

§ 8. Комплексные числа

Здесь я кратко рассказываю о том, как возникли и утвердились в математике комплексные числа. Затем даю их определение, описываю правила действий над ними, употребляя при этом их геометрическую интерпретацию, но не пользуюсь при этом как известными тригонометрическими формулами косинуса суммы и синуса суммы.

Историческая справка

Комплексные числа были введены в математику для того, чтобы сделать возможной операцию извлечения квадратного корня из любого действительного числа. Это, однако, не является достаточным основанием для того, чтобы вводить в математику новые числа. Оказалось, что если производить вычисления по обычным правилам над выражениями, в которых встречается квадратный корень из отрицательного числа, то можно прийти к результату, уже не содержащему квадратный корень из отрицательного числа. В XVI в. Кардано нашел формулу для решения кубического уравнения. Оказалось, что именно в том случае, когда кубическое уравнение имеет три действительных корня, в формуле Кардано встречается квадратный корень из отрицательного числа. Поэтому квадратные корни из отрицательных чисел стали употреблять в математике и назвали их мнимыми числами — тем самым они как бы приобрели право на нелегальное существование. Полные гражданские права мнимым числам на грани XVIII—XIX столетий дал

Гаусс, который назвал их комплексными числами, дал им геометрическую интерпретацию и, что самое главное, доказал основную теорему алгебры, утверждающую, что каждый многочлен имеет хотя бы один действительный или комплексный корень.

Определение комплексных чисел

Мы будем исходить из того, что действительные числа нам известны. Мы знаем, что для них определены два основных действия — сложение и умножение — и имеются обратные к ним действия — вычитание и деление. Для этих действий выполняются хорошо известные правила, которые обычно употребляются совершенно автоматически — поэтому я их не буду здесь формулировать.

Поставим теперь перед собой задачу расширить понятие числа таким образом, чтобы уравнение

$$u^2+1=0$$

имело решение. Корень этого уравнения объявляется новым числом и обозначается через i. Таким образом, для i имеем

$$i^2=-1. \tag{1}$$

Так как в множестве новых чисел содержатся все действительные числа и элемент i и так как в нем возможны действия сложения и умножения, то в этом множестве должны содержаться всевозможные многочлены относительно i с действительными коэффициентами, в частности все многочлены первой степени, т. е. выражения вида

$$z=x+yi=x+iy,$$

где x и y — действительные числа. Эти выражения и называются *комплексными числами*. Действия над ними мы определим как действия над многочленами, учитывая при этом условие (1). Комплексные числа вида

$$z=x+0i=x$$

являются *действительными числами*. Комплексные числа вида

$$z=0+yi=yi$$

называются *чисто мнимыми числами*.

Пусть $z_1=x_1+y_1i$, $z_2=x_2+y_2i$ — два комплексных числа. Согласно высказанному правилу сумма и произведение этих комплексных чисел определяются равенствами

$$z_1+z_2=(x_1+y_1i)+(x_2+y_2i)=(x_1+x_2)+(y_1+y_2)i, \quad (2)$$

$$z_1z_2=(x_1+y_1i)(x_2+y_2i)=x_1x_2+(x_1y_2+y_1x_2)i+$$
$$+y_1y_2i^2=(x_1x_2-y_1y_2)+(x_1y_2+y_1x_2)i. \quad (3)$$

При получении последнего равенства мы использовали условие (1).

В случае, если число $z_1=x_1$ действительное, получаем

$$x_1z_2=x_1x_2+x_1y_2i. \quad (4)$$

Из формул (2) и (3) видно, что сумма и произведение двух комплексных чисел есть также комплексное число.

Для того чтобы убедиться, что действие вычитания, обратное действию сложения, существует, достаточно найти число $-z$, противоположное числу z, а для того чтобы убедиться в том, что возможно деление, достаточно для $z\neq 0$ указать число z^{-1}, обратное числу $z=x+yi$. Числа эти, как легко видеть, задаются формулами

$$-z=-x-yi,$$

$$z^{-1}=\frac{x}{x^2+y^2}-\frac{y}{x^2+y^2}i.$$

Таким образом, величина z^{-1}, обратная к z, существует всегда, когда $z\neq 0$.

Геометрическое изображение комплексных чисел

Обозначим через P плоскость нашего чертежа и выберем на ней прямоугольную систему координат (рис. 1). Комплексное число $z=x+yi$ мы поместим в точку $z=(x, y)$ с координатами x, y. Обозначим также через z вектор, идущий из начала координат O в точку z. Таким образом, буква z обозначает у нас одновременно комплексное число, точку z, изображающую это комплексное число, и вектор z, соответствующий этому комплексному числу. При этом изображении действительные числа попадают на ось абсцисс, поэтому ось абсцисс называется *действительной осью* плоскости P комплексной переменной, а чисто мнимые числа попадают на ось ординат, поэтому ось ординат называется *мнимой осью* плоскости P комплексной переменной. Нуль попадает в начало координат.

Длина вектора z называется *модулем* комплексного числа $z = x + yi$ и обозначается $|z|$:

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

Комплексные числа z, удовлетворяющие условию $|z| = 1$, составляют окружность радиуса 1 с центром в начале координат. На этой окружности лежит число 1. Из точки 1 отложим по окружности дугу заданной длины φ в направлении против часовой стрелки. Конец этой дуги обозначим через (φ). Если φ — отрицательное число, то для получения (φ) нужно отложить от точки 1 длину дуги $|\varphi|$ по часовой стрелке. Как известно, абсцисса точки (φ) есть $\cos\varphi$, а ее ордината $\sin\varphi$. Таким образом, комплексное число (φ) задается формулой

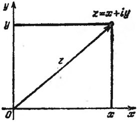


Рис. 1

$$(\varphi) = \cos\varphi + i\sin\varphi. \tag{5}$$

Итак, всякое комплексное число z, по модулю равное 1, записывается в виде (5). Если z — произвольное комплексное число, модуль которого $|z| = \rho$ отличен от 0, то число $\frac{z}{\rho}$ является комплексным числом, по модулю равным 1, и потому записывается в виде (5). Из равенства

$$\frac{z}{\rho} = \cos\varphi + i\sin\varphi$$

мы получаем

$$z = \rho(\cos\varphi + i\sin\varphi). \tag{6}$$

Запись (6) называется *тригонометрической формой* комплексного числа z. Число φ называется *аргументом* комплексного числа z. Если модуль ρ комплексного числа z отличен от нуля, то аргумент его определен с точностью до слагаемого $2k\pi$, где k — целое число. Если же модуль ρ комплексного числа равен 0, то формула (6) также имеет место, однако в этом случае аргумент комплексного числа вовсе не определен.

Числа ρ и φ называются *полярными координатами* точки z.

Геометрическое истолкование действий над комплексными числами

Из формул (2) и (4) следует, что комплексные числа складываются и умножаются на действительные числа, как векторы.

Геометрический смысл сложения комплексных чисел очевиден: вектор $z_1 + z_2$ — это диагональ параллелограмма, построенного на векторах z_1 и z_2. Отсюда вытекает важное неравенство

$$|z_1 + z_2| \leqslant |z_1| + |z_2|. \tag{7}$$

Для того чтобы дать геометрическое истолкование умножения комплексных чисел, нужно употребить операцию поворота вектора, или, что то же самое, комплексного числа. Повернув вектор z против часовой стрелки на угол α, мы получим некоторый новый вектор, который обозначим через $R_\alpha(z)$. Геометрически ясно, что операция поворота R_α имеет следующее свойство: если a — действительное число, то

$$R_{\alpha+\beta}(z) = R_\alpha(R_\beta(z)), \quad R_\alpha(az) = aR_\alpha(z),$$
$$R_\alpha(z_1 + z_2) = R_\alpha(z_1) + R_\alpha(z_2).$$

Из последних двух формул следует, что если a_1 и a_2 — два действительных числа, то имеет место соотношение

$$R_\alpha(a_1 z_1 + a_2 z_2) = a_1 R_\alpha(z_1) + a_2 R_\alpha(z_2). \tag{8}$$

Непосредственно ясно также, что

$$R_\alpha(1) = \cos\alpha + i\sin\alpha. \tag{9}$$

Докажем теперь, что поворот комплексного числа $z = x + yi$ на угол α равносилен умножению его на комплексное число $\cos\alpha + i\sin\alpha$, т. е. что

$$R_\alpha(z) = (\cos\alpha + i\sin\alpha)\,z. \tag{10}$$

Для этого рассмотрим сначала отдельно поворот на угол $\pi/2$. В этом случае $\cos\frac{\pi}{2} + i\sin\frac{\pi}{2} = i$ и равенство (10) принимает вид $R_{\pi/2}(z) = iz$. Проверим эту формулу. С одной стороны, геометрически очевидно, что $R_{\pi/2}(1) = i$, $R_{\pi/2}(i) = -1$. С другой стороны, $i1 = i$, $ii = -1$. Таким образом,

$$R_{\pi/2}(1) = i1, \quad R_{\pi/2}(i) = ii,$$

Из формулы (8) непосредственно вытекает

$$iz = i(x+iy) = xi + y(-1) = \\ = xR_{\pi/2}(1) + yR_{\pi/2}(1) = R_{\pi/2}(x1 + yi) = \\ = R_{\pi/2}(x+iy) = R_{\pi/2}(z).$$

Таким образом, формула (10) доказана для $\alpha = \pi/2$.

Пусть теперь α — произвольное действительное число. При $\hat{z} = \cos\alpha + i\sin\alpha$ получаем

$$\hat{z}i = i\hat{z} = R_{\pi/2}(\hat{z}) = R_{\pi/2}(\cos\alpha + i\sin\alpha) = \\ = \cos\left(\alpha + \frac{\pi}{2}\right) + i\sin\left(\alpha + \frac{\pi}{2}\right) = \\ = R_\alpha\left(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}\right) = R_\alpha(i). \quad (11)$$

Таким образом, формула (10) доказана при $z = i$.

Перейдем теперь к доказательству формулы (10) для произвольного комплексного числа

$$z = x + iy.$$

В силу формул (8), (9), (11) имеем

$$R_\alpha(z) = R_\alpha(x+iy) = xR_\alpha(1) + yR_\alpha(i) = \\ = x(\cos\alpha + i\sin\alpha) + y(\cos\alpha + i\sin\alpha) = \\ = (\cos\alpha + i\sin\alpha)(x + yi) = (\cos\alpha + i\sin\alpha)z.$$

Таким образом, формула (10) полностью доказана.

Применяя формулу (10) к комплексному числу $z = \cos\beta + i\sin\beta$, получаем

$$(\cos\alpha + i\sin\alpha)(\cos\beta + i\sin\beta) = R_\alpha(\cos\beta + i\sin\beta) = \\ = R_\alpha(R_\beta(1)) = R_{\alpha+\beta}(1) = \cos(\alpha+\beta) + i\sin(\alpha+\beta).$$

Таким образом,

$$(\cos\alpha + i\sin\alpha)(\cos\beta + i\sin\beta) = \cos(\alpha+\beta) + i\sin(\alpha+\beta).$$

Производя перемножение комплексных чисел, стоящих в левой части, по формуле (3), мы получим

$$(\cos\alpha + i\sin\alpha)(\cos\beta + i\sin\beta) = \\ = (\cos\alpha\cos\beta - \sin\alpha\sin\beta) + (\sin\alpha\cos\beta + \cos\alpha\sin\beta)i.$$

Значит, мы получили формулы для косинуса и синуса суммы:

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta,$$
$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta.$$

Для произвольных комплексных чисел, которые мы запишем в виде $r(\cos\alpha + i\sin\alpha)$, $s(\cos\beta + i\sin\beta)$ получаем

$$r(\cos\alpha + i\sin\alpha)\,s(\cos\beta + i\sin\beta) = \\ = rs(\cos(\alpha+\beta) + i\sin(\alpha+\beta)). \quad (12)$$

Таким образом, при перемножении двух комплексных чисел их модули перемножаются, а аргументы складываются.

Формулу (12) очевидным образом можно распространить на произвольное число сомножителей. Если все эти сомножители равны между собой и равны комплексному числу $r(\cos\alpha + i\sin\alpha)$, то мы получаем

$$(r(\cos\alpha + i\sin\alpha))^n = r^n(\cos n\alpha + i\sin n\alpha).$$

Эта формула очень интересна. Она дает возможность извлечь корень n-й степени из произвольного комплексного числа $\rho(\cos\varphi + i\sin\varphi)$. Именно, оказывается, что количество корней n-й степени из числа $\rho(\cos\varphi + i\sin\varphi) \neq 0$ равно n, причем корни эти расположены на окружности радиуса $\sqrt[n]{\rho}$ с центром в начале координат и составляют вершины правильного n-угольника. Это утверждение я предоставляю для доказательства читателям.

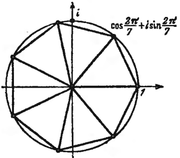


Рис. 2

В частности, корень n-й степени из единицы имеет n значений, которые являются вершинами правильного n-угольника, вписанного в единичный круг, причем одна из его вершин есть единица (рис. 2). В виде формулы эти корни записываются следующим образом:

$$\sqrt[n]{1} = \cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n}, \qquad k = 0, 1, 2, \ldots, n-1.$$

Комплексно сопряженные числа

При рассмотрении комплексных чисел важную роль играют так называемые комплексно сопряженные числа. Если $z = x + iy$ — некоторое комплексное число, то число $\bar{z}$, комплексно сопряженное с числом z, определяется формулой

$$\bar{z} = x - iy. \tag{13}$$

Таким образом, число $\bar{z}$, комплексно сопряженное с числом z, является зеркальным образом числа z относительно действительной оси, и мы имеем, в частности,

$$\bar{\bar{z}} = z.$$

Если

$$z = r(\cos\alpha + i\sin\alpha),$$

то

$$\bar{z} = r(\cos\alpha - i\sin\alpha) = r(\cos(-\alpha) + i\sin(-\alpha)),$$

и, следовательно, аргументы двух комплексно сопряженных чисел отличаются лишь знаком. Из формул (2), (3) следует, что

$$\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2. \tag{14}$$

Из формул (8), (13) следует, что

$$\overline{z_1 z_2} = \bar{z}_1 \bar{z}_2. \tag{15}$$

Заметим еще, что равенство

$$\bar{z} = z \tag{16}$$

имеет место тогда и только тогда, когда z есть действительное число (см. гл. 4, примеры 9—11).

§ 9. Основная теорема алгебры

Здесь будет доказана основная теорема алгебры, утверждающая, что всякий многочлен положительной степени с комплексными коэффициентами имеет по крайней мере один комплексный корень. При этом действительные числа считаются частным случаем комплексных чисел. Эта теорема впервые была доказана Гауссом в 1799 г. для частного случая многочленов с действительными коэффициентами. Гаусс показал, что всякий такой многочлен имеет по крайней мере один действительный или комплексный корень. Это значит, что, рассматривая корни алгебраических уравнений (т. е. корни многочленов), мы не можем прийти к новым числам. Здесь эта теорема доказывается для многочленов с комплексными коэффициентами.

Доказательство основной теоремы алгебры основано на конкретном рассмотрении комплексных чисел. Строгое доказательство основной теоремы алгебры, опирающееся на теорию функций комплексной переменной, приведено в моей книге «Анализ бесконечно малых» (см. пример на с. 246).

Здесь я привожу не строгое, но геометрически убедительное доказательство, основанное на рассмотрении путей в плоскости комплексной переменной и их деформаций. Доказательство это не только доказывает теорему, но до некоторой степени объясняет, почему она верна.

Как следствие основной теоремы алгебры мы дадим разложение многочлена с комплексными (в частности, действительными) коэффициентами на множители.

Пути в плоскости комплексной переменной

Если точка z в плоскости P комплексной переменной перемещается во времени, когда время t меняется в пределах $t_0 \leqslant t \leqslant t_1$, то мы считаем, что в плоскости P задан путь (рис. 3). Таким образом, путь есть функция

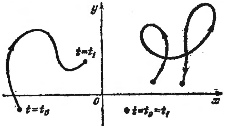


Рис. 3

$z(t)$ действительной переменной t, принимающая комплексные значения и заданная на отрезке $t_0 \leqslant t \leqslant t_1$:

$$z(t), \quad t_0 \leqslant t \leqslant t_1.$$

Формула эта задает путь. Речь идет здесь о движении точки, осуществляемом, естественно, без скачков, так что функция $z(t)$ является непрерывной. Мы не уточняем здесь понятие непрерывности, считая, что оно интуитивно ясно как движение точки. Следует отчетливо понимать, что путь есть процесс движения, а не та линия, которую описывает движущаяся точка. Одну и ту же линию можно описать разными способами.

В процессе движения точка $z(t)$ в разные моменты времени может попадать в одну и ту же точку плоскости, так что не исключается равенство

$$z(t_2)=z(t_3) \quad \text{при} \quad t_2 \neq t_3.$$

Таким образом, путь может иметь самопересечения. Он может даже состоять из одной точки, именно, в случае, когда точка $z(t)$ вовсе не перемещается при изменении t. В дальнейшем, если это не будет оговорено специально, мы всегда будем предполагать, что путь не проходит через начало координат, т. е. что величина $z(t)$ ни при каком значении t не обращается в нуль. Точка $z(t_0)=z_0$ называется *началом* пути, а точка $z(t_1)=z_1$ — его *концом*. Если имеет место равенство $z_0=z_1$, то путь называется *замкнутым* (рис. 4).

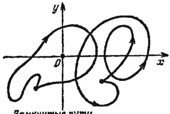


Рис. 4

Так как комплексное число $z(t)$ не обращается в нуль, то для всякого значения t определен аргумент $\varphi(t)$ комплексного числа $z(t)$, но он определен лишь с точностью до слагаемого $2k\pi$ (k — целое число). Эта неоднозначность для нас нежелательна. Для того чтобы освободиться от нее, мы выберем для начальной точки z_0 вполне определенный аргумент $\varphi_0=\varphi(t_0)$.

Затем по мере возрастания t будем выбирать аргумент $\varphi(t)$ точки $z(t)$ так, чтобы при малых изменениях t он менялся мало. Этим неоднозначность выбора аргумента будет устранена. Добавление к аргументу числа $2k\pi$ при $k\neq 0$ привело бы сразу к резкому изменению величины $\varphi(t)$. Выбрав начальное значение аргумента $\varphi(t_0)=\varphi_0$ и следя за тем, чтобы аргумент $\varphi(t)$ точки $z(t)$ менялся вместе с t непрерывно, мы получаем вполне определенную функцию $\varphi(t)$, меняющуюся непрерывно, т. е. без скачков. Если выбрать начальное значение аргумента φ_0 иначе, изменив его на $2k\pi$, то он будет отличаться от ранее выбранного ровно на $2k\pi$ на всем протяжении изменения t. Отсюда следует, что при таком способе построения функции $\varphi(t)$ величина

$$\varphi(t_1)-\varphi(t_0) \tag{17}$$

не зависит от случайно выбранного начального значения аргумента числа z_0.

Если путь замкнут, то точки z_0 и z_1 совпадают и, следовательно, их аргументы $\varphi(t_0)$ и $\varphi(t_1)$ могут отличаться лишь на $2k\pi$. Поэтому число (17) в случае замкнутого пути есть $2k\pi$. Целое число k называется *индексом* замкнутого пути в плоскости комплексной переменной z. Следует еще раз подчеркнуть, что индекс замкнутого пути можно определить лишь в том случае, когда путь не проходит через начало координат.

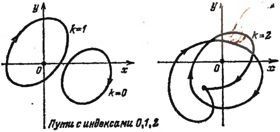


Рис. 5

Индекс k имеет простой геометрический смысл. Именно, он указывает, сколько раз точка $z(t)$, описывая замкнутый путь, обходит начало координат (рис. 5).

Рассмотрим простой пример. Путь

$$1 + r(\cos t + i \sin t), \qquad 0 \leqslant t \leqslant 2\pi, \tag{18}$$

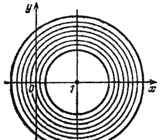
является замкнутым. Он описывает окружность с центром в точке 1 и радиуса r и проходит окружность с течением времени t равномерно против часовой стрелки. Если число r меньше 1, то окружность не содержит внутри себя начала координат и индекс пути равен 0. Если r больше 1, то окружность содержит внутри себя начало координат и индекс пути равен 1 (проверьте это самостоятельно). В случае $r = 1$ путь проходит через начало координат и его индекс не определен. Если число r меняется, то путь (18), как говорят, деформируется (рис. 6). Мы видим на этом примере, что во время деформации замкнутого пути его индекс не меняется, если только путь в какой-то момент деформации не проходит через начало координат.

Говоря, что путь описывается движением точки во времени, мы лишь хотели придать более интуитивный характер определению пути. В действительности же речь идет о зависимости комплексной переменной z от некоторого действительного параметра t (который можно обозначить и другой буквой). Так, например, путь (18) можно записать в виде

$$1 + r(\cos\alpha + i\sin\alpha),\quad 0 \leqslant \alpha \leqslant 2\pi, \qquad (19)$$

где параметром уже является не t, а α. Ясно, что путь (19) описывается точкой $1 + z$, когда точка z описывает путь

$$r(\cos\alpha + i\sin\alpha),\quad 0 \leqslant \alpha \leqslant 2\pi.$$



Деформация пути $1+r(\cos\alpha+i\sin\alpha)$ при изменении r

Рис. 6

Здесь r есть числовой параметр, от которого зависит сам путь. Говорят, что при изменении r путь (19) деформируется.

Дадим более сложный пример замкнутого пути, который мы обозначим через K_n. Его зададим формулой

$$z = (1 + s\cos t)(\cos nt + i\sin nt),\qquad 0 \leqslant t \leqslant 2\pi.$$

Здесь мы считаем, что $0 \leqslant s < 1$, и потому $|z| = 1 + s\cos t$ и положителен при произвольном t.

Путь K_n имеет самопересечения. При $n = 2$ одно самопересечение, при $n = 3$ два.

Введем понятие деформации пути. Будем считать, что путь деформируется, если он постепенно меняется без скачков в зависимости от некоторого параметра, который для пути (19) обозначен через r, а вообще может быть обозначен и другой буквой, например s (рис. 7). Таким образом, деформирующийся путь записывается формулой

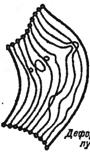

Деформация пути.

Рис. 7

$$z(\alpha, s),\quad \alpha_0 \leqslant \alpha \leqslant \alpha_1,\quad s_0 \leqslant s \leqslant s_1. \qquad (20)$$

Здесь при каждом фиксированном значении параметра s мы имеем определенный путь, описываемый во время изменения α от α_0 до α_1, а при изменении s сам путь меняется, деформируясь. Ясно, что если путь (20) замкнут, т. е. если при любом значении s имеет место равенство

$$z(\alpha_0, s) = z(\alpha_1, s),$$

то в течение деформации индекс пути должен меняться без скачков. А так как он есть целое число, то индекс этот остается постоянным. Конечно, это верно только в том случае, когда для произвольного значения S путь (20) не проходит через начало координат. В противном случае для этого значения s индекс пути не определен. Таким образом, мы можем высказать следующее утверждение.

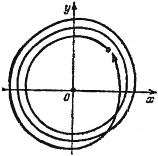


Рис. 8

Если замкнутый путь непрерывно деформируется, не проходя в процессе деформации через начало координат, то индекс его не меняется.

Дадим еще один пример замкнутого пути:

$$r^n(\cos n\alpha + i \sin n\alpha), \quad 0 \leqslant \alpha \leqslant 2\pi. \tag{21}$$

Ясно, что путь этот описывается точкой z^n, когда точка z описывает замкнутый путь:

$$r(\cos\alpha + i\sin\alpha), \quad 0 \leqslant \alpha \leqslant 2\pi.$$

Видно, что когда α меняется от 0 до 2π, аргумент точки z^n меняется от 0 до $2n\pi$. Таким образом, индекс пути (21) равен n (рис. 8, где схематически показан случай $n = 3$).

Комплексные функции комплексной переменной

Если числовое значение комплексной переменной величины w можно найти, зная числовое значение другой комплексной переменной величины z, то переменная величина w называется *функцией* переменной величины z, что записывается в форме

$$w = f(z).$$

Если комплексная функция $f(z)$ комплексной переменной z имеет производную, то она называется *аналитической функцией*. Теория аналитических функций является теперь одним из важнейших разделов математики. Здесь нас будут интересовать лишь аналитические функции очень частного вида, именно многочлены *).

Рассмотрим многочлен

$$w = f(z) = z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n, \qquad (22)$$

где $a_1, a_2, \ldots, a_n$ — комплексные коэффициенты, а n — неотрицательное целое число, которое называется *степенью* многочлена. Целью нашего исследования будет доказательство того, что многочлен (22) положительной степени имеет корень, т. е. что уравнение

$$f(z) = 0,$$

где $f(z)$ — многочлен (22) и $n > 0$, имеет решение. Для доказательства этого мы рассмотрим замкнутый путь

$$f(r(\cos\alpha + i\sin\alpha)), \quad 0 \leqslant \alpha \leqslant 2\pi, \qquad (23)$$

который описывает точка $f(z)$, когда точка z описывает замкнутый путь

$$r(\cos\alpha + i\sin\alpha), \quad 0 \leqslant \alpha \leqslant 2\pi.$$

Случай, когда свободный член a_n многочлена $f(z)$ равен 0, не требует рассмотрения, так как в этом случае многочлен $f(z)$ имеет очевидный корень $z = 0$. Поэтому мы будем считать, что $a_n \neq 0$. Замкнутый путь (23) зависит от параметра r и при изменении параметра r деформируется. При $r = 0$ число z равно 0, и путь $f(z)$ состоит из неподвижной точки a_n. Таким образом, его индекс при $r = 0$ равен 0. Мы докажем, что если взять r достаточно большим, то индекс пути (23) равен n. Но по предположению $n \neq 0$, поэтому при изменении числа r от большого значения к нулю путь (23), деформируясь, пройдет при каком-то значении r через начало координат, а это и значит, что при некотором значении z функция $f(z)$ обратится в нуль, т. е. корень у этого многочлена существует (рис. 9).

Для доказательства того, что при достаточно большом r замкнутый путь (23) имеет индекс, равный n,

---

*) Замечание о производной имеет целью лишь указать на существование теории аналитических функций. В дальнейшем производная функции комплексной переменной использоваться не будет.

продеформируем этот путь в более простой, индекс которого легко сосчитать.

Прежде всего мы разобьем многочлен $f(z)$ на сумму двух многочленов

$$f(z) = z^n + g(z),$$

где $g(z)$ задается формулой

$$g(z) = a_1 z^{n-1} + a_2 z^{n-2} + \ldots + a_{n-1} z + a_n.$$

Так как коэффициенты $a_1, a_2, \ldots, a_n$ многочлена $g(z)$ суть вполне определенные числа, то они все не превосходят по модулю некоторую константу c. Из неравенства (7), распространенного на произвольное число слагаемых, следует, что при $|z| > 1$

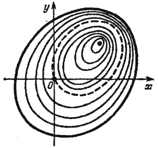


Рис. 9

$$|g(z)| \leqslant nc|z^{n-1}|. \quad (24)$$

Рассмотрим многочлен $f(z, s)$, зависящий от параметра s, $0 \leqslant s \leqslant 1$, задаваемый следующей формулой:

$$f(z, s) = z^n + sg(z).$$

Мы имеем равенство

$$z^n = f(z, s) - sg(z),$$

откуда

$$|z^n| \leqslant |f(z, s)| + |-sg(z)| \leqslant |f(z, s)| + nc|z^{n-1}|s \leqslant$$
$$\leqslant |f(z, s)| + nc|z^{n-1}|$$

(см. (24)). Отсюда следует

$$|f(z, s)| \geqslant |z^n| - nc|z^{n-1}|.$$

Обозначим $|z|$ через r, тогда последнее неравенство принимает вид

$$|f(z, s)| \geqslant r^n - ncr^{n-1} = r^{n-1}(r - cn).$$

Таким образом, при $r > cn$ правая часть предыдущего неравенства положительна. Следовательно, модуль функции $f(z, s)$ не обращается в нуль ни при каком значении s, если только $r > cn$.

Обратим внимание теперь на тот факт, что при $s = 0$ многочлен $f(z, s)$ превращается в известный нам многочлен z^n, а индекс пути z^n, когда z описывает окружность $r(\cos\alpha + i\sin\alpha)$, нами уже сосчитан (см. (21)). Он равен n. При $s = 1$ многочлен $f(z, s)$ превращается в многочлен $f(z)$, и определяемый им путь (23) имеет индекс, тоже равный n. Итак, мы доказали, что индекс пути (23), определяемый многочленом $f(z)$, при $r > cn$ равен n. Таким образом, при меняющемся r от 0 до $cn + \varepsilon$, где $\varepsilon > 0$, путь (25) деформируется, причем при $r = 0$ этот путь превращается в одну точку и его индекс равен 0, а при $r = cn + \varepsilon$ его индекс равен n. Из этого видно, что в процессе изменения r путь (23) при некотором значении r проходит через начало координат и, следовательно, многочлен $f(z)$ при некотором значении z, $|z| \leqslant cn + \varepsilon$, обращается в нуль.

Итак, основная теорема алгебры доказана (см. гл. 4, пример 12).

§ 10. Алгоритм Евклида

Деление многочленов

При делении целого положительного числа a на целое положительное число b мы приходим к равенству

$$a = bh + k, \tag{25}$$

где h и k — целые неотрицательные числа и $k < b$. Число h называется *частным*, а k — *остатком* при делении числа a на число b.

Эта формула получается в результате деления одного целого числа на другое целое число и описана в арифметике с использованием знака $\underline{|\quad}$: $256\,\underline{|\,21}$. Этот способ излагается в начальных классах средней школы. Точно так же со знаком $\underline{|\quad}$ можно делить друг на друга и многочлены: $x^3 + 4x^2 + 5x + 1\,\underline{|\,x^2 + 6x + 7}$. В результате получается следующее:

Будем исходить из двух многочленов

$$a(z) = a_0 z^p + a_1 z^{p-1} + \ldots + a_p,$$
$$b(z) = b_0 z^q + b_1 z^{q-1} + \ldots + b_q.$$

Мы будем предполагать здесь, что числа a_0 и b_0 не равны нулю, так что многочлен $a(z)$ имеет степень p, а многочлен $b(z)$ имеет степень q. В результате деления многочлена $a(z)$ на многочлен $b(z)$ при помощи $\underline{|\quad}$ придем к следующему равенству, аналогичному равенству (25):

$$a(z)=b(z)h(z)+k(z), \tag{26}$$

где степень многочлена $k(z)$ меньше q. Многочлены $h(z)$ и $k(z)$ называются соответственно *частным* и *остатком* при делении многочлена $a(z)$ на многочлен $b(z)$.

Если $k(z)=0$, то говорят, что многочлен $a(z)$ делится на многочлен $b(z)$, а $h(z)$ является частным от их деления.

Равенство (25) доказано в арифметике, а равенство (26) должно доказываться в алгебре. Но деление многочленов не входит в ныне действующую школьную программу. Чтобы доказать равенство (26), мы должны построить такие многочлены $h(z)$ и $k(z)$, которые удовлетворяют этому равенству. Процесс этого построения представляет собою очень важный алгоритм, осуществляемый при помощи знака $\underline{|\quad}$, так называемый *алгоритм Евклида*. Опишем его в общем виде.

Если $p<q$, то тогда $h(z)=0$, $k(z)=a(z)$, и равенство (26) выполнено.

Теперь мы будем строить многочлены $h(z)$ и $k(z)$ в предположении, что $p \geqslant q$. Сперва построим равенство

$$a(z)=b(z)h_1(z)+a_1(z), \tag{27}$$

в котором степень многочлена $a_1(z)$ меньше p. Для этого положим

$$h_1(z)=\frac{a_0}{b_0}z^{p-q}.$$

Тогда разность

$$a(z)-b(z)h_1(z)=a_1(z)$$

имеет степень, меньше p, так как в этом многочлене коэффициент при z^p равен нулю, а остальные степени z, входящие в этот многочлен, очевидно, меньше p. Таким образом, равенство (27) построено.

Если многочлен $a_1(z)$ имеет степень меньше q, то равенство (27) уже является равенством (26). В противоположном случае к многочлену $a_1(z)$ применим ту же процедуру, которая применена к многочлену $a(z)$ при построении равенства (27). И тогда мы получим для него равенство

$$a_1(z)=b(z)h_2(z)+a_2(z),$$

причем степень многочлена $a_2(z)$ уже меньше, чем степень многочлена $a_1(z)$. Если многочлен $a_2(z)$ уже имеет степень, меньшую чем q, то, подставляя $a_1(z)$ из последнего равенства в равенство (27), мы получим

$$a(z) = b(z)(h_1(z) + h_2(z)) + a_2(z),$$

которое уже является равенством (26). Если многочлен $a_2(z)$ тоже имеет степень, бо́льшую чем q, то мы продолжим наше построение дальше и в конце концов докажем нужное равенство (26).

Здесь мы описали процесс деления многочлена $a(z)$ на многочлен $b(z)$, т. е. нахождение многочленов $h(z)$ и $k(z)$, входящих в равенство (26). Докажем теперь, что эти многочлены $h(z)$ и $k(z)$ однозначно определены многочленами $a(z)$ и $b(z)$. Допустим, что наряду с равенством (26) имеет место равенство

$$a(z) = b(z)h_0(z) + k_0(z),$$

причем степень многочлена $k_0(z)$ меньше q. Вычитая это равенство из равенства (26), получим

$$b(z)(h(z) - h_0(z)) = k_0(z) - k(z).$$

Так как степень многочлена $b(z)$ равна q, а степень многочлена $k_0(z) - k(z)$ меньше q, то последнее равенство может иметь место лишь при условии $h(z) - h_0(z) \equiv 0$, так что и $k(z) - k_0(z) \equiv 0$.

Заметим, что при делении многочленов не могут возникнуть комплексные числа. Именно,

если многочлены $a(z)$ и $b(z)$ имеют действительные коэффициенты, то многочлены $h(z)$ и $k(z)$ также имеют действительные коэффициенты.

Разложение многочлена на множители

Теперь существующий по основной теореме алгебры корень многочлена

$$f_0(z) = z^n + a_1 z^{n-1} + \ldots + a_n, \quad n > 0,$$

с комплексными коэффициентами обозначим через α_1. Докажем, что многочлен $f_0(z)$ делится на двучлен $z - \alpha_1$. Применяя формулу (26), мы получим частное, которое обозначим через $f_1(z)$, и некоторый остаток в виде многочлена нулевой степени, т. е. число, которое мы обозначим через k. Таким образом, мы имеем

$$f_0(z) = f_1(z)(z - \alpha_1) + k.$$

Так как $f_0(\alpha_1) = 0$, то, полагая в этом равенстве $z = \alpha_1$, получаем $k = 0$.

Итак, многочлен $f_0(z)$ разделился на $z-\alpha_1$, и мы имеем

$$f_0(z) = (z-\alpha_1) f_1(z),$$

где $f_1(z)$ — многочлен степени $n-1$, который, очевидно, начинается с члена z^{n-1}. Если $n > 1$, то $n-1 > 0$; тогда многочлен $f_1(z)$ будет положительной степени и по доказанному ранее имеет некоторый корень α_2. Таким образом, по только что доказанному многочлен $f_1(z)$ разлагается на множители:

$$f_1(z) = (z-\alpha_2) f_2(z). \tag{28}$$

Продолжая этот процесс дальше, мы получим разложение $f_0(z)$ на n линейных множителей:

$$f_0(z) = (z-\alpha_1)(z-\alpha_2) \ldots (z-\alpha_n).$$

Числа $\alpha_1, \alpha_2, \ldots, \alpha_n$ являются корнями многочлена $f_0(z)$, и других корней многочлен $f_0(z)$, очевидно, не имеет. Может, однако, оказаться, что один и тот же корень встречается в этом разложении несколько раз. Группируя равные между собой корни, мы получаем разложение

$$f_0(z) = (z-\alpha_1)^{k_1}(z-\alpha_2)^{k_2} \ldots (z-\alpha_q)^{k_q}, \tag{29}$$

где все корни $\alpha_1, \alpha_2, \ldots, \alpha_q$ различны. Число k_1 называется кратностью корня α_1, число k_2 — кратностью корня α_2 и т. д., число k_q — кратностью корня α_q. Таким образом, число различных корней многочлена $f_0(z)$ может быть и меньше чем n. Однако если учитывать кратность каждого корня, то сумма кратностей в точности равна n. В этом смысле многочлен $f_0(z)$ имеет ровно n корней.

Рассмотрим теперь случай, когда все коэффициенты многочлена $f_0(z)$ — действительные числа. О корнях такого многочлена можно высказать некоторые весьма интересные дополнительные соображения.

Из формул (14)—(16) легко получить, что для многочлена $f_0(z)$ с действительными коэффициентами имеет место равенство

$$\bar{f}_0(z) = f_0(\bar{z}).$$

Из этого равенства следует, что если α_1 есть корень многочлена $f_0(z)$ с действительными коэффициентами, то $\bar{\alpha}_1$ есть также его корень. В случае, если α_1 — действительное число, то это утверждение бессодержательно. В случае, если α_1 не есть действительное число, то утверждение указывает на существование наряду с корнем α_1 отличного от него корня $\bar{\alpha}_1$. Таким образом, если α_1 — не действительное число, то многочлен $f_1(z)$ (см. (29)) имеет делителем $z-\bar{\alpha}_1$, и мы получаем разложение на множители

$$f_0(z)=(z-\alpha_1)(z-\bar{\alpha}_1)f_2(z).$$

Таким образом, мы выделили у многочлена $f_0(z)$ квадратичный множитель

$$g_2(z)=(z-\alpha_1)(z-\bar{\alpha}_1)=z^2-(\alpha_1+\bar{\alpha}_1)z+\alpha_1\bar{\alpha}_1.$$

Многочлен $g_2(z)$ очевидным образом имеет действительные коэффициенты. Отсюда следует, что $f_2(z)$ — также многочлен с действительными коэффициентами, так как он получается в результате деления многочлена $f_0(z)$ на действительный квадратичный трехчлен $g_2(z)$ (см. (28)).

Таким образом, если α_1 — действительный корень, то мы имеем разложение

$$f_0(z)=g_1(z)f_1(z),$$

где $g_1(z)=z-\alpha_1$, а если α_1 — не действительный корень, то мы имеем разложение

$$f_0(z)=g_2(z)f_1(z).$$

Таким образом, в обоих случаях $f_0(z)$ делится на действительный многочлен первой или второй степени.

§ 11. Наибольший общий делитель двух многочленов

Многочлен $b(z)$ считается *делителем* многочлена $a(z)$, если при делении $a(z)$ на $b(z)$ не получается остатка, т. е. в формуле (26) $k(z)\equiv 0$.

Многочлен $b(z)$ считается *общим делителем* двух многочленов $a_0(z)$ и $a_1(z)$, если он является делителем каждого из этих двух многочленов.

Общий делитель $c(z)$ двух многочленов $a_0(z)$ и $a_1(z)$ называется их *наибольшим общим* делителем, если он делится на каждый общий делитель $b(z)$ многочленов $a_0(z)$ и $a_1(z)$.

Существование наибольшего общего делителя двух многочленов не очевидно. Ниже оно будет доказано при помощи алгоритма Евклида, т. е. при помощи таких операций, которые совершаются без всяких трудностей, кроме громоздких вычислений. Но прежде всего мы докажем, что два наибольших общих делителя $c_0(z)$ и $c_1(z)$ двух многочленов $a_0(z)$ и $a_1(z)$ по существу совпадают. Именно, они могут отличаться друг от друга только числовым множителем, отличным от нуля. Докажем это.

Так как $c_0(z)$ является наибольшим общим делителем многочленов $a_0(z)$ и $a_1(z)$, а $c_1(z)$ — их общим делителем, то $c_0(z)$ делится на $c_1(z)$, и, следовательно, мы имеем тождество

$$c_0(z) = h_1(z)\, c_1(z). \tag{30}$$

Аналогично, мы имеем формулу

$$c_1(z) = h_0(z)\, c_0(z). \tag{31}$$

Подставляя выражение $c_1(z)$ из формулы (31) в формулу (30), мы получаем

$$c_0(z) = h_0(z)\, h_1(z)\, c_0(z).$$

А так как при перемножении многочленов степени их складываются, то из последней формулы следует, что многочлены $h_0(z)$ и $h_1(z)$ имеют степень нуль, т. е. являются числами.

Таким образом, единственность наибольшего общего делителя многочленов $a_0(z)$ и $a_1(z)$ нами доказана.

Перейдем теперь к построению наибольшего общего делителя $c(z)$ многочленов $a_0(z)$ и $a_1(z)$. Для этого произведем деление многочлена $a_0(z)$ на многочлен $a_1(z)$. Запишем формулу (26) для этих многочленов в виде

$$a_0(z) = a_1(z)\, h_1(z) + a_2(z), \tag{32}$$

т. е. обозначим остаток от этого деления через $a_2(z)$. Теперь подвергнем многочлен $a_1(z)$ делению на многочлен $a_2(z)$ и формулу (26) запишем в виде

$$a_1(z) = a_2(z)\, h_2(z) + a_3(z), \tag{33}$$

т. е. обозначим остаток от деления через $a_3(z)$. Теперь подвергнем делению многочлен $a_2(z)$ на многочлен $a_3(z)$ и запишем формулу (26) в виде

$$a_2(z) = a_3(z)\, h_3(z) + a_4(z).$$

Так как в процессе этого построения степень остаточного многочлена все время снижается, то мы дойдем до такого остатка, который будет иметь степень нуль, а при делении на многочлен степени нуль остаток, очевидно, равен нулю. Таким образом, в конце концов мы получим тождества

$$a_{n-2}(z) = a_{n-1}(z) h_{n-1}(z) + a_n(z), \qquad (34)$$

$$a_{n-1}(z) = a_n(z) h_n(z). \qquad (35)$$

Если теперь $b(z)$ есть общий делитель многочленов $a_0(z)$ и $a_1(z)$, то из формулы (32) следует, что он является делителем $a_2(z)$. Из формулы (33) следует, что многочлен $b(z)$ является делителем многочлена $a_3(z)$. Таким образом, последовательно мы докажем, что многочлен $b(z)$ является делителем всех построенных нами многочленов

$$a_0(z), \ a_1(z), \ \ldots, \ a_n(z), \qquad (36)$$

в частности многочлена $a_n(z)$. А из формулы (35) следует, что многочлен $a_n(z)$ является делителем $a_{n-1}(z)$, из формулы (34) следует, что он является делителем $a_{n-2}(z)$. В конце концов мы установим, что $a_n(z)$ является делителем многочленов $a_0(z)$ и $a_1(z)$. Таким образом, доказано, что $a_n(z)$ делится на любой делитель многочленов $a_0(z)$ и $a_1(z)$ и сам является их делителем. Следовательно, многочлен $a_n(z)$ является наибольшим общим делителем многочленов $a_0(z)$ и $a_1(z)$.

Докажем теперь следующий важный результат, имеющий многочисленные применения в алгебре.

А) Общий наибольший делитель $c(z)$ двух многочленов $a_0(z)$ и $a_1(z)$ может быть записан формулой

$$c(z) = p_0(z) a_0(z) + p_1(z) a_1(z), \qquad (37)$$

где $p_0(z)$, $p_1(z)$ — некоторые многочлены.

Докажем это утверждение. Подставляя выражение многочлена $a_2(z)$ из формулы (32) в формулу (33), мы получим формулу

$$a_3(z) = p_2(z) a_0(z) + q_0(z) a_1(z).$$

Продолжая этот процесс дальше, мы получим формулу (37). Таким образом, формула (37) доказана.

Устранение кратных корней

Нахождение корней многочлена является одной из наиболее старых и трудных задач алгебры. Первоначально ее пытались решить при помощи формулы в виде алгебраических выражений, включающих извлечение корней. Это очень удачно сделано для квадратного уравнения. Для кубического уравнения также имеется формула, но уже мало значительная в смысле приложений, для уравнений четвертой степени также есть формула, но лишенная всякого практического значения. Позже было доказано, что для уравнений выше четвертой степени общей формулы решения уравнений в радикалах не существует. Полная теория о возможностях нахождения корней многочлена при помощи радикалов была построена Галуа. На пути решения проблемы отыскания корней многочлена лежит следующая гораздо более простая задача: для многочлена $f(z)$, имеющего как простые так и кратные корни, найти такой многочлен $g(z)$, который имеет те же самые корни, что и $f(z)$, но только простые. Эта задача решается при помощи алгоритма Евклида, т. е. при помощи деления многочленов. Ее решение мы рассмотрим здесь.

Пусть $a(z)$ и $b(z)$ — два многочлена, а $p(z)$ — их произведение

$$p(z) = a(z)\, b(z).$$

Обозначим через

$$\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_q$$

совокупность всех чисел, каждое из которых является корнем хотя бы одного из многочленов $a(z)$ и $b(z)$. Тогда эти многочлены могут быть записаны в следующем виде (см. (30)):

$$a(z) = (z-\alpha_1)^{k_1}(z-\alpha_2)^{k_2}\ldots(z-\alpha_q)^{k_q}, \tag{38}$$

$$b(z) = (z-\alpha_1)^{l_1}(z-\alpha_2)^{l_2}\ldots(z-\alpha_q)^{l_q}. \tag{39}$$

В обоих этих разложениях некоторые показатели степеней могут равняться нулю и мы будем считать, что кратность соответствующих корней является нулем. Очевидно, мы имеем

$$p(z) = (z-\alpha_1)^{k_1+l_1}(z-\alpha_2)^{k_2+l_2}\ \ldots\ (z-\alpha_q)^{k_q+l_q}.$$

Таким образом, при перемножении многочленов кратности корней у них складываются.

Исходя из разложения многочленов $a(z)$, $b(z)$ на множители (см. (38) и (39)), легко найти их наибольший общий делитель $c(z)$. Для этого обозначим через m_1 наименьшее из чисел k_1 и l_1, через m_2 наименьшее из чисел k_2 и l_2 и т. д., через m_q наименьшее из чисел k_q и l_q. Тогда наибольший общий делитель многочленов $a(z)$ и $b(z)$ задается формулой

$$c(z) = (z-\alpha_1)^{m_1}(z-\alpha_2)^{m_2}\ \ldots\ (z-\alpha_q)^{m_q}.$$

Этот простой способ нахождения наибольшего общего делителя двух многочленов требует, однако, нахождения всех корней обоих многочленов, и поэтому он практически мало пригоден, так как сводит простую задачу к очень трудной. Между тем нахождение наибольшего общего делителя двух многочленов, как было показано ранее, представляет собой очень простую задачу, решающуюся при помощи алгоритма Евклида, т. е. при помощи деления многочленов.

Выясним теперь, как связаны между собой кратности корней многочленов $a_0(z)$ и $a_1(z) = a_0'(z)$, т. е. кратности корней многочлена $a_0(z)$ и его производной $a_0'(z)$.

Оказывается, что если корень α многочлена $a_0(z)$ имеет кратность k, то тот же корень для многочлена $a_1(z) = a_0'(z)$ имеет кратность $k-1$. Но, конечно, кроме корней, являющихся корнями многочлена $a_0(z)$, многочлен $a_0'(z)$ может иметь и другие корни.

Докажем, что если α — корень многочлена $a_0(z)$ кратности k, то тот же корень является корнем многочлена $a_1(z) = a_0'(z)$ кратности $k-1$.

Из разложения (30) многочлена $a_0(z)$ на множители следует, что если α — корень многочлена $a_0(z)$ кратности k, то

$$a_0(z) = (z-\alpha)^k b(z),$$

причем $b(z)$ уже не делится на $z-\alpha$ и, следовательно, не имеет α своим корнем. Дифференцируя последнее соотношение, получаем

$$a_0'(z) = a_1(z) = k(z-\alpha)^{k-1} b(z) + (z-\alpha)^k b'(z), \quad (40)$$

$$(z-\alpha)^{k-1} b(z) = \frac{1}{k} a_1(z) + \frac{1}{k}(z-\alpha)^k b'(z). \quad (41)$$

Из формулы (40) видно, что $a_1(z)$ делится на $(z-\alpha)^{k-1}$, а из формулы (41) следует, что $a_1(z)$ не делится на $(z-\alpha)^k$. В самом деле, если бы $a_1(z)$ делился на $(z-\alpha)^k$, то $b(z)$ делился бы на $z-\alpha$. Таким образом, кратность корня α для многочлена $a_1(z)$ есть $k-1$. В частности, если α — простой корень многочлена $a_0(z)$, то α не будет корнем многочлена $a_0'(z)$.

Таким образом, если $a_0(z)$ имеет разложение (30), то наибольшим общим делителем многочленов $a_0(z)$ и $a_1(z)$ является многочлен

$$c(z) = (z-\alpha_1)^{k_1-1}(z-\alpha_2)^{k_2-1}\ldots(z-\alpha_q)^{k_q-1}.$$

Отсюда видно, что многочлен $a_0(z)$ делится на многочлен $c(z)$, и частное имеет вид

$$\frac{a_0(z)}{c(z)} = (z - \alpha_1)(z - \alpha_2) \ldots (z - \alpha_q).$$

Таким образом, сформулированная в начале параграфа задача о нахождении многочлена $g(z)$, имеющего те же самые корни, что и многочлен $f(z)$, но только кратности 1, решена. Именно, для нахождения многочлена $g(z)$ следует сначала при помощи алгоритма Евклида найти общий наибольший делитель многочленов $f(z)$ и $f'(z)$ и затем разделить многочлен $f(z)$ на этот наибольший общий делитель.

Подсчет числа действительных корней многочлена на заданном отрезке

Нахождение комплексных корней многочлена является задачей более сложной, чем нахождение его действительных корней, так как в случае комплексного корня нам приходится фактически решать уравнение с двумя неизвестными, которыми являются действительная часть корня и его коэффициент при мнимой части.

Поскольку нами дан способ, позволяющий свести вычисление корней многочлена к вычислению корней многочлена, имеющего только простые корни, то мы займемся здесь именно этим случаем многочлена с простыми корнями. Здесь будет дан результат, позволяющий в значительной степени облегчить приближенное вычисление действительных корней многочлена.

Будем рассматривать многочлен $b_0(x)$ действительной переменной x, имеющий только простые корни, и дадим способ определения числа этих корней на заданном отрезке $\xi_0 \leqslant x \leqslant \xi_1$. Так как все корни многочлена $b_0(x)$ простые, то многочлены $b_0(x)$ и $b_1(x) = b_0'(x)$ не имеют общих корней и ни для какого значения x не могут одновременно обращаться в нуль.

Для решения этой задачи рассмотрим последовательность действительных чисел

$$b_0, b_1, \ldots, b_n. \tag{42}$$

Будем считать, что ни одно из этих чисел не обращается в нуль. Почти что само собой понятно, что значит число перемен знака в этой последовательности (42). Определим его формально. Мы будем считать, что между b_0 и

b_1 имеется перемена знака, если одно из этих чисел положительно, а другое отрицательно. Обозначим через p_1 в этом случае число 1. Если же числа b_0 и b_1 оба положительны, либо оба отрицательны, то будем считать, что перемены знака между ними нет. В этом случае обозначим через p_1 число, равное нулю. Точно так же определим число p_2, которое оценивает перемену знака между b_1 и b_2. Таким образом, мы получим последовательность нулей и единиц

$$p_1, p_2, \ldots, p_n.$$

Сумма всех чисел из этой последовательности есть число перемен знака в последовательности (42).

Исходя из многочленов $b_0(x)$ и $b_1(x) = b_0'(x)$, путем последовательного деления построим последовательность многочленов.

Разделим многочлен $b_0(x)$ на многочлен $b_1(x)$ и запишем формулу (26) в виде

$$b_0(x) = h_1(x)\, b_1(x) - b_2(x).$$

Разделим теперь многочлен $b_1(x)$ на многочлен $b_2(x)$ и запишем формулу (26) в виде

$$b_1(x) = h_2(x)\, b_2(x) - b_3(x).$$

На k-м шаге этого процесса мы получим тождество

$$b_{k-1}(x) = h_k(x)\, b_k(x) - b_{k+1}(x). \qquad (43)$$

Продолжая этот процесс до конца, мы придем к последовательности многочленов

$$b_0(x),\ b_1(x),\ \ldots,\ b_n(x). \qquad (44)$$

Процесс построения этой последовательности почти в точности совпадает с процессом нахождения наибольшего общего делителя двух многочленов, только многочлены ряда (44) могут отличаться от многочленов ряда (36) знаками. Так как многочлены $b_0(x)$ и $b_n(x)$ взаимно просты, то последний член последовательности, — именно, $b_n(x)$ — есть число.

Заметим прежде всего, что ни при каком значении $x = x_0$ два рядом стоящих многочлена последовательности (44) не могут обращаться в нуль. В самом деле, если $b_k(x_0) = b_{k+1}(x_0) = 0$, то из соотношения (43) следует, что $b_{k-1}(x_0) = 0$. Продолжая этот процесс, мы придем к выводу, что многочлены $b_0(x)$ и $b_1(x)$ обращаются в нуль при $x = x_0$, а это невозможно, так как они не имеют общих корней.

Пусть теперь x возрастает, начиная от значения ξ_0 и кончая ξ_1. Ясно, что при таком росте x число перемен знака последовательности чисел (44) может измениться. А это может произойти только при прохождении x через такое значение x_0, при котором один из членов последовательности (44) меняет знак, т. е. проходит через нуль. Допустим, что при $x = x_0$ мы имеем $b_k(x_0) = 0$ и при росте x вблизи точки x_0 знак функции $b_k(x)$ меняется. Полагая в соотношении (43) $x = x_0$, получим равенство

$$b_{k-1}(x_0) = -b_{k+1}(x_0).$$

Таким образом, числа $b_{k-1}(x_0)$ и $b_{k+1}(x_0)$ имеют противоположные знаки. Если при x, близком к x_0, число $b_k(x)$ имеет знак, совпадающий со знаком $b_{k-1}(x)$, то знаки чисел $b_k(x)$ и $b_{k+1}(x)$ различны. Таким образом, между $b_{k-1}(x)$ и $b_k(x)$ перемены знака нет, а между $b_k(x)$ и $b_{k+1}(x)$ перемена знака есть. Если при прохождении точки x через x_0 знак числа $b_k(x)$ меняется, то для этого x между $b_{k-1}(x)$ и $b_k(x)$ перемена знака есть, а между $b_k(x)$ и $b_{k+1}(x)$ перемены знака нет. Таким образом, устанавливается, что при прохождении точки x через x_0, при котором $b_k(x_0)$ обращается в нуль, число перемен знаков в трехчленной последовательности

$$b_{k-1}(x),\ b_k(x),\ b_{k+1}(x)$$

не меняется (см. табл. 1). Номер k обладает лишь тем свойством, что у него есть предыдущий номер $k-1$ и последующий $k+1$. Таким образом, число перемен знака в последовательности (44) может измениться лишь тогда, когда при прохождении x через x_0 меняет свой знак либо первый член последовательности (44), либо самый последний ее член. Но последний член не может менять знака, так как это есть число.

Таблица 1

	$b_{k-1}(x)$	$b_k(x)$	$b_{k+1}(x)$
При $x < x_0$	+	+	—
При $x > x_0$	+	—	—

Посмотрим теперь, как меняется число перемен знака в последовательности (44), когда в точке x_0 первый член последовательности (44) обращается в нуль. Если при x, близком к x_0, но меньшем x_0, $b_0(x) > 0$, то при прохождении x через точку x_0 функция $b_0(x)$ изменит знак. Но при $x = x_0$ число $b_1(x_0)$ будет отрицательным, так как при этом функция $b_0(x)$ убывает, и следовательно, производная ее $b_1(x)$ отрицательна. Таким образом, при x, близком к x_0, но меньшем x_0, между $b_0(x)$ и $b_1(x)$ имеется перемена знака, а при x, большем x_0, но близком к x_0, между $b_0(x)$ и $b_1(x)$ перемены знака нет. Поэтому при прохождении точки x через x_0, при котором функция $b_0(x)$ обращается в нуль и убывает, перемена знака между $b_0(x)$ и $b_1(x)$ исчезает. Точно так же устанавливается, что если при переходе через точку x_0 функция $b_0(x)$ возрастает, то перемена знака между $b_0(x)$ и $b_1(x)$ исчезает. Итак, установлено, что перемена знака в последовательности (44) при изменении x может произойти лишь тогда, когда x проходит через корень многочлена $b_0(x)$, при этом каждый раз одна перемена знака исчезает. Следовательно, когда x возрастает от ξ_0 до ξ_1, в последовательности (44) происходит потеря одной перемены знака при прохождении x через корень многочлена $b_0(x)$. Поэтому для нахождения числа корней многочлена $b_0(x)$ между ξ_0 и ξ_1 надо сравнить числовые последовательности

$$b_0(\xi_0),\ b_1(\xi_0),\ \ldots,\ b_n(\xi_0), \tag{45}$$

$$b_0(\xi_1),\ b_1(\xi_1),\ \ldots,\ b_n(\xi_1). \tag{46}$$

Число перемен знака в последовательности (45) может быть лишь больше, чем число перемен знака в последовательности (46).

Разность между числом перемен знака в последовательности (45) и числом перемен знака в последовательности (46) есть число корней многочлена $b_0(x)$, находящихся в промежутке $\xi_0 \leqslant x \leqslant \xi_1$.

Глава 3

ПРИВЕДЕНИЕ МАТРИЦ К КАНОНИЧЕСКОМУ ВИДУ

Теперь, когда построены комплексные числа, можно рассматривать комплексные векторы, т. е. векторы, координаты которых являются комплексными числами, и комплексные векторные пространства, состоящие из комплексных векторов. Линейные отображения комплексных векторных пространств друг в друга приведут нас к матрицам, элементы которых являются комплексными числами. Все результаты первой главы, верные для действительного случая, верны и для комплексного случая.

Если, однако, рассматривать задачу, сформулированную для действительного случая, а комплексные числа появляются в ней в результате каких-нибудь вычислений, то приходится делать различие между комплексными и действительными величинами. Для того чтобы выделить все действительные векторы в некотором комплексном векторном пространстве, надо задать в нем некоторый действительный базис

$$e_1, e_2, \ldots, e_n,$$

и тогда действительный вектор x в рассматриваемом пространстве запишется в виде

$$x = x^{\alpha} e_{\alpha}, \quad \alpha = 1, \ldots, n,$$

где $x_1, x_2, \ldots, x_n$ — действительные числа.

Если в том же базисе выбран комплексный вектор

$$x = x^{\alpha} e_{\alpha}, \quad \alpha = 1, \ldots, n,$$

где $x^1, x^2, \ldots, x^n$ — комплексные числа, то можно определить вектор $\bar{x}$, комплексно сопряженный x, положив

$$\bar{x} = \bar{x}^\alpha e_\alpha. \tag{1}$$

В дальнейшем, если, исходя из некоторого действительного векторного пространства A^n с базисом

$$e_1, e_2, \ldots, e_n,$$

мы будем рассматривать векторы с комплексными координатами, то придем к комплексному векторному пространству $\bar{A}^n$, которое естественно считать комплексным расширением действительного пространства A^n. В комплексном расширении $\bar{A}^n$ действительного векторного пространства A^n определено понятие комплексно сопряженного вектора (см. (1)). Если в действительном векторном пространстве A^n задана симметричная билинейная форма $f(x, y)$, то она записывается в виде

$$f(x, y) = a_{\alpha\beta} x^\alpha y^\beta, \qquad \alpha, \beta = 1, \ldots, n,$$

где x и y — действительные векторы, а коэффициенты $a_{ij} = a_{ji}$ — также действительные числа. Переходя к комплексному расширению $\bar{A}^n$ действительного векторного пространства A^n, мы можем задать комплексную симметричную форму

$$f(x, y) = a_{\alpha\beta} x^\alpha y^\beta, \tag{2}$$

где векторы x и y уже принадлежат пространству $\bar{A}^n$. Подставим теперь вместо вектора y в билинейную форму (2) вектор $\bar{x}$, т. е. положим $y = \bar{x}$. Оказывается тогда, что $f(x, \bar{x})$ есть действительное число. В самом деле, мы имеем

$$f(x, \bar{x}) = a_{\alpha\beta} x^\alpha \bar{x}^\beta,$$
$$\overline{f(x, \bar{x})} = \bar{a}_{\alpha\beta} \bar{x}^\alpha x^\beta = a_{\alpha\beta} x^\alpha \bar{x}^\beta.$$

Таким образом, мы получаем

$$\overline{f(x, \bar{x})} = f(x, \bar{x}),$$

а это и значит, что число $f(x, \bar{x})$ действительное. В частном случае, когда A^n есть действительное евклидово векторное пространство, мы приходим к выводу, что скалярное произведение $(x, \bar{x})$ вектора на сопряженный ему вектор есть действительное число:

$$\overline{(x, \bar{x})} = (x, \bar{x}).$$

§ 12. Связь между линейными отображениями и матрицами

В этом параграфе мы будем рассматривать линейное отображение φ векторного пространства A^n в себя. При выбранном базисе такому отображению φ соответствует вполне определенная квадратная матрица $A = \| a_j^i \|$. Однако при смене базиса матрица эта определенным образом меняется. Поскольку в первую очередь мы интересуемся линейными отображениями, то нас будут соответственно интересовать такие свойства соответствующих матриц, которые не меняются при смене базиса.

Преобразование матрицы при смене базиса

Пусть φ — линейное отображение векторного пространства в себя и

$$e_1, e_2, \ldots, e_n \tag{3}$$

— некоторый базис пространства A^n. При этом базисе отображению φ соответствует квадратная матрица

$$A = \| a_j^i \|, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n.$$

Элементы матрицы A определяются из формулы

$$\varphi(e_j) = a_j^\alpha e_\alpha.$$

Пусть теперь

$$f_1, f_2, \ldots, f_n \tag{4}$$

— некоторый другой базис пространства A^n. В этом новом базисе отображению φ соответствует матрица, которую мы обозначим через

$$B = \| b_j^i \|, \quad i, j = 1, 2, \ldots, n.$$

Установим связь между матрицами A и B.

А) Пусть

$$e_j = s_j^\alpha f_\alpha, \quad \alpha = 1, \ldots, n, \quad j = 1, \ldots, n, \tag{5}$$

— связь между базисами (3) и (4). Коэффициенты этого соотношения составляют матрицу

$$S = \| s_j^i \|, \quad i, j = 1, \ldots, n.$$

Элементы j-го столбца этой матрицы являются координатами вектора e_j в базисе (4). А так как векторы $e_1, e_2, \ldots, e_n$ линейно независимы и, следовательно, определитель матрицы S отличен от нуля, так что она имеет обратную матрицу S^{-1}. Оказывается, что матрицы A и B связаны соотношением

$$B = S^{-1}AS. \qquad (6)$$

Таким образом, B получается из A при помощи *трансформирования* матрицы A матрицей S.

Докажем соотношение (6). Пусть

$$S^{-1} = T = \| t_j^i \|, \quad i, j = 1, \ldots, n,$$

так что

$$t_i^\beta s_\beta^j = \delta_i^j.$$

Умножая соотношение (5) на t_i^j и суммируя полученный результат по j, мы получим

$$f_i = t_i^\beta e_\beta.$$

Таким образом, мы имеем

$$\varphi(f_i) = t_i^\beta \varphi(e_\beta) = t_i^\beta a_\beta^\alpha e_\alpha = t_i^\beta a_\beta^\alpha s_\alpha^\gamma f_\gamma.$$

Следовательно,

$$B = TAS = S^{-1}AS.$$

Собственные значения и векторы

Здесь нам придется употреблять тождественное отображение векторного пространства на себя. Его мы будем обозначать стандартным образом через θ, так что всегда имеет место соотношение

$$\theta x = x.$$

Тождественному отображению векторного пространства A^n на себя соответствует единичная матрица $E = \| \delta_j^i \|$.

В) Пусть A^n — n-мерное векторное пространство и

$$e_1, e_2, \ldots, e_n \qquad (7)$$

— некоторый его базис. Далее, пусть φ — некоторое линейное отображение пространства A^n в себя и

$$A = \| a_j^i \|$$

— матрица, соответствующая отображению φ в базисе (7). Число λ называется *собственным значением* отображения φ, а ненулевой вектор h — *собственным вектором* отображения φ, соответствующим этому собственному значению, если выполнено соотношение

$$\varphi h = \lambda h = \lambda \theta h. \tag{8}$$

Для того чтобы сделать попытку вычислить собственное значение λ и собственный вектор h, нужно перейти к координатной записи вектора h и вместо отображения φ употребить соответствующую ему матрицу A. Тогда соотношение (8) перепишется в виде

$$a^i_\alpha h^\alpha = \lambda h^i = \lambda \delta^i_\alpha h^\alpha.$$

Перенесем все члены этого соотношения в левую часть равенства. Тогда мы получим

$$\left(a^i_\alpha - \lambda \delta^i_\alpha\right) h^\alpha = 0. \tag{9}$$

Для того чтобы это уравнение имело ненулевое решение h, согласно следствию из теоремы 3 необходимо и достаточно, чтобы определитель матрицы $A - \lambda E$ был равен нулю. Таким образом, для числа λ мы получаем уравнение

$$D(A - \lambda E) = 0.$$

Левая часть этого уравнения представляет собой многочлен n-й степени относительно λ. Многочлен

$$\chi(z) = D(A - zE)$$

называется *характеристическим многочленом* матрицы A. Очевидно, что он имеет степень n. Таким образом, собственное значение является корнем многочлена $\chi(z)$ степени n. Если исходное векторное пространство A^n действительно, то матрица A тоже действительна и характеристический многочлен этой матрицы есть действительный многочлен, но он может не иметь действительных корней. Таким образом, собственное значение λ может оказаться комплексным. Вектор h, являющийся решением системы уравнений (9), тоже может быть комплексным. Таким образом, при нахождении собственных значений и собственных векторов нельзя ограничиться только действительными величинами, так как собственные значения и собственные векторы могут оказаться комплексными. Оказывается, что характеристический многочлен $\chi(z)$ матрицы A является в действительности характеристическим многочленом отображения φ, т. е. не зависит от случайного выбора базиса (7), а зависит только от отображения φ.

Докажем последнее утверждение. Перейдем от базиса ((7) или (3)) к базису (4). Тогда отображению φ будет соответствовать матрица $B = S^{-1}AS$ (см. А)), где матрица B зависит от базиса (4). Мы имеем

$$B - \lambda E = S^{-1}AS - \lambda S^{-1}ES = S^{-1}(A - \lambda E)S$$

и, далее,

$$D(B - \lambda E) = D(S^{-1}(A - \lambda E)S) =$$
$$= D(S^{-1})D(A - \lambda E)D(S) = D(A - \lambda E).$$

Таким образом, характеристические многочлены матриц A и B совпадают и, следовательно, характеристический многочлен зависит от отображения φ, а не от выбранного в пространстве A^n базиса.

Собственные векторы с различными собственными значениями

С) Оказывается, что собственные векторы отображения φ с попарно различными собственными значениями линейно независимы.

Утверждение это будем доказывать индуктивно по числу векторов. Пусть

$$\lambda_1, \lambda_2, \ldots, \lambda_r$$

— попарно различные собственные значения отображения φ, а

$$h_1, h_2, \ldots, h_r, \tag{10}$$

— соответствующие этим собственным значениям собственные векторы. Когда $r = 1$, система эта линейно независима, так как собственный вектор по определению отличен от нуля. Допустим, что при $r = s - 1$ система эта линейно независима. Докажем, что система линейно независима при $r = s$. Допустим, что имеется соотношение

$$c^{\alpha}h_{\alpha} = 0, \quad \alpha = 1, \ldots, s. \tag{11}$$

Применяя к этому соотношению отображение φ, получаем соотношение

$$c^{\alpha}\lambda_{\alpha}h_{\alpha} = 0. \tag{12}$$

Умножая теперь соотношение (11) на λ_s и вычитая полученное из соотношения (12), мы придем к соотношению

$$c^{\alpha}(\lambda_\alpha - \lambda_s) h_\alpha + c^s(\lambda_s - \lambda_s) h_s = 0, \qquad \alpha = 1, 2, \ldots, s-1.$$

В этом соотношении содержится только $s-1$ собственных векторов и потому по предположению индукции все коэффициенты соотношения равны нулю, т. е. мы имеем

$$c^1(\lambda_1 - \lambda_s) = 0, \quad c^2(\lambda_2 - \lambda_s) = 0, \quad \ldots, \quad c^{s-1}(\lambda_{s-1} - \lambda_s) = 0,$$

откуда получаем $c^1 = 0$, $c^2 = 0, \ldots, c^{s-1} = 0$. В силу этого соотношение (11) приобретает вид

$$c^s h_s = 0$$

и, следовательно, $c^s = 0$. Таким образом, из соотношения (11) вытекает обращение в нуль всех коэффициентов c_i, а это означает, что система (10) при $r = s$ линейно независима.

D) Если характеристический многочлен $\chi(z)$ отображения φ не имеет кратных корней, то базис в пространстве A^n можно выбрать таким образом, что соответствующая отображению φ матрица A имеет диагональный вид, причем на диагонали стоят корни многочлена $\chi(z)$. Это значит, что если задана квадратная матрица A порядка n и корни ее характеристического многочлена $\chi(z)$ все попарно различны, то можно подобрать такую квадратную матрицу S с определителем, отличным от нуля, что матрица

$$S^{-1}AS$$

имеет диагональный вид.

Докажем это утверждение. Пусть

$$\lambda_1, \lambda_2, \ldots, \lambda_n \tag{13}$$

— корни многочлена $\chi(z)$. Согласно предположению все они различны, поэтому в силу предложения C) соответствующие собственным значениям (13) собственные векторы

$$h_1, h_2, \ldots, h_n$$

линейно независимы и их можно принять за базис пространства A^n, — именно, положить

$$f_i = h_i, \quad i = 1, 2, \ldots, n.$$

Тогда мы имеем

$$\varphi(f_i) = \lambda_i f_i.$$

Отсюда видно, что матрица B, соответствующая выбранному базису, имеет диагональный вид, причем на диагонали стоят числа $\lambda_1, \lambda_2, \ldots, \lambda_n$, т. е. корни характеристического многочлена $\chi(z)$ отображения φ.

Предложение D) трактуется как приведение матрицы A к диагональному виду при помощи трансформации ее невырожденной матрицей S. Такой канонический вид для матрицы A мы получили в предположении, что все корни характеристического многочлена $\chi(z)$ матрицы A различны между собой. Но если характеристический многочлен $\chi(z)$ матрицы A имеет кратные корни, то матрицу, вообще говоря, нельзя привести к диагональному виду. При наличии кратных корней характеристического многочлена $\chi(z)$ матрицу A можно привести к более сложному, так называемому жорданову виду. Для доказательства этого результата требуется более сложная теория, которая будет изложена в следующих параграфах.

§ 13. Многочлены от матриц и отображений

Пусть R^n — некоторое векторное пространство размерности n, в котором задан некоторый фиксированный базис

$$e_1, e_2, \ldots, e_n.$$

Тогда каждому линейному отображению φ пространства R^n в себя соответствует некоторая определенная матрица A (см. A) § 2). Если φ и ψ — два отображения пространства R^n в себя, которым соответствуют матрицы A и B, то можно составить произведение $\varphi\psi$ отображений φ и ψ как последовательное проведение отображений $\varphi\psi x$, и этому произведению отображений $\varphi\psi$ соответствует произведение матриц A и B. Таким образом, можно определить любую степень φ^n отображения φ и этому отображению φ^n соответствует квадратная матрица A^n. Поскольку отображения пространства R^n в себя можно складывать и умножать на любое число (см. § 2), причем этим операциям соответствуют те же операции над соответствующими матрицами, то можно составить любой многочлен относительно отображения φ. Именно, если

$$f(z) = a_0 z^m + a_1 z^{m-1} + \ldots + a_m \tag{14}$$

— некоторый многочлен относительно переменной z, то вместо z в него можно подставить отображение φ, т. е.

получить многочлен

$$f(\varphi) = a_0\varphi^m + a_1\varphi^{m-1} + \ldots + a_m\theta,$$

где θ — тождественное отображение пространства R^n в себя. В формулу (14) вместо z можно поставить также матрицу A, которая соответствует отображению φ, и тогда мы получим матрицу

$$f(A) = a_0A^m + a_1A^{m-1} + \ldots + a_mE.$$

При этом отображению $f(\varphi)$ соответствует матрица $f(A)$. Результат применения отображения φ к вектору x мы будем обозначать через φx. Тогда можно записать

$$f(\varphi)\,x,$$

причем $f(\varphi)x$ означает применение отображения $f(\varphi)$ к вектору x.

Минимальный аннулирующий многочлен

Теперь нам нужно обратить особое внимание на нулевое отображение, которое переводит каждый вектор x пространства R^n в нуль и соответствующую ему нулевую матрицу. То и другое мы будем обозначать одним знаком 0.

А) Многочлен $f(z)$ называется *аннулирующим многочленом* отображения φ, если $f(\varphi)$ — нулевое отображение и соответственно $f(A)$ — нулевая матрица. Оказывается, что для каждого отображения φ существует аннулирующий многочлен. Именно, характеристический многочлен $\chi(z)$ отображения φ аннулирует отображение φ, т. е. мы имеем

$$\chi(\varphi) = 0, \tag{15}$$

или, что то же,

$$\chi(A) = 0. \tag{16}$$

Докажем эквивалентные между собой соотношения (15) и (16). Для этого выпишем соответствующее соотношение, определяющее матрицу $A = \| a_j^i \|$. Имеем

$$\varphi e_i = a_i^\alpha e_\alpha, \quad i = 1, 2, \ldots, n,$$

или, иначе,

$$(a_i^\alpha\theta - \delta_i^\alpha\varphi)\, e_\alpha = 0. \tag{17}$$

Рассмотрим матрицу

$$L = \| l_i^j \|,$$

элементы которой определяются формулой

$$l_j^i = a_j^i\theta - \delta_j^i\varphi.$$

Элементы матрицы L являются многочленами нулевой и первой степени относительно отображения φ и поэтому сами являются отображениями. Соотношение (17) при помощи матрицы L записывается в виде

$$l_i^\alpha e_\alpha = 0. \tag{18}$$

Минор элемента l_i^j матрицы L, взятый с надлежащим знаком, обозначим через M_i^j, так что мы имеем

$$M_j^\beta l_\beta^i = D(L)\,\delta_j^i \tag{19}$$

(см. § 4 гл. 1). Умножая соотношение (18) на M_j^i и суммируя полученный результат по i, получим

$$M_j^\beta l_\beta^\alpha e_\beta = 0. \tag{20}$$

В силу соотношения (19) левая часть последнего равенства переписывается в виде $D(L)\,\delta_j^\alpha e_\alpha$ и равенство (20) получает вид

$$D(L)\,\delta_j^\alpha e_\alpha = D(L)\,e_j = 0.$$

$D(\| a_j^i - \delta_j^i z \|)$ есть характеристический многочлен $\chi(z)$ отображения φ. Поэтому последнее соотношение переписывается в виде

$$\chi(\varphi)\,e_j = 0$$

и, таким образом, характеристический многочлен $\chi(\varphi)$ отображает в нуль каждый базисный элемент векторного пространства R^n, и, следовательно, любой вектор x пространства R^n. Итак, мы имеем

$$\chi(\varphi)\,x = 0,$$

и потому многочлен $\chi(z)$ является аннулирующим многочленом отображения φ.

Из предложения А) следует, что всякое отображение φ векторного пространства R^n в себя имеет аннулирующий многочлен, — именно, многочлен $\chi(z)$. Но, конечно, это не единственный аннулирующий многочлен для отображения φ, так как, умножив $\chi(z)$ на некоторый многочлен, мы получим снова аннулирующий многочлен. Среди всех аннулирующих многочленов отображения φ выделяется один, — именно, минимальный аннулирующий многочлен, — играющий важную роль для дальнейшего.

В) Оказывается, что все аннулирующие многочлены отображения φ минимальной степени — обозначим ее через r — отличаются лишь числовыми множителями. Тот из этих многочленов, у которого коэффициент при старшей степени r равен единице, мы будем обозначать через $\Delta(z)$ и назовем его *минимальным аннулирующим многочленом*, так что он оказывается единственным. Оказывается далее, что всякий аннулирующий многочлен отображения φ делится на минимальный аннулирующий многочлен $\Delta(z)$.

Докажем предложение В). Заметим прежде всего, что если $f(z)$ и $g(z)$ — два аннулирующих многочлена отображения φ, то их наибольший общий делитель $h(z)$ также есть аннулирующий многочлен. В силу предложения А) § 11 мы имеем

$$h(z) = a(z) f(z) + b(z) g(z), \tag{21}$$

где $a(z)$ и $b(z)$ — надлежащим образом подобранные многочлены. Подставляя в последнее соотношение вместо z отображение φ, мы видим, что $h(z)$ есть аннулирующий многочлен отображения φ. Допустим теперь, что многочлены $f(z)$ и $g(z)$ имеют минимальную степень r. Так как $h(z)$ — тоже аннулирующий многочлен, то он имеет степень, не меньшую чем r. А так как он является делителем каждого из многочленов $f(z)$ и $g(z)$, то степень его равна ровно r, и потому многочлены $f(z)$ и $g(z)$ отличаются от $h(z)$ только числовым множителем, а следовательно, и между собой они также отличаются числовым множителем. Этим первая часть предложения В) доказана.

Пусть

$$f(z) = \Delta(z)$$

— минимальный аннулирующий многочлен отображения φ, а $g(z)$ — произвольный аннулирующий многочлен отображения φ. Из формулы (21) следует, что многочлен $h(z)$ является аннулирующим многочленом отображения φ и потому имеет степень не меньше чем r. А так как $\Delta(z)$ имеет степень r и делится на $h(z)$, то $h(z)$ и $\Delta(z)$ отличаются лишь числовым множителем. Из того, что $g(z)$ делится на $h(z)$, следует теперь, что $g(z)$ делится и на $\Delta(z)$.

Итак, предложение В) доказано.

Разложение минимального аннулирующего многочлена на взаимно простые множители

C) Допустим, что векторное пространство R^n разлагается в сумму своих подпространств R_1 и R_2 размерности p и q соответственно, и допустим, что разложение это является инвариантным относительно отображения φ пространства R^n в себя, т. е.

$$\varphi x \in R_1 \text{ при } x \in R_1,$$
$$\varphi x \in R_2 \text{ при } x \in R_2.$$

Иначе это записывается так:

$$\varphi R_1 \subset R_1, \quad \varphi R_2 \subset R_2.$$

Выберем теперь базис пространства R^n:

$$e_1, e_2, \ldots, e_p, \quad e_{p+1}, \ldots, e_{p+q}, \qquad p+q=n, \tag{22}$$

так что первые p векторов этого базиса составляют базис пространства R_1, а оставшиеся q векторов составляют базис пространства R_2. Матрица

$$A = \| a_j^i \|,$$

соответствующая отображению φ в этом базисе, имеет тогда специальный вид,— именно, ее элементы удовлетворяют условиям:

$$\text{при } i<p, \ j>p \text{ имеем } a_j^i = 0,$$
$$\text{при } i>p, \ j<p \text{ имеем } a_j^i = 0.$$

$$A = \left\| \begin{matrix} a_1^1 & a_2^1 & \ldots & a_p^1 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_1^p & a_2^p & \ldots & a_p^p & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & a_{p+1}^{p+1} & \ldots & a_n^{p+1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 0 & a_{p+1}^n & \ldots & a_n^n \end{matrix} \right\| = \left\| \begin{matrix} A_1 & 0 \\ 0 & A_2 \end{matrix} \right\|.$$

Матрица A, как говорят, разбивается на два блока. В левом верхнем углу стоит квадратная матрица порядка p, которую мы обозначим через A_1, а в правом нижнем углу — квадратная матрица порядка q, которую мы обозначим через A_2. Как легко видеть, изучение матрицы A полностью сводится к изучению ее блоков A_1, A_2. При этом матрица A_1 соответствует отображению φ, действующему в пространстве R_1, а матрица A_2 соответствует отображению φ, действующему в пространстве R_2. В частности, имеет место нижеследующее предложение D).

Утверждение C) сводится к описанию структуры матрицы A в специальном базисе (22) и это нетрудно проверить. Я привожу утверждение C) без доказательства.

D) Если матрица A разбивается на блоки A_1 и A_2 (см. C)), то имеет место соотношение

$$D(A) = D(A_1)\,D(A_2). \tag{23}$$

Из этого, в частности, следует, что характеристический многочлен $\chi(z)$ матрицы A разлагается на множители:

$$\chi(z) = \chi_1(z)\,\chi_2(z), \tag{24}$$

где $\chi_1(z)$ — характеристический многочлен матрицы A_1, а $\chi_2(z)$ — характеристический многочлен матрицы A_2. Действительно $\chi(z) = D(A - zE)$, а матрица $A - zE$ разбивается на два блока. Отсюда и из формулы (23) следует формула (24).

Для доказательства предложения D) рассмотрим матрицы

$$A_1' = \begin{Vmatrix} A_1 & 0 \\ 0 & E_q \end{Vmatrix}, \quad A_2' = \begin{Vmatrix} E_p & 0 \\ 0 & A_2 \end{Vmatrix},$$

где матрицы A_1' и A_2' разбиваются на блоки, причем блок E_q есть единичная матрица порядка q, а E_p — единичная матрица порядка p. Разлагая определитель матрицы A_1' по элементам последней строки, мы докажем, что $D(A_1') = D\begin{Vmatrix} A_1 & 0 \\ 0 & E_{q-1} \end{Vmatrix}$. Продолжая этот процесс дальше, мы убедимся, что $D(A_1') = D(A_1)$. Точно так же доказывается, что $D(A_2') = D(A_2)$. Легко проверяется, что

$$A_1'A_2' = \begin{Vmatrix} A_1 & 0 \\ 0 & A_2 \end{Vmatrix} = A.$$

В силу теоремы 4 и последней формулы имеем $D(A) = D(A_1'A_2') = D(A_1)\,D(A_2)$ и доказательство предложения D) завершено. Таким образом, формула (23) доказана.

E) Пусть φ — линейное отображение векторного пространства R^n самого в себя и $\Delta(z)$ — минимальный аннулирующий многочлен отображения φ. Допустим, что многочлен $\Delta(z)$ разлагается на два взаимно простых множителя:

$$\Delta(z) = \Delta_1(z)\Delta_2(z).$$

И так как $\Delta_1(z)$ и $\Delta_2(z)$ по предположению взаимно простые, то мы имеем

$$1 = p_2(z)\Delta_2(z) + p_1(z)\Delta_1(z), \tag{25}$$

где $p_1(z)$ и $p_2(z)$ — два надлежащим образом выбранных многочлена (см. (37) § 11). Обозначим через R_1 векторное подпространство пространства R^n, состоящее из всех векторов пространства R^n, обращающихся в нуль при действии отображения $\Delta_1(\varphi)$, и через R_2 векторное подпространство пространства R^n, состоящее из всех векторов, обращающихся в нуль при действии отображения $\Delta_2(\varphi)$. Оказывается тогда, что пространство R^n разлагается в прямую сумму своих векторных подпространств R_1 и R_2. При этом разложение пространства R^n в сумму двух подпространств инвариантно относительно отображения φ, так что φ отображает подпространство R_1 в себя и подпространство R_2 в себя. Рассматривая отображение φ на подпространстве R_1, мы обозначим его через φ_1, а рассматривая отображение φ на R_2, мы обозначим его через φ_2. Оказывается тогда, что многочлен $\Delta_1(z)$ есть минимальный аннулирующий многочлен отображения φ_1, а $\Delta_2(z)$ есть минимальный аннулирующий многочлен отображения φ_2.

Докажем утверждение Е). Подставляя в соотношение (25) вместо z отображение φ, получаем для произвольного вектора x из пространства R^n равенство

$$x = p_2(\varphi)\Delta_2(\varphi)x + p_1(\varphi)\Delta_1(\varphi)x. \tag{26}$$

Положим

$$x_1 = p_2(\varphi)\Delta_2(\varphi)x, \quad x_2 = p_1(\varphi)\Delta_1(\varphi)x. \tag{27}$$

Тогда

$$\Delta_1(\varphi)x_1 = p_2(\varphi)\Delta_1(\varphi)\Delta_2(\varphi)x = p_2(\varphi)\Delta(\varphi)x = 0. \tag{28}$$

Точно так же имеем

$$\Delta_2(\varphi)x_2 = p_1(\varphi)\Delta_1(\varphi)\Delta_2(\varphi)x = p_1(\varphi)\Delta(\varphi)x_2 = 0. \tag{29}$$

Здесь имеется коммутативность между $\Delta_1(\varphi)$ и $p_2(\varphi)$, $\Delta_2(\varphi)$ и $p_1(\varphi)$, так как оба множителя являются многочленами относительно φ.

Таким образом, мы установили, что

$$x_1 \in R_1, \quad x_2 \in R_2. \tag{30}$$

Из формул (26), (27) и (30) следует, что произвольный вектор x пространства R^n разлагается в сумму

$$x = x_1 + x_2,$$

где

$$x_1 \in R_1, \quad x_2 \in R_2.$$

Покажем теперь, что если имеют место одновременно два соотношения

$$x \in R_1, \quad x \in R_2,$$

то $x = 0$.

Из соотношений (28), (29) следует, что $\Delta_1(\varphi)x = 0$, $\Delta_2(\varphi)x = 0$. Таким образом, оба слагаемых в правой части равенства (26) равны нулю и, следовательно, $x = 0$. Итак, установлено, что R^n разложено в прямую сумму своих подпространств R_1 и R_2. Покажем теперь, что это разложение в прямую сумму инвариантно относительно φ. Для этого достаточно показать, что

$$\varphi R_1 \subset R_1, \quad \varphi R_2 \subset R_2. \qquad (31)$$

Для того чтобы убедиться в этом, применим к векторному пространству φR_1 отображение $\Delta_1(\varphi)$. Тогда мы получим

$$\Delta_1(\varphi)\varphi R_1 = \varphi\Delta_1(\varphi) R_1 = 0.$$

Точно так же получаем

$$\Delta_2(\varphi)\varphi R_2 = \varphi\Delta_2(\varphi) R_2 = 0.$$

Здесь имеется коммутативность между φ и $\Delta_1(\varphi)$, φ и $\Delta_2(\varphi)$, так как оба множителя являются многочленами относительно φ. Таким образом, соотношение (31) доказано. Доказано уже также, что $\Delta_1(z)$ есть аннулирующий многочлен отображения φ_1, а $\Delta_2(z)$ — аннулирующий многочлен отображения φ_2. Это вытекает из самого определения подпространств R_1 и R_2. Докажем, что многочлены эти являются минимальными аннулирующими многочленами отображений φ_1 и φ_2.

Допустим, что минимальными аннулирующими многочленами отображений φ_1 и φ_2 являются соответственно многочлены $\hat{\Delta}_1(z)$ и $\hat{\Delta}_2(z)$. Положим $\hat{\Delta}(z) = \hat{\Delta}_1(z)\hat{\Delta}_2(z)$. Применяя $\hat{\Delta}\varphi$ к соотношению

$$x = x_1 + x_2,$$

где $x_1 \in R_1$, $x_2 \in R_2$, мы приходим к соотношению

$$\hat{\Delta}(\varphi)x = 0$$

при произвольном $x \in R^n$. Таким образом, $\widehat{\Delta}(\varphi)$ есть аннулирующий многочлен отображения φ. Если бы $\Delta_1(z)$ не был минимальным аннулирующим многочленом отображения φ_1, то $\widehat{\Delta}_1(z)$ имел бы степень, меньшую чем $\Delta_1(z)$. Точно так же, если бы $\Delta_2(z)$ не был минимальным аннулирующим многочленом отображения φ_2, то многочлен $\widehat{\Delta}_2(z)$ имел бы степень, меньшую чем $\Delta_2(z)$. В обоих этих случаях произведение $\widehat{\Delta}_1(z)\widehat{\Delta}_2(z)=\widehat{\Delta}(z)$ имело бы степень, меньшую чем $\Delta(z)$, и многочлен $\Delta(z)$ не был бы минимальным аннулирующим многочленом отображения φ.

Итак, предложение E) полностью доказано.

Из предложений E) и C) следует, что если минимальный аннулирующий многочлен $\Delta(z)$ некоторой матрицы A разлагается на два взаимно простых множителя: $\Delta(z)=\Delta_1(z)\Delta_2(z)$, то при надлежащем выборе базиса матрица A разбивается на два блока A_1 и A_2, причем $\Delta_1(z)$ есть минимальный аннулирующий многочлен матрицы A_1, а $\Delta_2(z)$ — минимальный аннулирующий многочлен матрицы A_2. Это вытекает из того, что разбиение матрицы на два блока происходит при специальном выборе базиса (см. C)). Будем говорить просто, что матрица A, минимальный аннулирующий многочлен которой разлагается на два взаимно простых множителя, разбивается на два блока, не упоминая при этом о специальном выборе базиса. Такое разбиение матрицы A на блоки можно вести до тех пор, пока мы не дойдем до блоков с минимальными аннулирующими многочленами, которые уже нельзя далее разложить на взаимно простые множители. В дальнейшем мы сосредоточили свое внимание на изучении такой матрицы A, минимальный аннулирующий многочлен которой уже нельзя разложить на два взаимно простых множителя. Многочлены, которые нельзя разложить на взаимно простые множители, являются многочленами вида $(z-\lambda)^k$. Именно на такие множители разлагается многочлен $\Delta(z)$, причем λ является корнем многочлена $\Delta(z)$. Нижеследующее предложение F) дает ответ на вопрос: какие корни имеет многочлен $\Delta(z)$?

F) Корнями минимального аннулирующего многочлена $\Delta(z)$ отображения φ являются собственные значения матрицы A, соответствующей отображению φ.

Докажем предложение F).

Пусть λ — собственное значение отображения φ с собственным вектором $h \neq 0$. Тогда мы имеем

$$\varphi h = \lambda h. \tag{32}$$

Применяя к этому соотношению операцию φ, получаем

$$\varphi^2 h = \lambda^2 h.$$

Продолжая этот процесс дальше, докажем, что

$$\varphi^k h = \lambda^k h.$$

Из этого легко следует, что для произвольного многочлена $f(z)$ имеет место соотношение

$$f(\varphi)\, h = f(\lambda)\, h.$$

Подставляя в это соотношение вместо многочлена $f(z)$ многочлен $\Delta(z)$, получаем

$$\Delta(\varphi)\, h = \Delta(\lambda)\, h.$$

В левой части этого равенства стоит нуль, а в правой части множитель h отличен от нуля, поэтому множитель $\Delta(\lambda)$ обращается в нуль, т. е. мы имеем $\Delta(\lambda) = 0$, и установлено, что собственное значение λ отображения φ является корнем многочлена $\Delta(z)$.

Допустим теперь, что λ есть корень многочлена $\Delta(z)$. Тогда

$$\Delta(z) = (z - \lambda)\, g(z). \tag{33}$$

Так как многочлен $g(z)$ имеет степень меньшую чем $\Delta(z)$, то он не является аннулирующим многочленом отображения φ, и потому существует такой вектор $\hat{h}$, что

$$g(\varphi)\, \hat{h} = h \neq 0. \tag{34}$$

Подставляя в соотношение (33) вместо z отображение φ, получаем (см. (34))

$$(\varphi - \lambda\theta)\, h = 0,$$

откуда следует

$$\varphi h = \lambda h,$$

т. е. λ является собственным значением отображения φ.

Итак, предложение F) полностью доказано.

G) Пусть φ — линейное отображение векторного пространства R^n в себя и A — соответствующая ему матрица. Пусть, далее,

$$\chi(z) = (z - \lambda_1)^{k_1} (z - \lambda_2)^{k_2} \ldots (z - \lambda_r)^{k_r} \tag{35}$$

— разложение характеристического многочлена $\chi(z)$ отображения φ на множители и

$$\Delta(z) = (z-\lambda_1)^{q_1}(z-\lambda_2)^{q_2} \ldots (z-\lambda_r)^{q_r}$$

— разложение минимального аннулирующего многочлена $\Delta(z)$ отображения φ на множители (см. F). При повторном применении предложения E) матрица A разбивается на блоки $A_1, A_2, \ldots, A_r$, причем минимальным аннулирующим многочленом блока A_i является $(z-\lambda_i)^{q_i}$. В силу предложения F) матрица A_i имеет единственное собственное значение λ_i. Если матрица A разбивается на блоки, то характеристический многочлен матрицы A является произведением характеристических многочленов соответствующих блоков (см. D)). Из этого разложения следует

$$\chi(z) = (z-\lambda_1)^{p_1}(z-\lambda_2)^{p_2} \ldots (z-\lambda_r)^{p_r}, \qquad (36)$$

где p_i — порядок матрицы A_i. Сравнивая соотношения (35) и (36), мы видим, что $p_i = k_i$. Таким образом, порядок блока A_i матрицы A равен кратности собственного значения λ_i отображения φ.

§ 14. Жорданова форма матрицы

Здесь мы будем рассматривать линейное отображение φ с единственным собственным значением λ. Согласно доказанному ранее минимальный аннулирующий многочлен отображения φ имеет вид

$$(z-\lambda)^k.$$

Положим

$$\psi = \varphi - \lambda\theta.$$

A) Пусть h_1 — вектор, удовлетворяющий условиям

$$\psi^l h_1 = 0, \quad \psi^{l-1} h_1 \neq 0. \qquad (37)$$

Положим

$$h_1 = h_1, \quad h_2 = \psi h_1, \quad \ldots, \quad h_l = \psi h_{l-1}, \quad h_{l+1} = \psi h_l.$$

Последовательность векторов $h_1, h_2, \ldots, h_l$ назовем *жордановой*. Из соотношений (37) следует, что $h_l \neq 0$, $h_{l+1} = 0$. Таким образом,

$$(\varphi - \lambda\theta) h_1 = h_2,$$

или, что то же,

$$\varphi h_1 = \lambda h_1 + h_2.$$

Далее,

$$(\varphi - \lambda\theta) h_2 = h_3$$

или, что то же,

$$\varphi h_2 = \lambda h_2 + h_3 \text{ и т. д.,}$$

$$(\varphi - \lambda\theta) h_l = 0$$

или, что то же,

$$\varphi h_l = \lambda h_l.$$

Если теперь $\hat{R}$ — векторное подпространство пространства R^n с жордановым базисом $h_1, h_2, \ldots, h_l$, то матрица A, соответствующая отображению φ, на пространстве $\hat{R}$ имеет вид

$$A = \begin{Vmatrix} \lambda & 1 & 0 & \ldots & 0 \\ 0 & \lambda & 1 & \ldots & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & \ldots & \lambda \end{Vmatrix}.$$

Матрица эта называется *жордановым блоком* с собственным значением λ.

Теперь мы докажем, что матрица A с единственным собственным значением λ при надлежащем выборе базиса разбивается на жордановы блоки.

В) Пусть R — векторное пространство размерности n и S — его векторное подпространство. Система векторов

$$x_1, x_2, \ldots, x_r \tag{38}$$

называется *линейно зависимой относительно подпространства* S, если существуют такие константы $c^1, c^2, \ldots, c^r$, не все равные нулю, что вектор $c^\alpha x_\alpha$ принадлежит S, в противном случае система (38) называется *линейно независимой относительно* S. Линейно независимая система (38) пространства R относительно S называется *базисом пространства* R *относительно* S, если для каждого вектора x пространства R найдутся такие константы $c^1, \ldots, c^r$, что вектор

$$x - c^\alpha x_\alpha$$

принадлежит пространству S. Ясно, что базис пространства R относительно S всегда существует и что всякую линейно независимую относительно S систему векторов пространства R можно дополнить до базиса в пространстве R относительно подпространства S.

Теорема 6. *Пусть φ — линейное отображение векторного пространства R^n в себя и λ — единственное соб-*

ственное значение отображения φ. *Тогда при надлежащем выборе базиса в пространстве* R^n *матрица* A, *соответствующая отображению* φ, *разбивается на жордановы блоки* (см. А)), *стоящие вдоль диагонали матрицы* A. *При этом каждый блок имеет собственное значение* λ.

Доказательство. Пусть $(z-\lambda)^k$ — минимальный аннулирующий многочлен отображения φ. Положим

$$\psi = \varphi - \lambda\theta.$$

Обозначим через S_l множество всех векторов x из R^n, удовлетворяющих условию

$$\psi^l x = 0.$$

Множество S_l представляет собой векторное подпространство пространства R^n. При этом имеют место следующие включения:

$$R^n = S_k \supset S_{k-1} \supset \ldots \supset S_1 \supset S_0 = 0.$$

Пусть B_1 — конечная совокупность векторов пространства S_k, составляющая его базис относительно подпространства S_{k-1}. Докажем, что векторы ψB_1, входящие в пространство S_{k-1}, линейно независимы относительно подпространства S_{k-2}. Пусть

$$e_1, e_2, \ldots, e_r$$

— совокупность всех векторов, входящих в B_1. Допустим, что имеет место соотношение

$$c^\alpha \psi e_\alpha \in S_{k-2}.$$

Тогда вектор

$$\psi c^\alpha e_\alpha \in S_{k-2}.$$

А это значит, что вектор

$$c^\alpha e_\alpha \in S_{k-1}.$$

Это возможно лишь при условии, что все коэффициенты $c^i = 0$.

Поскольку векторы конечной совокупности ψB_1 векторного пространства S_{k-1} линейно независимы относительно S_{k-2}, конечную совокупность ψB_1 можно дополнить до базиса B_2 пространства S_{k-1} относительно пространства S_{k-2}. Точно так же устанавливается, что векторы конечной совокупности ψB_2 векторного пространства S_2 линейно независимы относительно S_{k-3}, и потому конечную совокупность ψB_2 можно дополнить до базиса B_3 пространства S_2 относительно подпространства S_{k-3}. Рассуждая так дальше, мы дойдем в конце концов до базиса B_k в пространстве S_1 относительно пространства S_0, и потому B_k есть обычный базис в пространстве S_1. Докажем теперь, что совокупность векторов, принадлежащих ко всем базисам

$$B_1, B_2, \ldots, B_k,$$

составляет базис B пространства R^n.

Допустим, что имеет место линейная зависимость между векторами совокупности B:

$$c^{\alpha}e_{\alpha}+\ldots=0, \tag{39}$$

причем на первом месте выписаны слагаемые, принадлежащие B_1. Применяя к этому соотношению отображение ψ^{k-1}, мы получаем соотношение

$$c^{\alpha}\psi^{k-1}e_{\alpha}=0.$$

Но это значит, что конечная совокупность векторов $\psi^{k-1}B_1$, составляющая часть базиса B_{k-1} пространства B_1, линейно зависима, а потому все коэффициенты $c^1, c^2, \ldots, c^r$ равны нулю. После того как это доказано, мы докажем точно так же, что все коэффициенты при элементах B_2, входящих в соотношение (39), также равны нулю, и тем самым в конце концов установим, что совокупность векторов B линейно независима в пространстве R^n.

Докажем теперь, что произвольный вектор x_1 в пространстве $S_k=R^n$ линейно выражается через векторы, принадлежащие B. Так как векторы конечной совокупности B_1 составляют базис в пространстве S_k относительно S_{k-1}, то вектор x_1 может быть записан в виде

$$x_1=c^{\alpha}e_{\alpha}+x_2,$$

где $x_2\in S_{k-1}$. Таким образом, вектор x_1 выражается линейно через векторы, принадлежащие конечной совокупности B_1 и еще добавочный вектор x_2, принадлежащий пространству S_{k-1}. Продолжая этот процесс дальше, мы докажем в конце концов, что вектор x_1 линейно выражается через векторы, принадлежащие B. Таким образом, установлено, что B есть базис в пространстве $S^k=R^n$.

Разобьем теперь конечную совокупность векторов B на попарно непересекающиеся жордановы серии (см. А)).

Для этого выделим в B подмножество $\hat{B}$ всех векторов, которые будут служить начальными векторами h_1 жордановых серий (см. А)). К множеству $\hat{B}$ отнесем прежде всего все векторы совокупности B_1, затем все векторы совокупности B_2, не принадлежащие к ее части ψB_1, затем все векторы совокупности B_3, не принадлежащие к ее части ψB_2. Продолжая этот процесс дальше, мы дойдем до векторов совокупности B_k и включим в $\hat{B}$ те векторы из B_k, которые не принадлежат совокупности ψB_{k-1}.

Выберем начальный вектор h_1 нашей жордановой серии из конечной совокупности B_l, причем h_1 не принадлежит к совокупности ψB_{l-1}. Выбранный вектор h_1 принадлежит векторному пространству S_{k-l+1} и не принадлежит векторному пространству S_{k-l}. Итак,

$$\psi^{k-l} h_1 \neq 0,$$

$$\psi^{k-l+1} h_1 = 0.$$

Таким образом, мы имеем последовательность векторов

$$h_1, \ h_2 = \psi h_1, \ h_3 = \psi h_2, \ \ldots, \ h_{k-l+1} = \psi h_{k-l}, \ \psi h_{k-l+1} = 0,$$

и векторы

$$h_1, \ h_2, \ \ldots, \ h_{k-l+1} \tag{40}$$

составляют жорданову серию длины $k - l + 1$.

Жордановы серии типа (40) исчерпывают всю совокупность B и не пересекаются между собой.

Таким образом, мы нашли совокупность жордановых серий, составляющих вместе базис пространства R^n, и каждой жордановой серии соответствует жорданов блок матрицы A. Следовательно, теорема 6 доказана (см. гл. 4, пример 13).

С) Пусть исходное векторное пространство R^n является действительным и отображение φ тоже действительное, т. е. переводит каждый действительный вектор в действительный. Тогда характеристический многочлен $\chi(z)$ отображения φ имеет действительные коэффициенты и, следовательно, наряду с каждым комплексным собственным значением λ имеется сопряженное ему комплексное собственное значение $\bar{\lambda}$. Как было показано ранее (см. G) § 13), собственному значению λ в пространстве R^n соответствует некоторое векторное подпространство R' такое, что матрица A', соответствующая отображению φ пространства R' в себя, имеет минимальный аннулирующий многочлен $(z-\lambda)^{k'}$. В силу теоремы 6 базис в пространстве R', состоящий из жордановых серий векторов, является комплексным базисом. Как это видно из построения жордановых серий векторов в предложении A), отображение $\psi = \varphi - \lambda\theta$ является комплексным. В векторном подпространстве R'' векторного пространства R^n $(R'' = \bar{R}')$ можно взять базис, составленный из жордановых серий, имеющихся в пространстве R', но сопряженных с ними. Оказывается, что минимальный аннулирующий многочлен $(z - \bar{\lambda})^{k''}$ отображения φ на подпространстве R'' имеет тот же показатель k'', что и аннулирующий многочлен $(z - \lambda)^{k'}$ на пространстве R'. В самом деле ясно, что $(z - \bar{\lambda})^{k'}$ является аннулирующим многочленом отображения φ на подпространстве R''. Если бы этот аннулирующий многочлен $(z - \bar{\lambda})^{k'}$ не был минимальным, а минимальным был многочлен $(z - \bar{\lambda})^{k''}$, где $k'' < k'$, то, переходя от собственного значения $\bar{\lambda}$ к сопряженному ему собственному значению λ, мы пришли бы к заключению, что $(z - \lambda)^{k''}$ является аннулирующим многочленом отображения φ на подпространстве R'. Таким образом, оба показателя k' и k'' должны быть равны между собой. Таким образом, для двух комплексно сопряженных собственных значений λ и $\bar{\lambda}$ определены два векторных подпространства R' и R'', причем пространство R'' сопряжено пространству R' в том смысле, что каждый вектор из R'' сопряжен некоторому вектору из R' и наоборот.

Предложение C) имеет значение при отыскании действительных решений действительных обыкновенных дифференциальных уравнений.

§ 15. Квадратичные формы

Пусть E^n — действительное евклидово векторное пространство размерности n, $f(x, y)$ — действительная симметричная билинейная форма, заданная на этом пространстве, и $f(x, x)$ — соответствующая ей квадратичная форма. В координатном виде при выборе в E^n некоторого ортонормального базиса билинейная форма $f(x, y)$ записывается так:

$$f(x, y) = a_{\alpha\beta} x^\alpha y^\beta$$

(см. D) § 2). Здесь

$$A = \| a_{ij} \|$$

представляет собой квадратную матрицу порядка n. Эта запись отличается от обычной формы записи матриц, привычной нам, тем, что оба индекса стоят внизу, но мы будем считать, что первый индекс i указывает номер строки, а второй индекс j — номер столбца. Квадратичная форма $f(x, x)$ записывается в виде

$$f(x, x) = a_{\alpha\beta}x^{\alpha}x^{\beta}.$$

Основной задачей настоящего параграфа является выбор такого ортонормального базиса в пространстве E^n, при котором квадратичная форма записывается в наиболее простом виде.

А) Пусть

$$A = \| a_{ij} \|$$

— матрица, соответствующая действительной квадратичной форме $f(x, x)$, заданной на E^n при некотором ортонормальном базисе

$$e_1, e_2, \ldots, e_n. \tag{41}$$

Квадратичной форме $f(x, x)$ соответствует линейное отображение φ пространства E^n в себя с матрицей $\hat{A} = \| a_j^i \|$, которое в координатной форме при базисе (41) задается формулой

$$a_j^i = a_{ij}. \tag{42}$$

Матрицы A и $\hat{A}$, выписанные как квадратные таблицы, полностью совпадают. Они отличаются только тем, что элементы их обозначены по-разному. Оказывается, что так заданное отображение φ при помощи базиса (41) не зависит от случайно выбранного базиса, но определяется самой квадратичной формой $f(x, x)$. Собственные значения отображения φ называются *собственными значениями квадратичной формы* $f(x, x)$, а собственные векторы его называются *собственными векторами квадратичной формы*. Оказывается, что все собственные значения квадратичной формы являются действительными числами, а при надлежащим образом выбранном ортонормальном базисе

$$h_1', h_2', \ldots, h_n' \tag{43}$$

в пространстве E^n матрицы $\hat{A}$ и A имеют диагональную форму, причем на диагонали стоят собственные значения квадратичной формы $f(x, x)$, а векторы, составляющие базис (43), являются собственными векторами с этими собственными значениями.

Докажем предложение А). Докажем прежде всего, что отображение φ, заданное в координатной форме при помощи базиса (41) формулой (42), в действительности определяется самой квадратичной формой $f(x, x)$ и не зависит от выбора ортонормального базиса (41). Квадратичная форма $f(x, x)$ однозначно определяет симметричную билинейную форму $f(x, y)$ (см. Е) § 2). Далее, билинейная форма $f(x, y)$, рассматриваемая как функция вектора x, однозначно определяет такой вектор $u(y)$, что

$$f(x,\ y)=(u(y),\ x)$$

(см. В) § 7), где вектор $u(y)$, естественно, зависит от y. При базисе (41) мы имеем

$$f(x,\ y)=a_{\alpha\beta}x^{\alpha}x^{\beta}=a_{\alpha\beta}y^{\beta}x^{\alpha}=(u(y),\ x),$$

где

$$u(y)=(a_{1\beta}y^{\beta},\ a_{2\beta}y^{\beta},\ \ldots,\ a_{n\beta}y^{\beta}).$$

Таким образом, координаты вектора $u(y)$ в квадратичной форме определяются формулой

$$u^{i}=a_{i\beta}y^{\beta}.$$

Отсюда видно, что вектор $u(y)$ задается формулой

$$u(y)=\varphi y,$$

причем линейное отображение φ однозначно определяется симметричной билинейной формой $f(x, y)$, которая в свою очередь однозначно определяется квадратичной формой $f(x, x)$. Таким образом, мы доказали, что линейное отображение φ однозначно определяется квадратичной формой $f(x, x)$.

Теперь докажем, что любое собственное значение λ отображения φ действительно. Собственное значение λ отображения φ определяется соотношением

$$a_{i\alpha}h^{\alpha}=\lambda h^{i}=\lambda\delta_{i\alpha}h^{\alpha}, \tag{44}$$

где $h=(h^1, h^2, \ldots, h^n)$ — отличный от нуля вектор. Умножая это соотношение на $\bar{h}^i$ и суммируя по i, получаем

$$a_{\alpha\beta}h^{\alpha}\bar{h}^{\beta}=f(h,\ \bar{h})=\lambda(h,\ \bar{h}).$$

Согласно доказанному ранее (см. введение к гл. 3) билинейная форма $f(h, \bar{h})$, равно как и скалярное произведение $(h, \bar{h})$, являются действительными числами, причем $(h, \bar{h}) \neq 0$. Отсюда следует, что λ есть действительное число.

Доказательство того, что матрица A при надлежащим образом выбранном ортонормальном базисе (43) в пространстве E^n имеет диагональную форму, причем на диагонали стоят собственные значения квадратичной формы, соответствующие собственным векторам (43), будем вести индуктивно. При $n = 1$ утверждение очевидно, так как тогда матрица A является матрицей первого порядка и представляет собой действительное число.

При $n > 1$ собственный вектор h_1, соответствующий некоторому собственному значению $\lambda = \lambda_1$, выберем за первый элемент ортонормального базиса пространства E^n. Пусть

$$h_1, h_2, \ldots, h_n \tag{45}$$

— некоторый ортонормальный базис пространства E^n, начинающийся с собственного вектора h_1. Обозначим через E^{n-1} векторное подпространство пространства E^n с базисом

$$h_2, h_3, \ldots, h_n.$$

Обозначим через B матрицу, соответствующую квадратичной форме $f(x, x)$ в базисе (45). Согласно формуле (31) § 2 матрица $B = \|b_{ij}\|$ определяется формулой

$$b_{ij} = f(h_i, h_j).$$

Соотношение (44) для матрицы B записывается в виде

$$b_{i\alpha} h_1^\alpha = \lambda_1 h_1^i.$$

Умножая это соотношение скалярно на вектор h_j, $j > 1$, получаем

$$b_{\beta\alpha} h_1^\alpha h_j^\beta = \lambda_1 \sum_{\beta=1}^{n} h_1^\beta h_j^\beta = \lambda_1 \delta_{1j}.$$

Правая часть последнего соотношения ввиду ортонормальности базиса (45) обращается в нуль при $j \neq 1$, так что мы имеем

$$b_{1j} = f(h_1, h_j) = \lambda_1 \delta_{1j} = 0 \quad \text{при} \quad j > 1.$$

Таким образом, $f(x, y)$ в новом ортонормальном базисе (45) записывается в виде

$$f(x, y) = \lambda_1 x^1 y^1 + b_{\alpha\beta} x^\alpha y^\beta, \qquad \alpha, \beta = 2, \ldots, n.$$

Матрица

$$\widehat{B}=\|b_{ij}\|, \quad i, j=2, \ldots, n,$$

соответствует билинейной симметричной форме в подпространстве E^{n-1}, в которую превращается исходная билинейная форма $f(x, y)$, когда оба ее аргумента x и y принадлежат $f(x, y)$. Следовательно, по предположению индукции утверждение A) для нее верно. Таким образом, окончательно билинейная форма $f(x, y)$ в надлежащим образом выбранном ортонормальном базисе (43) приобретает вид

$$f(x, y)=\lambda_1 x^1 y^1+\lambda_2 x^2 y^2+\ldots+\lambda_n x^n y^n,$$

причем собственный вектор h_1' соответствует собственному значению λ_1, собственный вектор h_2' соответствует собственному значению λ_2 и т. д. Квадратичная форма $f(x, x)$ принимает вид

$$f(x, x)=\lambda_1 (x^1)^2+\lambda_2 (x^2)^2+\ldots+\lambda_n (x^n)^2. \tag{46}$$

Итак, предложение A) полностью доказано.

Говорят, что квадратичная форма $f(x, x)$ в надлежащим образом выбранном ортонормальном базисе может быть записана в виде суммы квадратов координат вектора, причем каждый квадрат берется с действительным коэффициентом. Коэффициент этот может быть как положительным, так и отрицательным числом, а также нулем.

Закон инерции

Мы доказали, что квадратичную форму $f(x, x)$ можно привести к виду (46), если взять в евклидовом векторном пространстве E^n ортонормальный базис. Если отказаться от ортонормальности базиса, то квадратичную форму $f(x, x)$ можно привести к еще более простому виду.

D) Квадратичную форму $f(x, x)$, заданную в действительном векторном пространстве E^n, надлежащим выбором базиса можно привести к следующему виду:

$$f(x, x)=\varepsilon_1 (x^1)^2+\varepsilon_2 (x^2)^2+\ldots+\varepsilon_n (x^n)^2, \tag{47}$$

где числа $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ принимают значения ± 1 и 0.

Докажем предложение D), воспользовавшись результатом предложения A). Если число λ_i положительно, то положим

$$\mu_i = \sqrt{\lambda_i}.$$

Если число λ_i отрицательно, то положим

$$\mu_i = \sqrt{-\lambda_i}$$

и, наконец, если λ_i равно нулю, то положим

$$\mu_i = 1.$$

Вместо координат $x^1, x^2, \ldots, x^n$, в которых записана квадратичная форма (46), введем новые координаты $\xi^1, \xi^2, \ldots, \xi^n$, положив

$$x^i = \frac{\xi^i}{\mu_i}.$$

Подставляя значения x^i в выражение квадратичной формы (46), получим

$$f(x, x) = \varepsilon_1 (\xi^1)^2 + \varepsilon_2 (\xi^2)^2 + \ldots + \varepsilon_n (\xi^n)^2.$$

Все числа $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ принимают значения ± 1 и 0. Таким образом, предложение D) доказано.

Предложение D) было доказано в предположении, что квадратичная форма $f(x, x)$ задана в евклидовом векторном пространстве E^n, так как при этом в качестве промежуточного результата использовалась каноническая форма (46) квадратичной формы. При доказательстве существенную роль играла евклидовость пространства E^n. В действительности для установления канонического вида (47) квадратичной формы $f(x, x)$ не требуется евклидовость пространства E^n. В пространстве R^n, которое не является евклидовым, можно произвольным образом ввести скалярное произведение и при помощи этого скалярного произведения доказать промежуточное предложение (46).

Сейчас мы займемся каноническим видом (47) квадратичной формы $f(x, x)$, причем квадратичная форма $f(x, x)$ задана в произвольном векторном пространстве R^n.

E) Пусть R^n — n-мерное векторное пространство, а

$$e_1, e_2, \ldots, e_n \qquad (48)$$

и

$$f_1, f_2, \ldots, f_n \qquad (49)$$

— два базиса в пространстве R^n. Пусть, далее,

$$x = x^\alpha e_\alpha, \quad y = y^\alpha f_\beta$$

— векторы, заданные в пространстве R^n. Допустим, что в базисе (48) квадратичная форма $f(x, x)$ записывается в виде

$$f(x,\ x) = \varepsilon_1 (x^1)^2 + \varepsilon_2 (x^2)^2 + \ldots + \varepsilon_p (x^p)^2 + \\ + \varepsilon_{p+1} (x^{p+1})^2 + \ldots + \varepsilon_{p+q} (x^{p+q})^2 + \ldots + \varepsilon_n (x^n)^2. \quad (50)$$

При этом базисные векторы $e_1, e_2, \ldots, e_n$ занумерованы так, что

$$\varepsilon_1 = \varepsilon_2 = \ldots = \varepsilon_p = 1,$$
$$\varepsilon_{p+1} = \varepsilon_{p+2} = \ldots = \varepsilon_{p+q} = -1,$$
$$\varepsilon_{p+q+1} = \varepsilon_{p+q+2} = \ldots = \varepsilon_n = 0.$$

Допустим, далее, что та же самая квадратичная форма $f(y, y)$ в базисе (49) записывается в виде

$$f(y,\ y) = \varepsilon_1' (y^1)^2 + \varepsilon_2' (y^2)^2 + \ldots + \varepsilon_{p'}' (y^{p'})^2 + \\ + \varepsilon_{p'+1}' (y^{p'+1})^2 + \ldots + \varepsilon_{p'+q'}' (y^{p'+q'})^2 + \ldots \\ \ldots + \varepsilon_n' (y^n)^2, \quad (51)$$

причем

$$\varepsilon_1' = \ldots = \varepsilon_{p'}' = 1,$$
$$\varepsilon_{p'+1}' = \ldots = \varepsilon_{p'+q'}' = -1,$$
$$\varepsilon_{p'+q'+1}' = \ldots = \varepsilon_n' = 0.$$

Оказывается, что имеет место следующая теорема инвариантности:

$$p = p', \quad q = q'. \quad (52)$$

Докажем соотношения (52). Доказательство будем вести от противного. Допустим, что

$$p \neq p',$$

и предположим для определенности, что

$$p' < p. \quad (53)$$

Обозначим через R^+ векторное подпространство пространства R^n с базисом

$$e_1,\ e_2,\ \ldots,\ e_p,$$

и через R^- векторное подпространство пространства R^n с базисом

$$f_{p'+1},\ \ldots,\ f_{p'+q'},\ \ldots,\ f_n.$$

Запись (50) квадратичной формы показывает, что если вектор $x \in R^+$ и $x \neq 0$, то $f(x, x) > 0$. Точно так же запись (51) квадратичной формы $f(y, y)$ показывает, что если вектор $y \in R^-$, то $f(y, y) \leqslant 0$. Ввиду неравенства (53) сумма размерностей пространств R^+ и R^- больше n, и потому пространства эти пересекаются по ненулевому вектору. Докажем это. Пусть

$$e_1, e_2, \ldots, e_p, f_{p'+1}, \ldots, f_n$$

— последовательность векторов. Ввиду неравенства (53) общее число этих векторов больше n, и потому они линейно зависимы. Запишем их линейную зависимость в виде:

$$a^\alpha e_\alpha = b^\beta f_\beta, \quad \alpha = 1, \ldots, p, \quad \beta = p' + 1, \ldots, n, \quad (54)$$

при этом коэффициенты a^i и b^j, вместе взятые, не могут все обращаться в нуль. Справа стоят элементы базиса (49), поэтому они линейно независимы, и, следовательно, коэффициенты a^i не могут все одновременно обращаться в нуль. Мы нашли отличный от нуля вектор x, стоящий в левой части соотношения (54), который принадлежит обоим пространствам R^+ и R^-. На этом векторе квадратичная форма $f(x, x)$ положительна, поскольку $x \in R^+$, и она неположительна в силу того, что $x \in R^-$. Таким образом, мы пришли к противоречию и неравенство $p' < p$ исключается. Точно так же доказывается, что неравенство $p < p'$ невозможно. Таким образом, доказано, что $p = p'$. Применяя полученный результат к квадратичной форме $-f(x, x)$, мы приходим к выводу, что $q = q'$. Итак, доказано, что в квадратичной форме $f(x, x)$ число положительных коэффициентов и число отрицательных коэффициентов в каноническом виде инварианты. Но так как число нулевых коэффициентов равно $n - p - q$, то этим самым доказана инвариантность числа нулевых коэффициентов. Мы доказали, следовательно, так называемый *закон инерции* для квадратичных форм.

§ 16. Экспонента квадратной матрицы

Этот параграф не вполне подходит под заголовок гл. 3, так как здесь не приводится к каноническому виду никакая матрица, однако результаты этого параграфа нужны для теории обыкновенных дифференциальных уравнений и их нужно куда-то поместить. По своему духу этот параграф ближе всего примыкает к гл. 3.

Экспонентой $\exp z$ называется функция

$$\exp z = e^z,$$

которая задается рядом

$$e^z = 1 + \frac{z}{1} + \frac{z^2}{2!} + \frac{z^3}{3!} + \ldots + \frac{z^k}{k!} + \ldots \tag{55}$$

Если вместо z в правую часть этого равенства подставить квадратную матрицу A и если окажется, что матричный ряд сходится, то мы определим функцию e^A. Оказывается, что, подставляя в ряд (55) произвольную квадратную матрицу порядка n, мы всегда получаем сходящийся ряд и функция e^A всегда определена. Сходимость ряда, составленного из квадратных матриц порядка n, определяется как сходимость ряда для каждого отдельного элемента матрицы, стоящего на определенном месте.

Для доказательства этого нужно произвести несложные оценки.

А) Пусть

$$B = \| b^i_j \|, \quad A = \| a^i_j \|$$

— две квадратные матрицы порядка n, элементы которых имеют оценки

$$| b^i_j | \leqslant b, \quad | a^i_j | \leqslant a,$$

где a и b — некоторые положительные числа. Оказывается тогда, что для матрицы

$$BA = C = \| c^i_j \|$$

имеют место оценки

$$| c^i_j | \leqslant nba. \tag{56}$$

Докажем последнее соотношение. Мы имеем

$$c^i_j = b^i_\alpha a^\alpha_j, \quad \alpha = 1, \ \ldots, \ n.$$

В правой части этого равенства стоят n слагаемых, каждое из которых по модулю не превосходит произведение ba, и, таким образом, неравенство (56) доказано.

В) Пусть

$$A = \| a^i_j \|$$

— квадратная матрица порядка n, для элементов которой имеются оценки

$$| a^i_j | \leqslant a,$$

где a — некоторое положительное число. Если k — целое неотрицательное число, то для матрицы

$$A^k = |c_j^i|$$

имеют место оценки

$$|c_j^i| \leqslant (na)^k.$$

Докажем это утверждение. Доказательство будем вести индуктивно по числу k. Так как $A^0 = E$, то для $k = 0$ утверждение верно. Положим теперь

$$A^{k-1} = B = |b_j^i|.$$

Тогда по предположению индукции имеем

$$|b_j^i| \leqslant (na)^{k-1}.$$

Далее,

$$A^k = BA.$$

Так как для матриц A и B уже имеются оценки, то в силу предложения А) мы получаем для элементов матрицы A^k оценку, указанную в формулировке предложения В).

С) Матричный ряд e^A, где A — квадратная матрица произвольного порядка n, всегда сходится.

Для доказательства этого утверждения рассмотрим член получаемого ряда

$$\frac{A^k}{k!} = |c_j^i|.$$

В силу предложения В)

$$|c_j^i| \leqslant \frac{(na)^k}{k!}.$$

При суммировании правых частей этого соотношения мы получим сходящийся ряд. Таким образом, матричный ряд e^A всегда сходится.

D) Пусть t — действительное число. Определим матричную функцию $f(t)$, положив

$$f(t) = e^{tA},$$

где A — некоторая квадратная матрица порядка n. Тогда оказывается, что

$$\frac{df(t)}{dt} = Af(t) = f(t)A. \qquad (57)$$

Для доказательства последнего соотношения выпишем функцию e^{tA} в виде ряда

$$f(t) = A^0 + \frac{tA}{1} + \frac{t^2A^2}{2!} + \ldots + \frac{t^kA^k}{k!} + \ldots$$

В силу оценок, полученных ранее (см. C)), ряд этот можно дифференцировать по t. Осуществляя дифференцирование, мы получаем соотношение (57).

Глава 4
ПРИМЕРЫ

Здесь будут даны примеры, иллюстрирующие и разъясняющие содержание трех первых теоретических глав.

К § 1

Пример 1.

Определение 7. *Вектором* называется направленный отрезок, т. е. такой отрезок, о котором известно, в каком его конце находится начало вектора и в каком его конце конец вектора.

Таково геометрическое определение вектора. Если отрезок рассматривается как принадлежащий плоскости E^2, то он двумерен. Если отрезок рассматривается как принадлежащий трехмерному пространству E^3, то он трехмерен. Вводятся операции сложения векторов и умножения их на действительные числа. Далее определяется равенство векторов и доказывается, что для каждого вектора из E^n $(n=2, 3)$ существует равный ему вектор, начинающийся в заданной точке пространства E^n.

Пусть u — вектор пространства E^n $(n=2, 3)$. Тогда существует равный ему вектор v, начинающийся в начале координат O. Обозначим через b конец вектора v. Координаты точки b являются числами. Их два или соответственно три в зависимости от n. Они являются координатами вектора u или вектора v. Таким образом, между векторами из E^n и парами или тройками чисел (в зависимости от n) устанавливается соответствие. Каждому геометрическому вектору ставится в соответствие вектор пространства A^n, заданный определением 1.

К § 2

Пусть A^n — n-мерное векторное пространство (см. определение 1) и φ — отображение пространства A^n в себя. Рассмотрим несколько типичных примеров отображения φ.

Пусть

$$e_1,\ e_2,\ \ldots,\ e_n$$

— базис пространства A^n. Зададим отображение φ соотношением

$$\varphi e_i = a_i^\alpha e_\alpha.$$

Таким образом, отображение φ задается матрицей $A = \| a_i^j \|$.

Пример 2. Допустим, что матрица A, задающая отображение φ, является диагональной и что на диагонали стоит одно и то же число λ. Тогда в силу формулы (23) гл. 1 для каждого вектора x отображение φ определяется формулой

$$\varphi x = \lambda x. \tag{1}$$

Такое отображение φ является растяжением пространства A^n с центром в начале координат с коэффициентом λ. Если $|\lambda| < 1$, то фактически отображение φ является сжатием. Заданное формулой (1) отображение φ называется *подобием* с центром в начале координат.

Пример 3. Допустим теперь, что матрица A, определяющая отображение φ, является диагональной и что на диагонали первый элемент равен нулю, а остальные равны единице. Тогда отображение φ представляет собой проектирование пространства A^n на его подпространство A^{n-1} с базисом $e_2, \ldots, e_n$ в направлении вектора e_1.

Пример 4. Пусть матрица A, задающая отображение φ, является диагональной, причем первый диагональный элемент равен -1, а остальные диагональные элементы равны $+1$. Тогда отображение φ является зеркальным отображением пространства A^n на себя относительно его подпространства A^{n-1} с базисом $e_2, \ldots, e_n$ в направлении e_1.

К §§ 3, 4

Пример 5. Рассмотрим систему двух линейных уравнений относительно двух неизвестных x^1, x^2:

$$\begin{aligned} a_1^1 x^1 + a_2^1 x^2 &= c^1, \\ a_1^2 x^1 + a_2^2 x^2 &= c^2. \end{aligned} \tag{2}$$

Исключим неизвестное x^2 из системы уравнений (2), умножив первое из уравнений системы (2) на a_2^2, а второе на a_2^1 и вычтя одно уравнение из другого. Тогда мы получим

$$(a_1^1 a_2^2 - a_2^1 a_1^2)\, x^1 = a_2^2 c^1 - a_2^1 c^2.$$

Таким образом, мы имеем

$$x^1 = \frac{a_2^2 c^1 - a_2^1 c^2}{a_1^1 a_2^2 - a_2^1 a_1^2}. \tag{3}$$

Аналогично получаем

$$x^2 = \frac{a_1^1 c^1 - a_1^2 c^1}{a_1^1 a_2^2 - a_2^1 a_1^2}. \tag{4}$$

Обе дроби (3) и (4) имеют один и тот же знаменатель

$$D(A) = a_1^1 a_2^2 - a_2^1 a_1^2.$$

Этот знаменатель называется определителем матрицы $A = \| a_j^i \|$, $i = 1, 2$; $j = 1, 2$, второго порядка. Таким образом, мы пришли к представлению об определителе квадратной матрицы второго порядка.

Положим

$$A_1 = \begin{Vmatrix} c_1 & a_2^1 \\ c_2 & a_2^2 \end{Vmatrix}, \quad A_2 = \begin{Vmatrix} a_1^1 & c^1 \\ a_1^2 & c_2 \end{Vmatrix}.$$

Тогда для x^1 и x^2 формулы (3) и (4) можно переписать в виде

$$x^1 = \frac{D(A_1)}{D(A)}, \quad x^2 = \frac{D(A_2)}{D(A)}.$$

В следующих параграфах эта конструкция будет обобщаться на случай решения системы n уравнений с n неизвестными при помощи определителей.

К § 5

Пример 6. Пусть $A = \| a_j^i \|$, $B = \| b_j^i \|$ — две треугольные матрицы порядка n. Это значит, что при $j > i$ имеем $a_j^i = 0$, $b_j^i = 0$. Докажем, что матрица $C = AB = \| c_j^i \|$ также является треугольной. Мы имеем

$$c_k^i = a_\alpha^i b_k^\alpha.$$

Докажем, что при $k > i$ имеем $c_k^i = 0$. В самом деле, если $\alpha > i$, то $a_\alpha^i = 0$. Если $\alpha \leqslant i$, то $\alpha < k$ и потому $b_k^\alpha = 0$. Следовательно, $c_k^i = 0$ при $k > i$.

К § 6

Пример 7. Пусть

$$x = (x^1, \ldots, x^p), \quad y = (y^1, \ldots, y^q)$$

— два отличных от нуля вектора размерностей p и q соответственно. Определим матрицу $A = \| a_j^i \|$ формулой

$$a_j^i = x^i y^j.$$

Тогда ранг матрицы A равен единице.

К § 7

Пример 8. Пусть E^n — n-мерное евклидово пространство. Отображение φ пространства E^n на себя называется *вращением*, если оно сохраняет длины всех векторов, т. е. если $|\varphi x| = |x|$.

Дадим пример вращения пространства E^2. Пусть $x = (x^1, x^2)$. Зададим отображение φ формулами

$$\begin{aligned} (\varphi x)^1 &= x^1 \cos \alpha + x^2 \sin \alpha, \\ (\varphi x)^2 &= -x^1 \sin \alpha + x^2 \cos \alpha. \end{aligned} \tag{5}$$

Легко проверяется, что заданное формулами (5) отображение φ является вращением. В действительности φ есть поворот плоскости E^2 вокруг начала координат на угол α.

К § 8

Появление комплексных чисел побудило математиков сделать попытку дальнейшего обобщения действительных чисел, присоединив к ним не одну мнимую единицу, а несколько. Эта попытка имела ограниченный успех. Были построены кватернионы с тремя мнимыми единицами, но при этом пришлось отказаться от коммутативности умножения. Дадим определение кватернионов.

Пример 9. В основу кватернионов положены три мнимые единицы, которые обозначаются через

$$i,\ j,\ k, \tag{6}$$

так что каждый кватернион записывается в виде

$$x = x_0 + x_1 i + x_2 j + x_3 k.$$

Здесь x_0, x_1, x_2, x_3 — действительные числа, коммутативные по умножению с единицами (6). А единицы (6) перемножаются по следующим правилам:

$$i^2 = j^2 = k^2 = -1, \quad ij = k, \quad jk = i, \quad ki = j.$$

Кватернион $\bar{x}$, сопряженный к кватерниону x, задается формулой

$$\bar{x} = x_0 - x_1 i - x_2 j - x_3 k.$$

Легко проверяется, что

$$x\bar{x} = |x|^2 = x_0^2 + x_1^2 + x_2^2 + x_3^2.$$

Из этого следует, что кватернион x^{-1}, обратный к кватерниону x по умножению, задается формулой

$$x^{-1} = \frac{\bar{x}}{|x|^2}.$$

Легко проверяется, что

$$\overline{xy} = \bar{y}\bar{x}.$$

Кватернион x, по модулю равный единице, т. е. удовлетворяющий условию $|x| = 1$, записывается в виде

$$x = \cos\alpha + u \sin\alpha,$$

где u — чисто мнимый кватернион, т. е. $u = u_1 i + u_2 j + u_3 k$, а α — угол.

Ввиду отсутствия коммутативности умножения попытки построить теорию кватернионных функций не имели успеха, но кватернионы имеют красивые геометрические применения. Укажем их.

Пример 10. Совокупность всех кватернионов K представляет собой четырехмерное евклидово векторное пространство. Длина каждого вектора-кватерниона x определяется как $|x|$. Четырехмерное векторное пространство K всех кватернионов разлагается в прямую сумму одномерного векторного подпространства действительных кватернионов D, имеющих вид $x = x_0$, и трехмерного векторного пространства чисто мнимых кватернионов I, имеющих вид

$$u = u_1 i + u_2 j + u_3 k.$$

Каждому кватерниону a, по модулю равному единице, поставим в соответствие линейное отображение f_a евклидова векторного пространства K на себя, задаваемое формулой

$$f_a x = axa^{-1}.$$

Легко проверяется, что линейное отображение f_a переводит каждое линейное подпространство D и I само на себя. Далее, оказывается, что отображение f_a является вращением евклидова векторного пространства K (см. пример 8). Докажем это. Для этого подсчитаем модуль кватерниона axa^{-1}. Имеем

$$|axa^{-1}|^2 = axa^{-1}\overline{axa^{-1}} =$$

$$= axa^{-1}\overline{a^{-1}}\bar{x}\bar{a} = axa^{-1}a\bar{x}a^{-1} = ax\bar{x}a^{-1} = |x|^2.$$

Пример 11. Оказывается, что при помощи отображения f_a можно получить любое вращение евклидова векторного пространства K, получающееся в результате непрерывного изменения из тождественного, причем такого, что в процессе изменения все время мы имеем вращение и пространство D переводится в себя. Таким образом, отображение f_a описывает все вращения пространства I, получающиеся непрерывным образом из тождественного. Доказательство этого нетривиального утверждения я предоставляю читателю.

К § 9

В 1545 году Кардано дал формулу решения кубического уравнения в радикалах. Дадим вывод его формулы.

Пример 12. Рассмотрим кубическое уравнение

$$z^3 + az + b = 0. \qquad (7)$$

К такому виду легко привести любое кубическое уравнение. Представим теперь z в виде

$$z = x + y.$$

Тогда уравнение (7) запишется в виде

$$x^3 + y^3 + 3xy(x + y) + a(x + y) + b = 0. \qquad (8)$$

Наложим на x и y дополнительное условие

$$xy = -\frac{a}{3}. \tag{9}$$

Тогда уравнение (8) запишется в виде

$$x^3 + y^3 = -b. \tag{10}$$

Так как имеет место соотношение (9), то

$$x^3 y^3 = -\frac{a^3}{27}. \tag{11}$$

Из соотношений (10) и (11) следует, что x^3 и y^3 являются корнями квадратного уравнения

$$u^2 + bu - \frac{a^3}{27} = 0.$$

Таким образом,

$$x^3 = -\frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}, \quad y^3 = -\frac{b}{2} - \sqrt{\frac{b^2}{4} + \frac{a^3}{27}},$$

и для z получаем формулу Кардано

$$z = \sqrt[3]{-\frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}} + \sqrt[3]{-\frac{b}{2} - \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}}.$$

Ясно, что, применяя формулу Кардано, нельзя брать произвольные значения кубических корней, а надо брать такие пары кубических корней, чтобы их произведение равнялось $-a/3$. Оказывается, что величина $\frac{b^2}{4} + \frac{a^3}{27}$ является отрицательной тогда и только тогда, когда уравнение (7) имеет три действительных корня. Это утверждение приводится здесь без доказательства. Указание на доказательство имеются в § 4 книги «Математический анализ для школьников». Таким образом, в формуле Кардано встречается извлечение корня из отрицательного числа именно в том случае, когда корни кубического уравнения действительны. В этом случае два кубических корня, стоящих в формуле Кардано, извлекаются из комплексно сопряженных величин и при извлечении кубических корней следует также взять комплексно сопряженные величины. Тогда сумма кубических корней, входящих в формулу Кардано, оказывается действительной. То обстоятельство, что именно в случае действительных корней кубического уравнения (7) в формуле Кардано появляются мнимые числа, очень сильно содействовало введению комплексных чисел.

Проиллюстрируем доказательство теоремы 6.

Пример 13. Выпишем базис пространства $R^n = S_k$, разбив этот базис на строки. В качестве первой строки выпишем базис пространства S_k относительно его подпространства S_{k-1}, т. е. векторы системы B_1. Пусть он состоит из векторов $e_1, e_2, \ldots, e_r$. Во вторую строку выпишем векторы системы B_2 — сначала векторы $\psi e_1, \ldots, \psi e_r$, а затем векторы $e_{r+1}, \ldots, e_{r+s_1}$, причем под вектором e_i первой строки выпишем вектор ψe_i, а далее векторы $e_{r+1}, \ldots, e_{r+s_1}$, не входящие в ψB_1. В третью строку выпишем векторы системы B_3 — сначала векторы $\psi^2 e_1, \ldots, \psi^2 e_r$, далее $\psi e_{r+1}, \ldots, \psi e_{r+s_1}$, а затем векторы $e_{r+s_1+1}, \ldots, e_{r+s_1+s_2}$, не входящие в ψB_2. Продолжая этот процесс дальше, мы получим последовательность строк, причем первая состоит из векторов системы B_1, вторая из векторов системы B_2, третья из векторов системы B_3 и т. д. Каждая из таких строк справа может выступать из-под предыдущей строки на несколько элементов. Таким образом, эти строки образуют лесенку со ступеньками, расположенными справа. В качестве верхней площадки лесенки взята система B_1. Взяв вектор h_1 из строки B_i, находящийся на ступеньке этой строки и выступающий из-под предыдущей строки, мы получим жорданову серию $h_1, \psi h_1, \ldots, \psi^{k-i+1} h_1$:

$$
\begin{array}{lllllllll}
B_1 & e_1 & e_2 \ldots & e_r & & & & & \\
B_2 & \psi e_1 & \psi e_2 \ldots & \psi e_r & e_{r+1} \ldots & e_{r+s_1} & & & \\
B_3 & \psi^2 e_1 & \psi^2 e_2 \ldots & \psi^2 e_r & \psi e_{r+1} \ldots & \psi e_{r+s_1} & e_{r+s_1+1} \ldots & e_{r+s_1+s_2} & \\
\end{array}
$$

. . . .

Академик Л. С. Понтрягин написал четыре небольших популярных книги для издания под общим названием «ЗНАКОМСТВО С ВЫСШЕЙ МАТЕМАТИКОЙ»:

- **МЕТОД КООРДИНАТ**
- **АНАЛИЗ БЕСКОНЕЧНО МАЛЫХ**
- **АЛГЕБРА**
- **ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ И ИХ ПРИЛОЖЕНИЯ**

«Предполагается, что в них будут даны важнейшие результаты классической высшей математики. Выбор материала и порядок его изложения не соответствует никакой учебной программе. Книги будут отражать мои личные вкусы и взгляды на математику, сложившиеся за много лет работы. Кроме того, они будут учитывать мои юношеские воспоминания о возможностях восприятия молодого человека, с тем, чтобы нынешнее поколение молодых людей, начиная со школьников старших классов, могло знакомиться по ним с высшей математикой и приобретать правильный здоровый вкус к ней. Внимание читателя должно быть направлено не на изощренности типа теории множеств, теории пределов и т. п., а на главные математические результаты, сложившиеся в течение тысячелетий».

ИЗДАТЕЛЬСТВО УРСС
НАУЧНОЙ И УЧЕБНОЙ ЛИТЕРАТУРЫ
E-mail: URSS@URSS.ru
Каталог изданий
в *Internet:* http://URSS.ru
Тел./факс: 7 (095) 135–44–23
Тел./факс: 7 (095) 135–42–46

2275 ID 18642

9 785354 006144