С.Н. Ильин

ЭЛЕМЕНТЫ АЛГЕБРЫ: МАТРИЦЫ, КОМПЛЕКСНЫЕ ЧИСЛА, СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ, МНОГОЧЛЕНЫ

Казань — 2018

КАЗАНСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
ИНСТИТУТ МАТЕМАТИКИ И МЕХАНИКИ
им. Н.И. ЛОБАЧЕВСКОГО

С.Н. ИЛЬИН

ЭЛЕМЕНТЫ АЛГЕБРЫ: МАТРИЦЫ, КОМПЛЕКСНЫЕ ЧИСЛА, СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ, МНОГОЧЛЕНЫ

УЧЕБНОЕ ПОСОБИЕ

Казань – 2018

УДК 512

Печатается по решению
Учебно-методической комиссии
Института математики и механики им. Н.И. Лобачевского КФУ

Рецензент:
кандидат физико-математических наук, доцент Абызов А.Н.

Ильин С.Н.

Элементы алгебры: матрицы, комплексные числа, системы линейных уравнений, многочлены. Учебное пособие. — Казань: Казанский (Приволжский) федеральный университет, 2018. — 86 с.

Учебное пособие предназначено для студентов I курса Института математики и механики им. Н.И. Лобачевского КФУ.

1 Начальные сведения о матрицах

При решении многих задач высшей математики приходится оперировать с таблицами, составленными из объектов некоторого фиксированного типа. Такие таблицы называют матрицами. В данном разделе приведены начальные сведения о матрицах и действиях с ними.

Пусть X — непустое множество. *Матрицей* над X называется составленная из элементов множества X прямоугольная таблица, содержащая m строк и n столбцов:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Числа m и n называются размерами матрицы A. Элементы матрицы нумеруются парами индексов: a_{ij} — элемент матрицы A, находящийся в i-й строке и j-м столбце. Если $m = n$, то матрица A называется *квадратной*, или еще говорят, что A — матрица *порядка* n. Набор $(a_{11}, a_{22}, \dots, a_{nn})$ элементов такой матрицы называется ее *главной диагональю*, набор $(a_{1n}, a_{2,n-1}, \dots, a_{n1})$ — *побочной диагональю*. Квадратная матрица, в которой все элементы, находящиеся вне главной диагонали, равны 0, называется *диагональной*. В случае, когда равны 0 все элементы, расположенные ниже (выше) главной диагонали, матрица называется *верхнетреугольной* (*нижнетреугольной*).

1.1 Действия с матрицами

В дальнейшем нас, как правило, будут интересовать только матрицы, составленные из чисел; для определенности все рассматриваемые ниже матрицы (если не сказано иное) будем считать вещественными, то есть состоящими из вещественных чисел. Заметим, что всякое число можно рассматривать как матрицу порядка 1, тем самым, матрицы являются естественным обобщением чисел. Как известно, с числами можно производить различные арифметические действия — сложение, вычитание, умножение и т. п. Эти действия переносятся и на матрицы.

1. *Сложение матриц. Суммой* $m \times n$*-матриц* $A = (a_{ij})$ и $B = (b_{ij})$ называется матрица $C = (c_{ij})$ тех же размеров, где $c_{ij} = a_{ij} + b_{ij}$ для

всех i, j, другими словами, сложение матриц выполняется поэлементно. Очевидны следующие свойства сложения матриц:

1.1. $A + B = B + A$ (коммутативность).

1.2. $(A + B) + C = A + (B + C)$ (ассоциативность).

Матрица, все элементы которой равны нулю, называется *нулевой* и обозначается символом 0; матрица, составленная из элементов $-a_{ij}$ для всех i, j, называется *противоположной* к A и обозначается $-A$. Вполне очевидны свойства:

1.3. $A + 0 = 0 + A = A$ для любой матрицы A.

1.4. $A + (-A) = -A + A = 0$.

2. *Умножение матрицы на число.* Пусть $A = (a_{ij})$ — $m \times n$-матрица, λ — число. Под λA понимается матрица $B = (b_{ij})$ тех же размеров, где $b_{ij} = \lambda a_{ij}$ при всех i, j. Таким образом, чтобы умножить матрицу A на λ, нужно умножить на λ каждый ее элемент. Легко проверяются свойства:

2.1. $(\lambda\mu)A = \lambda(\mu A)$.

2.2. $\lambda(A + B) = \lambda A + \lambda B$.

2.3. $(\lambda + \mu)A = \lambda A + \mu A$.

3. *Умножение матриц.* Пусть $A = (a_{ij})$ — $m \times n$-матрица, $B = (b_{ij})$ — $n \times k$-матрица. *Произведением* матриц A и B называется $m \times k$-матрица $C = (c_{ij})$, элементы которой при всех i, j вычисляются по правилу:

$$c_{ij} = \sum_{l=1}^{n} a_{il} b_{lj}.$$

Умножение матриц обладает следующими свойствами:

3.1. $(AB)C = A(BC)$.

3.2. $(A + B)C = AC + BC$.

3.3. $A(B + C) = AB + AC$.

3.4. Не всегда $AB = BA$.

Таким образом, умножение матриц ассоциативно, дистрибутивно справа и слева относительно сложения, но не всегда коммутативно.

Доказательство. Проверим свойство 3.1. Пусть A имеет размеры $m \times n$, B — $n \times k$, C — $k \times l$. Тогда и $(AB)C$, и $A(BC)$ имеют размеры $m \times l$. Проверим равенство соответствующих элементов:

$$\big((AB)C\big)_{ij} = \sum_s (AB)_{is} c_{sj} = \sum_s \Big(\sum_t a_{it} b_{ts}\Big) c_{sj} = \sum_{s,\,t} a_{it} b_{ts} c_{sj}.$$

Аналогично,

$$\big(A(BC)\big)_{ij} = \sum_{s,\,t} a_{it} b_{ts} c_{sj},$$

следовательно, $(AB)C = A(BC)$. Схожим образом устанавливается справедливость свойств 3.2 и 3.3. Наконец, равенства $AB = BA$ в 3.4 может не быть, например, из-за несоответствия размеров, а именно, если A — $m \times n$-матрица, B — $n \times k$-матрица и $m \neq k$, то произведение AB определено, а BA — нет. Если же $m = k$, но $m \neq n$, то определены оба произведения AB и BA, но они имеют разные размеры — $m \times m$ и $n \times n$. И даже если $m = n = k$ (то есть A, B, AB, BA — квадратные матрицы одного порядка), то AB необязательно совпадает с BA, в чем легко убедиться, перемножив, например, матрицы $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ и $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. В самом деле, непосредственные вычисления показывают, что $AB = B$, а $BA = 0$.□

Обозначим через $E_n = (\delta_{ij})$ диагональную матрицу порядка n, все диагональные элементы которой равны 1. Такая матрица называется *единичной*. (В случае, когда порядок фиксирован, нижний индекс у единичной матрицы обычно опускают и вместо E_n пишут просто E.) Числа $\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$ называют *символами Кронекера*. Вполне очевидно свойство

3.5. Если A — $m \times n$-матрица, то $E_m A = A E_n = A$.

В дальнейшем нам понадобятся *матричные единицы* — матрицы E_{ij}, устроенные следующим образом: (i, j)-элемент матрицы E_{ij} равен 1, а все остальные элементы — нулевые. Легко проверяется правило умножения матричных единиц:

$$E_{ij} E_{kl} = \begin{cases} E_{il}, & \text{если } j = k, \\ 0, & \text{в противном случае.} \end{cases}$$

4. *Транспонирование. Транспонированной* к $m \times n$-матрице $A = (a_{ij})$ называется $n \times m$-матрица $A^t = (a'_{ij})$, где $a'_{ij} = a_{ji}$ для всех i, j. Таким образом, строки матрицы A являются столбцами матрицы A^t и наоборот. Свойства операции транспонирования:

4.1. $(A^t)^t = A$.

4.2. $(A + B)^t = A^t + B^t$.

4.3. $(\lambda A)^t = \lambda A^t$.

4.4. $(AB)^t = B^t A^t$.

Доказательство. Свойства 4.1–4.3 вполне очевидны. Докажем 4.4. Для всех i, j имеем:

$$\big((AB)^t\big)_{ij} = (AB)_{ji} = \sum_k a_{jk} b_{ki} = \sum_k (B^t)_{ik} (A^t)_{kj} = \big(B^t A^t\big)_{ij},$$

что и требовалось. □

1.2 Элементарные преобразования и элементарные матрицы

Каждую матрицу можно рассматривать как упорядоченный набор строк и/или столбцов. Среди всевозможных способов преобразований этих наборов особо выделяют так называемые *элементарные преобразования*. Ниже будут рассматриваться преимущественно элементарные преобразования строк, проведение аналогичных рассуждений о преобразованиях столбцов оставляется в качестве упражнения.

1. Пусть $s \neq t$. Обозначим через $\mathcal{F}_{st}$ преобразование, меняющее местами s-ю и t-ю строки матрицы. Такое преобразование называют элементарным преобразованием I-го рода.

2. Пусть $s \neq t$, $\lambda \in K$. Прибавим к t-й строке матрицы s-ю строку, предварительно умноженную на λ. Такое преобразование обозначается через $\mathcal{F}_{st}(\lambda)$ и называется элементарным преобразованием II-го рода.

3. Умножим все элементы s-й строки на $\lambda \neq 0$. Назовем такое преобразование $\mathcal{F}_s(\lambda)$ элементарным преобразованием III-го рода.

Применим перечисленные выше преобразования к матрице A, действуя по следующему алгоритму.

Шаг 1: Двигаясь сверху вниз, ищем в первом столбце отличный от нуля элемент. Если его нет, то повторяем алгоритм для матрицы, полученной из исходной вычеркиванием первого столбца. Если просмотренный нулевой столбец оказался последним, то алгоритм завершен. Теперь предположим, что найден ненулевой элемент a_{i1}. Поменяем местами 1-ю и i-ю строки (то есть, применим $\mathcal{F}_{1i}$). Получим матрицу $\tilde{A} = (\tilde{a}_{ij})$, где $\tilde{a}_{11} = a_{i1} \neq 0$.

Шаг 2: Для каждого $i > 1$ последовательно применим преобразование $\mathcal{F}_{1i}\,(-\tilde{a}_{i1}/\tilde{a}_{11})$. В результате получим матрицу, в которой все элементы первого столбца (кроме, разумеется, первого элемента) равны 0. Теперь

мысленно вычеркиваем из матрицы первую строку и первый столбец и, если еще остались строки и столбцы, повторяем алгоритм с шага 1 для оставшейся части матрицы; если же строк и столбцов не осталось, то работа алгоритма закончена.

Нетрудно понять, что после таких преобразований матрица примет вид

$$\begin{pmatrix} 0 & \dots & \underline{|\,\bar{a}_{1j_1}} & \dots & \bar{a}_{1j_2} & \dots & \bar{a}_{1j_r} & \dots \\ 0 & \dots & 0 & \dots & \underline{|\,\bar{a}_{2j_2}} & \dots & \bar{a}_{2j_r} & \dots \\ & & \dots & & & \dots & & \\ 0 & \dots & 0 & \dots & 0 & \dots & \underline{|\,\bar{a}_{rj_r}} & \dots \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots \\ & & \dots & & & \dots & & \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots \end{pmatrix}, \qquad (1)$$

называемый *ступенчатым*, а сам процесс преобразования матрицы называют *приведением к ступенчатому виду*. Отметим, что в матрице ступенчатого вида нулевых столбцов слева и нулевых строк внизу может не быть. Подчеркнем также, что элементы $\bar{a}_{1j_1}, \bar{a}_{2j_2}, \dots, \bar{a}_{rj_r}$, находящиеся в вершинах "ступенек", отличны от нуля. Итак, доказана

Лемма 1.1 *Любая матрица элементарными преобразованиями строк I-го и II-го рода может быть приведена к ступенчатому виду.*

Каждому элементарному преобразованию сопоставим *элементарную матрицу*:

1. $F_{st} = E - E_{ss} - E_{tt} + E_{st} + E_{ts}$ — I-го рода,
2. $F_{st}(\lambda) = E + \lambda E_{ts}$ — II-го рода,
3. $F_s(\lambda) = E + (\lambda - 1)E_{ss}$ — III-го рода.

Непосредственно проверяется

Лемма 1.2 *Выполнение элементарного преобразования строк матрицы равносильно ее домножению слева на соответствующую элементарную матрицу.*

1.3 Обратимые матрицы

Элементарные преобразования и элементарные матрицы имеют многочисленные применения. Одно из них — проверка обратимости матрицы и нахождение обратной матрицы.

Матрица A называется *обратимой справа* (*слева*), если существует такая матрица B, что $AB = E$ ($BA = E$), при этом B называют *правой* (*левой*) *обратной* к A матрицей. Матрица *обратима*, если она одновременно обратима справа и слева. Отметим, что в последнем случае правая обратная и левая обратная матрицы совпадают. В самом деле, если $AB = E$ и $CA = E$, то $B = EB = CAB = CE = C$. Поэтому матрицу B называют просто *обратной* к A матрицей и обозначают A^{-1}.

Упражнение 1.1

1. Если A обратима, то $(A^t)^{-1} = (A^{-1})^t$.
2. Если A и B обратимы, то $(AB)^{-1} = B^{-1}A^{-1}$.

Лемма 1.3 *Каждая элементарная матрица обратима.*

Доказательство. Непосредственно проверяется, что $F_{st}^{-1} = F_{st}$, $(F_{st}(\lambda))^{-1} = F_{st}(-\lambda)$, $(F_s(\lambda))^{-1} = F_s(\lambda^{-1})$.$\square$

Предложение 1.4 *Если матрица A обратима справа, то в ней элементарными преобразованиями строк невозможно получить нулевую строку.*

Доказательство. Если A обратима справа, то для некоторой матрицы B верно равенство $AB = E$. Следовательно, в A нет нулевых строк, поскольку в противном случае произведение AB также содержало бы нулевую строку. Предположим теперь, что некоторыми элементарными преобразованиями строк A приведена к матрице $\tilde{A}$, содержащей нулевую строку. В силу лемм 1.2 и 1.3 найдется такая обратимая матрица F, что $\tilde{A} = FA$. Тогда $\tilde{A}(BF^{-1}) = FABF^{-1} = FEF^{-1} = FF^{-1} = E$, так что содержащая нулевую строку матрица $\tilde{A}$ обратима справа, но выше было показано, что это невозможно. Следовательно, в A нулевую строку получить нельзя.$\square$

Следствие 1.5 *Если $m \times n$-матрица A обратима, то $m = n$.*

Доказательство. Матрица A обратима, следовательно, она обратима справа. Воспользовавшись леммой 1.1, приведем A к ступенчатому виду $\tilde{A}$. В силу предложения 1.4 матрица $\tilde{A}$ не содержит нулевых строк, но с учетом строения матрицы ступенчатого вида (см. (1)) это возможно только при $m \leq n$. С помощью аналогичных рассуждений об обратимой (см. упражнение 1.1) матрице A^t получаем неравенство $n \leq m$.$\square$

Теорема 1.6 *Квадратная матрица A обратима справа тогда и только тогда, когда она обратима слева.*

Доказательство. Пусть A обратима справа, то есть для некоторой матрицы B верно равенство $AB = E$. Как и в доказательстве следствия 1.5 заметим, что матрица A после приведения к ступенчатому виду $\tilde{A}$ не содержит нулевых строк. Но для квадратной матрицы это возможно только тогда, когда вершины всех ступенек находятся на главной диагонали. Таким образом, все диагональные элементы $\tilde{a}_{11},\ldots,\tilde{a}_{nn}$ отличны от нуля. Следовательно, элементарными преобразованиями строк II-го рода матрицу $\tilde{A}$ можно привести к диагональной матрице $\operatorname{diag}[\tilde{a}_{11},\ldots,\tilde{a}_{nn}]$. (Сначала с помощью преобразований $\mathcal{F}_{ni}(-\tilde{a}_{in}/\tilde{a}_{nn})$ обнуляем первые $n-1$ элементов последнего столбца, затем, используя предпоследнюю строку, обнуляем первые $n-2$ элементов предпоследнего столбца и так далее.) Наконец, с помощью элементарных преобразований III-го рода диагональная матрица превращается в единичную.

Итак, обратимую справа квадратную матрицу A элементарными преобразованиями строк можно привести к единичной матрице E. Ввиду леммы 1.2 это означает, что существует такая матрица F, что $FA = E$, следовательно, A обратима слева.

Обратно, если A обратима слева, то $BA = E$ для некоторой матрицы B. Ясно, что B обратима справа. Тогда в силу первой части доказательства матрица B обратима слева, а значит, обратима, так что $A = B^{-1}$ и, в частности, $AB = E$. Таким образом, A обратима справа. $\square$

Подчеркнем важное обстоятельство: в процессе доказательства теоремы 1.6 было установлено, что обратная к A матрица является произведением элементарных матриц, соответствующих элементарным преобразованиям строк, приводящих A к единичной матрице. А именно, если $\mathcal{F}_1,\ldots,\mathcal{F}_k$ — соответствующая последовательность элементарных преобразований и $F_1,\ldots,F_k$ — последовательность отвечающих этим преобразованиям элементарных матриц, то $A^{-1} = F_k\ldots F_1$. На применении последней формулы основан следующий

Способ нахождения обратной матрицы.

Приписав справа к матрице A единичную матрицу того же порядка, составим $n\times 2n$-матрицу $(A|E)$. С помощью элементарных преобразований

строк приведем ее к ступенчатому виду. Если при этом в матрице A появляется нулевая строка, то ввиду предложения 1.4 матрица A не имеет обратной. Если же нулевых строк нет, то согласно первой части доказательства теоремы 1.6 матрицу $(A|E)$ можно привести к такому виду, чтобы в первых ее n столбцах получилась матрица E. Тогда последние n столбцов образуют матрицу A^{-1}.

Обоснование. Элементы n первых и n последних столбцов $n \times 2n$-матрицы $(A|E)$ меняются по одним и тем же правилам, поэтому если $F_k \dots F_1 A = E$, то $F_k \dots F_1 E = F_k \dots F_1 = A^{-1}$.$\square$

В заключение отметим, что обратные матрицы можно использовать для решения матричных уравнений. В самом деле, пусть требуется решить матричное уравнение $AX = B$, где A и B — заданные матрицы, а X — неизвестная матрица. Очевидно, что если матрица A обратима, то единственным решением будет матрица $X = A^{-1}B$. Аналогично, единственным решением уравнения $YA = B$ в случае обратимости матрицы A является матрица $Y = BA^{-1}$.

Полученные формулы можно использовать и для решения систем линейных уравнений при некоторых дополнительных условиях. Пусть дана система

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \dots \quad \dots \quad \dots & \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n \end{cases}.$$

Легко видеть, что она может быть переписана в матричном виде $Ax = b$, где

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

Если A обратима, то система имеет единственное решение $x = A^{-1}b$.

1.4 Решение систем линейных уравнений методом Гаусса

Рассмотрим систему линейных уравнений общего вида

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \quad\cdots \quad\quad \cdots \quad\quad \cdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{cases} . \tag{2}$$

Как и в предыдущем пункте, систему (2) можно переписать в матричном виде $Ax = b$, где $A = (a_{ij})$ — матрица системы, x — столбец неизвестных, b — столбец свободных членов. Приписав справа к матрице A столбец b, получаем *расширенную* матрицу $\bar{A} = (A|b)$ системы.

Лемма 1.7 *Множество решений системы $Ax = b$ не меняется при элементарных преобразованиях строк матрицы $\bar{A}$.*

Доказательство. В силу лемм 1.2 и 1.3 элементарные преобразования строк матрицы $\bar{A}$ равносильны ее домножению слева на некоторую обратимую матрицу F, поэтому после элементарных преобразований система $Ax = b$ примет вид $FAx = Fb$. Ясно, что если x_0 — некоторое решение системы $Ax = b$, то $Ax_0 = b$, откуда $FAx_0 = Fb$, то есть x_0 — решение преобразованной системы.

Обратно, если $FAx_0 = Fb$, то ввиду обратимости матрицы F получаем $Ax_0 = F^{-1}FAx_0 = F^{-1}Fb = b$, то есть x_0 — решение исходной системы $Ax = b$. Следовательно, множества решений исходной и преобразованной систем совпадают. □

Опишем теперь способ решения системы (2), известный как метод Гаусса.

1. Приведем расширенную матрицу $\bar{A}$ системы к ступенчатому виду. Согласно лемме 1.7 множество решений системы при этом не изменится. Если в получившейся матрице есть строка вида $(0\ldots0\,|\,b)$, где $b \neq 0$, то этой строке отвечает не имеющее решений уравнение $0 = b$, следовательно, и вся система не имеет решений. Если же таких строк нет, то с точностью до перенумерации неизвестных (или, эквивалентно, с точностью до перестановки столбцов матрицы) можно считать, что вершины

"ступенек" в матрице вида (1) находятся в первых ее r столбцах. Следовательно, преобразованная система примет вид

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1r}x_r + \cdots + a_{1n}x_n &= b_1 \\ a_{22}x_2 + \cdots + a_{2r}x_r + \cdots + a_{2n}x_n &= b_2 \\ \cdots \qquad \cdots \qquad \cdots & \\ a_{rr}x_r + \cdots + a_{rn}x_n &= b_r, \end{cases} \tag{3}$$

где $a_{11} \dots a_{rr} \neq 0$. (Для упрощения обозначений мы используем одинаковые буквы для элементов как исходной, так и преобразованной матриц.)

2. Объявим неизвестные $x_1, \dots, x_r$ главными, а прочие (если они есть) — свободными. Перенесем свободные неизвестные в правые части уравнений. Получим

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1r}x_r &= b_1 - a_{1,r+1}x_{r+1} - \cdots - a_{1n}x_n \\ a_{22}x_2 + \cdots + a_{2r}x_r &= b_2 - a_{2,r+1}x_{r+1} - \cdots - a_{2n}x_n \\ \cdots \qquad \cdots & \cdots \\ a_{rr}x_r &= b_r - a_{r,r+1}x_{r+1} - \cdots - a_{rn}x_n \end{cases} \tag{4}$$

Двигаясь по системе (4) снизу вверх и обращаясь с правыми частями уравнений как с буквенными выражениями, последовательно выражаем главные неизвестные через свободные: $x_r = f_r(x_{r+1}, \dots, x_n), \dots, x_1 = f_1(x_{r+1}, \dots, x_n)$. В результате получим общее решение системы (2):

$$X_{\text{общ}} = (f_1(x_{r+1}, \dots, x_n), \dots, f_r(x_{r+1}, \dots, x_n), x_{r+1}, \dots, x_n).$$

Всякое частное решение системы получается из ее общего решения подстановкой вместо свободных переменных произвольных чисел.

2 Поле комплексных чисел

В цепочке $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$ известных из курса школьной математики числовых множеств каждое последующее множество обладает более "хорошими" алгебраическими свойствами по сравнению с предыдущим: натуральные числа можно только складывать и умножать, целые — еще и вычитать, рациональные — делить (если делитель отличен от 0), из неотрицательных вещественных чисел можно извлекать арифметические корни. Как уже упоминалось в предыдущем разделе, обобщением вещественных чисел являются вещественные матрицы, однако, при этом

некоторые полезные свойства операций с вещественными числами — коммутативность умножения, обратимость любого ненулевого числа — для матриц перестают быть верными. Естественно задаться вопросом: нельзя ли расширить множество вещественных чисел с сохранением всех свойств операций так, чтобы при этом корни можно было извлекать из всех чисел? Построению и изучению свойств таких чисел, называемых комплексными, посвящен данный раздел.

2.1 Начальные сведения об алгебраических системах

Декартовым произведением множеств A и B называется множество $A \times B$, состоящее из упорядоченных пар, первый элемент которых лежит в A, а второй — в B, то есть $A \times B = \{(a, b) : a \in A, b \in B\}$. Данное определение легко распространяется на любое конечное число сомножителей $A_1, \ldots, A_k$, что позволяет определить декартово произведение $A_1 \times \cdots \times A_k$. В случае, когда $A_1 = \cdots = A_k = A$, такое произведение называется *декартовой степенью* множества A и обозначается A^k.

Отображением из множества A в множество B называется соответствие φ, которое каждому элементу $a \in A$ сопоставляет некоторый элемент $\varphi(a) \in B$. Отображение $*$ множества $A \times A$ в A называется *бинарной алгебраической операцией* на множестве A. Образ пары $(a, b) \in A \times A$ при этом записывают обычно в виде $a * b$. Говорят, что непустое подмножество $X \subseteq A$ *замкнуто* относительно операции $*$, если $x * y \in X$ для всех $x, y \in X$, то есть применение операции $*$ к любой паре элементов множества X снова дает элемент из X.

Естественными примерами алгебраических операций могут служить обычные операции сложения "$+$" и умножения "$\cdot$" чисел. Легко видеть, что вычитание будет алгебраической операцией на множествах $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$, но не на множестве $\mathbb{N}$, поскольку $\mathbb{N}$ не является замкнутым относительно вычитания.

Основными объектами изучения в алгебре являются *алгебраические системы* — множества с заданными на них алгебраическими операциями, при этом главное значение имеет не природа самих множеств, а свойства алгебраических операций. Наиболее важными примерами алгебраических систем являются группы, кольца и поля.

Система $(G, *)$ называется *группой*, если выполняются следующие условия:

1) для всех $a, b, c \in G$ верно $(a*b)*c = a*(b*c)$ (ассоциативность);

2) существует такой элемент $e \in G$, что $e * a = a * e = a$ для всех $a \in G$ (существование единичного элемента);

3) для любого $a \in G$ существует такой элемент $a' \in G$, что $a * a' = a' * a = e$ (существование обратного элемента).

Если дополнительно выполняется условие

4) для всех $a, b \in G$ верно $a * b = b * a$ (коммутативность),

то группа G называется *абелевой*.

В зависимости от выбора знака алгебраической операции различают мультипликативную и аддитивную терминологии. Различия между ними приведены в таблице 1. Аддитивная терминология применяется, как правило, для абелевых групп.

терминология	знак	e	a'
мультипликативная	$\cdot$	единичный элемент, единица, e, 1	обратный элемент, a^{-1}
аддитивная	$+$	нейтральный элемент, нуль, 0	противоположный элемент, $-a$

Таблица 1

Легко видеть, что множества $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ образуют абелевы группы относительно сложения, а множества $\mathbb{Q}^* = \mathbb{Q}\backslash\{0\}$ и $\mathbb{R}^* = \mathbb{R}\backslash\{0\}$ — относительно умножения.

Система $(R, +, \cdot)$ называется *кольцом*, если

1) $(R, +)$ — абелева группа;

2) для всех $a, b, c \in R$ верно $(a + b)c = ac + bc$ и $a(b + c) = ab + ac$ (дистрибутивность).

Если умножение в R обладает дополнительными свойствами, например, ассоциативностью, коммутативностью или в кольце существует единица, то говорят, что кольцо R ассоциативно, коммутативно или, соответственно, обладает единицей. Если $(R\backslash\{0\}, \cdot)$ — абелева группа, то кольцо R называется *полем*.

Непосредственно проверяется, что относительно обычных операций сложения и умножения множество $\mathbb{Z}$ образует коммутативное ассоциативное кольцо с единицей, множество $2\mathbb{Z}$ четных чисел — коммутативное ассоциативное кольцо без единицы, а $\mathbb{Q}$ и $\mathbb{R}$ являются полями. Примером некоммутативного ассоциативного кольца с единицей может служить множество $M_n(\mathbb{R})$ вещественных матриц порядка $n \geq 2$ относительно матричных операций сложения и умножения.

2.2 Построение поля комплексных чисел. Алгебраическая форма комплексного числа

Рассмотрим в кольце $M_2(\mathbb{R})$ подмножество $\mathbb{C}$ всех матриц вида $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$.

Теорема 2.1 $(\mathbb{C}, +, \cdot)$ — *поле.*

Доказательство. Обозначим через i матрицу $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathbb{C}$. Заметим, что всякая матрица $A \in \mathbb{C}$ имеет вид $A = aE + bi$, причем $i^2 = -E$. При $a = b = 0$ получаем, что в $\mathbb{C}$ содержится нулевая матрица, а при $a = 1,\ b = 0$ — единичная. Легко также видеть, что $\mathbb{C}$ замкнуто относительно сложения и умножения, причем умножение коммутативно, и что $\mathbb{C}$ содержит противоположную матрицу $-A$ ко всякой матрице $A \in \mathbb{C}$. Кроме того, в предыдущем разделе было показано, что сложение матриц коммутативно и ассоциативно, а умножение ассоциативно и дистрибутивно слева и справа относительно сложения. Таким образом, остается лишь показать, что любая ненулевая матрица $A \in \mathbb{C}$ обратима. В самом деле, если $A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, $A \neq 0$, то $a \neq 0$ или $b \neq 0$, поэтому $a^2 + b^2 \neq 0$. Нетрудно видеть, что матрица $B = \dfrac{1}{a^2 + b^2} \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ лежит в $\mathbb{C}$ и удовлетворяет равенству $AB = E$, следовательно, $B = A^{-1}$.$\square$

Построенное поле $\mathbb{C}$ называется полем *комплексных* чисел. Заметим, что лежащие в нем *скалярные* матрицы вида aE, $a \in \mathbb{R}$, складываются и перемножаются точно так же, как и вещественные числа:

$aE + bE = (a+b)E$, $aE \cdot bE = abE$. Поэтому для упрощения обозначений можно отождествить такие матрицы с вещественными числами и вместо aE писать просто a. В этом смысле можно считать, что $\mathbb{R} \subset \mathbb{C}$, а упомянутое в доказательстве теоремы 2.1 представление комплексного числа $z = aE + bi$ тогда принимает вид $z = a + bi$, $i^2 = -1$, называемый *алгебраической формой* числа z. Вещественные числа a и b называются *вещественной* и *мнимой* частями числа z и обозначаются $\operatorname{Re} z$ и $\operatorname{Im} z$, соответственно. Таким образом, число z является вещественным тогда и только тогда, когда его мнимая часть равна 0. Если же у ненулевого числа z равна нулю вещественная часть, то такое число называют *чисто мнимым*.

Каждому комплексному числу $z = a + bi$ можно сопоставить *комплексно-сопряженное* число $\bar{z} = a - bi$. В качестве упражнения докажите следующие свойства:

1°. $\bar{\bar{z}} = z$;

2°. $\bar{z} = z \Leftrightarrow \operatorname{Im} z = 0$;

3°. $\bar{z} = -z \Leftrightarrow \operatorname{Re} z = 0$;

4°. $z + \bar{z} = 2 \operatorname{Re} z \in \mathbb{R}$;

5°. $z\bar{z} = (\operatorname{Re} z)^2 + (\operatorname{Im} z)^2 \geq 0$, причем $z\bar{z} = 0 \Leftrightarrow z = 0$.

Последнее свойство объясняет вид числа z^{-1} в доказательстве теоремы 2.1:

$$z^{-1} = \frac{1}{z} = \frac{\bar{z}}{z\bar{z}} = \frac{a - bi}{a^2 + b^2} = \frac{a}{a^2 + b^2} + \frac{-b}{a^2 + b^2} i.$$

Приведем еще несколько свойств операции комплексного сопряжения:

6°. $\overline{z_1 \pm z_2} = \bar{z}_1 \pm \bar{z}_2$;

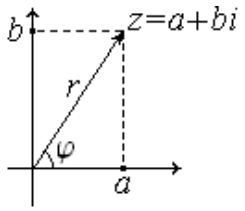
7°. $\overline{z_1 z_2} = \bar{z}_1 \bar{z}_2$;

8°. $\left(\overline{\frac{z_1}{z_2}}\right) = \frac{\bar{z}_1}{\bar{z}_2}$.

2.3 Тригонометрическая форма комплексных чисел

Как известно, вещественные числа принято изображать точками на вещественной прямой. Для комплексных чисел естественно дополнить вещественную ось абсцисс мнимой осью ординат и изображать число $z = a + bi$ точкой на плоскости с координатами (a, b).

Каждое комплексное число $z = a + bi$ можно также отождествить с вектором, выходящим из начала координат и заканчивающимся в точке



(a, b). Легко видеть, что сложение/вычитание комплексных чисел согласуется со сложением/вычитанием соответствующих векторов.

Кроме декартовой системы координат на плоскости существует также полярная система координат, в которой положение точки $z = (a, b)$ характеризуется расстоянием r от начала координат и (в случае $z \neq 0$) углом φ между положительной полуосью и соответствующим числу z вектором. Используя методы школьной геометрии, нетрудно вывести равенства

$$r = \sqrt{a^2 + b^2}, \quad \cos\varphi = \frac{a}{r}, \quad \sin\varphi = \frac{b}{r}.$$

Тогда число z можно записать в виде

$$z = a + bi = r\Big(\frac{a}{r} + \frac{b}{r}\,i\Big) = r(\cos\varphi + i\sin\varphi),$$

который называется *тригонометрической формой* числа z. Число r называется *модулем* числа z и обозначается $|z|$, а угол φ — *аргументом* ($\arg z$). Следует подчеркнуть, что аргумент существует только у ненулевых чисел и находится из условий

$$\cos\varphi = \frac{a}{r},\ \sin\varphi = \frac{b}{r}$$

с точностью до угла, кратного 2π.

Приведем очевидные свойства модуля:

1°. $|z| = \sqrt{z\bar{z}}$;

2°. $|z| = 0 \ \Leftrightarrow \ z = 0$;

3°. $|z| = |-z|$;

4°. $|z| = |\bar{z}|$.

Пусть z_1, z_2 — ненулевые комплексные числа. Запишем их в тригонометрической форме: $z_k = r_k(\cos\varphi_k + i\sin\varphi_k)$, $k = 1, 2$. Тогда

$$z_1 z_2 = r_1 r_2(\cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2 + i(\cos\varphi_1\sin\varphi_2 + \sin\varphi_1\cos\varphi_2)) = r_1 r_2(\cos(\varphi_1 + \varphi_2) + i\sin(\varphi_1 + \varphi_2)).$$

Ясно, что

$$|z_1 z_2| = r_1 r_2 = |z_1|\,|z_2|, \quad \arg(z_1 z_2) = \varphi_1 + \varphi_2 = \arg\, z_1 + \arg\, z_2,$$

другими словами, *при умножении комплексных чилел их модули перемножаются, а аргументы складываются.* Применяя данное правило к произведению одинаковых сомножителей, получаем **формулу Муавра**:

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

Если $z = r(\cos\varphi + i\sin\varphi)$, то

$$z^{-1} = \frac{\bar z}{z\bar z} = \frac{r(\cos\varphi - i\sin\varphi)}{r^2} = \frac{1}{r}(\cos(-\varphi) + i\sin(-\varphi)),$$

значит, $|z^{-1}| = |z|^{-1}$, $\arg\, z^{-1} = -\arg\, z$, так что

$$\left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|}, \quad \arg\Big(\frac{z_1}{z_2}\Big) = \arg\, z_1 - \arg\, z_2,$$

то есть *при делении комплексных чисел модуль первого числа делится на модуль второго и из аргумента первого числа вычитается аргумент второго.*

Напомним, что $|z|$ — это длина вектора, соответствующего числу z, поэтому приведенные выше свойства модуля можно дополнить **неравенством треугольника**:

$$\Big||z_1| - |z_2|\Big| \le |z_1 \pm z_2| \le |z_1| + |z_2|.$$

В самом деле, неравенство $|z_1 + z_2| \le |z_1| + |z_2|$ есть переформулировка известного геометрического факта, согласно которому длина любой стороны треугольника не превосходит суммы длин двух других его сторон. Воспользовавшись этим неравенством, получаем

$$|z_1| = |(z_1 + z_2) - z_2| \le |z_1 + z_2| + |-z_2| = |z_1 + z_2| + |z_2|,$$

откуда

$$|z_1| - |z_2| \le |z_1 + z_2|.$$

Аналогично доказывается неравенство $|z_2| - |z_1| \le |z_1 + z_2|$, так что

$$\Big||z_1| - |z_2|\Big| \le |z_1 + z_2| \le |z_1| + |z_2|.$$

Для завершения доказательства осталось заменить в последней цепочке неравенств z_2 на $-z_2$.$\square$

2.4 Извлечение корней из комплексных чисел

Как было показано выше, тригонометрическая форма комплексных чисел удобна для их умножения, деления и возведения в степень. То же относится и к извлечению корней.

Пусть $n \in \mathbb{N}$, $z \in \mathbb{C}$. Обозначим через $\sqrt[n]{z}$ множество $\{w \in \mathbb{C} : w^n = z\}$ *корней* степени n из числа z. Выясним, как оно устроено.

Пусть $w \in \sqrt[n]{z}$. Представим z и w в тригонометрической форме:

$$z = r(\cos\varphi + i\sin\varphi), \quad w = \rho(\cos\psi + i\sin\psi).$$

Тогда по формуле Муавра имеем $z = w^n = \rho^n(\cos(n\psi) + i\sin(n\psi))$, следовательно,

$$r = \rho^n, \quad n\psi = \varphi + 2\pi k, \qquad k \in \mathbb{Z},$$

откуда

$$\rho = \sqrt[n]{r} \text{ (— арифметический корень(!))}, \qquad \psi = \frac{\varphi + 2\pi k}{n}.$$

Таким образом, каждое входящее в $\sqrt[n]{z}$ число имеет вид

$$w_k = \sqrt[n]{r}\left(\cos\frac{\varphi + 2\pi k}{n} + i\sin\frac{\varphi + 2\pi k}{n}\right), \ k \in \mathbb{Z}. \tag{5}$$

Заметим, что

$$w_{n+k} = \sqrt[n]{r}\left(\cos\frac{\varphi + 2\pi(k+n)}{n} + i\sin\frac{\varphi + 2\pi(k+n)}{n}\right) =$$

$$\sqrt[n]{r}\left(\cos\frac{\varphi + 2\pi k}{n} + i\sin\frac{\varphi + 2\pi k}{n}\right) = w_k$$

при всех $k \in \mathbb{Z}$, так что числа w_k циклически повторяются через каждые n шагов, следовательно, окончательно имеем

$$\sqrt[n]{z} = \{w_0, w_1, \ldots, w_{n-1}\}. \tag{6}$$

В частности, при $z = 1$ получаем формулу для нахождения комплексных корней из единицы: $\sqrt[n]{1} = \{\epsilon_0, \epsilon_1, \ldots, \epsilon_{n-1}\}$, где

$$\epsilon_k = \cos\frac{2\pi k}{n} + i\sin\frac{2\pi k}{n} \ \text{ для всех } k.$$

Множество $\sqrt[n]{1}$ обычно обозначают через $\mathbf{U}_n$. Используя правила арифметических действий с комплексными числами в тригонометрической

форме и учитывая доказанную выше циклическую повторяемость корней ($\epsilon_k = \epsilon_l$, если $k - l$ кратно n), нетрудно проверить справедливость следующих свойств:

1°. $\epsilon_k = \epsilon_1^k$.

2°. $\epsilon_k \epsilon_l = \epsilon_{k+l}$.

3°. $\epsilon_k^{-1} = \epsilon_{-k}$.

4°. $(\mathbf{U}_n, \cdot)$ — абелева группа.

3 Матрицы и системы линейных уравнений

В разделе 1 уже были изложены некоторые начальные сведения о матрицах и их применении к решению систем линейных уравнений. В данном разделе вводится одна из важнейших характеристик матрицы — ее ранг, что позволяет получить дальнейшие результаты о матрицах и системах линейных уравнений.

3.1 Арифметические векторные пространства

Пусть K — поле, $n \in \mathbb{N}$. *Арифметическим векторным пространством* называется множество $K^n = \{x = (x_1, \ldots, x_n) : x_1, \ldots, x_n \in K\}$ всех строк (или, эквивалентно, столбцов) длины n, составленных из элементов поля K, с обычными матричными операциями сложения и умножения на элементы из K:

$$x + y = (x_1 + y_1, \ldots, x_n + y_n), \quad \lambda x = (\lambda x_1, \ldots, \lambda x_n).$$

Элементы векторного пространства называют *векторами*, а элементы поля K — *скалярами* (или просто числами). *Линейной комбинацией* векторов $a_1, \ldots, a_k \in K^n$ называется всякий вектор $b = \lambda_1 a_1 + \ldots + \lambda_k a_k$, где $\lambda_1, \ldots, \lambda_k \in K$; говорят также, что b *линейно выражается* через векторы $a_1, \ldots, a_k$.

Система векторов $\{a_1, \ldots, a_k\}$ называется *линейно зависимой*, если существуют такие числа $\lambda_1, \ldots, \lambda_k \in K$, не все равные нулю, что $\lambda_1 a_1 + \ldots + \lambda_k a_k = 0$, в противном случае система называется *линейно независимой*. Другими словами, система $\{a_1, \ldots, a_k\}$ линейно независима, если линейная комбинация ее векторов дает нулевой вектор только в том случае, когда все коэффициенты равны нулю.

Упражнение 3.1

1. Система $\{a\}$ линейно зависима тогда и только тогда, когда $a = 0$.

2. Система, содержащая линейно зависимую подсистему, сама линейно зависима.

3. Если система $\{a_1, \ldots, a_k\}$ линейно независима, а система $\{a_1, \ldots, a_k, b\}$ линейно зависима, то b линейно выражается через векторы $a_1, \ldots, a_k$.

4. Система линейно зависима в точности тогда, когда хотя бы один из ее векторов линейно выражается через остальные векторы системы.

Пусть X — произвольный (возможно, бесконечный) набор векторов. Линейно независимая система $\{a_1, \ldots, a_k\}$ векторов из X называется *максимально линейно независимой* в X, если добавление к ней любого вектора из X превращает ее в линейно зависимую систему. Линейно независимая система $\{a_1, \ldots, a_k\}$ называется *базисом* для X, если любой вектор из X через нее линейно выражается. С учетом пп. 3–4 упражнения 3.1 нетрудно показать, что всякая максимально линейно независимая система является базисом, и наоборот.

Легко видеть, что во всяком арифметическом пространстве K^n существует так называемый *стандартный* базис:

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

Для того, чтобы установить существование базиса для любого набора векторов пространства K^n, сначала докажем вспомогательное

Предложение 3.1 *Пусть $\{a_1, \ldots, a_m\}$ — линейно независимая система, каждый вектор которой линейно выражается через векторы системы $\{b_1, \ldots, b_n\}$. Тогда $m \leq n$.*

Доказательство. Обозначим через l количество векторов, входящих одновременно в обе системы. Если $l > 0$, то перенумеруем векторы систем так, чтобы первые номера получили "общие" векторы, то есть считаем, что $a_1 = b_1, \ldots, a_l = b_l$. Покажем, что если $l < m$, то $l < n$ и для некоторого индекса j, где $j > l$, каждый вектор системы $\{a_1, \ldots, a_m\}$ линейно выражается через векторы системы, полученной из $\{b_1, \ldots, b_n\}$ заменой b_j на a_{l+1}.

Действительно, согласно условию при подходящих коэффициентах $\lambda_1, \ldots, \lambda_n$ имеем равенство $a_{l+1} = \lambda_1 b_1 + \ldots + \lambda_n b_n$. Ясно, что $l < n$,

поскольку в противном случае вектор a_{l+1} линейно выражался бы через векторы $b_1 = a_1, \ldots, b_l = a_l$ в противоречие с линейной независимостью системы $\{a_1, \ldots, a_m\}$. По той же причине среди $\lambda_{l+1}, \ldots, \lambda_n$ есть хотя бы одно число $\lambda_j \neq 0$, что позволяет выразить вектор b_j через векторы системы $\{b_1, \ldots, b_{j-1}, a_{l+1}, b_{j+1}, \ldots, b_n\}$. В результате получаем, что все векторы системы $\{a_1, \ldots, a_m\}$ линейно выражаются через векторы системы $\{b_1, \ldots, b_n\}$, которые, в свою очередь, линейно выражаются через векторы системы $\{b_1, \ldots, b_{j-1}, a_{l+1}, b_{j+1}, \ldots, b_n\}$, следовательно, все векторы первой системы линейно выражаются через векторы последней системы.

Итак, предположив, что $l < m$, мы построили систему, состоящую из n векторов, через которые линейно выражаются все векторы системы $\{a_1, \ldots, a_m\}$, при этом новая система имеет с системой $\{a_1, \ldots, a_m\}$ на один "общий" вектор больше, чем система $\{b_1, \ldots, b_n\}$. Поскольку описанный выше процесс можно применять до тех пор, пока количество общих векторов строго меньше m, то через конечное число шагов будет получена состоящая из n векторов система, содержащая все векторы системы $\{a_1, \ldots, a_m\}$, следовательно, $m \leq n$.□

Следствие 3.2 *Пусть X — произвольный набор векторов пространства K^n. Любую линейно независимую систему в X (в том числе, пустое множество векторов) можно дополнить до базиса набора X, следовательно, базис существует.*

Доказательство. В силу предложения 3.1 количество векторов, образующих линейно независимую систему, не превосходит числа n векторов в стандартном базисе пространства K^n. Пусть $0 \leq k \leq n$ и $\{a_1, \ldots, a_k\}$ — линейно независимая система векторов из X. Ясно, что либо система $\{a_1, \ldots, a_k\}$ максимально линейно независима, либо найдется такой вектор $a_{k+1} \in X$, что система $\{a_1, \ldots, a_k, a_{k+1}\}$ линейно независима. В последнем случае снова либо система $\{a_1, \ldots, a_k, a_{k+1}\}$ максимально линейно независима, либо для некоторого вектора $a_{k+2} \in X$ система $\{a_1, \ldots, a_k, a_{k+1}, a_{k+2}\}$ линейно независима и так далее. Заметим, что ввиду упомянутого предложения 3.1 увеличение числа линейно независимых векторов из X не может продолжаться бесконечно. Следовательно, через конечное число шагов будет получена максимально линейно независимая в X система, то есть базис.□

Отметим также, что в силу предложения 3.1 число векторов в любом базисе набора X не превосходит числа векторов в любом другом его базисе, поэтому справедливо

Следствие 3.3 *Любые два базиса набора X содержат одно и то же число векторов.*

В случае, когда набор X совпадает с K^n или является подпространством (см. ниже), число векторов в базисе X называется *размерностью* X и обозначается $\dim X$; для конечных же наборов векторов, как правило, используется термин *ранг* и обозначение $\operatorname{rk} X$.

Лемма 3.4 *Если каждый вектор набора X линейно выражается через векторы набора Y, то $\operatorname{rk} X \leq \operatorname{rk} Y$.*

Доказательство. Очевидно, из условия леммы вытекает, что каждый базисный вектор набора X линейно выражается через базисные векторы набора Y, следовательно, остается лишь воспользоваться предложением 3.1.$\square$

Непустое подмножество $U \subseteq K^n$ называется *подпространством*, если оно замкнуто относительно сложения и умножения на элементы поля K, то есть для всех $x, y \in U$, $\lambda \in K$ верно $x + y, \lambda x \in U$. Простейшими примерами подпространств являются нулевое подпространство $\{0\}$ и само пространство K^n.

Общий способ построения подпространств состоит в следующем: каждому непустому набору X векторов сопоставим множество $\langle X \rangle$ всевозможных конечных линейных комбинаций векторов из X, иначе говоря, $a \in \langle X \rangle$ в точности тогда, когда $a = \lambda_1 x_1 + \ldots + \lambda_k x_k$ для некоторых $x_1, \ldots, x_k \in X$, $\lambda_1, \ldots, \lambda_k \in K$. Множество $\langle X \rangle$ называется *линейной оболочкой* набора X. Легко проверяется, что $\langle X \rangle$ есть наименьшее (по включению) подпространство, содержащее все векторы из X. В частности, X есть подпространство ровно тогда, когда $X = \langle X \rangle$. Поскольку все векторы из $\langle X \rangle$ линейно выражаются через векторы набора X, и наоборот, то из леммы 3.4 вытекает

Следствие 3.5 *Справедливо равенство $\dim\langle X \rangle = \operatorname{rk} X$.*

Следствие 3.6 *Пусть $\operatorname{rk} X = r$. Если система $\{a_1, \ldots, a_r\}$ векторов из X линейно независима, то она является базисом.*

Доказательство. С учетом предложения 3.1 число линейно независимых векторов в X не может быть больше r, откуда вытекает, что система $\{a_1, \ldots, a_r\}$ максимально линейно независима и, значит, является базисом. $\square$

3.2 Ранг матрицы

Как уже отмечалось ранее, каждую $m \times n$-матрицу A можно рассматривать как систему $\{A_{(1)}, \ldots, A_{(m)}\}$ строк — векторов n-мерного арифметического пространства. Ранг этой системы называется *рангом* матрицы A *по строкам* и обозначается $\mathrm{rk}_{\text{г}}(A)$. Ранг $\mathrm{rk}_{\text{в}}(A)$ системы столбцов $\{A^{(1)}, \ldots, A^{(n)}\}$ матрицы A называется ее *рангом по столбцам.*

Предложение 3.7 *Величины* $\mathrm{rk}_{\text{г}}(A)$ *и* $\mathrm{rk}_{\text{в}}(A)$ *не меняются при элементарных преобразованиях строк матрицы* A.

Доказательство. Достаточно ограничиться случаем, когда матрица $\tilde{A}$ получена из A с помощью одного элементарного преобразования.

1. Сначала докажем равенство $\mathrm{rk}_{\text{г}}(A) = \mathrm{rk}_{\text{г}}(\tilde{A})$. Возможны три случая.

1) К A было применено преобразование $\mathcal{F}_{st}$ I-го рода. В этом случае изменился лишь порядок векторов системы, но не сама система, следовательно, не изменился и ее ранг.

2) К A было применено преобразование $\mathcal{F}_{st}(\lambda)$ II-го рода. Системы строк матриц A и $\tilde{A}$ линейно выражаются друг через друга:

$$\begin{array}{ll}
\tilde{A}_{(1)} = A_{(1)} & A_{(1)} = \tilde{A}_{(1)} \\
\ldots & \ldots \\
\tilde{A}_{(s)} = A_{(s)} & A_{(s)} = \tilde{A}_{(s)} \\
\ldots & \ldots \\
\tilde{A}_{(t)} = A_{(t)} + \lambda A_{(s)} & A_{(t)} = \tilde{A}_{(t)} - \lambda \tilde{A}_{(s)} \\
\ldots & \ldots \\
\tilde{A}_{(n)} = A_{(n)} & A_{(n)} = \tilde{A}_{(n)}
\end{array}$$

следовательно, по лемме 3.4 имеем

$$\mathrm{rk}_{\text{г}}(A) = \mathrm{rk}\{A_{(1)}, \ldots, A_{(n)}\} = \mathrm{rk}\{\tilde{A}_{(1)}, \ldots, \tilde{A}_{(n)}\} = \mathrm{rk}_{\text{г}}(\tilde{A}).$$

3) К A было применено преобразование $\mathcal{F}_s(\lambda)$ III-го рода. Как и в случае 2), системы строк матриц A и $\tilde{A}$ линейно выражаются друг через друга:

$$\begin{array}{ll} \tilde{A}_{(1)} = A_{(1)} & A_{(1)} = \tilde{A}_{(1)} \\ \dots & \dots \\ \tilde{A}_{(s)} = \lambda A_{(s)} & A_{(s)} = \lambda^{-1}\tilde{A}_{(s)} \\ \dots & \dots \\ \tilde{A}_{(n)} = A_{(n)} & A_{(n)} = \tilde{A}_{(n)} \end{array}$$

следовательно, снова $\mathrm{rk}_{\varepsilon}(A) = \mathrm{rk}\{A_{(1)}, \dots, A_{(n)}\} = \mathrm{rk}\{\tilde{A}_{(1)}, \dots, \tilde{A}_{(n)}\} = \mathrm{rk}_{\varepsilon}(\tilde{A})$.

2. Для совпадения рангов по столбцам матриц A и $\tilde{A}$ достаточно, чтобы равенство нулю произвольной линейной комбинации столбцов матрицы A выполнялось тогда и только тогда, когда равна нулю точно такая же линейная комбинация столбцов матрицы $\tilde{A}$ с теми же номерами, то есть

$$\sum_i \lambda_i A^{(i)} = 0 \;\Leftrightarrow\; \sum_i \lambda_i \tilde{A}^{(i)} = 0. \tag{7}$$

(Докажите, что если верно (7), то каждой максимально линейно независимой системе столбцов матрицы A отвечает максимально линейно независимая система столбцов матрицы $\tilde{A}$ с теми же номерами.) Составим из коэффициентов $\lambda_1, \dots, \lambda_n$ столбец $\bar{\lambda}$. Тогда условие (7) примет более простой вид:

$$A\bar{\lambda} = 0 \;\Leftrightarrow\; \tilde{A}\bar{\lambda} = 0. \tag{8}$$

Матрица $\tilde{A}$ получена из A элементарным преобразованием строк, следовательно, $\tilde{A} = FA$ для некоторой элементарной матрицы F. Теперь с учетом леммы 1.3 выполнение условия (8) почти очевидно: если $A\bar{\lambda} = 0$, то $\tilde{A}\bar{\lambda} = FA\bar{\lambda} = F0 = 0$, и обратно, если $\tilde{A}\bar{\lambda} = 0$, то $A\bar{\lambda} = F^{-1}FA\bar{\lambda} = F^{-1}\tilde{A}\bar{\lambda} = F^{-1}0 = 0$, что и требовалось. $\square$

Теорема 3.8 $\mathrm{rk}_{\varepsilon}(A) = \mathrm{rk}_{\varepsilon}(A)$.

Доказательство. Ввиду леммы 1.1 и предложения 3.7 можно считать, что матрица A имеет ступенчатый вид (1). Более того, с помощью элементарных преобразований строк II-го и III-го рода можно привести матрицу к такому виду, чтобы ее столбцы с номерами $j_1, \dots, j_r$, проходящие через вершины "ступенек", имели вид $(1, 0, \dots, 0)^t$, $(0, 1, \dots, 0)^t$ и так далее, то есть совпадали с первыми r столбцами единичной матрицы.

Очевидно, такие столбцы образуют базис системы столбцов матрицы A, так что $\mathrm{rk}_{\text{в}}(A) = r$. С другой стороны, не менее очевидно, что в матрице указанного вида первые r строк также образуют базис системы всех строк матрицы, поэтому $\mathrm{rk}_{\text{г}}(A) = r$. Следовательно, $\mathrm{rk}_{\text{г}}(A) = \mathrm{rk}_{\text{в}}(A)$.□

Итак, ранг по строкам любой матрицы A равен ее рангу по столбцам, поэтому данную величину естественно называть просто *рангом*. Обозначение ранга матрицы: $\mathrm{rk}\, A$.

Отметим, что доказательство теоремы 3.8 дает способ нахождения ранга матрицы: *ранг равен количеству ненулевых строк, остающихся в матрице после приведения ее к ступенчатому виду.*

Ранг является одной из наиболее важных характеристик матрицы. В частности, зная ранг квадратной матрицы, легко решить вопрос о ее обратимости.

Теорема 3.9 (критерий обратимости матрицы в терминах ранга) *Матрица A порядка n обратима тогда и только тогда, когда* $\mathrm{rk}\, A = n$.

Доказательство. Если матрица A обратима, то после приведения к ступенчатому виду в ней не должно быть нулевых строк, так что $\mathrm{rk}\, A = n$. Обратно, приведение к ступенчатому виду квадратной матрицы A ранга n дает верхнетреугольную матрицу с ненулевыми элементами по главной диагонали, а из такой матрицы с помощью элементарных преобразований строк можно получить единичную матрицу. Следовательно, матрица A обратима.□

Теорема 3.10 *Справедливо неравенство*

$$\mathrm{rk}\,(AB) \le \min\{\mathrm{rk}\, A, \mathrm{rk}\, B\}. \tag{9}$$

Доказательство. Рассмотрим строки матрицы $C = AB$. Непосредственно из правила умножения матриц получаем:

$$C_{(i)} = A_{(i)}B = \sum_j a_{ij}B_{(j)},$$

то есть каждая строка матрицы C является линейной комбинацией строк матрицы B, следовательно,

$$\mathrm{rk}\, C \le \mathrm{rk}\, B. \tag{10}$$

Аналогично, для столбцов матрицы C имеем:

$$C^{(j)} = AB^{(j)} = \sum_i A^{(i)} b_{ij},$$

то есть столбцы матрицы C линейно выражаются через столбцы матрицы A, откуда

$$\operatorname{rk} C \le \operatorname{rk} A. \tag{11}$$

Одновременное выполнение неравенств (10) и (11) дает (9).□

Следствие 3.11 *Если матрицы B и C обратимы, то* $\operatorname{rk}(BAC) = \operatorname{rk} A$.

Доказательство. Согласно теореме 3.10 имеем $\operatorname{rk}(BAC) \le \operatorname{rk}(BA) \le \operatorname{rk} A$. С другой стороны, $A = B^{-1}(BAC)C^{-1}$, поэтому $\operatorname{rk} A = \operatorname{rk}(B^{-1}(BAC)C^{-1}) \le \operatorname{rk}(BAC)$.□

3.3 Классификация систем линейных уравнений. Однородные системы

Системы линейных уравнений классифицируются по нескольким признакам. Система, имеющая хотя бы одно решение, называется *совместной*, если же решений нет — *несовместной*. Совместная система, имеющая ровно одно решение, называется *определенной*, в противном случае — *неопределенной*. Критерий совместности системы дает

Теорема 3.12 (Кронекер – Капелли) *Система $Ax = b$ совместна тогда и только тогда, когда ранги основной и расширенной матриц системы совпадают.*

Доказательство. Пусть система $Ax = b$ совместна, то есть существует решение x_0. Из равенства $Ax_0 = b$ вытекает, что столбец b является линейной комбинацией столбцов $A^{(1)}, \dots, A^{(n)}$ матрицы A, следовательно, $\operatorname{rk} A = \operatorname{rk}\{A^{(1)}, \dots, A^{(n)}\} = \operatorname{rk}\{A^{(1)}, \dots, A^{(n)}, b\} = \operatorname{rk} \bar{A}$.

Обратно, пусть $\operatorname{rk} A = \operatorname{rk} \bar{A}$. Тогда с учетом следствия 3.6 произвольный базис $\{A^{(i_1)}, \dots, A^{(i_r)}\}$ системы столбцов матрицы A одновременно является базисом системы столбцов расширенной матрицы. Следовательно, столбец b является линейной комбинацией столбцов $A^{(i_1)}, \dots, A^{(i_r)}$ и тем более — всех столбцов матрицы A. Очевидно, коэффициенты соответствующей линейной комбинации всех столбцов матрицы A дают решение системы $Ax = b$.□

Система $Ax = b$ называется *однородной*, если ее столбец свободных членов — нулевой, то есть $b = 0$. В противном случае система называется *неоднородной*. Однородная система $Ax = 0$ имеет тривиальное решение $x = 0$, поэтому она всегда совместна.

Пусть A — $m \times n$-матрица. Легко проверяется, что множество V_A решений системы $Ax = 0$ образует подпространство в пространстве K^n всех n-мерных столбцов.

Теорема 3.13 *Пусть* $s = \dim V_A$, $r = \operatorname{rk} A$. *Тогда* $r + s = n$.

Доказательство. Выберем в V_A базис $\{X_1, \ldots, X_s\}$ и дополним его до базиса $\{X_1, \ldots, X_s, \ldots, X_n\}$ пространства V. Любой столбец $X \in V$ линейно выражается через векторы базиса:

$$X = \lambda_1 X_1 + \cdots + \lambda_s X_s + \cdots + \lambda_n X_n,$$

откуда

$$AX = \lambda_1 AX_1 + \cdots + \lambda_s AX_s + \cdots + \lambda_n AX_n = \lambda_{s+1} AX_{s+1} + \cdots + \lambda_n AX_n,$$

поскольку $AX_1 = \cdots = AX_s = 0$. Таким образом, любой столбец вида AX есть линейная комбинация столбцов $\tilde{X}_1 = AX_{s+1}, \ldots, \tilde{X}_{n-s} = AX_n$.

Покажем, что система $\{\tilde{X}_1, \ldots, \tilde{X}_{n-s}\}$ линейно независима. В самом деле,

$$0 = \lambda_1 \tilde{X}_1 + \cdots + \lambda_{n-s} \tilde{X}_{n-s} = \lambda_1 AX_{s+1} + \cdots + \lambda_{n-s} AX_n = A(\lambda_1 X_{s+1} + \cdots + \lambda_{n-s} X_n)$$

влечет $\lambda_1 X_{s+1} + \cdots + \lambda_{n-s} X_n \in V_A$, следовательно,

$$\lambda_1 X_{s+1} + \cdots + \lambda_{n-s} X_n = \mu_1 X_1 + \cdots + \mu_s X_s$$

для некоторых $\mu_1, \ldots, \mu_s$, откуда

$$\mu_1 X_1 + \cdots + \mu_s X_s - \lambda_1 X_{s+1} - \cdots - \lambda_{n-s} X_n = 0.$$

Но $\{X_1, \ldots, X_s, \ldots, X_n\}$ — базис V, поэтому $\mu_1 = \cdots = \mu_s = \lambda_1 = \cdots = \lambda_{n-s} = 0$.

Заметим, что если $X = (0 \ldots \overset{i}{\underset{\downarrow}{1}} \ldots 0)^t$, то $AX = A^{(i)}$, поэтому любой столбец матрицы A (как столбец вида AX) линейно выражается через столбцы $\tilde{X}_1, \ldots, \tilde{X}_{n-s}$, следовательно, с учетом леммы 3.4 имеем $\operatorname{rk} A \leq$

$\operatorname{rk}\{\tilde{X}_1, \ldots, \tilde{X}_{n-s}\} = n - s$. С другой стороны, непосредственно из определения операции умножения матриц следует, что любой столбец вида AX является линейной комбинацией столбцов матрицы A, так что в силу той же леммы 3.4 получаем $\operatorname{rk}\{\tilde{X}_1, \ldots, \tilde{X}_{n-s}\} \leq \operatorname{rk}\{A^{(1)}, \ldots, A^{(n)}\} = \operatorname{rk} A$. В итоге, $r = \operatorname{rk} A = n - s$, откуда $r + s = n$.□

Следствие 3.14 *Однородная система имеет только нулевое решение тогда и только тогда, когда ранг ее матрицы совпадает с количеством неизвестных.*

Доказательство. Ввиду теоремы 3.13 имеем $V_A = \{0\} \Leftrightarrow n - r = \dim V_A = 0 \Leftrightarrow n = r$.□

Итак, задача о нахождении общего решения однородной системы сводится к поиску базиса подпространства V_A (он называется *фундаментальным набором решений* (ФНР) системы): если $\{X_1, \ldots, X_{n-r}\}$ — ФНР, то

$$X^{\text{одн}}_{\text{общ}} = C_1 X_1 + \cdots + C_{n-r} X_{n-r}, \tag{12}$$

где $C_1, \ldots, C_{n-r}$ — произвольные числовые коэффициенты.

Существуют различные способы нахождения ФНР. Опишем один из них.

Рассмотрим однородную систему $Ax = 0$. По аналогии с рассуждениями, изложенными при описании метода Гаусса (см. п. 1.4), приведем матрицу системы к ступенчатому виду, выделим главные и свободные неизвестные и перепишем систему в виде

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1r}x_r &= -a_{1,r+1}x_{r+1} - \cdots - a_{1n}x_n \\ a_{22}x_2 + \cdots + a_{2r}x_r &= -a_{2,r+1}x_{r+1} - \cdots - a_{2n}x_n \\ \cdots \qquad\qquad \cdots &\qquad\qquad \cdots \\ a_{rr}x_r &= -a_{r,r+1}x_{r+1} - \cdots - a_{rn}x_n \end{cases} \tag{13}$$

Каждому набору $(c_1, \ldots, c_{n-r})$ значений свободных неизвестных отвечает частное решение $X_0 = (x_1, \ldots, x_r, c_1, \ldots, c_{n-r})$ исходной системы (его удобно искать, двигаясь по системе (13) снизу вверх — из последнего уравнения находим значение неизвестной x_r, подставляем его в предыдущее уравнение, находим значение x_{r-1} и так далее). Тогда решения X_1, X_2,..., X_{n-r}, отвечающие наборам $(1, 0, \ldots, 0)$, $(0, 1, \ldots, 0)$,..., $(0, 0, \ldots, 1)$ значений свободных неизвестных, образуют ФНР. В самом деле, векторы $X_1, \ldots, X_{n-r}$ линейно независимы, так как составленная

из их координат матрица имеет ранг $n - r$ в силу того, что последние ее $n - r$ столбцов образуют единичную матрицу. Следовательно, $\mathrm{rk}\{X_1, \ldots, X_{n-r}\} = n - r$ и осталось заметить, что количество векторов $X_1, \ldots, X_{n-r}$ совпадает с размерностью пространства решений V_A.

3.4 Неоднородные системы

Вернемся к неоднородным системам общего вида. Мы уже знаем способ проверки совместности систем с помощью теоремы Кронекера–Капелли. Получим теперь формулу общего решения неоднородной системы в случае ее совместности.

Итак, пусть $Ax = b$ — совместная система линейных уравнений. Следовательно, существует удовлетворяющий ей вектор X'. Тогда для любого частного решения X_0 однородной системы $Ax = 0$ имеем

$$A(X' + X_0) = AX' + AX_0 = b + 0 = b,$$

то есть $X' + X_0$ — снова решение системы $Ax = b$. Обратно, пусть X'' — некоторое решение системы $Ax = b$. Тогда

$$A(X'' - X') = AX'' - AX' = b - b = 0,$$

значит, вектор $X_0 = X'' - X'$ есть решение однородной системы, а решение X'' исходной системы можно представить в виде $X'' = X' + X_0$. Тем самым получена формула общего решения для неоднородных систем

$$X^{\text{неодн}}_{\text{общ}} = X^{\text{неодн}}_{\text{част}} + X^{\text{одн}}_{\text{общ}}. \tag{14}$$

Следствие 3.15 *Совместная система $Ax = b$ является определенной в точности тогда, когда система $Ax = 0$ имеет только нулевое решение.*

Итак, общее решение неднородной системы есть сумма ее частного решения и общего решения соответствующей однородной системы. О решении однородных систем было рассказано выше, следовательно, осталось указать способ нахождения частного решения совместной неоднородной системы $Ax = b$.

Как и в случае однородных систем, приведя матрицу к ступенчатому виду, выделив главные и свободные неизвестные и перенумеровав их при необходимости, можно считать, что система имеет вид (4). Придав свободным неизвестным произвольные значения (как правило, для

простоты вычислений эти значения полагают равными 0) и решив получившуюся определенную систему из r уравнений, получим требуемое частное решение исходной системы.

Суммируя приведенные выше результаты, получаем следующий АЛГОРИТМ РЕШЕНИЯ НЕОДНОРОДНЫХ СИСТЕМ:

Шаг 1. С помощью теоремы Кронекера–Капелли проверяем совместность системы. Если система несовместна, то решений нет, в случае совместности переходим к шагу 2.

Шаг 2. Находим ФНР соответствующей однородной системы и по формуле (12) получаем ее общее решение $X^{\text{одн}}_{\text{общ}}$.

Шаг 3. Находим частное решение $X^{\text{неодн}}_{\text{част}}$ неоднородной системы и согласно (14) получаем ее общее решение $X^{\text{неодн}}_{\text{общ}}$.

Следующая таблица отражает зависимость числа решений системы m уравнений с n неизвестными от ее типа.

Система	Неоднородная		Однородная	
	Общая	$m < n$	Общая	$m < n$
Число решений	0,1,∞	0,∞	1,∞	∞

Таблица 2

4 Перестановки

Перейдем к изучению еще одного важного класса алгебраических объектов — перестановок. Сначала дадим несколько общих определений.

Отображение $\varphi : A \to B$ называется *инъективным* (или *инъекцией*), если $a \neq b$ влечет $\varphi(a) \neq \varphi(b)$ для всех $a, b \in A$. Другими словами, отображение инъективно, если образы различных элементов различны.

Отображение $\varphi : A \to B$ называется *сюръективным* (или *сюръекцией*), если $\varphi(A) = B$, то есть для каждого элемента $b \in B$ найдется такой элемент $a \in A$, что $\varphi(a) = b$. Инъективное и сюръективное отображение называют *биективным* (или *биекцией*). *Тождественное* отображение $id_A : A \to A$, переводящее каждый элемент из A в себя, очевидно, является биективным.

Если $\varphi : A \to B$ и $\psi : B \to C$ — отображения, то соответствие $a \mapsto \psi(\varphi(a))$ задает отображение $\psi \circ \varphi : A \to C$, называемое *композицией* отображений φ и ψ.

Лемма 4.1 *Пусть* $\alpha : A \to B$, $\beta : B \to C$ *и* $\gamma : C \to D$ *— отображения. Тогда* $(\gamma \circ \beta) \circ \alpha = \gamma \circ (\beta \circ \alpha)$.

Доказательство. Для каждого $a \in A$ имеем

$$((\gamma \circ \beta) \circ \alpha)(a) = (\gamma \circ \beta)(\alpha(a)) = \gamma(\beta(\alpha(a))) = \gamma((\beta \circ \alpha)(a)) = (\gamma \circ (\beta \circ \alpha))(a).\square$$

Если для отображения $\varphi : A \to B$ существует такое отображение $\psi : B \to A$, что $\psi \circ \varphi = id_A$ и $\varphi \circ \psi = id_B$, то ψ называется *обратным* к φ и обозначается φ^{-1}.

Упражнение 4.1

1. Приведите примеры отображений, которые инъективны, но не сюръективны; сюръективны, но не инъективны.

2. Докажите, что композиция инъективных (сюръективных) отображений инъективна (сюръективна).

3. Докажите, что отображение обладает обратным тогда и только тогда, когда оно биективно.

4. Докажите, что если множество A конечно, то инъективность отображения $\alpha : A \to A$ эквивалентна его сюръективности.

4.1 Группа перестановок

Пусть Ω_n — множество, состоящее из n элементов. *Перестановкой* на Ω_n называется произвольная биекция $\alpha : \Omega_n \to \Omega_n$. Обозначим через S_n множество всех перестановок на Ω_n. Композицию $\alpha \circ \beta$ перестановок $\alpha, \beta \in S_n$ будем называть их *произведением* и обозначать через $\alpha\beta$.

Для дальнейших рассуждений природа элементов множества Ω_n не имеет никакого значения, поэтому для определенности будем считать, что Ω_n состоит из чисел $1, 2, \ldots, n$. Каждую перестановку $\alpha \in S_n$ удобно записывать в виде таблицы, содержащей две строки и n столбцов:

$$\alpha = \begin{pmatrix} 1 & 2 & \ldots & n \\ \alpha(1) & \alpha(2) & \ldots & \alpha(n) \end{pmatrix},$$

где в первой строке перечислены элементы множества Ω_n, а во второй — их образы при действии α. Именно с такими таблицами будем в дальнейшем связывать термин "перестановка". Следует подчеркнуть, что порядок расположения столбцов таблицы, задающей перестановку, может быть произвольным.

Теорема 4.2 $(S_n, \cdot)$ — *группа.*

Доказательство. Ассоциативность произведения перестановок вытекает из леммы 4.1. Непосредственно проверяется, что единицей в S_n является тождественная перестановка $e = \begin{pmatrix} 1 & 2 & \dots & n \\ 1 & 2 & \dots & n \end{pmatrix}$, а обратным к α элементом — перестановка $\alpha^{-1} = \begin{pmatrix} \alpha(1) & \alpha(2) & \dots & \alpha(n) \\ 1 & 2 & \dots & n \end{pmatrix}$.□

Заметим, что группа S_n при $n \geq 3$ не является коммутативной. В самом деле, например, для $\alpha = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$ и $\beta = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ имеем $\alpha\beta = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$, но $\beta\alpha = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$.

Изучим более подробно строение элементов группы S_n. Каждую перестановку $\alpha \in S_n$ можно возвести в степень, причем, полагая $\alpha^0 = e$ и $\alpha^{-k} = (\alpha^{-1})^k$, можно рассматривать степени с любым целым показателем. Вполне очевидны обычные свойства степеней: $\alpha^k\alpha^l = \alpha^{k+l}$, $(\alpha^k)^l = \alpha^{kl}$.

Фиксируем перестановку $\alpha \in S_n$ и число $i \in \Omega_n$. Так как множество Ω_n конечно, то в последовательности $i, \alpha(i), \alpha^2(i), \dots$ в какой-то момент впервые возникнет число $\alpha^k(i)$, уже встречавшееся ранее, другими словами, числа $i, \alpha(i), \dots, \alpha^{k-1}(i)$ попарно различны и $\alpha^k(i) = \alpha^j(i)$ при некотором $j < k$. Заметим, что при $j > 0$ различные элементы $\alpha^{j-1}(i)$ и $\alpha^{k-1}(i)$ переходили бы под действием α в одно и то же число, что противоречит биективности α. Следовательно, $j = 0$, так что $\alpha^k(i) = \alpha^0(i) = i$. Таким образом, в последовательности $i, \alpha(i), \alpha^2(i), \dots$ числа циклически повторяются через каждые k шагов. Перестановки, циклически переставляющие заданные k чисел и не меняющие остальные числа множества Ω_n, называют *циклическими* или просто *циклами*; число k называется *длиной* цикла. Следовательно, ограничение перестановки α на множество $[i] = \{i, \alpha(i), \dots, \alpha^{k-1}(i)\}$ есть цикл длины k; обозначим его через α_i. Ясно, что если $[i] = \Omega_n$, то $\alpha = \alpha_i$.

Пусть теперь $[i] \subset \Omega_n$. Тогда фиксируем любое число $j \in \Omega_n$, не входящее в $[i]$, и построим множество $[j] = \{j, \alpha(j), \ldots, \alpha^{l-1}(j)\}$, элементы которого также циклически переходят друг в друга под действием α, и поэтому $[i] \cap [j] = \emptyset$. Соответствующий цикл длины l обозначим через α_2. Легко понять, что если $[i] \cup [j] = \Omega_n$, то $\alpha = \alpha_i \alpha_j$. В противном случае фиксируем не входящее в $[i] \cup [j]$ число $m \in \Omega_n$ и, повторяя описанный выше процесс, построим не пересекающееся с $[i] \cup [j]$ множество $[m]$ и цикл α_m и так далее. Ясно, что в силу конечности множества Ω_n через конечное число шагов будет получено разбиение $\Omega_n = [i] \cup [j] \cup \ldots \cup [t]$ и разложение $\alpha = \alpha_i \alpha_j \ldots \alpha_t$. Поскольку подмножества $[i], [j], \ldots, [t]$ попарно не пересекаются, циклы $\alpha_i, \alpha_j, \ldots, \alpha_t$ называют *независимыми*. Таким образом, верна

Теорема 4.3 *Любая перестановка представима в виде произведения независимых циклов.*

Цикл длины k, переводящий j_1 в j_2, j_2 в j_3, $\ldots$, j_{k-1} в j_k и j_k снова в j_1, удобно записывать в виде $(j_1, j_2, \ldots, j_k)$. Циклы длины 2 называют *транспозициями*.

Следствие 4.4 *Любая перестановка представима в виде произведения транспозиций.*

Доказательство. С учетом теоремы 4.3 достаточно разложить в произведение транспозиций произвольный цикл. Для упрощения обозначений можно считать, что циклически переставляются числа 1,2,$\ldots$, k. Непосредственно проверяется, что

$$(1, 2, \ldots, k) = (1, k)(1, k-1) \ldots (1, 3)(1, 2). \square$$

4.2 Действие перестановок на функциях n переменных. Четность перестановки

Пусть X — непустое множество. Для каждой функции $f : X^n \to \mathbb{R}$ и каждой перестановки $\alpha \in S_n$ определим новую функцию $\alpha \circ f$ по правилу: $(\alpha \circ f)(x_1, \ldots, x_n) = f(x_{\alpha(1)}, \ldots, x_{\alpha(n)})$. Легко проверяются свойства:

1. $e \circ f = f$, где $e \in S_n$ — тождественная перестановка.
2. $\alpha \circ (\beta \circ f) = (\alpha\beta) \circ f$ для всех $\alpha, \beta \in S_n$.

(Используя терминологию теории групп, говорят, что группа перестановок *действует* на множестве функций.)

Функция f называется *кососимметрической*, если $\alpha \circ f = -f$ для любой транспозиции α вида $(i, i+1)$, иначе говоря, кососимметрическая функция при перестановке соседних аргументов меняет знак. Простейшим примером кососимметрической функции является функция, тождественно равная нулю. Следующий пример показывает, что существуют и нетривиальные кососимметрические функции.

Пример 4.5 Рассмотрим функцию $\Delta(x_1, \dots, x_n) = \prod\limits_{1 \le i < j \le n} (x_j - x_i)$, где $x_i \in \mathbb{R}$ для всех i. Фиксируем k в интервале от 1 до $n-1$ и разобьем произведение $\frac{n(n-1)}{2}$ входящих в Δ разностей на блоки:

$$\Delta(x_1, \dots, x_n) =$$

$$\overbrace{\prod_{1 \le i < j < k} (x_j - x_i)}^{A} \overbrace{\prod_{i=1}^{k-1} (x_k - x_i)}^{B} \overbrace{\prod_{i=1}^{k-1} (x_{k+1} - x_i)}^{C} (x_{k+1} - x_k) \overbrace{\prod_{\substack{1 \le i < j \le n, \\ j > k+1}} (x_j - x_i)}^{D}.$$

Очевидно, что действие транспозиции $(k, k+1)$ не изменит блоки A и D, блок B переведет в C и наоборот, а у разности $x_{k+1} - x_k$ поменяет знак. Следовательно, $(k, k+1) \circ f = -f$, так что функция Δ — кососимметрическая. Ясно, что $\Delta(x_1, \dots, x_n) \neq 0$, в случае, когда значения всех ее аргументов различны.

Лемма 4.6 *Если функция f — кососимметрическая, то $\alpha \circ f = -f$ для любой транспозиции $\alpha \in S_n$.*

Доказательство. Пусть $\alpha = (i, i+k)$ — произвольная транспозиция. Проведем доказательство индукцией по разности переставляемых чисел.

При $k = 1$ доказываемое утверждение совпадает с определением кососимметрической функции. Предположим, что утверждение уже доказано для всех транспозиций с разностью, строго меньшей k. Тогда

$$(\alpha \circ f)(\dots, x_i, x_{i+1}, \dots, x_{i+k}, \dots) = f(\dots, x_{i+k}, x_{i+1}, \dots, x_i, \dots) =$$
$$-f(\dots, x_{i+1}, x_{i+k}, \dots, x_i, \dots) = f(\dots, x_{i+1}, x_i, \dots, x_{i+k}, \dots) =$$
$$-f(\dots, x_i, x_{i+1}, \dots, x_{i+k}, \dots). \square$$

Теорема 4.7 *Каждой перестановке $\alpha \in S_n$ отвечает такое число $\varepsilon_\alpha = \pm 1$, называемое ее четностью, что $\alpha \circ f = \varepsilon_\alpha f$ для любой кососимметрической функции f от n переменных. При этом $\varepsilon_{\alpha\beta} = \varepsilon_\alpha \varepsilon_\beta$.*

Доказательство. Фиксируем неравную тождественно нулю кососимметрическую функцию f от n переменных. (Существование таких функций установлено в примере 4.5.) Разложим перестановку $\alpha \in S_n$ в произведение транспозиций: $\alpha = \tau_1 \dots \tau_k$. Тогда с учетом леммы 4.6 имеем $\alpha \circ f = \varepsilon_\alpha f$, где $\varepsilon_\alpha = (-1)^k$. Если $\alpha = \sigma_1 \dots \sigma_m$ — еще одно разложение в произведение транспозиций, то $\alpha \circ f = (-1)^m f$, следовательно, $(\alpha \circ f)(x_1, \dots, x_n) = (-1)^m f(x_1, \dots, x_n)$ для всех $x_1, \dots, x_n$. Выберем значения $a_1, \dots, a_n$ переменных так, чтобы $f(a_1, \dots, a_n) \neq 0$. Тогда

$$(-1)^m f(a_1, \dots, a_n) = (\alpha \circ f)(a_1, \dots, a_n) = \varepsilon_\alpha f(a_1, \dots, a_n),$$

откуда, сокращая на $f(a_1, \dots, a_n)$, получаем $(-1)^m = \varepsilon_\alpha$. Тем самым доказано, что число ε_α однозначно определяется самой перестановкой α и не зависит от способа ее разложения в произведение транспозиций. В частности,

$$\varepsilon_{\alpha\beta} f = (\alpha\beta) \circ f = \alpha \circ (\beta \circ f) = \alpha \circ (\varepsilon_\beta f) = \varepsilon_\alpha \varepsilon_\beta f,$$

так что $\varepsilon_{\alpha\beta} = \varepsilon_\alpha \varepsilon_\beta$. □

Следствие 4.8 *Если $\alpha = \alpha_1 \dots \alpha_k$ — разложение в произведение независимых циклов, то $\varepsilon_\alpha = (-1)^{\sum_{s=1}^{k} (l_s - 1)} = (-1)^{\sum_{s=1}^{k} l_s - k}$, где $l_1, \dots, l_k$ — длины циклов.*

Доказательство вытекает из предыдущей теоремы и указанного в доказательстве следствия 4.4 способа разложения цикла в произведение транспозиций. □

Приведем еще один способ вычисления четности перестановки α. Говорят, что числа i и j *образуют инверсию* относительно α, если $i < j$, но $\alpha(i) > \alpha(j)$.

Теорема 4.9 *Для любой перестановки α верно $\varepsilon_\alpha = (-1)^k$, где k — число всех инверсий относительно α.*

Доказательство. Применим метод математической индукции по числу k инверсий.

Если $k = 0$, то α — тождественная перестановка и утверждение тривиально верно.

Пусть $k \geq 1$ и для перестановок с менее чем k инверсиями, теорема уже доказана. Покажем сначала, что среди чисел, образующих инверсии,

есть соседние числа. В самом деле, пусть инверсию образуют числа i и $i+l$. Если $l=1$, то нужные соседние числа найдены. При $l>1$ рассмотрим число $\alpha(i+1)$. Если $\alpha(i)>\alpha(i+1)$, то инверсию образуют соседние числа i и $i+1$. Если же $\alpha(i)<\alpha(i+1)$, то $\alpha(i+1)>\alpha(i)>\alpha(i+l)$, следовательно, инверсию образуют числа $i+1$ и $i+l$, находящиеся на одну позицию ближе друг к другу, чем i и $i+l$. Ясно, что через конечное число шагов нужная пара соседних чисел будет найдена.

Итак, пусть i и $i+1$ образуют инверсию. Рассмотрим перестановку $\alpha'=\tau\alpha$, где $\tau=(\alpha(i),\alpha(i+1))$. Нетрудно видеть, что α' содержит ровно на одну инверсию меньше, чем α, значит, по предположению индукции верно $\varepsilon_{\alpha'}=(-1)^{k-1}$. Но тогда из равенства $\alpha=\tau\alpha'$ выводим $\varepsilon_\alpha=\varepsilon_\tau\varepsilon_{\alpha'}=(-1)(-1)^{k-1}=(-1)^k$.$\square$

5 Определители

Пусть $A=(a_{ij})$ — матрица порядка n над полем K. Сопоставим ей элемент поля, вычисляемый по формуле

$$\det A=\sum_{\alpha\in S_n}\varepsilon_\alpha a_{1\alpha(1)}a_{2\alpha(2)}\dots a_{n\alpha(n)} \tag{15}$$

и называемый ее *определителем*. Также для обозначения определителя используется запись матрицы в прямых скобках. Выпишем формулу (15), называемую *формулой полного развертывания*, в явном виде при малых значениях n:

$n=1$. $\det A=a_{11}$.

$$n=2.\begin{vmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\end{vmatrix}=a_{11}a_{22}-a_{12}a_{21}.$$

$$n=3.\begin{vmatrix} a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{vmatrix}=\begin{array}{l} a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-\\ \qquad a_{13}a_{22}a_{31}-a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}.\end{array}$$

При $n>3$ количество слагаемых в правой части (15) быстро возрастает (при $n=4$ формула содержит 24 слагаемых, при $n=5$ — уже 120), поэтому для вычисления определителей больших порядков, как правило, используются специальные приемы, опирающиеся на различные свойства определителей, о которых пойдет речь ниже.

5.1 Определитель транспонированной матрицы

Следующая теорема показывает, что определитель матрицы не меняется при ее транспонировании:

Теорема 5.1 $\det A^t = \det A$.

Доказательство. Пусть $A = (a_{ij})$, $A^t = (a'_{ij})$, где $a'_{ij} = a_{ji}$ при всех i, j. Тогда согласно (15) имеем

$$\det A^t = \sum_{\alpha \in S_n} \varepsilon_\alpha a'_{1\alpha(1)} a'_{2\alpha(2)} \dots a'_{n\alpha(n)} = \sum_{\alpha \in S_n} \varepsilon_\alpha a_{\alpha(1)1} a_{\alpha(2)2} \dots a_{\alpha(n)n}.$$

Заметим, что пары индексов элементов, участвующих в произведении $a_{\alpha(1)1} a_{\alpha(2)2} \dots a_{\alpha(n)n}$, отвечают перестановке, обратной к α, поэтому с учетом коммутативности умножения элементов поля получаем $a_{\alpha(1)1} a_{\alpha(2)2} \dots a_{\alpha(n)n} = a_{1\alpha^{-1}(1)} a_{2\alpha^{-1}(2)} \dots a_{n\alpha^{-1}(n)}$, откуда

$$\det A^t = \sum_{\alpha \in S_n} \varepsilon_\alpha a_{1\alpha^{-1}(1)} a_{2\alpha^{-1}(2)} \dots a_{n\alpha^{-1}(n)}. \tag{16}$$

Воспользуемся теперь тем, что отображение $\alpha \mapsto \alpha^{-1}$ является биекцией группы перестановок S_n в себя (проверьте это в качестве упражнения), следовательно, если перестановка α пробегает всю группу, то всю группу пробегает и перестановка α^{-1}, поэтому в (16) можно заменить индекс суммирования:

$$\sum_{\alpha \in S_n} \varepsilon_\alpha a_{1\alpha^{-1}(1)} a_{2\alpha^{-1}(2)} \dots a_{n\alpha^{-1}(n)} = \sum_{\alpha^{-1} \in S_n} \varepsilon_\alpha a_{1\alpha^{-1}(1)} a_{2\alpha^{-1}(2)} \dots a_{n\alpha^{-1}(n)}.$$

Поскольку $\alpha\alpha^{-1} = e$, то $1 = \varepsilon_e = \varepsilon_{\alpha\alpha^{-1}} = \varepsilon_\alpha \varepsilon_{\alpha^{-1}}$, откуда с учетом $\varepsilon_\alpha = \pm 1$ выводим $\varepsilon_\alpha = \varepsilon_{\alpha^{-1}}$. Обозначив α^{-1} через β, окончательно получаем

$$\det A^t = \sum_{\beta \in S_n} \varepsilon_\beta a_{1\beta(1)} a_{2\beta(2)} \dots a_{n\beta(n)} = \det A. \square$$

5.2 Определитель, как полилинейная кососимметрическая функция строк (столбцов) матрицы

Как уже не раз отмечалось, каждую матрицу A порядка n можно рассматривать как набор $(A_{(1)}, \dots, A_{(n)})$ n-мерных векторов — строк матрицы, либо как набор $(A^{(1)}, \dots, A^{(n)})$ векторов — столбцов. Поэтому

на определитель матрицы можно смотреть как на функцию от n аргументов. При транспонировании матрицы ее строки становятся столбцами и наоборот, столбцы — строками, определитель же согласно теореме 5.1 не меняется. Следовательно, любое свойство, доказанное для определителя, рассматриваемого как функция строк, автоматически будет верным и в том случае, когда определитель рассматривается в качестве функции столбцов.

Пусть V — векторное пространство над полем K. (Будем считать, что K — одно из полей $\mathbb{Q}$, $\mathbb{R}$ или $\mathbb{C}$.) Функция $f : V^n \to K$ называется *полилинейной*, если для каждого $i = 1, \dots, n$ справедливо равенство

$$f(a_1, \dots, \lambda a'_i + \mu a''_i, \dots, a_n) = \lambda f(a_1, \dots, a'_i, \dots, a_n) + \mu f(a_1, \dots, a''_i, \dots, a_n),$$

то есть функция f линейна по каждому аргументу.

D1. $\det(A_{(1)}, \dots, A_{(n)})$ — *полилинейная функция.*

Доказательство вытекает из цепочки равенств

$$\det(A_{(1)}, \dots, \lambda A'_{(i)} + \mu A''_{(i)}, \dots, A_{(n)}) = \sum_{\alpha \in S_n} \varepsilon_\alpha a_{1\alpha(1)} \dots (\lambda a'_{i\alpha(i)} + \mu a''_{i\alpha(i)}) \dots a_{n\alpha(n)} =$$
$$\lambda \sum_{\alpha \in S_n} \varepsilon_\alpha a_{1\alpha(1)} \dots a'_{i\alpha(i)} \dots a_{n\alpha(n)} + \mu \sum_{\alpha \in S_n} \varepsilon_\alpha a_{1\alpha(1)} \dots a''_{i\alpha(i)} \dots a_{n\alpha(n)} =$$
$$\lambda \det(A_{(1)}, \dots, A'_{(i)}, \dots, A_{(n)}) + \mu \det(A_{(1)}, \dots, A''_{(i)}, \dots, A_{(n)}). \square$$

D2. $\det(A_{(1)}, \dots, A_{(n)})$ — *кососимметрическая функция.*

Доказательство. Обозначим через τ транспозицию $(i, i{+}1)$. Для любой перестановки α верно $\varepsilon_{\alpha\tau} = \varepsilon_\alpha \varepsilon_\tau = -\varepsilon_\alpha$. Следовательно,

$$\det(\dots, A_{(i+1)}, A_{(i)}, \dots) = \sum_{\alpha \in S_n} \varepsilon_\alpha a_{1,\,\alpha(1)} \dots a_{i,\,\alpha(i+1)} a_{i+1,\,\alpha(i)} \dots a_{n,\,\alpha(n)} =$$

$$- \sum_{\alpha \in S_n} \varepsilon_{\alpha\tau} a_{1,(\alpha\tau)(1)} \dots a_{i,(\alpha\tau)(i)} a_{i+1,(\alpha\tau)(i+1)} \dots a_{n,(\alpha\tau)(n)}.$$

Легко проверяется, что отображение $\alpha \mapsto \alpha\tau$ является биекцией группы S_n в себя, поэтому в последнем выражении можно заменить индекс суммирования α на $\beta = \alpha\tau$, так что

$$\det(\dots, A_{(i+1)}, A_{(i)}, \dots) = - \sum_{\beta \in S_n} \varepsilon_\beta a_{1,\,\beta(1)} \dots a_{i,\,\beta(i)} a_{i+1,\,\beta(i+1)} \dots a_{n,\,\beta(n)} =$$

$$-\det(\dots, A_{(i)}, A_{(i+1)}, \dots).\square$$

Замечание. Ввиду леммы 4.6 свойство D2 означает, что если в матрице поменять местами две строки, то ее определитель поменяет знак.

D3. $\det E = 1$.

Доказательство. Согласно (15) имеем

$$\det E = \sum_{\alpha \in S_n} \varepsilon_\alpha \delta_{1\alpha(1)} \dots \delta_{n\alpha(n)}.$$

Если $\alpha(i) \neq i$ для некоторого i, то $\delta_{i\alpha(i)} = 0$, следовательно, равно нулю и все содержащее $\delta_{i\alpha(i)}$ произведение, поэтому под знаком суммы в последнем выражении остается только произведение, отвечающее тождественной перестановке, то есть $\det E = \varepsilon_e \delta_{11} \dots \delta_{nn}$. Но $\varepsilon_e = 1$ и $\delta_{ii} = 1$ при любом i, значит, $\det E = 1.\square$

5.3 Свойства полилинейных кососимметрических функций

Пусть $\mathcal{D}$ — произвольная полилинейная кососимметрическая функция строк матрицы A, то есть функция, обладающая свойствами D1 и D2.

D4. $\mathcal{D}(\lambda A) = \lambda^n \mathcal{D}(A)$.

Доказательство. Воспользовавшись n раз свойством D1, получаем

$$\mathcal{D}(\lambda A) = \mathcal{D}(\lambda A_{(1)}, \dots, \lambda A_{(n)}) = \lambda^n \mathcal{D}(A_{(1)}, \dots, A_{(n)}) = \lambda^n \mathcal{D}(A).\square$$

D5. *Если A содержит нулевую строку, то $\mathcal{D}(A) = 0$.*

Доказательство. Пусть $A_{(i)} = 0$. Тогда $A_{(i)} = 0 \cdot A_{(i)}$, следовательно, в силу D1 имеем

$$\mathcal{D}(A) = \mathcal{D}(A_{(1)}, \dots, A_{(i)}, \dots, A_{(n)}) = \mathcal{D}(A_{(1)}, \dots, 0 \cdot A_{(i)}, \dots, A_{(n)}) = 0 \cdot \mathcal{D}(A) = 0.\square$$

D6. *Если матрица A содержит две одинаковых строки, то $\mathcal{D}(A) = 0$.*

Доказательство. Пусть $i \neq j$ и $A_{(i)} = A_{(j)}$. Согласно D2, если в матрице A поменять местами i-ю и j-ю строки, то $\mathcal{D}(A)$ поменяет знак. С другой стороны, от перестановки одинаковых строк матрица A не изменится. Следовательно, $\mathcal{D}(A) = -\mathcal{D}(A)$, откуда, $\mathcal{D}(A) = 0.\square$

D7. *При элементарных преобразованиях строк II-го рода матрицы A значение $\mathcal{D}(A)$ не меняется.*

Доказательство. Воспользуемся свойствами D1 и D6:

$$\mathcal{D}(\dots, A_{(i)}, \dots, \lambda A_{(i)} + A_{(j)}, \dots) = \lambda \mathcal{D}(\dots, A_{(i)}, \dots, A_{(i)}, \dots) +$$

$$\mathcal{D}(\dots, A_{(i)}, \dots, A_{(j)}, \dots) = \lambda \cdot 0 + \mathcal{D}(A) = \mathcal{D}(A). \square$$

Подчеркнем, что свойства D4–D7 справедливы для всех полилинейных кососимметрических функций строк матриц, а значит, справедливы для определителя как частного случая таких функций. На самом деле между определителями и полилинейными кососимметрическими функциями существует и обратная связь: каждую полилинейную кососимметрическую функцию строк матрицы можно выразить через ее определитель. Предварительно докажем вспомогательную лемму.

Лемма 5.2 *Если $A = (a_{ij})$ — верхнетреугольная матрица порядка n, то $\mathcal{D}(A) = a_{11}a_{22}\dots a_{nn}\mathcal{D}(E)$.*

Доказательство. Последняя строка матрицы A имеет вид $A_{(n)} = (0, 0, \dots, a_{nn})$. Поскольку $A_{(n)} = a_{nn}(0, 0, \dots, 1)$, то согласно D1

$$\mathcal{D}(A) = a_{nn}\mathcal{D}\begin{pmatrix} a_{11} & \dots & a_{1,n-1} & a_{1n} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & \dots & 0 & 1 \end{pmatrix}. \qquad (17)$$

Легко видеть, что элементарными преобразованиями II-го рода строк полученной в (17) матрицы ее последний столбец можно привести к виду $(0, \dots, 0, 1)^t$, поэтому в силу D7

$$\mathcal{D}(A) = a_{nn}\mathcal{D}\begin{pmatrix} a_{11} & \dots & a_{1,n-1} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & a_{n-1,n-1} & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}.$$

Повторив аналогичные преобразования матрицы для строк с номерами $n-1$,..., 1, получим требуемое равенство. $\square$

Теорема 5.3 *Если $\mathcal{D}$ — полилинейная кососимметрическая функция строк матрицы A, то $\mathcal{D}(A) = \mathcal{D}(E)\det A$.*

Доказательство. С помощью элементарных преобразований I-го и II-го рода приведем матрицу A к ступенчатому виду $\tilde{A}$. Пусть при этом было выполнено k преобразований I-го рода. В силу D7 преобразования II-го рода не меняют значения $\mathcal{D}(A)$, а каждое преобразование I-го рода согласно D2 меняет знак, так что

$$\mathcal{D}(A) = (-1)^k \mathcal{D}(\tilde{A}).$$

Матрица $\tilde{A} = (\tilde{a}_{ij})$ имеет ступенчатый вид (см. (1)), следовательно, она является верхнетреугольной матрицей. Тогда по лемме 5.2 имеем $\mathcal{D}(\tilde{A}) = \tilde{a}_{11} \dots \tilde{a}_{nn} \mathcal{D}(E)$, откуда

$$\mathcal{D}(A) = \mathcal{D}(E)(-1)^k \tilde{a}_{11} \dots \tilde{a}_{nn}. \tag{18}$$

Заметим, что приведенные выше рассуждения остаются верными, если функцию $\mathcal{D}$ заменить определителем, при этом (18) с учетом D3 примет вид

$$\det A = (-1)^k \tilde{a}_{11} \dots \tilde{a}_{nn},$$

следовательно, $\mathcal{D}(A) = \mathcal{D}(E)(-1)^k \tilde{a}_{11} \dots \tilde{a}_{nn} = \mathcal{D}(E) \det A$. $\square$

В качестве применения теоремы 5.3 докажем два следующих утверждения об определителях.

Теорема 5.4 *Пусть* $A = \begin{pmatrix} B & D \\ 0 & C \end{pmatrix}$ *— блочно-верхнетреугольная матрица, причем блоки* B *и* C *— квадратные. Тогда* $\det A = \det B \, \det C$.

Доказательство. Фиксируем блоки B, D и рассмотрим функцию $\mathcal{D}(C) = \begin{vmatrix} B & D \\ 0 & C \end{vmatrix}$. Обозначим через k и l порядки блоков B и C соответственно. Очевидно, если $C_{(i)} = \lambda C'_{(i)} + \mu C''_{(i)}$, то $A_{(k+i)} = \lambda A'_{(k+i)} + \mu A''_{(k+i)}$, поэтому

$$\begin{gathered}
\mathcal{D}(C_{(1)}, \dots, \lambda C'_{(i)} + \mu C''_{(i)}, \dots, C_{(l)}) = \\
\det(A_{(1)}, \dots, \lambda A'_{(k+i)} + \mu A''_{(k+i)}, \dots, A_{(k+l)}) = \\
\lambda \det(A_{(1)}, \dots, A'_{(k+i)}, \dots, A_{(k+l)}) + \mu \det(A_{(1)}, \dots, A''_{(k+i)}, \dots, A_{(k+l)}) = \\
\lambda \mathcal{D}(C_{(1)}, \dots, C'_{(i)}, \dots, C_{(l)}) + \mu \mathcal{D}(C_{(1)}, \dots, C''_{(i)}, \dots, C_{(l)}),
\end{gathered}$$

следовательно, функция $\mathcal{D}$ — полилинейная. Аналогично,

$$\mathcal{D}(\dots, C_{(i+1)}, C_{(i)}, \dots) = \det(\dots, A_{(k+i+1)}, A_{(k+i)}, \dots) =$$

$$-\det(\dots, A_{(k+i)}, A_{(k+i+1)}, \dots) = -\mathcal{D}(\dots, C_{(i)}, C_{(i+1)}, \dots),$$

так что $\mathcal{D}$ — кососимметрическая. Тогда по теореме 5.3

$$\mathcal{D}(C) = \mathcal{D}(E)\det C.$$

Осталось показать, что $\mathcal{D}(E) = \det B$. Положим $\tilde{A} = \begin{pmatrix} B & D \\ 0 & E \end{pmatrix}$. Имеем

$$\mathcal{D}(E) = \det \tilde{A} = \sum_{\alpha \in S_{k+l}} \varepsilon_\alpha \tilde{a}_{1,\alpha(1)} \dots \tilde{a}_{k+l,\alpha(k+l)}. \tag{19}$$

Заметим, что $\tilde{a}_{ij} = \delta_{ij}$ при $i > k$, поэтому можно считать, что суммирование в (19) ведется только по тем перестановкам α, для которых $\alpha(k+1) = k+1, \dots, \alpha(k+l) = k+l$, то есть, фактически, по перестановкам β чисел 1, 2,..., k. Полагая $\beta(i) = \alpha(i)$ при $i \le k$, с учетом $\tilde{a}_{ij} = b_{ij}$ $(i, j \le k)$ получаем

$$\sum_{\alpha \in S_{k+l}} \varepsilon_\alpha \tilde{a}_{1,\alpha(1)} \dots \tilde{a}_{k+l,\alpha(k+l)} = \sum_{\beta \in S_k} \varepsilon_\beta \tilde{a}_{1,\beta(1)} \dots \tilde{a}_{k,\beta(k)} \delta_{k+1,k+1} \dots \delta_{k+l,k+l} =$$

$$\sum_{\beta \in S_k} \varepsilon_\beta b_{1,\beta(1)} \dots b_{k,\beta(k)} = \det B.$$

Подставив полученное выражение в (19), выводим требуемое равенство $\mathcal{D}(E) = \det B$.□

Теорема 5.5 *Пусть A, B — квадратные матрицы одного порядка. Тогда* $\det(AB) = \det A \det B$.

Доказательство. Фиксируем матрицу B и рассмотрим функцию $\mathcal{D}(A) = \det(AB)$. Легко видеть, что $(AB)_{(i)} = A_{(i)}B$ для всех i. Тогда

$$\mathcal{D}(A_{(1)}, \dots, \lambda A'_{(i)} + \mu A''_{(i)}, \dots, A_{(n)}) =$$
$$\det(A_{(1)}B, \dots, \lambda A'_{(i)}B + \mu A''_{(i)}B, \dots, A_{(n)}B) =$$
$$\lambda \det(A_{(1)}B, \dots, A'_{(i)}B, \dots, A_{(n)}B) + \mu \det(A_{(1)}B, \dots, A''_{(i)}B, \dots, A_{(n)}B) =$$
$$\lambda \mathcal{D}(A_{(1)}, \dots, A'_{(i)}, \dots, A_{(n)}) + \mu \mathcal{D}(A_{(1)}, \dots, A''_{(i)}, \dots, A_{(n)}),$$

так что функция $\mathcal{D}$ — полилинейная. Аналогично,

$$\mathcal{D}(\dots, A_{(i+1)}, A_{(i)}, \dots) = \det(\dots, A_{(i+1)}B, A_{(i)}B, \dots) =$$
$$-\det(\dots, A_{(i)}B, A_{(i+1)}B, \dots) = -\mathcal{D}(\dots, A_{(i)}, A_{(i+1)}, \dots),$$

значит, $\mathcal{D}$ — кососимметрическая. Ввиду теоремы 5.3

$$\det(AB) = \mathcal{D}(A) = \mathcal{D}(E)\det A = \det(EB)\det A = \det A \det B.\square$$

5.4 Разложение определителя по строке (столбцу)

Пусть $A = (a_{ij})$ матрица порядка n. *Дополняющим минором* M_{ij} элемента a_{ij} называется определитель матрицы, полученной вычеркиванием из матрицы A i-й строки и j-го столбца. Число $A_{ij} = (-1)^{i+j} M_{ij}$ называется *алгебраическим дополнением* элемента a_{ij}.

Теорема 5.6 *Пусть $A = (a_{ij})$ — матрица порядка n. Для всех i справедливы равенства*

$$\det A = \sum_{k=1}^{n} (-1)^{i+k} a_{ik} M_{ik} = \sum_{k=1}^{n} a_{ik} A_{ik}, \tag{20}$$

$$\det A = \sum_{k=1}^{n} (-1)^{k+i} a_{ki} M_{ki} = \sum_{k=1}^{n} a_{ki} A_{ki}, \tag{21}$$

называемые формулами разложения определителя по строке и, соответственно, столбцу.

Доказательство. Сначала докажем (21). Представим i-й столбец матрицы A в виде суммы n столбцов:

$$A^{(i)} = (a_{1i}\, 0 \ldots 0)^t + \cdots + (0\, 0 \ldots a_{ni})^t.$$

Тогда

$$\det A = \sum_{k=1}^{n} \begin{vmatrix} a_{11} & \ldots & 0 & \ldots & a_{1n} \\ & \ldots & & \ldots & \\ a_{k1} & \ldots & a_{ki} & \ldots & a_{kn} \\ & \ldots & & \ldots & \\ a_{n1} & \ldots & 0 & \ldots & a_{nn} \end{vmatrix}.$$

Вычислим отдельно k-е слагаемое получившейся суммы. Последовательно меняя местами i-й столбец с каждым предыдущим столбцом, переставим его на место первого столбца. При этом мы $i-1$ раз применили элементарное преобразование I-го рода, поэтому

$$\begin{vmatrix} a_{11} & \ldots & 0 & \ldots & a_{1n} \\ & \ldots & & \ldots & \\ a_{k1} & \ldots & a_{ki} & \ldots & a_{kn} \\ & \ldots & & \ldots & \\ a_{n1} & \ldots & 0 & \ldots & a_{nn} \end{vmatrix} = (-1)^{i-1} \begin{vmatrix} 0 & a_{11} & \ldots & a_{1n} \\ & \ldots & \ldots & \\ a_{ki} & a_{k1} & \ldots & a_{kn} \\ & \ldots & \ldots & \\ 0 & a_{n1} & \ldots & a_{nn} \end{vmatrix}.$$

Теперь аналогичным образом переместим k-ю строку на место первой строки:

$$(-1)^{i-1}\begin{vmatrix} 0 & a_{11} & \dots & a_{1n} \\ & \dots & \dots & \\ a_{ki} & a_{k1} & \dots & a_{kn} \\ & \dots & \dots & \\ 0 & a_{n1} & \dots & a_{nn} \end{vmatrix} = (-1)^{i-1+k-1}\begin{vmatrix} a_{ki} & a_{k1} & \dots & a_{kn} \\ 0 & a_{11} & \dots & a_{1n} \\ & \dots & \dots & \\ 0 & a_{n1} & \dots & a_{nn} \end{vmatrix} =$$

$$(-1)^{k+i}\begin{vmatrix} a_{ki} & a_{k1} & \dots & a_{kn} \\ 0 & a_{11} & \dots & a_{1n} \\ & \dots & \dots & \\ 0 & a_{n1} & \dots & a_{nn} \end{vmatrix}.$$

Заметим, что матрица, стоящая под знаком определителя в правой части последней цепочки равенств, имеет блочно-верхнетреугольный вид $\begin{pmatrix} B & D \\ 0 & C \end{pmatrix}$, где $B = (a_{ki})$ — матрица порядка 1, а блок C получается вычеркиванием из матрицы A k-й строки и i-го столбца. Следовательно, согласно теореме 5.4

$$\begin{vmatrix} a_{ki} & a_{k1} & \dots & a_{kn} \\ 0 & a_{11} & \dots & a_{1n} \\ & \dots & \dots & \\ 0 & a_{n1} & \dots & a_{nn} \end{vmatrix} = \det B \, \det C = a_{ki}M_{ki},$$

так что окончательно получаем

$$\det A = \sum_{k=1}^{n}\begin{vmatrix} a_{11} & \dots & 0 & \dots & a_{1n} \\ & \dots & & \dots & \\ a_{k1} & \dots & a_{ki} & \dots & a_{kn} \\ & \dots & & \dots & \\ a_{n1} & \dots & 0 & \dots & a_{nn} \end{vmatrix} = \sum_{k=1}^{n}(-1)^{k+i}a_{ki}M_{ki} = \sum_{k=1}^{n} a_{ki}A_{ki},$$

то есть, верно (21).

Легко видеть, что если в матрице A вычеркнуть i-ю строку и j-й столбец, а в матрице A^t — j-ю строку и i-й столбец, то получившиеся матрицы порядка $n-1$ являются транспонированными друг к другу, следовательно, их определители равны. Таким образом, минор M_{ij} матрицы

A равен минору M'_{ji} матрицы $A^t = (a'_{ij})$. Применив (21) к определителю матрицы A^t, получаем

$$\det A = \det A^t = \sum_{k=1}^{n} (-1)^{k+i} a'_{ki} M'_{ki} = \sum_{k=1}^{n} (-1)^{i+k} a_{ik} M_{ik} = \sum_{k=1}^{n} a_{ik} A_{ik},$$

то есть справедливо (20).□

5.5 Применение определителей

Воспользуемся полученными выше результатами об определителях для решения следующих задач.

1. *Критерий обратимости матрицы в терминах определителя.*

Пусть $A = (a_{ij})$ — матрица порядка n. Для каждого элемента матрицы A найдем его алгебраическое дополнение и составим из них *присоединенную* матрицу

$$A^{\vee} = \begin{pmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{pmatrix}.$$

Лемма 5.7 $AA^{\vee} = (\det A)E$.

Доказательство. С учетом формулы (20) имеем

$$(AA^{\vee})_{ii} = \sum_{k=1}^{n} a_{ik} A_{ik} = \det A.$$

Пусть теперь $i \neq j$. Обозначим через $\bar{A}$ матрицу, которая получается, если в A j-ю строку заменить на i-ю. Так как $\bar{A}$ содержит две одинаковых строки, ее определитель равен 0. С другой стороны, разложив $\det \bar{A}$ по j-й строке, получаем

$$\det \bar{A} = \sum_{k=1}^{n} \bar{a}_{jk} A_{jk} = \sum_{k=1}^{n} a_{ik} A_{jk} = (AA^{\vee})_{ij},$$

так что $(AA^{\vee})_{ij} = 0$ при $i \neq j$. Следовательно, $AA^{\vee} = (\det A)E$.□

Теорема 5.8 (критерий обратимости) *Матрица A обратима тогда и только тогда, когда* $\det A \neq 0$. *Если* $\det A \neq 0$, *то*

$$A^{-1} = \frac{1}{\det A} A^{\vee}. \tag{22}$$

Доказательство. Если матрица A обратима, то $AB = E$ для некоторой матрицы B. Тогда в силу теоремы 5.5

$$1 = \det E = \det(AB) = \det A \det B,$$

следовательно, $\det A \neq 0$.

Теперь докажем обратное утверждение. Пусть $\det A \neq 0$. Положим $B = \dfrac{1}{\det A} A^{\vee}$. Тогда с учетом леммы 5.7

$$AB = A\Big(\frac{1}{\det A} A^{\vee}\Big) = \frac{1}{\det A}(AA^{\vee}) = E,$$

то есть A обратима справа. Согласно теореме 1.6 для квадратной матрицы одно- и двусторонняя обратимость равносильны, поэтому A обратима и $A^{-1} = B = \dfrac{1}{\det A} A^{\vee}$.$\square$

Из теорем 5.8 и 3.9 немедленно вытекает

Следствие 5.9 *Пусть A — матрица порядка n. Тогда* $\det A \neq 0 \Leftrightarrow \operatorname{rk} A = n$.

2. *Формулы Крамера.*

Рассмотрим систему линейных уравнений, записанную в матричном виде $Ax = b$. Как известно (см. п. 1.3), единственное решение данной системы в случае обратимости матрицы A находится по формуле $x = A^{-1}b$. Воспользовавшись формулой (22) для обратной матрицы, выводим

$$x = \frac{1}{\det A} A^{\vee} b. \tag{23}$$

Обозначим через $\triangle$ определитель матрицы A, а через $\triangle_i$ $(i = 1, \ldots, n)$ — определители матриц, которые получаются, если в матрице A заменить i-й столбец столбцом b. В частности, производя разложение определителя $\triangle_i$ по i-му столбцу, имеем

$$\triangle_i = \sum_k b_k A_{ki}.$$

Переходя в (23) к поэлементной записи, получаем

$$x_i = \frac{1}{\triangle} \sum_k b_k A_{ki} = \frac{\triangle_i}{\triangle}, \quad i = 1, \ldots, n. \tag{24}$$

Равенства (24) называются **формулами Крамера**.

3. *Метод окаймляющих миноров.*

Пусть $A = (a_{ij})$ — $m \times n$-матрица. Фиксируем индексы $1 \le i_1 \le i_2 \le \cdots \le i_k \le m$ и $1 \le j_1 \le j_2 \le \cdots \le j_k \le n$. Определитель

$$\begin{vmatrix} a_{i_1 j_1} & \dots & a_{i_1 j_k} \\ \vdots & \ddots & \vdots \\ a_{i_k j_1} & \dots & a_{i_k j_k} \end{vmatrix}$$

называется *минором* порядка k матрицы A. Минор $\tilde{M}$ порядка $k+1$ называется *окаймляющим* для минора M, если M получается из $\tilde{M}$ вычеркиванием одной крайней (первой или последней) строки и одного крайнего столбца.

Алгоритм нахождения ранга матрицы.

Шаг 1. Ищем в матрице A ненулевой элемент a_{ij}. Если такого нет, то $\operatorname{rk} A = 0$, в противном случае найден отличный от нуля минор $M = a_{ij}$ порядка 1.

Шаг 2. Пусть уже найден минор $M \neq 0$ порядка k. Ищем отличный от нуля минор $\tilde{M}$ порядка $k+1$ среди миноров, окаймляющих M. Если такой минор есть, то повторяем шаг 2 для минора $\tilde{M}$, а если все окаймляющие миноры равны нулю, то $\operatorname{rk} A = k$.

Обоснование алгоритма. Ясно, что алгоритм может остановиться на шаге 1 только в том случае, когда матрица A — нулевая и, следовательно, ее ранг равен 0.

Пусть теперь $A \neq 0$. Покажем сначала, что алгоритм не может работать бесконечно долго. В самом деле, в таком случае из строк и столбцов матрицы A можно было бы составлять ненулевые миноры сколь угодно большого порядка. Но матрица A имеет m строк, поэтому любой ее минор порядка $m+1$ содержит по крайней мере две одинаковых строки и, следовательно, равен нулю. (Аналогично показывается, что порядок отличного от нуля минора матрицы не может быть больше количества ее столбцов, поэтому, если k — порядок ненулевого минора матрицы A, то $k \le \min(m, n)$.)

Итак, через конечное число шагов алгоритм завершит работу и будет найден минор $M \neq 0$ порядка r, все окаймляющие миноры которого равны 0. Докажем, что $\operatorname{rk} A = r$. Для простоты обозначений будем

считать, что минор M образован элементами первых r строк и первых r столбцов матрицы A. (Такого расположения минора M всегда можно добиться подходящими перестановками строк и столбцов матрицы, при этом ее ранг не меняется.) Разобьем матрицу A на блоки: $A = \begin{pmatrix} B & C \\ D & F \end{pmatrix}$, где блок B — матрица порядка r (размеры остальных блоков легко вычисляются, исходя из размеров B и A, в частности, если $r = m$, то $A = (B|C)$). Ясно, что $\det B = M \neq 0$, откуда ввиду следствия 5.9 выводим $\operatorname{rk} B = r$. Тогда и $\operatorname{rk}(B|C) = r$, так как $r = \operatorname{rk} B \leq \operatorname{rk}(B|C) = \operatorname{rk}\{A_{(1)}, \ldots, A_{(r)}\} \leq r$. Если $r = m$, то требуемое равенство $\operatorname{rk} A = r$ получено.

Пусть теперь $r < m$. Фиксируем номера i, j строки и столбца матрицы A и рассмотрим окаймляющий M минор

$$\tilde{M} = \begin{vmatrix} a_{11} & \ldots & a_{1r} & a_{1j} \\ \vdots & \ddots & \vdots & \vdots \\ a_{r1} & \ldots & a_{rr} & a_{rj} \\ a_{i1} & \ldots & a_{ir} & a_{ij} \end{vmatrix}.$$

Согласно предположению $\tilde{M} = 0$, значит, его столбцы $\tilde{M}_1, \ldots, \tilde{M}_r, \tilde{M}_{r+1}$ линейно зависимы (предположение о том, что $\operatorname{rk}\{\tilde{M}_1, \ldots, \tilde{M}_r, \tilde{M}_{r+1}\} = r + 1$ ввиду следствия 5.9 означало бы, что $\tilde{M} \neq 0$). В то же время первые r столбцов минора $\tilde{M}$ линейно независимы, так как $r \geq \operatorname{rk}\{\tilde{M}_1, \ldots, \tilde{M}_r\} \geq \operatorname{rk} B = r$. Воспользовавшись свойством, данным в п. 3 упражнения 3.1, получаем, что столбец $\tilde{M}_{r+1} = (a_{1j}, \ldots, a_{rj}, a_{ij})^t$ линейно выражается через первые r столбцов минора $\tilde{M}$. В силу произвольности номера j заключаем, что через линейно независимые столбцы $\tilde{M}_1 = (a_{11}, \ldots, a_{r1}, a_{i1})^t, \ldots, \tilde{M}_r = (a_{1r}, \ldots, a_{rr}, a_{ir})^t$ линейно выражаются все столбцы матрицы $\tilde{A} = \begin{pmatrix} a_{11} & \ldots & a_{1r} & \ldots & a_{1n} \\ & \ldots & & \ldots & \\ a_{r1} & \ldots & a_{rr} & \ldots & a_{rn} \\ a_{i1} & \ldots & a_{ir} & \ldots & a_{in} \end{pmatrix}$, следовательно, ее ранг равен r. Но $\tilde{A}$ составлена из строк $A_{(1)}, \ldots, A_{(r)}, A_{(i)}$ матрицы A, поэтому $\operatorname{rk}\{A_{(1)}, \ldots, A_{(r)}, A_{(i)}\} = \operatorname{rk} \tilde{A} = r$, откуда, в частности, вытекает линейная зависимость последней системы строк. Заметим, что полученное ранее равенство $\operatorname{rk}(B|C) = r$ означает, что строки $A_{(1)}, \ldots, A_{(r)}$ линейно

независимы, следовательно, строка $A_{(i)}$ является их линейной комбинацией. Ввиду произвольности номера i приходим к выводу, что все строки матрицы A линейно выражаются через первые ее r линейно независимых строк, так что $\operatorname{rk} A = r$.$\square$

Следствие 5.10 *Ранг матрицы равен наибольшему из порядков ее ненулевых миноров.*

Доказательство. Пусть $\operatorname{rk} A = r$ и k — наибольший из порядков ненулевых миноров матрицы A. Если минор $M \neq 0$ порядка k составлен из элементов строк $A_{(i_1)}, \dots, A_{(i_k)}$, то эти строки линейно независимы, откуда $r \geq k$. С другой стороны, применение метода окаймляющих миноров дает ненулевой минор порядка r, следовательно, $k \geq r$.$\square$

6 Многочлены

В предыдущих параграфах были рассмотрены примеры различных алгебраических систем, такие как поле комплексных чисел, кольцо матриц, группа перестановок. Данный параграф посвящен еще одному важному классу алгебраических объектов — колец многочленов.

6.1 Построение кольца многочленов. Степень многочлена

Пусть K — поле. Рассмотрим множество P всех последовательностей элементов поля K, в которых все члены, начиная с некоторого номера, равны 0, то есть $P = \{(f_0, f_1, f_2, \dots) : \exists n \, \forall k > n \ f_k = 0\}$. Зададим на P операции сложения и умножения:

$$(f_0, f_1, f_2, \dots) + (g_0, g_1, g_2, \dots) = (f_0 + g_0, f_1 + g_1, f_2 + g_2, \dots),$$

$$(f_0, f_1, f_2, \dots)(g_0, g_1, g_2, \dots) = (h_0, h_1, h_2, \dots),$$

где $h_k = \sum\limits_{i+j=k} f_i g_j$ при всех k. Например, $h_0 = f_0 g_0$, $h_1 = f_0 g_1 + f_1 g_0$, $h_2 = f_0 g_2 + f_1 g_1 + f_2 g_0$ и так далее.

Теорема 6.1 $(P, +, \cdot)$ — *коммутативное ассоциативное кольцо с единицей.*

Доказательство. Пусть $f, g \in P$. Тогда $f_k = 0$ при $k > n$ для некоторого n и $g_k = 0$ при $k > m$ для некоторого m. Ясно, что $(f + g)_k = 0$

при $k > \max(n, m)$, значит, $f + g \in P$. Если $i + j > n + m$, то $i > n$ или $j > m$, следовательно, $f_i g_j = 0$, поскольку по крайней мере один из элементов f_i и g_j равен нулю. Поэтому $(fg)_k = \sum\limits_{i+j=k} f_i g_j = 0$ при $k > n+m$, значит, $fg \in P$. Таким образом, P замкнуто относительно введенных на нем операций сложения и умножения.

Перейдем к проверке аксиом кольца. Поскольку сложение последовательностей из P производится поэлементно, оно наследует коммутативность и ассоциативность сложения в K, нулем в P будет последовательность $(0, 0, 0, \dots)$, противоположным к $(f_0, f_1, f_2, \dots)$ элементом — последовательность $(-f_0, -f_1, -f_2, \dots)$. Таким образом, $(P, +)$ — абелева группа. Коммутативность умножения в P вытекает из коммутативности умножения в K:

$$(fg)_k = \sum_{i+j=k} f_i g_j = \sum_{j+i=k} g_j f_i = (gf)_k, \quad k = 0, 1, 2, \dots$$

Проверим дистрибутивность:

$$[(f+g)h]_k = \sum_{i+j=k} (f+g)_i h_j = \sum_{i+j=k} (f_i + g_i) h_j = \sum_{i+j=k} f_i h_j + \sum_{i+j=k} g_i h_j =$$
$$(fh)_k + (gh)_k = (fh + gh)_k, \quad k = 0, 1, 2, \dots$$

следовательно, $(f + g)h = fh + gh$. Воспользовавшись коммутативностью умножения, получаем и второй закон дистрибутивности: $f(g+h) = (g+h)f = gf + hf = fg + fh$. Теперь убедимся в ассоциативности умножения:

$$[(fg)h]_k = \sum_{i+j=k} (fg)_i h_j = \sum_{i+j=k} \Big(\sum_{s+t=i} f_s g_t \Big) h_j = \sum_{s+t+j=k} f_s g_t h_j.$$

Аналогичные выкладки дают то же выражение и для $[f(gh)]_k$. Наконец, непосредственная проверка показывает, что единицей в P является последовательность $(1, 0, 0, \dots)$. $\square$

Заметим, что с последовательностями вида $(a, 0, 0, \dots)$ операции сложения и умножения выполняются так же, как и с элементами поля K: $(a, 0, 0, \dots) + (b, 0, 0, \dots) = (a + b, 0, 0, \dots)$, $(a, 0, 0, \dots)(b, 0, 0, \dots) = (ab, 0, 0, \dots)$, кроме того, $(a, 0, 0, \dots)(f_0, f_1, f_2, \dots) = (af_0, af_1, af_2, \dots)$, поэтому такие элементы кольца P удобно отождествлять с элементами из K и вместо $(a, 0, 0, \dots)$ писать просто a. В частности, единица $(1, 0, 0, \dots)$ кольца P при этом записывается в привычной форме — 1.

Обозначим элемент $(0,1,0,0,0,\dots) \in P$ через X. Имеем:

$$\begin{aligned} X &= (0,1,0,0,0,\dots) \\ X^2 &= (0,0,1,0,0,\dots) \\ X^3 &= (0,0,0,1,0,\dots) \end{aligned}$$

и так далее. Удобно также считать, что $X^0 = 1$. Тогда

$$(f_0, f_1, \dots, f_n, 0, 0, \dots) =$$

$$(f_0,0,\dots,0,0,0,\dots)+(0,f_1,\dots,0,0,0,\dots)+\dots+(0,0,\dots,f_n,0,0,\dots) =$$
$$f_0+f_1X+\dots+f_nX^n = \sum_k f_kX^k.$$

Элемент X называется *переменной* или *неизвестной*, выражение $f = f_0 + f_1X + \dots + f_nX^n$ — *многочленом* (*полиномом*) от X, элементы $f_0, f_1, \dots, f_n \in K$ — *коэффициентами* многочлена f, а кольцо P обозначается через $K[X]$ и называется *кольцом многочленов* от переменной X над полем K.

Легко видеть, что введенные выше операции сложения и умножения многочленов при переходе к новой форме записи соответствуют обычным правилам сложения и умножения выражений, содержащих переменную X. В последних порядок нумерации коэффициентов не имеет существенного значения. Договоримся в дальнейшем нумеровать коэффициенты многочленов в порядке убывания степеней переменной X: $f = a_0X^n + a_1X^{n-1} + \dots + a_{n-1}X + a_n$ $(a_0 \neq 0)$. Наибольшая из степеней переменной X, коэффициент при которой отличен от 0, называется *степенью* многочлена f и обозначается $\deg f$, коэффициент a_0 при X^n, где $n = \deg f$, называется *старшим* коэффициентом, а коэффициент a_n — *свободным членом*. Степень нулевого многочлена обозначается через $-\infty$. Множество степеней многочленов естественным образом упорядочено: $-\infty < 0 < 1 < 2 < \dots$, при этом считается, что $-\infty + \deg f = -\infty$ для любого $f \in K[X]$. Вполне очевидны следующие

Свойства степеней:

1°. $\deg(f \pm g) \leq \max(\deg f, \deg g)$.

2°. $\deg(fg) = \deg f + \deg g$.

Замечание. По аналогии с приведенным выше построением кольца многочленов над полем можно определить кольцо $R[X]$ многочленов над произвольным кольцом R. Разумеется, свойства кольца $R[X]$, в частности,

коммутативность умножения, свойства степеней многочленов и др., существенно зависят от свойств кольца R.

Говорят, что кольцо не содержит *делителей нуля*, если для любых его элементов a, b из $ab = 0$ следует $a = 0$ или $b = 0$. Коммутативное кольцо без делителей нуля называется *целостным*. Очевидно, любое поле целостно.

Предложение 6.2 *Кольцо $R[X]$ многочленов над целостным кольцом R целостно.*

Доказательство. Коммутативность кольца $R[X]$ немедленно вытекает из коммутативности кольца R.

Пусть теперь $f, g \in R[X]$, $f, g \neq 0$. Тогда для некоторых $n, m \geq 0$ имеем $f = a_0X^n + \ldots + a_n$, $g = b_0X^m + \ldots + b_m$ при подходящих $a_0, \ldots, a_n, b_0, \ldots, b_m \in R$, где $a_0, b_0 \neq 0$. Поскольку R целостно, старшим коэффициентом многочлена fg будет элемент $a_0b_0 \neq 0$, следовательно, $fg \neq 0$. Получаем, что $R[X]$ не содержит делителей нуля. $\square$

6.2 Деление многочленов с остатком

Пусть K — поле, $f, g \in K[X]$. Если $f = qg + r$, где $q, r \in K[X]$, $\deg r < \deg g$, то говорят, что f *делится на g с остатком r*.

Теорема 6.3 (о делении с остатком). *Пусть $f, g \in K[X]$, $g \neq 0$. Тогда существует единственная пара многочленов $q, r \in K[X]$ такая, что*

$$f = qg + r, \quad \deg r < \deg g.$$

Доказательство. Сначала докажем существование пары (q, r), воспользовавшись методом математической индукции по степени n многочлена f.

Основание индукции. Согласно условию теоремы $g \neq 0$, так что $\deg g > -\infty$, поэтому $f = 0 \cdot g + f$ — требуемое представление f при любом $n < \deg g$.

Пусть для всех многочленов, степень которых строго меньше n, утверждение уже доказано. Обозначим степень g через m. С учетом сказанного выше, можно считать, что $m \leq n$. Запишем многочлены f

и g в явном виде:

$$f = a_0X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n,$$
$$g = b_0X^m + b_1X^{m-1} + \cdots + b_{m-1}X + b_m.$$

Рассмотрим многочлен $\tilde{g} = a_0b_0^{-1}X^{n-m}g$. Очевидно, что $\deg \tilde{g} = n$ и что старшие члены многочленов f и $\tilde{g}$ совпадают, поэтому степень многочлена $\tilde{f} = f - \tilde{g}$ строго меньше n. По предположению индукции найдутся многочлены $\tilde{q}, \tilde{r}$ такие, что $\tilde{f} = \tilde{q}g + \tilde{r}$, $\deg \tilde{r} < \deg\, g$. Тогда

$$f = \tilde{g} + \tilde{f} = a_0b_0^{-1}X^{n-m}g + \tilde{q}g + \tilde{r} = (a_0b_0^{-1}X^{n-m} + \tilde{q})g + \tilde{r}$$

— требуемое представление f.

Докажем единственность. Пусть $f = q_ig + r_i$, $\deg r_i < \deg\, g$ $(i = 1, 2)$ — два представления многочлена f. Тогда $(q_1 - q_2)g = r_2 - r_1$. Если $q_1 - q_2 \neq 0$, то $\deg(q_1 - q_2) \geq 0$, следовательно,

$$\deg\, g \leq \deg(q_1 - q_2) + \deg\, g = \deg((q_1 - q_2)g) = \deg(r_2 - r_1) \leq \max(\deg\, r_2, \deg\, r_1) < \deg\, g,$$

откуда $\deg\, g < \deg\, g$ — противоречие. Значит, $q_1 - q_2 = 0$, то есть $q_1 = q_2$. Тогда $q_1g = q_2g$, поэтому из равенств $q_1g + r_1 = f = q_2g + r_2$ выводим $r_1 = r_2$.$\square$

6.3 Делимость в кольце многочленов

Пусть R — целостное кольцо с единицей, $a, b \in R$. Говорят, что a *делит* b (обозначение: $a \mid b$), если $b = ac$ для некоторого $c \in R$. Выражение $a \nmid b$ означает, что a не делит b. Делители единицы называются *обратимыми* элементами.

Если многочлен $f \in K[X]$ обратим, то $fg = 1$ для некоторого $g \in K[X]$. Тогда $0 = \deg(fg) = \deg\, f + \deg\, g$, откуда $\deg\, f = \deg\, g = 0$, то есть $f, g \in K$. Следовательно, обратимыми элементами кольца $K[X]$ являются только ненулевые элементы поля K.

Если $a \mid b$ и $b \mid a$, то элементы $a, b \in R$ называются *ассоциированными*. Покажем, что в этом случае $b = ua$, где u — обратимый элемент. В самом деле, если $a = 0$, то $b = ac = 0c = 0$, так что $a = 1 \cdot b$, а если $a \neq 0$, то $a = bd = acd$, откуда $a(1 - cd) = 0$. Тогда в силу целостности R получаем $1 - cd = 0$, так что $cd = 1$ и, следовательно, c, d

— обратимые элементы. Очевидно, верно и обратное: любые два элемента, отличающиеся лишь обратимым множителем, ассоциированы.

Легко проверяются следующие

Свойства делимости:

1. $a \mid b,\ b \mid c \Rightarrow a \mid c$.

2. $c \mid a,\ c \mid b \Rightarrow c \mid (a \pm b)$.

3. $a \mid b \Rightarrow a \mid bc$.

4. $a \mid b_1, \ldots, a \mid b_k \Rightarrow a \mid (b_1c_1 + \cdots + b_kc_k)$ при любых $c_1, \ldots, c_k \in R$.

Элемент $p \in R,\ p \nmid 1$, называется *простым*, если любой его делитель с точностью до ассоциированности есть либо 1, либо p. Элемент $a \in R,\ a \nmid 1$, *неразложим*, если его нельзя представить в виде $a = bc$, где $b \nmid 1,\ c \nmid 1$. Заметим, что в целостном кольце понятия простоты и неразложимости эквивалентны. Действительно, всякий неразложимый элемент, очевидно, прост. Обратно, если $p \in R$ прост и для некоторых $a, b \in R$ верно $p = ab$, то хотя бы один из элементов a, b ассоциирован с p. Пусть, для определенности, $b = up$, где $u \mid 1$. Тогда $p = ab = aup$, откуда получаем $(1 - au)p = 0$. Последнее равенство ввиду целостности R влечет $1 = au$, следовательно, $a \mid 1$ и, значит, p неразложим.

Неразложимые элементы кольца многочленов обычно называют *неприводимыми* многочленами. Многочлен степени 1 всегда неприводим. В самом деле, если $p = ab$, то $1 = \deg p = \deg a + \deg b$, откуда либо $\deg a = 0$ и, следовательно, $a \mid 1$, либо $\deg b = 0$ и $b \mid 1$.

6.4 НОД и НОК в кольце многочленов

Пусть R — целостное кольцо с единицей. Говорят, что $d \in R$ является *наибольшим общим делителем* элементов $a, b \in R$ (обозначение: $d = \text{НОД}\,(a, b)$), если

1) $d \mid a,\ d \mid b$;

2) $c \mid a,\ c \mid b \Rightarrow c \mid d$.

Ясно, что НОД определяется с точностью до обратимого множителя. Справедливы следующие свойства:

1. $\text{НОД}\,(a, b) = a \Leftrightarrow a \mid b$.

2. $\text{НОД}\,(a, 0) = a$.

3. $\text{НОД}\,(ta, tb) = t\,\text{НОД}\,(a, b)$.

4. $\text{НОД}\,(\text{НОД}\,(a,b),c) = \text{НОД}\,(a,\text{НОД}\,(b,c))$.

Докажем свойство 3 (проверка остальных свойств предлагается в качестве упражнения). Введем обозначения: $\text{НОД}\,(a,b) = d$, $\text{НОД}\,(ta,tb) = f$. Так как $d \mid a$, $d \mid b$, то $td \mid ta$, $td \mid tb$. Следовательно, $td \mid f$, тем самым, $f = tdg$ для некоторого g. Тогда $tdg = f \mid ta$, что с учетом целостности влечет $dg \mid a$, аналогично, $tdg = f \mid tb$ влечет $dg \mid b$. Значит, $dg \mid \text{НОД}\,(a,b) = d$, откуда $g \mid 1$. Получаем, что f и td — ассоциированные элементы.$\square$

Если $\text{НОД}\,(a,b) = 1$, то элементы $a, b \in R$ называются *взаимно простыми.*

Двойственным образом к НОД определяется понятие *наименьшего общего кратного* (НОК): $m = \text{НОК}\,(a,b)$, если

1') $a \mid m$, $b \mid m$;

2') $a \mid c$, $b \mid c \Rightarrow m \mid c$.

Предложение 6.4 *Пусть $a, b \in R$ таковы, что для любого $c \in R$ существует НОД(ca, cb). Тогда существует НОК(a,b), при этом*

$$\text{НОД}\,(a,b)\,\text{НОК}\,(a,b) = ab.$$

Доказательство. Положим $d = \text{НОД}\,(a,b)$. Если $a = 0$ или $b = 0$, то $\text{НОК}\,(a,b) = 0$ и равенство $\text{НОД}\,(a,b)\text{НОК}\,(a,b) = ab$ тривиально верно, поэтому можно считать, что $a \neq 0$, $b \neq 0$ и, следовательно, $d \neq 0$.

Представив элементы $a, b \in R$ в виде $a = a'd$, $b = b'd$ при подходящих $a', b' \in R$, получаем $ab = (a'b'd)d$. Положим $m = a'b'd$ и докажем, что $m = \text{НОК}\,(a,b)$. Имеем: $a \mid ab' = a'b'd = m$ и, аналогично, $b \mid a'b = a'b'd = m$, тем самым выполнено условие 1') определения НОК.

Пусть теперь $a \mid c$ и $b \mid c$. Тогда с учетом свойства 3 наибольшего общего делителя из $ab \mid ca$, $ab \mid cb$ выводим $md = ab \mid \text{НОД}\,(ca,cb) = c\,\text{НОД}\,(a,b) = cd$. Таким образом, $md \mid cd$, следовательно, $m \mid c$, то есть для m выполнено и условие 2').$\square$

В кольце $K[X]$ для вычисления $\text{НОД}\,(a,b)$ обычно используют метод, называемый АЛГОРИТМОМ ЕВКЛИДА. В силу свойства 2 можно ограничиться случаем, когда a и b отличны от нуля.

Введем обозначения $a = r_{-1}$, $b = r_0$ и положим $k = 0$.

Шаг k: Поделим r_{k-1} с остатком на r_k. Получим

$$r_{k-1} = q_k r_k + r_{k+1}, \quad \deg r_{k+1} < \deg r_k.$$

Если $r_{k+1} = 0$, то алгоритм завершен и НОД $(a, b) = r_k$, в противном случае переходим к шагу $k + 1$.

Обоснование алгоритма. Поскольку последовательность степеней остатков r_0, r_1,... является строго убывающей, через некоторое число m шагов ($m \leq \deg r_0 + 1$) получим $\deg r_{m+1} = -\infty$, то есть $r_{m+1} = 0$, и алгоритм закончит свою работу. Покажем, что $r_m = \text{НОД}\,(a, b)$.

Имеем систему равенств:

$$\begin{aligned} a &= q_0 b + r_1, \\ b &= q_1 r_1 + r_2, \\ \dots &\quad \dots \\ r_{m-2} &= q_{m-1} r_{m-1} + r_m, \\ r_{m-1} &= q_m r_m. \end{aligned} \tag{25}$$

Очевидно, $r_m \mid r_m$. Согласно последнему равенству системы (25) верно $r_m \mid r_{m-1}$. Тогда из предпоследнего равенства вытекает $r_m \mid r_{m-2}$. Воспользовавшись третьим с конца равенством, выводим $r_m \mid r_{m-3}$ и так далее. В итоге получаем $r_m \mid b$ и $r_m \mid a$, тем самым для r_m выполнено условие 1) определения НОД.

Пусть теперь $c \mid a$ и $c \mid b$. Двигаясь по системе (25) сверху вниз, последовательно получаем $c \mid r_1$, $c \mid r_2$, ..., $c \mid r_m$, то есть верно и условие 2). Следовательно, $r_m = \text{НОД}\,(a, b)$.$\square$

Перепишем равенства системы (25) (кроме последнего) в виде

$$\begin{aligned} a - q_0 b &= r_1, \\ b - q_1 r_1 &= r_2, \\ \dots &\quad \dots \\ r_{m-2} - q_{m-1} r_{m-1} &= r_m. \end{aligned} \tag{26}$$

Двигаясь по системе (26) сверху вниз, замечаем, что r_1 выражается через многочлены a и b, многочлен r_2 выражается через b, r_1 и поэтому опять выражается через a и b, и так далее. В итоге, получим выражение многочлена r_m через a и b. Таким образом, опираясь дополнительно на предложение 6.4 и алгоритм Евклида, заключаем, что верна

Теорема 6.5 *В кольце $K[X]$ любые многочлены a и b имеют* НОД *и* НОК*, причем существуют многочлены $u, v \in K[X]$ такие, что*

$$НОД(a, b) = ua + vb.$$

В частности, a и b взаимно просты $\Leftrightarrow$ $ua + vb = 1$ для некоторых $u, v \in K[X]$.

Следствие 6.6 *Для всех $a, b, c \in K[X]$ справедливы импликации:*

1. $НОД(a, b) = 1,\ НОД(a, c) = 1 \Rightarrow НОД(a, bc) = 1.$

2. $a \mid bc,\ НОД(a, b) = 1 \Rightarrow a \mid c.$

3. $b \mid a,\ c \mid a,\ НОД(b, c) = 1 \Rightarrow bc \mid a.$

Доказательство. 1. Согласно теореме 6.5 имеем: $u_1a + v_1b = 1$, $u_2a + v_2c = 1$. Тогда $1 = (u_1a+v_1b)(u_2a+v_2c) = (u_1u_2a+u_1v_2c+v_1u_2b)a+(v_1v_2)bc$, следовательно, $НОД(a, bc) = 1$.

2. Если $a \mid bc$, $НОД(a, b) = 1$, то $bc = aw$, $ua + vb = 1$ для некоторых $u, v, w \in K[X]$. Тогда $c = c(ua + vb) = acu + bcv = acu + awv = a(cu + wv)$, откуда $a \mid c$.

3. Так как $b \mid a$ и $c \mid a$, то $НОК(b, c) \mid a$. Поэтому с учетом равенства $НОД(b, c) = 1$ и предложения 6.4 получаем $НОК(b, c) = НОК(b, c)НОД(b, c) = bc$, следовательно, $bc \mid a$. $\square$

6.5 Факториальность кольца $K[X]$

Пусть R — целостное кольцо с единицей. Говорят, что R — *факториально* или является кольцом *с однозначным разложением на простые множители*, если для любого $a \in R$ $(a \neq 0)$, существуют обратимый элемент $u \in R$ и (не обязательно различные) простые элементы $p_1, \ldots, p_r \in R$ такие, что

$$a = up_1 \ldots p_r, \qquad (*)$$

причем если $a = wq_1 \ldots q_s$ — еще одно разложение вида $(*)$, то $s = r$ и при подходящей нумерации q_i ассоциирован с p_i для всех i.

Замечание. Допуская случай $r = 0$, можно считать, что обратимые элементы тоже обладают представлением вида $(*)$.

Теорема 6.7 *Кольцо R с разложением на простые множители факториально тогда и только тогда, когда для всех $a, b \in R$ и для любого простого $p \in R$ верна импликация $p \mid ab \Rightarrow p \mid a$ или $p \mid b$.*

Замечание. Из присутствующего в формулировке теоремы 6.7 условия делимости простым элементом одного из двух сомножителей вытекает справедливость аналогичного условия для любого конечного числа сомножителей: *если $p \in R$ — прост и $p \mid \prod\limits_{i=1}^{k} a_k$, то $p \mid a_i$ для некоторого i.* Доказательство является несложным упражнением на применение метода математической индукции.

Доказательство теоремы. Пусть R факториально и простой элемент $p \in R$ делит ab, то есть $ab = pc$ для некоторого $c \in R$. Разложим элементы a, b, c на простые множители: $a = u\prod\limits_i a_i$, $b = v\prod\limits_j b_i$, $c = w\prod\limits_k c_k$. Тогда $(uv)\prod\limits_i a_i \prod\limits_j b_j = ab = wp\prod\limits_k c_k$, откуда ввиду факториальности R вытекает ассоциированность p с одним из простых множителей a_i или b_j. Следовательно, $p \mid a$ или $p \mid b$.

Для доказательства обратного утверждения достаточно установить единственность разложения вида $(*)$ для любого ненулевого $a \in R$. Применим метод математической индукции по количеству n участвующих в разложении простых сомножителей.

При $n = 0$ утверждение очевидно.

Предположим, что для всех элементов кольца, допускающих разложение с менее чем n простыми сомножителями, единственность разложения уже доказана. Пусть $a \neq 0$ и

$$a = u\prod_{i=1}^{n} p_i = v\prod_{j=1}^{m} q_j,$$

где все множители p_i, q_j — простые и $m \geq n$. В частности, p_n прост и $p_n \mid v\prod\limits_{j=1}^{m} q_j$. Следовательно, элемент p_n делит некоторый множитель q_j и поэтому, в силу простоты q_j, ассоциирован с ним, то есть $q_j = wp_n$ для некоторого обратимого w. Перенумеровав при необходимости сомножители второго разложения, можно считать, что $j = m$. Сократив на p_n, получаем

$$u\prod_{i=1}^{n-1} p_i = (vw)\prod_{j=1}^{m-1} q_j.$$

Количество простых сомножителей в левой части последнего равенства меньше n, так что по предположению индукции $n-1=m-1$, то есть $n=m$ и при подходящей нумерации p_i ассоциирован с q_i для всех i.$\square$

Лемма 6.8 *Каждый ненулевой многочлен кольца $K[X]$ обладает разложением на простые множители.*

Доказательство. Установим существование разложения вида $(*)$ для ненулевого многочлена $a \in K[X]$, воспользовавшись методом математической индукции по степени n многочлена.

Для многочленов нулевой степени, то есть обратимых элементов кольца $K[X]$, утверждение тривиально верно.

Пусть утверждение уже доказано для всех многочленов степени, меньшей n. Если a прост, то доказывать нечего, в противном случае $a = bc$ для некоторых необратимых $b, c \in K[X]$. Тогда $\deg a = \deg b + \deg c$, откуда $1 \leq \deg b, \deg c < \deg a = n$, следовательно, в силу предположения индукции b и c обладают разложением вида $(*)$, значит, таким разложением обладает и $bc = a$.$\square$

Теорема 6.9 *Кольцо $K[X]$ факториально.*

Доказательство. В силу леммы 6.8 и теоремы 6.7 достаточно показать, что для любого неприводимого многочлена $p \in K[X]$ верна импликация $p \mid ab \Rightarrow p \mid a$ или $p \mid b$. Справедливость данной импликации при $a = 0$ или $b = 0$ очевидна, поэтому можно считать, что $ab \neq 0$.

Обозначим НОД многочленов a и p через d. Так как p неприводим, то либо $d = p$, либо $d = 1$. В первом случае $p = d = \text{НОД}\,(p, a) \mid a$. Во втором случае согласно п. 2 следствия 6.6 из $p \mid ab$ и $\text{НОД}\,(p, a) = d = 1$ вытекает $p \mid b$.$\square$

Из факториальности кольца $K[X]$ следует, что любые два ненулевых многочлена $a, b \in K[X]$ обладают "общим" разложением

$$a = up_1^{k_1} \dots p_r^{k_r}, \quad b = vp_1^{l_1} \dots p_r^{l_r},$$

где некоторые неприводимые сомножители могут присутствовать с нулевыми степенями. Нетрудно убедиться в справедливости следующих утверждений (докажите их в качестве упражнения):

1. $a \mid b \Leftrightarrow k_i \leq l_i$ для всех i.
2. $\text{НОД}\,(a, b) = p_1^{s_1} \dots p_r^{s_r}$, где $s_i = \min(k_i, l_i)$ для всех i.
3. $\text{НОК}\,(a, b) = p_1^{t_1} \dots p_r^{t_r}$, где $t_i = \max(k_i, l_i)$ для всех i.

6.6 Корни многочленов

Пусть $f \in K[X]$, $f = a_0X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n$. Многочлену f отвечает функция, действующая из K в K по правилу: $c \mapsto a_0c^n + a_1c^{n-1} + \cdots + a_{n-1}c + a_n$. Ее значение на элементе $c \in K$ обозначается через $f(c)$ и называется *значением многочлена f в точке c*. Следует подчеркнуть принципиальное различие алгебраической и функциональной точек зрения на многочлены, поскольку для некоторых полей разным многочленам может отвечать одна и та же функция. Однако, для многих полей, в том числе, для полей $\mathbb{Q}, \mathbb{R}, \mathbb{C}$, соответствие между многочленами и отвечающими им функциями является биективным. (Подробнее см. [1], гл. 5, 6.)

Говорят, что элемент $c \in K$ является *корнем* многочлена $f \in K[X]$, если $f(c) = 0$. Элемент $c \in K$ называют также *корнем уравнения* $f(x) = 0$.

Теорема 6.10 (Безу) *Элемент $c \in K$ есть корень многочлена $f \in K[X] \Leftrightarrow (X - c) \mid f$.*

Доказательство. Поделим f с остатком на $X - c$:

$$f = (X - c)q + r,$$

где $\deg r < \deg(X - c) = 1$, следовательно, $r \in K$.

Пусть c — корень f. Тогда $0 = f(c) = (c-c)q(c)+r = 0 \cdot q(c) + r = r$, так что $r = 0$ и, следовательно, $(X - c) \mid f$.

Обратно, если $(X-c) \mid f$, то $f = (X-c)q$ для некоторого $q \in K[X]$. Тогда $f(c) = 0 \cdot q(c) = 0$, так что c — корень. $\square$

Замечание. $f(c) = r$, то есть значение многочлена f на элементе c равно остатку от деления f на $X - c$.

Существует алгоритм "быстрого" деления f на $X-c$, известный как схема Горнера. Он состоит в следующем: пусть $f = a_0X^n+a_{n-1}X^{n-1}+\cdots+a_{n-1}X+a_0$, $q = b_0X^{n-1}+b_1X^{n-2}+\cdots+b_{n-2}X+b_{n-1}$ и $f = (X-c)q+r$, тогда $b_0 = a_0$, $b_k = a_k + b_{k-1}c$ при $k = 1, 2, \ldots, n-1$, $r = a_n + b_{n-1}c$. Эти вычисления удобно записывать в виде таблицы:

	a_0	a_1	$\ldots$	a_{n-1}	a_n
c	$b_0 = a_0$	$b_1 = a_1 + b_0c$	$\ldots$	$b_{n-1} = a_{n-1} + b_{n-2}c$	$r = a_n + b_{n-1}c$

Заполняется таблица так: в верхней строке (кроме первой ячейки) записываются коэффициенты многочлена f, в первой ячейке нижней строки пишется элемент c, в следующей ячейке — a_0, а все последующие ячейки заполняются слева направо по правилу: к содержимому верхней ячейки прибавляется содержимое предыдущей (левой) ячейки, умноженное предварительно на c. Корректность данного алгоритма легко установить, подставив выражения для коэффициентов многочлена q и остатка r в равенство $f = (X - c)q + r$ (проверьте это в качестве упражнения).

Элемент $c \in K$ называется *k-кратным корнем* многочлена $f \in K[X]$, если $(X - c)^k \mid f$ и $(X - c)^{k+1} \nmid f$. В частности, при $k = 1$ корень называется *простым*, при $k = 2$ — *двойным*, при $k = 3$ — *тройным* и так далее. Таким образом, $c \in K$ есть k-кратный корень многочлена f тогда и только тогда, когда $f = (X - c)^k g$, где $\text{НОД}\,(X - c, g) = 1$, то есть $g(c) \neq 0$. Отметим, что при этом $\deg f = k + \deg g$.

Теорема 6.11 *Пусть $f \in K[X]$, $\deg f \geq 1$ и $c_1, \ldots, c_r$ — корни f кратностей $k_1, \ldots, k_r$. Тогда найдется $g \in K[X]$ такой, что*

$$f = (X - c_1)^{k_1} \ldots (X - c_r)^{k_r} g, \text{ где } g(c_i) \neq 0 \quad (i = 1, \ldots, r).$$

В частности, $k_1 + \cdots + k_r \leq \deg f$, то есть сумма кратностей корней ненулевого многочлена не превосходит его степени.

Доказательство проведем индукцией по r. При $r = 1$ доказывать нечего, так как утверждение совпадает с определением k_1-кратного корня.

Предположим, что утверждение уже доказано для $r - 1$ корней, то есть существует $h \in K[X]$ такой, что $f = (X - c_1)^{k_1} \ldots (X - c_{r-1})^{k_{r-1}} h$ и $h(c_i) \neq 0$ при $i = 1, \ldots, r - 1$. Так как $c_r \neq c_i$ при $i < r$, то $\text{НОД}\,(X - c_r, X - c_i) = 1$, откуда ввиду пп. 1, 3 следствия 6.6 выводим $\text{НОД}\,((X - c_r)^{k_r}, (X - c_1)^{k_1} \ldots (X - c_{r-1})^{k_{r-1}}) = 1$. Учитывая последнее, а также то, что c_r является k_r-кратным корнем f, ввиду п. 2 следствия 6.6 получаем $(X - c_r)^{k_r} \mid h$. Следовательно, для некоторого $v \in K[X]$, $v(c_r) \neq 0$, имеем $h = (X - c_r)^{k_r} v$, поэтому

$$f = (X - c_1)^{k_1} \ldots (X - c_r)^{k_r} v.$$

Осталось заметить, что $v(c_i) \neq 0$ при $i < r$, поскольку $v \mid h$, а $h(c_i) \neq 0$ для всех $i = 1, \ldots, r - 1$.

Наконец, $\deg f = k_1 + \cdots + k_r + \deg v \geq k_1 + \cdots + k_r$.$\square$

Следствие 6.12 *Пусть* $f, g \in K[X]$, $\deg f$, $\deg g \leq n$. *Если существуют различные элементы* $c_0, c_1, \ldots, c_n \in K$ *такие, что* $f(c_i) = g(c_i)$ *для всех* $i = 0, 1, \ldots, n$, *то* $f = g$.

Доказательство. Обозначим $f - g$ через h. Тогда $\deg h \leq \max(\deg f, \deg g) \leq n$ и $h(c_i) = f(c_i) - g(c_i) = 0$ $(i = 0, 1, \ldots, n)$, то есть многочлен h степени не выше n имеет $n + 1$ корней. В силу второй части утверждения теоремы 6.11 это возможно только при $h = 0$. Следовательно, $f = g$.$\square$

Заметим, что в силу следствия 6.12 для всякого набора, состоящего из $n + 1$ попарно различных элементов поля K, существует не более одного многочлена степени, не превосходящей n, с заданными значениями на этих элементах. Естественно задаться вопросом: всегда ли существует такой многочлен, и если да, то как его найти? Более точно, рассмотрим следующую задачу: для заданных попарно различных элементов $c_0, c_1, \ldots, c_n \in K$, а также элементов $y_0, y_1, \ldots, y_n \in K$ требуется найти такой многочлен степени не выше, чем n, для которого справедливы равенства $f(c_i) = y_i$ при всех $i = 0, 1, \ldots, n$.

Эта задача всегда имеет единственное решение, которое может быть записано в виде многочлена

$$f = \sum_{i=0}^{n} y_i \prod_{j \neq i} \frac{(X - c_j)}{(c_i - c_j)},$$

называемого *интерполяционным многочленом Лагранжа.*

Также для нахождения многочлена f часто бывает удобно пользоваться следующей *интерполяционной формулой Ньютона*:

$$f = u_0 + u_1(X - c_0) + u_2(X - c_0)(X - c_1) + \ldots + u_n(X - c_0)(X - c_1)\ldots(X - c_{n-1}).$$

Здесь неизвестные коэффициенты $u_0, u_1, \ldots, u_n$ последовательно находятся из условий $f(c_0) = y_0$, $f(c_1) = y_1, \ldots$, $f(c_n) = y_n$.

6.7 Производная многочлена

Пусть $f \in K[X]$, $f = a_0X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n$. *Производной* многочлена f называется многочлен

$$f' = na_0X^{n-1} + (n-1)a_1X^{n-2} + \cdots + a_{n-1}.$$

Замечание. В произвольном поле K выражение na для $n \in \mathbb{N}$ и $a \in K$ понимается как сумма n одинаковых слагаемых: $a + \cdots + a$.

Свойства:

1. $(\alpha f + \beta g)' = \alpha f' + \beta g'$ для всех $\alpha, \beta \in K$ и $f, g \in K[X]$.
2. $(fg)' = f'g + fg'$ для всех $f, g \in K[X]$.
3. $(f^k)' = kf^{k-1}f'$.

Доказательство. Свойство 1 очевидно. Справедливость свойства 2 ввиду свойства 1 достаточно установить для многочленов вида $f = X^n$, $g = X^m$ $(n, m \geq 1)$. Имеем:

$$(fg)' = (X^{n+m})' = (n+m)X^{n+m-1} = nX^{n-1}X^m + X^n(mX^{m-1}) = f'g + fg'.$$

Свойство 3 с учетом свойства 2 легко доказывается индукцией по k.$\square$

В дальнейшем будем считать, что K — одно из полей $\mathbb{Q}$, $\mathbb{R}$ или $\mathbb{C}$. Заметим, что в этом случае $\deg f' = \deg f - 1$.

Согласно теореме 6.9 кольцо $K[X]$ факториально, следовательно, для всякого многочлена $f \in K[X]$, $\deg f \geq 1$, существует разложение на неприводимые множители: $f = p_1^{k_1} \dots p_r^{k_r}$. Неприводимые многочлены p_i, $(i = 1, \dots, r)$, будем называть *k_i-кратными множителями* многочлена f. Другими словами, p есть k-кратный множитель для f, если $p^k \mid f$, но $p^{k+1} \nmid f$.

Теорема 6.13 *Пусть неприводимый многочлен $p \in K[X]$ является k-кратным множителем многочлена f, $k \geq 1$. Тогда p есть $(k-1)$-кратный множитель многочлена f'. В частности, $p \nmid f'$ при $k = 1$.*

Доказательство. По условию теоремы имеем $f = p^k g$, где $p \nmid g$. Тогда

$$f' = kp^{k-1}p'g + p^k g' = p^{k-1}(kp'g + pg'),$$

откуда $p^{k-1} \mid f'$. Покажем, что $p^k \nmid f'$. Действительно, если $p^k \mid f'$, то $p \mid (kp'g + pg')$, значит, $p \mid kp'g$. Тогда в силу неприводимости p и факториальности $K[X]$ имеем $p \mid g$ или $p \mid (kp')$, но ни то, ни другое невозможно, так как $p \nmid g$ в силу выбора g и $p \nmid (kp')$, поскольку $\deg p > \deg(kp') > -\infty$.$\square$

Следствие 6.14 *Если $f = p_1^{k_1} \dots p_r^{k_r}$ — разложение f на неприводимые множители, то* $\text{НОД}\,(f, f') = p_1^{k_1-1} \dots p_r^{k_r-1}$.

Доказательство. По теореме 6.13 из соотношения $p_i^{k_i} \mid f$ вытекает $p_i^{k_i-1} \mid f'$. Следовательно, с учетом неприводимости множителей p_i имеем $f' = p_1^{k_1-1} \dots p_r^{k_r-1} g$ для некоторого $g \in K[X]$, где $\text{НОД}\,(p_i, g) = 1$ $(i = 1, \dots, r)$, откуда и получается требуемое утверждение. $\square$

Отметим, что с помощью следствия 6.14 можно найти многочлен h, являющийся произведением всех неприводимых множителей многочлена f, а именно,

$$h = \frac{f}{\text{НОД}\,(f, f')} = p_1 \dots p_r, \tag{27}$$

причем для нахождения h не обязательно знать исходные разложения f и f', достаточно лишь использовать алгоритм Евклида.

6.8 Неприводимость многочленов над $\mathbb{Q}$ и $\mathbb{Z}$

Пусть $f = a_0 X^n + a_1 X^{n-1} + \dots + a_n \in \mathbb{Z}[X]$, то есть все коэффициенты многочлена f — целые числа. Элемент $d(f) = \text{НОД}(a_0, a_1, \dots, a_n)$ называется *содержанием* многочлена f. Если $d(f) \mid 1$, то f называется *примитивным*.

Лемма 6.15 (Гаусс) *Для любых многочленов $f, g \in \mathbb{Z}[X]$ верно*

$$d(fg) \approx d(f)d(g).$$

(Знак $\approx$ понимается как равенство с точностью до ассоциированности.)

Доказательство. Покажем сначала, что если $d(f) \approx 1 \approx d(g)$, то и $d(fg) \approx 1$. Предположим, что многочлены $f = a_0 X^n + a_1 X^{n-1} + \dots + a_n$ и $g = b_0 X^m + b_1 X^{m-1} + \dots + b_m$ примитивны, а их произведение $fg = c_0 X^{n+m} + c_1 X^{n+m-1} + \dots + c_{n+m}$ — не примитивный многочлен. Тогда найдется простое число p, которое делит все коэффициенты многочлена fg. Заметим, что p не может делить все коэффициенты многочленов f и g в силу примитивности f и g. Следовательно, можно выбрать наименьшие индексы i и j такие, что $p \nmid a_i$ и $p \nmid b_j$. Выражая коэффициент c_{i+j} через коэффициенты многочленов f и g, получаем

$$c_{i+j} = a_i b_j + (a_{i-1} b_{j+1} + a_{i-2} b_{j+2} + \dots) + (a_{i+1} b_{j-1} + a_{i+2} b_{j-2} + \dots).$$

По предположению, p делит c_{i+j} равно как и все находящиеся в скобках слагаемые из правой части последнего равенства. Следовательно, $p \mid a_i b_j$,

откуда $p \mid a_i$ или $p \mid b_j$. Полученное противоречие с соотношениями $p \nmid a_i$ и $p \nmid b_j$ доказывает, что $d(fg) \approx 1$.

Пусть теперь f и g — произвольные многочлены. Тогда $f = d(f)f_0$ и $g = d(g)g_0$ для подходящих примитивных многочленов f_0 и g_0. Следовательно, $fg = d(f)d(g)f_0g_0$, причем многочлен f_0g_0 примитивен в силу первой части доказательства и, значит, $d(fg) \approx d(f)d(g)$.$\square$

Следствие 6.16 *Многочлен $f \in \mathbb{Z}[X]$, $\deg f \geq 1$, неприводим над $\mathbb{Z}$ тогда и только тогда, когда он неприводим над $\mathbb{Q}$.*

Доказательство. Очевидно, всякий неприводимый над $\mathbb{Q}$ многочлен с целыми коэффициентами неприводим и над $\mathbb{Z}$.

Предположим теперь, что $f \in \mathbb{Z}[X]$ приводим над $\mathbb{Q}$, то есть $f = gh$ для некоторых многочленов $g, h \in \mathbb{Q}[X]$, $\deg g, \deg h \geq 1$. Умножая равенство $f = gh$ на число $a \in \mathbb{Z}$, равное наименьшему общему кратному знаменателей всех коэффициентов многочленов g и h, получаем $af = \tilde{g}\tilde{h}$, где $\tilde{g}, \tilde{h} \in \mathbb{Z}[X]$. Ввиду равенств $\tilde{g} = d(\tilde{g})g_0$, $\tilde{h} = d(\tilde{h})h_0$ получаем $af = bg_0h_0$, где $b = d(\tilde{g})d(\tilde{h})$, а многочлены $g_0, h_0 \in \mathbb{Z}[X]$ — примитивны. Тогда по лемме Гаусса имеем $ad(f) = \pm b$, так что $af = \pm ad(f)g_0h_0$, откуда $f = \pm d(f)g_0h_0$. Таким образом, f приводим и над $\mathbb{Z}$.$\square$

Признак неприводимости (Эйзенштейн). *Пусть $f \in \mathbb{Z}[X]$, $f = a_0X^n + a_1X^{n-1} + \ldots + a_n$, и p — такое простое число, что*

1) $p \nmid a_0$;

2) $p \mid a_i$ для всех $i = 1, 2, \ldots, n$;

3) $p^2 \nmid a_n$.

Тогда f неприводим над $\mathbb{Q}$.

Доказательство. Предположим, что условия 1)–3) выполнены, но f приводим над $\mathbb{Q}$. Тогда в силу следствия из леммы Гаусса f приводим и над $\mathbb{Z}$, то есть найдутся такие многочлены $g, h \in \mathbb{Z}[X]$, $\deg g, \deg h \geq 1$, что $f = gh$.

Пусть $g = b_0X^s + b_1X^{s-1} + \ldots + b_s$, $h = c_0X^t + c_1X^{t-1} + \ldots + c_t$. Поскольку $p \mid a_n = b_sc_t$, то $p \mid b_s$ или $p \mid c_t$. Предположим для определенности, что $p \mid b_s$. Тогда $p \nmid c_t$, так как иначе $p^2 \mid b_sc_t = a_n$. Заметим также, что $p \nmid b_0$, поскольку в противном случае $p \mid b_0c_0 = a_0$. Следовательно, найдется такой индекс i, $0 \leq i < s$, что $p \nmid b_i$ и $p \mid b_j$ для всех $j > i$. Выражая коэффициент a_{t+i} через коэффициенты многочленов g

и h, получаем

$$a_{i+t} = b_i c_t + b_{i+1} c_{t-1} + b_{i+2} c_{t-2} + \dots$$

По предположению, $p \mid a_{i+t}$, $p \mid b_{i+1}$, $p \mid b_{i+2}$, $\dots$, значит, $p \mid b_i c_t$ и, следовательно, $p \mid b_i$ или $p \mid c_t$, но ни то, ни другое невозможно. Полученное противоречие завершает доказательство. $\square$

6.9 Формулы Виета

Пусть K, P — поля, причем $K \subseteq P$. Предположим, что многочлен $f \in K[X]$, $f = X^n + a_1 X^{n-1} + \dots + a_n$ разлагается в кольце $P[X]$ на линейные множители: $f = (X - c_1) \dots (X - c_n)$ (другими словами, $c_1, \dots, c_n \in P$ — все корни многочлена f с учетом их кратностей). Раскрывая скобки в произведении линейных множителей и приравнивая коэффициенты при соответствующих степенях переменной, получаем следующую систему равенств:

$$\begin{aligned}
a_1 &= -(c_1 + \dots + c_n) \\
a_2 &= c_1 c_2 + \dots + c_1 c_n + \dots + c_{n-1} c_n \\
\dots &\quad \dots \quad\quad \dots \\
a_k &= (-1)^k \sum_{i_1 < \dots < i_k} c_{i_1} \dots c_{i_k} \\
\dots &\quad \dots \quad\quad \dots \\
a_n &= (-1)^n c_1 \dots c_n
\end{aligned}$$

Эти равенства называются *формулами Виета*. Если $f = a_0 X^n + a_1 X^{n-1} + \dots + a_n$, то формулы Виета принимают вид

$$\frac{a_k}{a_0} = (-1)^k \sum_{i_1 < \dots < i_k} c_{i_1} \dots c_{i_k}, \quad k = 1, 2, \dots, n.$$

7 Многочлены от нескольких переменных

Пусть K — целостное кольцо. Тогда в силу предложения 6.2 кольцо $K[X]$ также будет целостным. Следовательно, для кольца $R = K[X]$ можно построить кольцо $R[Y] = (K[X])[Y]$. Каждый многочлен над R имеет вид

$$f = g_0(X) Y^n + g_1(X) Y^{n-1} + \dots + g_n(X).$$

Представим каждый из многочленов $g_0, g_1, \dots, g_n \in K[X]$ в стандартном виде: $g_k = a_{0k} X^{n_k} + a_{1k} X^{n_k - 1} + \dots + a_{n_k k}$, $k = 0, 1, \dots, n$. Затем подставим полученные выражения в представление многочлена f, раскроем

скобки и приведем подобные слагаемые при соответствующих степенях переменной X. Очевидно, f примет вид

$$f = h_0(Y)X^m + h_1(Y)X^{m-1} + \ldots + h_m(Y),$$

и можно считать, что f есть многочлен от переменной X над кольцом $K[Y]$, то есть $f \in (K[Y])[X]$. Таким образом, нами установлено естественное соответствие между элементами колец $(K[X])[Y]$ и $(K[Y])[X]$, поэтому такие кольца не отличают друг от друга и обозначают через $K[X,Y]$. Построенное кольцо называют *кольцом многочленов над K от коммутирующих переменных X и Y*.

Повторяя описанную выше процедуру произвольное конечное число раз, получаем кольцо $K[X_1, X_2, \ldots, X_n]$ многочленов от коммутирующих переменных $X_1, X_2, \ldots, X_n$. Ясно, что всякий многочлен из $K[X_1, X_2, \ldots, X_n]$ может быть представлен в виде суммы конечного числа слагаемых вида $a_{k_1 \ldots k_n} X_1^{k_1} \ldots X_n^{k_n}$, $k_i \geq 0$, $i = 1, \ldots, n$. Каждое такое слагаемое называется *одночленом* (или *мономом*), число $k_1 + \ldots + k_n$ называется *степенью* монома. *Степень* многочлена есть наибольшая из степеней входящих в него мономов. Если степени всех мономов многочлена совпадают друг с другом, то многочлен называется *однородным*.

Свойства:

1. $\deg(f \pm g) \leq \max\{\deg f, \deg g\}$.

2. $\deg(fg) = \deg f + \deg g$.

3. Если многочлены f, g однородны и $\deg f = \deg g$, то $f \pm g$ — однородный многочлен.

4. Если многочлены f, g однородны, то fg — однородный многочлен.

Доказательство. Установим справедливость свойства 2. Проверка остальных свойств оставляется в качестве упражнения.

Итак, пусть $f, g \in K[X_1, \ldots, X_n]$. Если $f = 0$ или $g = 0$, то доказываемое равенство тривиально верно, поэтому считаем, что $f, g \neq 0$. Очевидно также, что $\deg(fg) \leq \deg f + \deg g$.

Для доказательства обратного неравенства определим множества F_0, $F_1, \ldots$, F_n следующим образом: F_0 есть множество всех мономов многочлена f, степень которых равна $\deg f$, а при $i > 0$ полагаем, что F_i есть множество всех мономов из F_{i-1}, имеющих наибольшую степень по переменной X_i. Аналогично, из мономов многочлена g составим

множества G_0, G_1,..., G_n. Нетрудно понять, что множества F_n и G_n одноэлементны, причем произведение мономов из F_n и G_n не может сократиться ни с каким другим мономом произведения fg. Следовательно, $\deg(fg) \geq \deg f + \deg g$.$\square$

7.1 Симметрические многочлены

Многочлен $f \in K[X_1, \ldots, X_n]$ называется *симметрическим*, если $\alpha \circ f = f$ для всех $\alpha \in S_n$. (Напомним, что $(\alpha \circ f)(X_1, \ldots, X_n) = f(X_{\alpha(1)}, \ldots, X_{\alpha(n)})$.) В частности, симметрическими будут многочлены вида:

$$\sigma_k = \sum_{i_1 < \ldots < i_k} X_{i_1} \ldots X_{i_k}, \quad k = 1, 2, \ldots, n,$$

называемые *элементарными* симметрическими многочленами.

Нетрудно видеть, что если моном $aX_1^{k_1} \ldots X_n^{k_n}$ входит в симметрический многочлен $f \in K[X_1, \ldots, X_n]$, то в f входит и любой моном вида $aX_{\alpha(1)}^{k_1} \ldots X_{\alpha(n)}^{k_n}$, $\alpha \in S_n$. Следовательно, справедливо следующее

Предложение 7.1 *Множество всех симметрических многочленов от переменных X_1,..., X_n образует подкольцо в кольце $K[X_1, \ldots, X_n]$. В частности, сумма и произведение симметрических многочленов — снова симметрические многочлены.*

Опираясь на предложение 7.1, несложно понять, что для всякого многочлена $g \in K[Y_1, \ldots, Y_n]$ многочлен

$$f(X_1, \ldots, X_n) = g(\sigma_1(X_1, \ldots, X_n), \ldots, \sigma_n(X_1, \ldots, X_n))$$

является симметрическим. На самом деле, как будет показано ниже, любой симметрический многочлен может быть получен указанным способом. Предварительно дадим еще несколько определений.

Под *весом* монома $aY_1^{k_1} \ldots Y_n^{k_n}$, $a \neq 0$, понимается число $k_1 + 2k_2 + \ldots + nk_n$. *Весом многочлена* $g(Y_1, \ldots, Y_n)$ называется наибольший из весов входящих в g мономов.

Теорема 7.2 *Для любого симметрического многочлена $f \in K[X_1, \ldots, X_n]$ степени m существует единственный многочлен $g \in K[Y_1, \ldots, Y_n]$ веса m такой, что*

$$f(X_1, \ldots, X_n) = g(\sigma_1(X_1, \ldots, X_n), \ldots, \sigma_n(X_1, \ldots, X_n)).$$

При этом коэффициенты многочлена g являются целочисленными линейными комбинациями коэффициентов многочлена f.

Доказательство. Зададим на множестве всех мономов лексикографический порядок, считая моном $aX_1^{i_1}\dots X_n^{i_n}$ старше монома $bX_1^{j_1}\dots X_n^{j_n}$, $ab\neq 0$, если последовательность $i_1-j_1,\ i_2-j_2,\dots,i_n-j_n$ имеет вид $0,\dots,0,t_k,\dots,t_n$, где $t_k>0$. Самый старший из входящих в многочлен f мономов будем называть *высшим мономом* и обозначать $\mathrm{BM}(f)$. Кроме того, назовем моном $aX_1^{i_1}\dots X_n^{i_n}$ *монотонным*, если $i_1\geq i_2\geq\dots\geq i_n\geq 0$.

Лемма 7.3 *Если $f\in K[X_1,\dots,X_n]$ — симметрический многочлен, то $\mathrm{BM}(f)$ — монотонный.*

Доказательство. Предположим, что $\mathrm{BM}(f)=aX_1^{i_1}\dots X_n^{i_n}$ и существует такое k, что $i_k<i_{k+1}$. В силу симметричности многочлена f он содержит моном $h=aX_1^{i_1}\dots X_k^{i_{k+1}}X_{k+1}^{i_k}\dots X_n^{i_n}$. Но тогда h старше, чем $\mathrm{BM}(f)$, — противоречие.$\square$

Лемма 7.4 *Для любых многочленов $f_1,\dots,f_r\in K[X_1,\dots,X_n]$ верно*

$$\mathrm{BM}(f_1\dots f_r)=\prod_{j=1}^{r}\mathrm{BM}(f_j).$$

Доказательство. Очевидно, достаточно ограничиться случаем, когда $r=2$, другими словами, требуется доказать равенство $\mathrm{BM}(fg)=\mathrm{BM}(f)\mathrm{BM}(g)$. Применим метод математической индукции по количеству n переменных. При $n=1$ утверждение тривиально. Предположим, что для многочленов от $n-1$ переменных лемма уже доказана.

Рассматривая fg как многочлен от переменной X_1 над кольцом $K[X_2,\dots,X_n]$, представим его в виде

$$fg=X_1^s h_0(X_2,\dots,X_n)+X_1^{s-1}h_1(X_2,\dots,X_n)+\dots+h_s(X_2,\dots,X_n).$$

Ясно, что $\mathrm{BM}(fg)=X_1^s\,\mathrm{BM}(h_0)$. Аналогичным образом представим в виде многочленов от X_1 над кольцом $K[X_2,\dots,X_n]$ многочлены f и g. Пусть $h_f(X_2,\dots,X_n)$ и $h_g(X_2,\dots,X_n)$ — их старшие коэффициенты. Нетрудно видеть, что $h_0=h_fh_g$, и поэтому $\mathrm{BM}(h_0)=\mathrm{BM}(h_f)\mathrm{BM}(h_g)$ в силу предположения индукции, поскольку многочлены h_0,h_f,h_g содержат $n-1$ переменных. Таким образом, окончательно получаем

$$\mathrm{BM}(fg)=X_1^s\,\mathrm{BM}(h_0)=X_1^{s_1}\,\mathrm{BM}(h_f)X_1^{s_2}\,\mathrm{BM}(h_g)=\mathrm{BM}(f)\mathrm{BM}(g).\square$$

Существование. Пусть $\mathrm{BM}(f) = aX_1^{i_1} \dots X_n^{i_n}$. В силу леммы 7.3 имеем: $i_1 \geq i_2 \geq \dots \geq i_n \geq 0$. Рассмотрим симметрический многочлен $f_1 = f - a\sigma_1^{i_1-i_2}\sigma_2^{i_2-i_3} \dots \sigma_n^{i_n}$. Легко видеть, что $\mathrm{BM}(f) = \mathrm{BM}(a\sigma_1^{i_1-i_2}\sigma_2^{i_2-i_3} \dots \sigma_n^{i_n})$, и поэтому $\mathrm{BM}(f_1)$ строго меньше, чем $\mathrm{BM}(f)$. Кроме того, заметим, что $\deg(a\sigma_1^{i_1-i_2}\sigma_2^{i_2-i_3} \dots \sigma_n^{i_n}) \leq m$, что с учетом свойства 1 степеней многочленов влечет $\deg(f_1) \leq m$, а вес монома $aY_1^{i_1-i_2}Y_2^{i_2-i_3} \dots Y_n^{i_n}$ равен $i_1 - i_2 + 2(i_2 - i_3) + \dots + ni_n = i_1 + i_2 + \dots + i_n = \deg(f)$ и поэтому не превосходит m. Наконец, отметим, что коэффициенты многочлена f_1 являются целочисленными линейными комбинациями коэффициентов многочлена f.

Повторяя описанную выше процедуру для симметрического многочлена f_1, найдем симметрический многочлен $f_2 = f_1 - b\sigma_1^{j_1-j_2}\sigma_2^{j_2-j_3} \dots \sigma_n^{j_n}$, где $bX_1^{j_1} \dots X_n^{j_n} = \mathrm{BM}(f_1)$ и так далее. Ясно, что на каждом шаге получается симметрический многочлен, высший моном которого строго меньше, чем у предыдущего многочлена. Поскольку всех (а тем более, монотонных) мономов степени не выше m имеется с точностью до коэффициента лишь конечное число, то через некоторое конечное число s шагов получим

$$0 = f_s = f - a\sigma_1^{i_1} \dots \sigma_n^{i_n} - b\sigma_1^{j_1} \dots \sigma_n^{j_n} - \dots - c\sigma_1^{k_1} \dots \sigma_n^{k_n},$$

откуда

$$f = a\sigma_1^{i_1} \dots \sigma_n^{i_n} + b\sigma_1^{j_1} \dots \sigma_n^{j_n} + \dots + c\sigma_1^{k_1} \dots \sigma_n^{k_n} = g(\sigma_1, \dots, \sigma_n),$$

причем вес многочлена $g(Y_1, \dots, Y_n)$ не превосходит m.

Единственность. Предположим, что существуют два различных многочлена $g_1, g_2 \in K[Y_1, \dots, Y_n]$, для которых верно $g_1(\sigma_1, \dots, \sigma_n) = f = g_2(\sigma_1, \dots, \sigma_n)$. Тогда $g = g_1 - g_2$ — ненулевой многочлен от переменных $Y_1, \dots, Y_n$, для которого $g(\sigma_1, \dots, \sigma_n) = 0$. Пусть $aY_1^{\alpha_1} \dots Y_n^{\alpha_n}$ — содержащийся в g моном. Положим $i_1 = \alpha_1 + \dots + \alpha_n$, $i_2 = \alpha_2 + \dots + \alpha_n, \dots, i_n = \alpha_n$. Тогда $aY_1^{\alpha_1} \dots Y_n^{\alpha_n} = aY_1^{i_1-i_2} \dots Y_n^{i_n}$ и $\mathrm{BM}(a\sigma_1^{i_1-i_2} \dots \sigma_n^{i_n}) = aX_1^{i_1} \dots X_n^{i_n}$. Таким образом, каждому входящему в g моному от переменных $Y_1, \dots, Y_n$ отвечает высший моном многочлена от переменных $X_1, \dots, X_n$, причем разным мономам из g соответствуют разные высшие мономы, среди которых найдется самый высший. Но тогда при переходе от $g(Y_1, \dots, Y_n)$ к $g(\sigma_1, \dots, \sigma_n)$ этот самый высший моном ни с каким другим мономом не сократится, так что многочлен $g(\sigma_1, \dots, \sigma_n)$ ненулевой, — противоречие.

Теорема полностью доказана.□

Говорят, что поле P является *расширением* поля K, если $K \subseteq P$.

Следствие 7.5 *Пусть P — расширение поля K, $c_1, \ldots, c_n \in P$ — все корни многочлена $f \in K[X]$ с учетом их кратностей и $h \in K[X_1, \ldots, X_n]$ — произвольный симметрический многочлен. Тогда $h(c_1, \ldots, c_n) \in K$.*

Доказательство. Пусть $f = a_0X^n + a_1X^{n-1} + \ldots + a_n$. Согласно теореме 7.2 найдется многочлен $g \in K[Y_1, \ldots, Y_n]$ такой, что $h = g(\sigma_1, \ldots, \sigma_n)$. Применив формулы Виета, получаем $a_k/a_0 = (-1)^k \sum\limits_{i_1<\ldots<i_k} c_{i_1} \ldots c_{i_k} = (-1)^k \sigma_k(c_1, \ldots, c_n)$, $k = 1, \ldots, n$. Тогда

$$h(c_1, \ldots, c_n) = g(-a_1/a_0, \ldots, (-1)^n a_n/a_0) \in K.\square$$

7.2 Степенные суммы. Формулы Ньютона

Помимо элементарных симметрических многочленов от переменных $X_1, \ldots, X_n$, иногда бывает удобно пользоваться симметрическими многочленами следующего вида:

$$p_k = \sum_{i=1}^{n} X_i^k, \qquad k = 0, 1, 2, \ldots$$

Справедливы следующие соотношения между многочленами p_k и элементарными симметрическими многочленами:

$$\begin{array}{ll} p_k - p_{k-1}\sigma_1 + \ldots + (-1)^{k-1}p_1\sigma_{k-1} + (-1)^k k\sigma_k = 0, & 1 \le k \le n, \\ p_k - p_{k-1}\sigma_1 + \ldots + (-1)^n p_{k-n}\sigma_n = 0, & k > n. \end{array} \tag{28}$$

Эти соотношения называют *формулами Ньютона*. Докажем их. Для этого в кольце многочленов $K[X_1, \ldots, X_n, Y]$ рассмотрим многочлен

$$h = (Y - X_1)(Y - X_2)\ldots(Y - X_n)$$

и разложим его по степеням переменной Y. Согласно формулам Виета получим

$$h = Y^n - \sigma_1(X_1, \ldots, X_n)Y^{n-1} + \ldots + (-1)^n\sigma_n(X_1, \ldots, X_n).$$

Очевидно, при любом $i = 1, 2, \ldots, n$ подстановка X_i вместо Y обращает многочлен h в ноль, поэтому верны равенства

$$0 = X_i^n - \sigma_1(X_1, \ldots, X_n)X_i^{n-1} + \ldots + (-1)^n\sigma_n(X_1, \ldots, X_n).$$

Умножая обе части на X_i^{k-n} (здесь предполагается, что $k \geq n$) и суммируя получающиеся равенства при всех значениях i от 1 до n, выводим

$$0 = p_k - p_{k-1}\sigma_1 + p_{k-2}\sigma_2 + \ldots + (-1)^{n-1}p_{k-n+1}\sigma_{n-1} + (-1)^n p_{k-n}\sigma_n,$$

так что формулы Ньютона справедливы при любом $k > n$, а также при $k = n$, поскольку $p_0 = \sum\limits_{i=1}^{n} X_i^0 = 1 + \ldots + 1 = n$.

Предположим теперь, что $k \leq n$. Рассмотрим следующие однородные симметрические многочлены степени не выше k:

$$f_{k,n} = p_k - p_{k-1}\sigma_1 + p_{k-2}\sigma_2 + \ldots + (-1)^{k-1}p_1\sigma_{k-1} + (-1)^k k\sigma_k.$$

Индукцией по $r = n - k$ покажем, что $f_{k,n} = 0$. При $r = 0$ (то есть при $k = n$) равенство $f_{n,n} = 0$ уже доказано. Пусть $r > 0$ и $f_{k,m} = 0$ при $m - k < r$.

Очевидно, $(p_j)_0 = p_j(X_1, \ldots, X_{n-1}, 0) = p_j(X_1, \ldots, X_{n-1})$ и $(\sigma_j)_0 = \sigma_j(X_1, \ldots, X_{n-1}, 0) = \sigma_j(X_1, \ldots, X_{n-1})$ при любом $j = 1, \ldots, k$, поэтому

$$f_{k,n}(X_1, \ldots, X_{n-1}, 0) = (p_k)_0 - (p_{k-1})_0(\sigma_1)_0 + \ldots + (-1)^k k(\sigma_k)_0 =$$

$$f_{k,n-1}(X_1, \ldots, X_{n-1}) = 0$$

в силу предположения индукции, поскольку $n-1-k < n-k = r$. Таким образом, если представить $f_{k,n}$ в виде многочлена от X_n над кольцом $K[X_1, \ldots, X_{n-1}]$, то его свободный член будет равен нулю и, следовательно, всякий входящий в $f_{k,n}$ моном содержит переменную X_n. Но тогда с учетом того, что многочлен $f_{k,n}$ — симметрический, получаем, что всякий его моном содержит каждую из переменных $X_1, \ldots, X_n$ и, значит, содержит произведение $X_1 \ldots X_n$. В итоге, многочлен $f_{k,n}$ степени не выше k делится на многочлен $X_1 \ldots X_n$ степени $n > k$, что, очевидно, возможно только при $f_{k,n} = 0$.

Заметим, что формулы Ньютона позволяют явным образом выразить степенные суммы через элементарные симметрические многочлены, и наоборот. В самом деле, при $k = 1$ получаем $p_1 - \sigma_1 = 0$, так что $p_1 = \sigma_1$ (впрочем, справедливость последнего равенства совершенно очевидна в силу определения многочленов p_1 и σ_1). При $k = 2$ имеем равенство $p_2 - p_1\sigma_1 + 2\sigma_2 = 0$, из которого (ввиду $p_1 = \sigma_1$) выводим $p_2 = \sigma_1^2 - 2\sigma_2$ и $\sigma_2 = \frac{1}{2}(p_1^2 - p_2)$. Аналогично, пользуясь уже полученными равенствами

и формулами Ньютона при $k = 3$, можно выразить p_3 через σ_1, σ_2 и σ_3, а σ_3 — через p_1, p_2 и p_3, и так далее. В частности, с учетом теоремы 7.2 получаем

Следствие 7.6 *Для любого симметрического многочлена* $f \in K[X_1, \ldots, X_n]$ *существует такой многочлен* $h \in K[X_1, \ldots, X_n]$, *что*

$$f(X_1, \ldots, X_n) = h(p_1(X_1, \ldots, X_n), \ldots, p_n(X_1, \ldots, X_n)).$$

7.3 Алгебраическая замкнутость поля $\mathbb{C}$

В предыдущем разделе нами уже рассматривались неприводимые многочлены над кольцом $\mathbb{Z}$ целых чисел и полем $\mathbb{Q}$ рациональных чисел. В отличие от них неприводимые многочлены над полями $\mathbb{R}$ и $\mathbb{C}$ можно полностью описать, и для этого нам понадобится результат, долгое время носивший название "основной теоремы алгебры". Эта теорема утверждает, что всякий многочлен степени большей или равной 1 с комплексными коэффициентами обладает корнем в $\mathbb{C}$; известны несколько вариантов ее доказательства, и каждый из них в большей или меньшей степени использует те или иные факты, излагаемые в курсах математического анализа, теории функций комплексного переменного и др. Мы приводим здесь один из наиболее "алгебраических" вариантов доказательства, с которым можно ознакомиться также, например, по книге [1, гл. 6].

Предварительно докажем, что всякое поле можно расширить так, чтобы оно содержало все корни заданного многочлена. А именно, справедлива

Теорема 7.7 *Для каждого поля* F *и каждого многочлена* $f \in F[X]$, $\deg f \geq 1$, *существует такое расширение* K, *что* f *разлагается в кольце* $K[X]$ *на линейные множители.*

Доказательство проведем в два этапа.

1. Фиксируем произвольный неприводимый многочлен $g \in F[X]$ и построим для F расширение L, в котором у g есть хотя бы один корень.

Ясно, что если $\deg g = 1$, то в качестве L можно взять само F, поэтому далее полагаем, что $\deg g = n > 1$. Элементами поля L будем считать все многочлены из $F[X]$, степень которых строго меньше n.

Заметим, что имеется естественное сюръективное отображение $F[X] \ni f \mapsto \bar{f} \in L$, где $\bar{f}$ есть остаток от деления f на g. Зададим

на L операции сложения и умножения по правилам: $\bar{f}_1 + \bar{f}_2 = \overline{f_1 + f_2}$ и $\bar{f}_1 \cdot \bar{f}_2 = \overline{f_1 f_2}$. Прежде всего необходимо убедиться в том, что эти операции определены корректно, то есть их результат не зависит от выбора конкретных многочленов f_1 и f_2, дающих при делении на g остатки $\bar{f}_1$ и $\bar{f}_2$.

В самом деле, пусть $h_1, h_2 \in F[X]$ таковы, что $\bar{h}_i = \bar{f}_i$, $i = 1, 2$. Тогда для подходящих $p_1, q_1 \in F[X]$ имеем $f_1 - p_1 g = h_1 - q_1 g$, откуда $h_1 = f_1 + r_1 g$, где $r_1 = q_1 - p_1$; аналогично, $h_2 = f_2 + r_2 g$ для некоторого $r_2 \in F[X]$. Следовательно, верны равенства $h_1 + h_2 = f_1 + f_2 + (r_1 + r_2)g$ и $h_1 h_2 = f_1 f_2 + (f_1 r_2 + f_2 r_1 + r_1 r_2 g)g$, из которых немедленно вытекает, что $\overline{h_1 + h_2} = \overline{f_1 + f_2}$ и $\overline{h_1 h_2} = \overline{f_1 f_2}$.

Заметим, что теперь проверка аксиом, показывающих, что L образует коммутативное ассоциативное кольцо с единицей, не представляет никаких трудностей, поскольку таковым является кольцо $F[X]$ (см. п. 6.1). Например, коммутативность умножения в L следует из цепочки равенств $\bar{f}_1 \cdot \bar{f}_2 = \overline{f_1 f_2} = \overline{f_2 f_1} = \bar{f}_2 \cdot \bar{f}_1$; аналогично проверяются остальные аксиомы. Следовательно, проверка аксиом поля для L сводится лишь к проверке обратимости всякого отличного от $\bar{0}$ элемента $\bar{f} \in L$. Последнее же легко вытекает из следующих рассуждений: так как g неприводим и $g \nmid f$, то $\text{НОД}\,(f, g) = 1$, поэтому ввиду теоремы 6.5 для некоторых $u, v \in F[X]$ верно $1 = uf + vg$, откуда $\bar{1} = \overline{uf} = \bar{u} \cdot \bar{f}$ и, значит, элемент $\bar{f}$ обратим.

Очевидно, элементы поля F под действием отображения $f \mapsto \bar{f}$ не меняются, поэтому вполне естественно для всякого $a \in F$ вместо $\bar{a}$ писать просто a и тем самым считать, что поле F содержится в L. Учитывая последнее, для элемента $\bar{X} \in L$ имеем $g(\bar{X}) = \bar{g} = \bar{0} = 0$, следовательно, $\bar{X}$ является корнем для g.

2. Пусть теперь $f \in F[X]$ — произвольный многочлен степени большей или равной 1. Разложим его на неприводимые множители. Если все они линейны, то в качестве искомого поля K можно взять само F. Если же среди неприводимых множителей есть некоторый многочлен g степени строго большей 1, то согласно первой части доказательства для F найдется расширение L_1, в котором у g есть корень, так что по теореме Безу g приводим в кольце $L_1[X]$. Ясно, что все остальные неприводимые над F множители многочлена f в самом худшем случае останутся

неприводимыми и над L_1, но даже тогда за счет выделения у g линейного множителя общее количество неприводимых множителей в разложении f над L_1 увеличится. Если среди них еще остались многочлены степени строго большей 1, то фиксируем какой-либо из них и, вновь воспользовавшись первой частью доказательства, построим для L_1 расширение L_2 и так далее. Поскольку количество неприводимых множителей в разложении многочлена f не может превышать его степени, через конечное число шагов будет построено такое поле K, что все неприводимые множители в разложении f над K окажутся линейными.$\square$

Лемма 7.8 *Всякий многочлен $f \in \mathbb{R}[X]$ нечетной степени обладает корнем в $\mathbb{R}$.*

Доказательство. Без ограничения общности можно считать многочлен f нормализованным, то есть $f = X^n + a_1X^{n-1} + \ldots + a_n$. Фиксируем число $M = \max\{1, |a_1| + \ldots + |a_n|\}$. Для любого $i > 0$ имеем $M^{n-i} \leq M^{n-1}$, откуда с учетом неравенства треугольника (см. п. 2.3) выводим

$$|a_1M^{n-1} + \ldots + a_n| \leq |a_1|M^{n-1} + \ldots + |a_n| \leq (|a_1| + \ldots + |a_n|)M^{n-1} \leq M^n,$$

следовательно, в силу нечетности n получаем $f(-M) \leq 0$ и $f(M) \geq 0$. Тогда по теореме Коши о промежуточном значении непрерывная функция $f(x)$ обращается в ноль в некоторой точке $c \in [-M, M]$, так что c — корень f.$\square$

Теперь мы можем доказать упомянутую выше ОСНОВНУЮ ТЕОРЕМУ АЛГЕБРЫ:

Теорема 7.9 *Любой многочлен $f \in \mathbb{C}[X]$, $\deg f \geq 1$, обладает корнем в $\mathbb{C}$.*

Доказательство. Пусть $f = a_0X^n + a_1X^{n-1} + \ldots + a_n$. Рассмотрим сначала следующий частный случай:

1. $f \in \mathbb{R}[X]$. Представим число n в виде $n = 2^km$, где $k \geq 0$ и m нечетно, и с помощью индукции по k докажем, что у f есть корень в $\mathbb{C}$. При $k = 0$ это верно ввиду леммы 7.8. Пусть утверждение уже доказано для всех многочленов из $\mathbb{R}[X]$, степени которых содержат множитель 2 с показателем, строго меньшим k.

По теореме 7.7 для $\mathbb{R}$ существует расширение K, содержащее все корни $u_1, \ldots, u_n$ многочлена f. Фиксируем произвольное число $a \in \mathbb{R}$, введем набор переменных $\{X_1, \ldots, X_n\}$ и для всех индексов i, j, где $1 \leq i < j \leq n$, положим $Y_{ij} = X_i X_j + a(X_i + X_j)$. Рассмотрим в кольце $K[X, X_1, \ldots, X_n]$ многочлен

$$g = \prod_{1 \leq i < j \leq n} (X - Y_{ij}). \tag{29}$$

Очевидно, степень g (как многочлена от X) равна $s = n(n-1)/2$.

Нетрудно заметить, что всякая транспозиция τ, действующая на множестве $\{X_1, \ldots, X_n\}$, индуцирует некоторую перестановку элементов $Y_{12}, \ldots, Y_{1n}, \ldots, Y_{n-1,n}$. Согласно формулам Виета коэффициенты $b_1, \ldots, b_s \in K[X_1, \ldots, X_n]$ нормализованного многочлена g с точностью до знака совпадают со значениями соответствующих элементарных симметрических многочленов на элементах $Y_{12}, \ldots, Y_{1n}, \ldots, Y_{n-1,n}$ и, значит, не меняются под действием τ. Тогда ввиду следствия 4.4 коэффициенты $b_1, \ldots, b_s$ не меняются при любой перестановке переменных $X_1, \ldots, X_n$, то есть являются симметрическими. С учетом следствия 7.5 получаем, что числа $b_1(u_1, \ldots, u_n), \ldots, b_n(u_1, \ldots, u_n)$ вещественны, так что $h(X) = g(X, u_1, \ldots, u_n)$ — многочлен степени s с вещественными коэффициентами. Поскольку $\deg h = s = n(n-1)/2 = 2^k m(2^k m - 1)/2 = 2^{k-1}t$, где $t = m(2^k m - 1)$ — нечетно, то по предположению индукции у h есть корень $c \in \mathbb{C}$. С другой стороны, ввиду (29) все корни h содержатся среди чисел вида $a(u_i + u_j) + u_i u_j$, следовательно, $a(u_i + u_j) + u_i u_j \in \mathbb{C}$ для некоторых индексов i, j.

Напомним, что число $a \in \mathbb{R}$ было выбрано произвольным образом. Ясно, что для любого другого числа $a' \in \mathbb{R}$ соответствующие индексы i' и j', для которых верно $a'(u_{i'} + u_{j'}) + u_{i'} u_{j'} \in \mathbb{C}$, вообще говоря, не обязаны совпадать с i и j. Но поскольку для пар индексов имеется лишь s возможностей, а вещественных чисел бесконечно много, то найдутся такие индексы i, j и числа $b, b' \in \mathbb{R}$, $b \neq b'$, что

$$b(u_i + u_j) + u_i u_j,\ b'(u_i + u_j) + u_i u_j \in \mathbb{C},$$

откуда легко выводятся включения $u_i + u_j, u_i u_j \in \mathbb{C}$. Но тогда $u_i - u_j \in \mathbb{C}$, поскольку $(u_i - u_j)^2 = (u_i + u_j)^2 - 4u_i u_j \in \mathbb{C}$. Очевидно, из включений

$u_i + u_j, u_i - u_j \in \mathbb{C}$ вытекает, что числа u_i, u_j также лежат в $\mathbb{C}$. Таким образом, среди корней $u_1, \ldots, u_n$ многочлена f есть хотя бы одно комплексное число.

2. Рассмотрим теперь общий случай, когда $f \in \mathbb{C}[X]$. Обозначим через $\bar{f}$ многочлен, получающийся из f заменой всех коэффициентов на комплексно-сопряженные к ним числа, и положим $g = f\bar{f}$. Очевидно, $\bar{g} = g$, следовательно, $g \in \mathbb{R}[X]$. Согласно первой части доказательства у g есть корень $c \in \mathbb{C}$, так что $f(c)\bar{f}(c) = g(c) = 0$. Если $f(c) = 0$, то c — искомый корень f, а если $\bar{f}(c) = 0$, то корнем f будет число $\bar{c}$, поскольку тогда $f(\bar{c}) = \overline{\bar{f}(c)} = 0$.□

Следствие 7.10 *Всякий многочлен $f \in \mathbb{C}[X]$, $\deg f \geq 1$, разлагается над $\mathbb{C}$ на линейные множители.*

Доказательство. Согласно теореме 7.9 у f есть корень $c_1 \in \mathbb{C}$, так что по теореме Безу $f = (X - c_1)f_1$ для некоторого $f_1 \in \mathbb{C}[X]$. Если $\deg f_1 \geq 1$, то у f_1 есть корень $c_2 \in \mathbb{C}$, поэтому $f = (X - c_1)(X - c_2)f_2$ для некоторого $f_2 \in \mathbb{C}[X]$ и так далее. Ясно, что через конечное число шагов f будет разложен на линейные множители.□

Напомним, что согласно теореме 6.11 количество всех корней многочлена с учетом их кратностей не может превышать его степени, поэтому следствие 7.10 означает, что в поле $\mathbb{C}$ лежит любой корень любого многочлена с коэффициентами из $\mathbb{C}$, иначе говоря, поле $\mathbb{C}$ *алгебраически замкнуто.*

7.4 Неприводимые многочлены над полями $\mathbb{R}$ и $\mathbb{C}$

Доказанное в предыдущем пункте следствие 7.10, а также установленная ранее неприводимость многочленов степени 1 над произвольным полем (см. п. 6.3) позволяют дать описание всех неприводимых многочленов с комплексными коэффициентами:

Следствие 7.11 *Многочлен $f \in \mathbb{C}[X]$ неприводим $\Leftrightarrow \deg f = 1$.*

Для многочленов с вещественными коэффициентами ситуация немногим сложнее. Разумеется, неприводимыми будут все многочлены степени 1. Кроме того, неприводимым будет любой многочлен вида $f = aX^2 + bX + c$, где $b^2 - 4ac < 0$, поскольку в этом случае уравнение

$f(x) = 0$ не имеет вещественных корней и, следовательно, f невозможно разложить в произведение многочленов степени 1. Покажем, что других неприводимых многочленов над $\mathbb{R}$ не существует.

Прежде всего заметим, что для любого $f \in \mathbb{R}[X]$ и всякого $c \in \mathbb{C}$ числа $f(c)$ и $f(\bar{c})$ комплексно сопряжены друг к другу, следовательно, верна

Лемма 7.12 *Если $c \in \mathbb{C}$ — корень многочлена $f \in \mathbb{R}[X]$, то $\bar{c}$ — тоже корень f.*

Предложение 7.13 *Многочлен $f \in \mathbb{R}[X]$ неприводим тогда и только тогда, когда либо* $\deg f = 1$*, либо f — многочлен степени 2, не имеющий вещественных корней.*

Доказательство требуется лишь для части "тогда". Ясно, что если f имеет вещественный корень c, то неприводимым он может быть только в случае ассоциированности с многочленом $X - c$, значит, $\deg f = 1$.

Предположим теперь, что вещественных корней у f нет. Очевидно, тогда $\deg f \geq 2$. Поскольку $\mathbb{R} \subset \mathbb{C}$, на f можно смотреть, как на многочлен над полем $\mathbb{C}$, следовательно, у f есть комплексный корень $c = \alpha + i\beta$ ($\beta \neq 0$). Применяя лемму 7.12, получаем, что $\bar{c} = \alpha - i\beta$ также является корнем f. В силу теоремы Безу f делится на $X - c$ и на $X - \bar{c}$, откуда ввиду взаимной простоты этих линейных многочленов выводим, что f делится на их произведение. Но $g = (X - c)(X - \bar{c}) = (X - \alpha - i\beta)(X - \alpha + i\beta) = X^2 - 2\alpha X + (\alpha^2 + \beta^2)$ — многочлен с вещественными коэффициентами, поэтому из неприводимости f и полученного соотношения $g \mid f$ вытекает ассоциированность f и g.□

7.5 Дискриминант многочлена

Согласно примеру 4.5 функция $\Delta(x_1, \ldots, x_n) = \prod\limits_{1 \leq i < j \leq n} (x_j - x_i)$ является кососимметрической, следовательно, ее квадрат есть симметрическая функция, которой отвечает симметрический многочлен $\Delta^2 = \prod\limits_{1 \leq i < j \leq n} (X_j - X_i)^2$. В силу теоремы 7.2 многочлен Δ^2 можно единственным образом выразить через элементарные симметрические многочлены: $\Delta^2 = \mathrm{Dis}(\sigma_1, \ldots, \sigma_n)$. Многочлен Dis называется *дискриминантом семейства* $\{X_1, \ldots, X_n\}$.

Нетрудно видеть, что многочлен Δ можно записать с помощью *определителя Вандермонда*:

$$\Delta(X_1,\ldots,X_n)=\begin{vmatrix} 1 & 1 & \ldots & 1 \\ X_1 & X_2 & \ldots & X_n \\ X_1^2 & X_2^2 & \ldots & X_n^2 \\ & \ldots & \ldots & \\ X_1^{n-1} & X_2^{n-1} & \ldots & X_n^{n-1} \end{vmatrix}.$$

Обозначив стоящую под знаком определителя матрицу через A и пользуясь свойствами определителя (см. теоремы 5.1 и 5.5), получаем $\mathrm{Dis}(\sigma_1,\ldots,\sigma_n)=(\det A)^2=\det A\cdot\det A^t=\det(AA^t)$, следовательно,

$$\mathrm{Dis}(\sigma_1,\ldots,\sigma_n)=\begin{vmatrix} n & p_1 & \ldots & p_{n-1} \\ p_1 & p_2 & \ldots & p_n \\ p_2 & p_3 & \ldots & p_{n+1} \\ & \ldots & \ldots & \\ p_{n-1} & p_n & \ldots & p_{2n-2} \end{vmatrix},$$

где, напомним, $p_k=\sum_i X_i^k$ — степенные суммы. Применив формулы Ньютона (28) и выразив суммы p_k через элементарные симметрические многочлены, придем к явному выражению для дискриминанта.

Пусть $f=X^n+a_1X^{n-1}+\ldots+a_n$ — нормализованный многочлен над некоторым полем F. По теореме 7.7 для F найдется расширение K, содержащие все корни $c_1,\ldots,c_n$ многочлена f. Тогда с учетом формул Виета имеем:

$$\mathrm{Dis}(\sigma_1(c_1,\ldots,c_n),\ldots,\sigma_n(c_1,\ldots,c_n))=\mathrm{Dis}(-a_1,\ldots,(-1)^na_n).$$

Правая часть последнего равенства называется *дискриминантом многочлена* f и обозначается $\mathrm{D}(f)$. Таким образом, $\mathrm{D}(f)=\Delta^2(c_1,\ldots,c_n)$. Ясно, что $\mathrm{D}(f)=0$ тогда и только тогда, когда среди корней $c_1,\ldots,c_n$ есть хотя бы два одинаковых числа; другими словами, $\mathrm{D}(f)=0\Leftrightarrow$ *у* f *есть кратный корень*.

7.6 Результант многочленов

Пусть даны два многочлена: $f=a_0X^n+a_1X^{n-1}+\ldots+a_n$ и $g=b_0X^m+b_1X^{m-1}+\ldots+b_m$, где $a_0,b_0\neq 0$. Их *результантом* называется

определитель

$$\mathrm{Res}(f,g)=\left|\begin{array}{cccccc}
a_0 & a_1 & \dots & a_n & & \\
 & a_0 & a_1 & \dots & a_n & \\
 & & \dots & & \dots & \\
 & & & a_0 & a_1 & \dots\ a_n \\
b_0 & b_1 & \dots & b_m & & \\
 & b_0 & b_1 & \dots & b_m & \\
 & & \dots & & \dots & \\
 & & & b_0 & b_1 & \dots\ b_m
\end{array}\right\begin{array}{l}\left.\vphantom{\begin{array}{c}a\\a\\a\\a\end{array}}\right\} m \text{ строк}\\ \left.\vphantom{\begin{array}{c}a\\a\\a\\a\end{array}}\right\} n \text{ строк}\end{array}$$

Теорема 7.14 *Следующие условия эквивалентны:*

1) у многочленов f и g есть общий корень;

2) существуют такие многочлены $f_1, g_1 \neq 0$, что $\deg f_1 < n$, $\deg g_1 < m$ и $fg_1 + f_1g = 0$;

3) $\mathrm{Res}(f,g) = 0$.

Доказательство. Введем следующие обозначения: $h = \text{НОД}(f,g)$, $f_1 = c_0X^{n-1} + \ldots + c_{n-1}$ и $g_1 = d_0X^{m-1} + \ldots + d_{m-1}$ — многочлены в условии 2), $w = (d_0, \ldots, d_{m-1}, c_0, \ldots, c_{n-1})^t$ — составленный из их коэффициентов столбец, A — матрица, стоящая под знаком определителя в условии 3).

1) $\Rightarrow$ 2): Так как у f и g есть общий корень, то $\deg h \geq 1$. Поделив f и $-g$ на h, получим $f = f_1h$, $-g = g_1h$, где $\deg f_1 < n$, $\deg g_1 < m$. Очевидно, $fg_1 + f_1g = 0$.

2) $\Rightarrow$ 1): Если $h = 1$, то $fg_1 = -f_1g$ влечет $f \mid f_1$, что невозможно, поскольку $\deg f = n > \deg f_1 \geq 0$. Следовательно, $\deg h \geq 1$, значит, у f и g есть общий корень.

2) $\Rightarrow$ 3): Приравнивая к нулю коэффициенты многочлена fg_1+f_1g, получаем $n+m$ равенств

$$\begin{array}{l}
a_0d_0 + b_0d_0 = 0\\
a_1d_0 + a_0d_1 + b_1c_0 + b_0c_1 = 0\\
a_2d_0 + a_1d_1 + a_0d_2 + b_2c_0 + b_1c_1 + b_0c_2 = 0\\
\quad\dots \qquad\qquad \dots\\
a_nd_{m-2} + a_{n-1}d_{m-1} + b_mc_{n-2} + b_{m-1}c_{n-1} = 0\\
a_nd_{m-1} + b_mc_{n-1} = 0
\end{array}$$

которые можно переписать в матричном виде $A^t w = 0$. Так как $f_1, g_1 \neq 0$, то w есть ненулевое решение однородной системы $A^t x = 0$, следовательно, ее матрица вырождена, поэтому $\operatorname{Res}(f,g) = \det A = \det A^t = 0$.

3) $\Rightarrow$ 2): В силу равенства $\operatorname{Res}(f,g) = 0$ система $A^t x = 0$ имеет некоторое решение $v \neq 0$. Учитывая линейную независимость первых m строк матрицы A и линейную независимость ее последних n строк, получаем, что ненулевые числа есть как среди первых m элементов столбца v, так и среди последних его n элементов. Следовательно, первые m и последние n элементов столбца v можно взять в качестве коэффициентов искомых многочленов g_1 и f_1, соответственно.$\square$

7.7 Метод Штурма отделения вещественных корней

Пусть $f \in \mathbb{R}[X]$, $a, b \in \mathbb{R}$, $a < b$. Рассмотрим следующий вопрос: сколько различных корней многочлена f находятся в интервале $[a,b]$? Без ограничения общности далее считаем, что f не имеет кратных корней, поскольку ввиду равенства (27) задачу всегда можно свести к этому случаю.

Система ненулевых многочленов $\{f_0 = f, f_1, \ldots, f_s\}$ с вещественными коэффициентами называется *системой Штурма* для f в интервале $[a,b]$, если

(i) f_s не имеет корней в $[a,b]$;

(ii) $f_0(a)f_0(b) \neq 0$;

(iii) если $f_k(c) = 0$ для $c \in [a,b]$ и $0 < k < s$, то $f_{k-1}(c)f_{k+1}(c) < 0$;

(iv) если $f_0(c) = 0$ для $c \in (a,b)$, то для некоторого $\varepsilon > 0$ функция $f_0 f_1$ принимает строго отрицательные значения в интервале $(c-\varepsilon, c)$ и строго положительные — в интервале $(c, c+\varepsilon)$.

Предложение 7.15 *Если $f(a)f(b) \neq 0$, то для f в $[a,b]$ существует система Штурма.*

Доказательство. Построим искомую систему, слегка изменив алгоритм Евклида, примененный для нахождения НОД многочлена f и его производной.

Шаг 0: Положим $f_0 = f$, $f_1 = f'$, $k = 1$.

Шаг k: Поделим f_{k-1} на f_k с остатком: $f_{k-1} = q_k f_k - f_{k+1}$, $\deg f_{k+1} < \deg f_k$. Если $f_{k+1} = 0$, то алгоритм завершен, иначе переходим к шагу $k+1$.

По предположению у f нет кратных корней, следовательно, НОД$(f, f') = 1$, поэтому через конечное число шагов алгоритм завершит работу, при этом последний многочлен f_s построенной системы $\{f_0, f_1, \ldots, f_s\}$ будет многочленом степени 0, тем самым для системы выполнено условие (i). Поскольку $f_0 = f$, то выполнено и условие (ii).

Пусть теперь $f_k(c) = 0$ для $c \in [a, b]$ и $0 < k < s$. Согласно алгоритму имеем $f_{k-1} = q_k f_k - f_{k+1}$, следовательно, $f_{k-1}(c) = -f_{k+1}(c)$, так что $f_{k-1}(c) f_{k+1}(c) \leq 0$. Заметим, что равенство $f_{k-1}(c) f_{k+1}(c) = 0$ означало бы, что в системе есть два многочлена с некоторыми номерами $i-1$ и i, для которых c является корнем. Но тогда c является корнем для $f_{i+1}, \ldots, f_s$ в противоречие с условием (i). Следовательно, $f_{k-1}(c) f_{k+1}(c) < 0$, то есть выполнено (iii).

Наконец, пусть $f_0(c) = 0$ для $c \in (a, b)$. Тогда $f_1(c) \neq 0$, поскольку иначе снова приходим к равенствам $f_2(c) = \ldots = f_s(c) = 0$, последнее из которых противоречит условию (i). Следовательно, функция $f' = f_1$ не меняет знак в некоторой окрестности $U = (c - \varepsilon, c + \varepsilon)$ точки c. Если $f_1(c) > 0$, то $f_0 = f$ возрастает на U и поэтому принимает строго отрицательные значения в интервале $(c - \varepsilon, c)$ и строго положительные — в интервале $(c, c + \varepsilon)$. Тогда точно так же ведет себя на U и $f_0 f_1$, поскольку f_1 строго положительна на U. Аналогично рассматривается случай, когда $f_1(c) < 0$.

Итак, все требуемые условия проверены, следовательно, $\{f_0, f_1, \ldots, f_s\}$ есть система Штурма.$\square$

Числом перемен знаков в последовательности $a_1, \ldots, a_n$ ненулевых вещественных чисел называется количество пар ее соседних чисел, имеющих разные знаки. Если в последовательности есть числа, равные 0, то под числом перемен знаков в ней понимается соответствующее число для последовательности, полученной из исходной вычеркиванием всех нулей.

Пусть $\{f_0, f_1, \ldots, f_s\}$ — система Штурма для f в интервале $[a, b]$. Сопоставим каждой точке $c \in [a, b]$ число V_c перемен знаков в последовательности $f_0(c), f_1(c), \ldots, f_s(c)$.

Теорема 7.16 (Штурм) *Количество различных корней многочлена* f

в интервале $[a, b]$ *равно* $V_a - V_b$.

Доказательство. Множество всех чисел, являющихся корнями хотя бы одного из многочленов системы Штурма $\{f_0, f_1, \ldots, f_s\}$, разбивает интервал $[a, b]$ на подынтервалы $[a = a_0, a_1]$, $[a_1, a_2]$,..., $[a_{t-1}, a_t = b]$.

Фиксируем $c \in (a_0, a_1)$. Ясно, что все числа $f_0(c), f_1(c), \ldots, f_s(c)$ отличны от нуля, и если при этом для некоторого k верно $f_k(a) \neq 0$, то знаки чисел $f_k(a)$ и $f_k(c)$ совпадают друг с другом. Если же найдется такое k, что $f_k(a) = 0$, то $0 < k < s$ ввиду условий (i) и (ii), следовательно, числа $f_{k-1}(a)$ и $f_{k+1}(a)$ имеют разные знаки согласно условию (iii), так что в подпоследовательности $f_{k-1}(a), f_k(a), f_{k+1}(a)$ имеется ровно одна перемена знака. Поскольку $f_{k-1}(a), f_{k+1}(a) \neq 0$, числа $f_{k-1}(c)$ и $f_{k+1}(c)$ тоже имеют разные знаки. Легко понять, что при этом и в подпоследовательности $f_{k-1}(c), f_k(c), f_{k+1}(c)$ имеется ровно одна перемена знака вне зависимости от числа $f_k(c)$. В итоге получаем $V_a = V_c$. Аналогично доказывается равенство $V_{c'} = V_b$ для любого $c' \in [a_{t-1}, a_t]$.

Пусть теперь $0 < i < t$, $c' \in (a_{i-1}, a_i)$, $c \in (a_i, a_{i+1})$. При $f_0(a_i) \neq 0$ по аналогии с приведенными выше рассуждениями получаем $V_c = V_{a_i} = V_{c'}$. Наконец, рассмотрим оставшийся случай, когда $f_0(a_i) = 0$. С учетом условия (iv) получаем, что числа $f_0(c')$ и $f_1(c')$ имеют разные знаки, а знаки чисел $f_0(c)$ и $f_1(c)$ совпадают друг с другом. Равенство же чисел перемен знаков в подпоследовательностях $f_1(c'), \ldots, f_s(c')$ и $f_1(c), \ldots, f_s(c)$ ввиду неравенства $f_1(a_i) \neq 0$ устанавливается рассуждениями, аналогичными приведенным выше. Таким образом, $V_{c'} - V_c$ равно 1 при $f_0(a_i) = 0$ и равно 0 при $f_0(a_i) \neq 0$.

Фиксируем точки $c_1 \in (a_0, a_1)$,..., $c_t \in (a_{t-1}, a_t)$. Имеем:

$$V_a - V_b = (V_a - V_{c_1}) + (V_{c_1} - V_{c_2}) + \ldots + (V_{c_t} - V_b).$$

Первая и последняя разности в правой части предыдущего равенства — нулевые, а каждая разность вида $V_{c_i} - V_{c_{i+1}}$ равна 1, если a_i — корень f, и равна 0 в противном случае. Следовательно, $V_a - V_b$ в точности совпадает с числом всех корней f, лежащих в интервале $[a, b]$. $\square$

Замечание. Пусть у f есть кратные корни и $d = \text{НОД}(f, f')$. Применение описанного в доказательстве предложения 7.15 алгоритма дает систему многочленов $\{f_0, f_1, \ldots, f_s\}$, которая, разумеется, не будет системой Штурма для f, но таковой будет система $\{h_0 = f_0/d, h_1 =$

$f_1/d, \ldots, h_s = f_s/d\}$ для многочлена $h = f/d$, имеющего ровно те же корни, что и f. Нетрудно понять, что если $f(c) \neq 0$ для некоторого $c \in \mathbb{R}$, то числа перемен знаков в последовательностях $f_0(c), f_1(c), \ldots, f_s(c)$ и $h_0(c), h_1(c), \ldots, h_s(c)$ равны друг другу, поэтому описанный выше способ вычисления количества различных вещественных корней можно применять и для многочленов, имеющих кратные корни.

Рекомендуемая литература

1. Кострикин, А.И. Введение в алгебру. Ч.1. Основы алгебры. – М.: МЦНМО, 2009.
2. Кострикин, А.И. Введение в алгебру. Ч.2. Линейная алгебра. – М.: МЦНМО, 2009.
3. Винберг, Э.Б. Курс алгебры. – М.: МЦНМО, 2013.
4. Сборник задач по алгебре (под редакцией А.И. Кострикина). – М.: МЦНМО, 2009.
5. Курош, А.Г. Курс высшей алгебры. – СПб.: Лань, 2013.
6. Проскуряков, И.В. Сборник задач по линейной алгебре. – СПб.: Лань, 2010.
7. Фаддеев, Д.К., Соминский, И.С. Задачи по высшей алгебре. – СПб.: Лань, 2008.

Содержание