Министерство образования и науки Российской Федерации
Федеральное агентство по образованию

Ивановский государственный университет

Н. И. Яцкин

АЛГЕБРА

ТЕОРЕМЫ И АЛГОРИТМЫ

***Допущено учебно-методическим советом
по математике и механике
Учебно-методического объединения
по классическому университетскому образованию РФ
в качестве учебного пособия
для студентов высших учебных заведений,
обучающихся по направлению подготовки ВПО
010300.62 «Математика. Компьютерные науки»***

Иваново
Издательство «Ивановский государственный университет»
2006

ББК 22.14
Я 936

Яцкин, Н. И.
Алгебра : Теоремы и алгоритмы : учеб. пособие / Н. И. Яцкин. — Иваново : Иван. гос. ун-т, 2006. — 506 с.

Излагаются основы теории и приводятся указания к практическим и лабораторным занятиям по курсу алгебры и геометрии в рамках следующих тем: введение в линейную алгебру, алгебра комплексных чисел и алгебра многочленов.

Пособие предназначено для студентов вузов, обучающихся по направлению «Математика. Компьютерные науки».

Печатается по решению редакционно-издательского совета Ивановского государственного университета

Рецензенты:

доктор физико-математических наук, профессор **В. Г. Дурнев**
(Ярославский государственный университет)

доктор физико-математических наук, профессор **Б. Я. Солон**
(Ивановский государственный химико-технологический университет)

ISBN 5-7807-0562-3

© Яцкин Н. И., 2006

ОГЛАВЛЕНИЕ

Предисловие 13

Глава 1. СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ И АЛГЕБРА МАТРИЦ 19

§ 1. Системы линейных уравнений и их решения. Матрицы и действия над ними 19
1.1. Развернутая запись системы линейных уравнений 19
1.2. Матрицы 20
1.3. Матрицы-столбцы (арифметические векторы) 22
1.4. Матричная запись для с.л.у. Множество решений с.л.у. 23
1.5. Однородные с.л.у. 25

§ 2. Законы матричной алгебры 25
2.1. Аксиомы поля 25
2.2. Алгебраическая система матриц. Операция транспонирования . . 27
2.3. Законы для алгебраических операций над матрицами 28

§ 3. Свойства решений систем линейных уравнений 36
3.1. Свойства решений однородных и неоднородных с.л.у. 36
3.2. Линейные подпространства пространства $\mathbb{R}^n$ 38
3.3. Подмножества решений однородных и неоднородных с.л.у. 39

§ 4. Равносильные системы линейных уравнений. Элементарные преобразования. Понятие о методе Гаусса 41
4.1. Равносильные с.л.у. 41
4.2. Элементарные преобразования с.л.у. 41
4.3. Расширенная матрица с.л.у. Матричное выражение элементарных преобразований над с.л.у. 42
4.4. Идея метода Жордана — Гаусса (на примере) 43

§ 5. Метод Жордана — Гаусса для матриц 51
5.1. Матрицы ступенчатого вида, вида Жордана — Гаусса, модифицированного вида Жордана — Гаусса, скелетного вида 51
5.2. Теорема Жордана — Гаусса для матриц 53

§ 6. Метод Жордана — Гаусса для систем линейных уравнений . . 56
6.1. Теорема Жордана — Гаусса для с.л.у. 56
6.2. Случай однородной с.л.у. 60

6.3. Случай квадратной с.л.у. Альтернатива Фредгольма 61

§ 7. Некоторые типовые задачи: системы линейных уравнений с параметром, линейные матричные уравнения 62
7.1. С.л.у. с параметром 62
7.2. Линейные матричные уравнения 69

Глава 2. АРИФМЕТИЧЕСКИЕ ЛИНЕЙНЫЕ ПРОСТРАНСТВА И ИХ ПОДПРОСТРАНСТВА. БАЗИСЫ И РАЗМЕРНОСТИ . . 75

§ 8. Системы векторов в пространстве $\mathbb{R}^n$ и их линейные оболочки . 75
8.1. Конечные системы арифметических векторов и соответствующие матрицы 75
8.2. Линейная оболочка конечной с.в. 77
8.3. Критерий принадлежности вектора линейной оболочке системы векторов 79
8.4. Примеры линейных оболочек 81

§ 9. Линейно зависимые и линейно независимые системы векторов 82
9.1. Понятие линейно зависимой (линейно независимой) с.в. 82
9.2. Критерий линейной зависимости (линейной независимости) с.в. . . 83
9.3. Свойства линейно зависимых (линейно независимых) с.в. 84
9.4. Примеры линейно независимых с.в. 86

§ 10. Базисы в линейных подпространствах пространства $\mathbb{R}^n$ 87
10.1. Понятие базиса в линейном подпространстве арифметического линейного пространства 87
10.2. Свойство единственности для разложения по базису 88
10.3. Теорема существования базиса 89
10.4. Свойство продолжения базисов 92

§ 11. Равномощность базисов в подпространстве. Понятие размерности подпространства. Ступенчатый ранг матрицы 93
11.1. Теорема о равномощности всех базисов в данном подпространстве 93
11.2. Понятие размерности для линейного подпространства в пространстве $\mathbb{R}^n$ 96
11.3. Размерность подпространства решений однородной с.л.у. 97
11.4. Ступенчатый ранг матрицы 97

§ 12. Столбцовый и строчный ранги матрицы 98
12.1. Ранг системы векторов и его свойства 98
12.2. Столбцовый и строчный ранги матрицы 99
12.3. Инвариантность столбцового (строчного) ранга при элементарных преобразованиях над строками (столбцами) 100
12.4. Первая теорема о ранге матрицы 102
12.5. Ранг матрицы и исследование с.л.у. (теорема Кронекера — Капелли) 103

§ 13. Алгоритмы построения базисов и вычисления размерностей и рангов 104
13.1. Два способа задания линейных подпространств в пространстве $\mathbb{R}^n$ 104
13.2. Базис и размерность для нуль-пространства матрицы 105
13.3. Базис и размерность для линейной оболочки столбцов матрицы . 107
13.4. Переход от второго способа задания линейных подпространств к первому 110
13.5. Вычисление ранга матрицы, зависящей от параметра 114
13.6. Решение задач с арифметическими векторами средствами системы Maple 116

§ 14. Обратимые квадратные матрицы 119
14.1. Кольцо квадратных матриц заданного размера 119
14.2. Группа обратимых квадратных матриц 121
14.3. Элементарные матрицы и их связь с элементарными преобразованиями 123
14.4. Обратимость элементарных матриц. Выражение с помощью элементарных матриц приводимости прямоугольной матрицы к скелетному виду 127
14.5. Невырожденные матрицы. Первая теорема об условиях обратимости матрицы 128
14.6. Алгоритм Жордана — Гаусса вычисления обратной матрицы . . . 130

§ 15. Линейные операторы в арифметических линейных пространствах 134
15.1. Линейные отображения арифметических линейных пространств и алгебраические действия над ними 134
15.2. Матрица линейного оператора (относительно естественных базисов в арифметических линейных пространствах) 139
15.3. Теорема об изоморфизме для алгебраической системы линейных операторов в арифметических линейных пространствах и алгебраической системы прямоугольных матриц 141
15.4. Законы для алгебраических действий над линейными операторами 145
15.5. Линейные операторы в одном пространстве. Обратимые линейные операторы и обратимые матрицы 147
15.6. Линейные эпиморфизмы и линейные мономорфизмы. Равносильность мономорфности и эпиморфности для эндоморфизмов . . . 148

Глава 3. ТЕОРИЯ ПЕРЕСТАНОВОК 157

§ 16. Перестановки и алгебраические действия над ними 157
16.1. Биекции конечного множества 157
16.2. Умножение перестановок 160
16.3. Тождественная (единичная) перестановка 160
16.4. Обратная перестановка 161
16.5. Группа перестановок 161
16.6. Область действия перестановки. Независимые перестановки . . . 162
16.7. Коммутирующие перестановки 162

16.8. Отображения левого (правого) сдвига и обращения на группе перестановок . . . 163

§ 17. Циклические перестановки. Разложение перестановки в произведение независимых циклов . . . 164
17.1. Циклы и частичные циклы . . . 164
17.2. Теорема о разложении перестановки на независимые циклы . . . 166
17.3. Декремент перестановки . . . 170

§ 18. Степени перестановки. Порядок перестановки . . . 171
18.1. Целые степени перестановки и их свойства . . . 171
18.2. Порядок перестановки . . . 173
18.3. Степени и порядок для циклической перестановки . . . 175
18.4. Вычисление порядка перестановки с помощью ее разложения на независимые циклы . . . 176

§ 19. Разложение перестановки в произведение транспозиций . . . 179
19.1. Разложение циклической перестановки в произведение транспозиций . . . 179
19.2. Разложение произвольной перестановки в произведение транспозиций . . . 180
19.3. Коммутационные соотношения для транспозиций . . . 181
19.4. Теорема о разложениях перестановки на транспозиции . . . 182

§ 20. Знак и четность перестановки . . . 185
20.1. Знак перестановки . . . 185
20.2. Свойства знака . . . 186
20.3. Четность перестановки. Подгруппа четных перестановок . . . 188

§ 21. Число инверсий в перестановке. Второй способ определения знака перестановки . . . 191
21.1. Инверсии в перестановке . . . 191
21.2. Второй способ определения знака перестановки . . . 192

§ 22. Вычисления с перестановками в системе Maple . . . 194
22.1. Форматы задания перестановок . . . 194
22.2. Умножение перестановок в пакете group . . . 196
22.3. Обращение перестановок в пакете group . . . 197
22.4. Вычисление порядка перестановки . . . 197
22.5. Вычисление знака перестановки . . . 198

Глава 4. ТЕОРИЯ ОПРЕДЕЛИТЕЛЕЙ . . . 199

§ 23. Определение определителя квадратной матрицы. Определитель треугольной матрицы. Определитель транспонированной матрицы . . . 199
23.1. Определение определителя . . . 199
23.2. Определители малых порядков . . . 202
23.3. Определитель треугольной матрицы . . . 203
23.4. Определитель транспонированной матрицы . . . 204

§ 24. Определитель квадратной матрицы как полилинейная и антисимметрическая функция ее столбцов (строк) 206
24.1. Функции от векторов-столбцов (векторов-строк) квадратной матрицы 206
24.2. Полилинейность и антисимметричность функции $A \mapsto \det(A)$ . . 208
24.3. Следствия из свойств полилинейности и антисимметричности определителя 211
24.4. Метод Гаусса вычисления определителей 212

§ 25. Вычисление определителя с помощью разложения по столбцу (строке) 213
25.1. Алгебраические дополнения к элементам квадратной матрицы . . 213
25.2. Теорема Лапласа о вычислении определителя разложением по строке (столбцу) 216
25.3. Еще одно свойство алгебраических дополнений 217
25.4. Индуктивный алгоритм вычисления определителя 218

§ 26. Описание всех полилинейных и антисимметрических функций от столбцов (строк) квадратной матрицы 220
26.1. Теорема о полилинейных и антисимметрических функциях от столбцов (строк) квадратной матрицы 220
26.2. Аксиоматический подход к определению определителя 222

§ 27. Определитель блочно-треугольной матрицы. Определитель произведения матриц 223
27.1. Определитель блочно-треугольной матрицы 223
27.2. Мультипликативное свойство определителя 226

§ 28. Присоединенная матрица. Выражение обратной матрицы через присоединенную 228
28.1. Определение и основное свойство присоединенной матрицы . . . 228
28.2. Неособые матрицы. Вторая теорема об условиях обратимости матрицы 229
28.3. Алгоритм вычисления обратной матрицы с помощью присоединенной 231

§ 29. Решение квадратных систем линейных уравнений с помощью обратной матрицы и по формулам Крамера 233
29.1. Решение квадратных с.л.у. с помощью обратной матрицы 233
29.2. Решение квадратных с.л.у. по формулам Крамера 234

§ 30. Минорный ранг матрицы. Вторая теорема о ранге матрицы . 236
30.1. Миноры матрицы 236
30.2. Минорный ранг матрицы 237
30.3. Вторая теорема о ранге матрицы 239
30.4. Метод окаймляющих миноров для вычисления ранга матрицы . . 241

§ 30a. Рекуррентности и определители 243
30a.1. Понятие о рекуррентностях 243
30a.2. Линейные однородные рекуррентности второго порядка 246
30a.3. Последовательность Фибоначчи 249
30a.4. Определители некоторых трехдиагональных матриц 250
30a.5. Определитель Вандермонда 252

Глава 5. ПОЛЕ КОМПЛЕКСНЫХ ЧИСЕЛ 255

§ 31. Векторная модель поля комплексных чисел 255
31.1. Интуитивное представление о комплексных числах. Квадратное уравнение с действительными коэффициентами (случай отрицательного дискриминанта) 255
31.2. Алгебраические действия над комплексными числами 259
31.3. Комплексные числа как двумерные действительные векторы . . . 262
31.4. Матричная модель для поля комплексных чисел 266
31.5. Вычисления с комплексными числами в алгебраической форме . 267

§ 32. Комплексные числа в тригонометрической форме 272
32.1. Геометрическое представление комплексного числа. Модуль, аргумент и тригонометрическая форма комплексного числа 272
32.2. Умножение и деление комплексных чисел в тригонометрической форме. Формула Муавра 281
32.3. Извлечение корней из комплексных чисел в тригонометрической форме 285

§ 33. Корни из единицы. Первообразные корни 292
33.1. Комплексные корни из единицы и их свойства 292
33.2. Первообразные (примитивные) корни из единицы 297
33.3. Использование корней из единицы при вычислении корней из других комплексных чисел 298

§ 34. Показательная функция комплексного аргумента. Показательная форма комплексного числа 299
34.1. Основные понятия математического анализа функций комплексного переменного 299
34.2. Показательная функция (экспонента) комплексного переменного . 303
34.3. Формулы Эйлера 305
34.4. Показательная форма комплексного числа 306
34.5. Натуральный логарифм комплексного числа 307

§ 35. Основная теорема алгебры 308
35.1. Корни многочленов (с комплексными коэффициентами) 308
35.2. Основная теорема алгебры (формулировка, идеи и этапы доказательства) 309
35.3. Основная теорема алгебры (подробное доказательство) 311

ГЛАВА 6. АЛГЕБРА МНОГОЧЛЕНОВ 318

§ 36. Векторная модель кольца многочленов (над полем) 318
36.1. Алгебраическое определение многочлена над полем 318
36.2. Линейное пространство многочленов 323
36.3. Кольцо многочленов 324
36.4. Группа обратимых элементов кольца многочленов 330
36.5. Отношение ассоциированности в коммутативном кольце. Ассоциированные (пропорциональные) многочлены. Нормализованные многочлены 331
36.6. Целостные коммутативные кольца. Целостность кольца многочленов 332

§ 37. Деление с остатком и отношение делимости в кольце многочленов (над полем) 334
37.1. Деление с остатком в кольце многочленов 334
37.2. Отношение делимости в целостном кольце. Делимость (нацело) для многочленов 338

§ 38. Алгоритм Евклида отыскания наибольшего общего делителя в кольце многочленов над полем 340
38.1. Понятие наибольшего общего делителя в целостном кольце. Условие Безу 340
38.2. Существование НОД и алгоритм Евклида для его отыскания в кольце целых чисел 343
38.3. Существование НОД и алгоритм Евклида для его отыскания в кольце многочленов над полем 346
38.4. Отыскание линейного представления для НОД в кольце многочленов методом неопределенных коэффициентов 348
38.5. Взаимно простые элементы в целостном кольце (с условием Безу) 354
38.6. Взаимно простые многочлены 356
38.7. Наименьшее общее кратное. Связь НОК и НОД 358
38.8. Понятие о евклидовых кольцах 362

§ 39. Многочлены и полиномиальные функции. Корни многочленов. Теорема Безу 363
39.1. Полиномиальные функции. Равенство многочленов и равенство полиномиальных функций 363
39.2. Корни многочленов и теорема Безу 368
39.3. Оценка количества (различных) корней многочлена 370
39.4. Многочлены и полиномиальные функции над бесконечным полем 371

§ 40. Кратность корня. Оценка суммы кратностей корней. Алгебраически замкнутые поля. Разложимость многочленов на линейные множители. Теорема Виета 372
40.1. Понятие кратности корня многочлена 372
40.2. Оценка суммы кратностей корней многочлена 373
40.3. Алгебраически замкнутые поля 375

40.4. Разложение многочлена над алгебраически замкнутым полем на линейные множители 376
40.5. Вычисление корней многочленов и разложение многочленов на множители средствами системы Maple 377
40.6. Теорема Виета 380

§ 41. Схема Горнера 383
41.1. Схема Горнера для вычисления значений многочленов 383
41.2. Разложение многочлена по степеням $x - c$ (формула Тейлора) . . 385
41.3. Определение кратности корня многочлена с помощью схемы Горнера 387

§ 42. Рациональные корни многочленов с рациональными коэффициентами 388
42.1. Многочлены с рациональными коэффициентами и многочлены с целыми коэффициентами 388
42.2. Рациональные корни многочлена с целыми коэффициентами . . . 389
42.3. Алгоритм отыскания всех рациональных корней для многочлена с целыми коэффициентами 391

§ 43. Многочлены с действительными коэффициентами и их разложение на линейные и квадратичные множители 395
43.1. Сопряженные многочлены для многочленов с комплексными коэффициентами 395
43.2. Комплексные корни для многочленов с действительными коэффициентами 396
43.3. Разложение многочлена с действительными коэффициентами на линейные и квадратичные множители 398
43.4. Примеры разложения многочленов над полем действительных чисел 399

§ 44. Неразложимые элементы в целостном кольце. Простые элементы. Неприводимые многочлены 401
44.1. Понятие неразложимого элемента в целостном кольце 401
44.2. Понятие простого элемента в целостном кольце 403
44.3. Канонические неразложимые (простые) элементы 405
44.4. Неприводимые многочлены 405
44.5. Неприводимые многочлены над алгебраически замкнутыми полями и над полем действительных чисел 407

§ 45. Факториальные кольца. Факториальность кольца целых чисел (основная теорема арифметики) и факториальность кольца многочленов (над полем) 408
45.1. Определение факториального кольца 408
45.2. Свойства факториальных колец 411
45.3. Достаточные условия факториальности кольца. Факториальность евклидовых колец 413
45.4. Основная теорема арифметики 416

45.5. Факториальность кольца многочленов над полем 417

§ 46. Факториальность кольца многочленов над факториальным кольцом. Неприводимые многочлены над полем $\mathbb{Q}$ и над кольцом $\mathbb{Z}$ 417
46.1. Содержание многочлена над факториальным кольцом. Примитивные многочлены 417
46.2. Неразложимые элементы в кольце многочленов над факториальным кольцом 420
46.3. Лемма Гаусса 420
46.4. Поле частных целостного кольца 423
46.5. Сохранение неприводимости многочленов над факториальным кольцом при переходе к полю частных 424
46.6. Факториальность кольца многочленов над факториальным кольцом 426
46.7. Признак Эйзенштейна неприводимости многочлена над факториальным кольцом 430

§ 47. Дифференцирование в кольце многочленов. Отделение кратных множителей 434
47.1. Понятие характеристики поля 434
47.2. Понятие (формальной) производной от многочлена 436
47.3. Высшие производные и формула Тейлора для многочленов . . . 441
47.4. Кратность корня многочлена и значения производных 443
47.5. Уменьшение кратности неприводимого множителя многочлена при дифференцировании 443
47.6. Отделение кратных неприводимых множителей (кратных корней) 444

§ 48. Первоначальные понятия теории многочленов от нескольких переменных 447
48.1. Мультииндексы и их лексикографическое упорядочение 447
48.2. Многочлены от нескольких переменных. Лексикографическое упорядочение одночленов 452
48.3. Лемма о высшем члене произведения многочленов 457
48.4. Однородные многочлены (формы) 458
48.5. Композиция многочленов (подстановка многочленов в многочлен) 459

§ 49. Симметрические многочлены 459
49.1. Определение симметрического многочлена 459
49.2. Лемма о высшем члене симметрического многочлена 462
49.3. Моногенные симметрические многочлены 463
49.4. Основная теорема о симметрических многочленах 464
49.5. Примеры выражения симметрических многочленов через элементарные симметрические 469
49.6. Значения симметрических многочленов от корней многочлена . . 472
49.7. Дискриминант многочлена (от одной переменной) 474

§ 50. Ферро, Тарталья, Кардано, Феррари и другие 477
50.1. Уравнения малых степеней (над полем $\mathbb{C}$) 477

50.2. Метод Ферро — Тартальи — Кардано решения уравнений третьей степени 479
50.3. Метод Феррари решения уравнений четвертой степени 487

Приложение 1. Рисунки к главе 5 494

Приложение 2. Таблицы к главе 6 499

Список рекомендуемой литературы 505

Предисловие

Перед вами учебное пособие по курсу алгебры, предназначенное для студентов-первокурсников математического факультета, обучающихся по направлению "Математика. Компьютерные науки" (бакалавриат). В этом пособии излагаются основы теории и даются указания к решению и образцы решения типовых задач по следующим темам (им соответствуют главы предлагаемого издания).

1. Системы линейных уравнений и алгебра матриц.
2. Арифметические линейные пространства.
3. Теория перестановок.
4. Теория определителей.
5. Поле комплексных чисел.
6. Алгебра многочленов.

Первая, вторая и четвертая из названных тем представляют собой введение в *линейную алгебру* — алгебраическую науку, одним из объектов изучения которой являются *системы линейных уравнений*. В первом семестре предпринимается "первый подступ" к этой большой и важной математической дисциплине.

Чтобы успешно решать задачи линейной алгебры, надо прежде всего научиться их правильно записывать, выработать адекватный язык, освоить специфические обозначения. С первых же занятий вы познакомитесь с *матрицами* и *арифметическими векторами*, без которых даже постановка указанных задач была бы затруднительной (во всяком случае — очень громоздкой).

В настоящем учебном пособии главным рабочим инструментом при изучении основ линейной алгебры будет так называемый *метод Гаусса* — одно из весьма значимых математических изобретений человечества. (Вооружившись этим методом, вы во втором семестре продолжите изучение предмета на новом уровне.)

Автор надеется, что читатели очень скоро осознают исключительную важность линейной алгебры для компьютерных приложений. К такому выводу их приведут не только математические курсы, но и специальные дисциплины. Невозможно, скажем, глубоко разобрать-

ся в алгоритмах компьютерной графики без основательных познаний в линейной алгебре и геометрии.

Третья, пятая и шестая из перечисленных выше тем относятся к другому разделу алгебраической науки, который принято называть *общей алгеброй*. В этом разделе изучаются множества, на которых заданы *алгебраические действия* (типа сложения и умножения; но только операции, имеющие вполне привычные названия, производятся над объектами более сложными и менее привычными, чем числа). В становлении математика-компьютерщика указанная проблематика призвана играть двойную роль: прежде всего методы общей алгебры облегчают понимание, привносят систематизацию в вопросы линейной алгебры, но, кроме того, знакомство с этими методами стимулирует "культурный рост" компьютерщиков в области теоретической математики. (Замыкаться в прикладных областях, как показывает исторический опыт, не плодотворно.)

В отличие от популярных учебников для первых курсов математических факультетов (см. список литературы) в настоящем издании бóльшее внимание уделяется алгоритмическим вопросам, организации и оформлению вычислений, таким "мелочам", для которых в стандартных курсах места обычно не остается. Автор позволяет себе не стремиться к лаконичности, полагая, что на начальном этапе обучения очень важно все подробно растолковать. Поэтому книжка получилась довольно объемной.

Но пусть данное обстоятельство вас не пугает. Это действительно *книга для чтения*. Она облегчит вам труд конспектирования (весьма напрягающий младшекурсников): теоретическую часть книги можно рассматривать как несколько расширенный *конспект лекций*. (Предостережение: однако все равно никакая книга не заменит живой рассказ лектора, предполагающий обратную связь со слушателями.)

Кроме того, изложение теоретических вопросов в пособии сочетается с разбором типовых задач. Таким образом, данное издание должно послужить не только дополнением к стандартным учебникам, но и *руководством к практикуму*.

Есть и еще одна цель (учитывающая изначальную "компьютерную ориентацию" потенциальных читателей) — подвести студентов к осознанию необходимости самостоятельного изучения и активного использования при решении как рутинных, так и более сложных, исследовательских задач мощных компьютерных математических систем Maple и MATLAB.

Осваивать эти системы необходимо самостоятельно, не дожидаясь, когда до их изучения дойдет черед в дисциплинах компьютерного цикла. (Там вы будете изучать их как профессионалы-программисты, а здесь пока вам предлагается выступать в качестве *квалифицированных пользователей*, разбирающихся в математической части изучаемых задач.)

Но ни в коем случае не следует думать, что достаточно будет научиться правильно ставить задачу и грамотно разбираться в выдаваемых системой ответах. Нет, вы должны будете "влезть внутрь" алгоритма, просчитать множество примеров вручную, чтобы быть способными в случае необходимости видоизменить и модифицировать этот алгоритм.

В списке литературы мы указываем только одну (хорошую) книгу [6] по названным системам. Подобные издания выходят из печати во все возрастающем количестве, что свидетельствует о стремительном прогрессе в развитии компьютерной математики. Вы сможете следить за новинками, посещая в Интернете русскоязычный сайт www.exponenta.ru .

Несколько слов о стиле изложения, принятом в математических учебниках, пособиях и т. п. Как правило, эти тексты содержат "выделенные утверждения": *теоремы*, *предложения*, *леммы*.

Студенты иногда спрашивают: "Какая разница между предложениями и теоремами?" Никакой принципиальной разницы между этими видами утверждений нет. Теоремы обычно являются более сложными или же более важными результатами по сравнению с рядовыми предложениями. Однако это не всегда так. Иногда по традиции и из уважения к их первооткрывателям теоремами именуют довольно простые утверждения. (В работах первооткрывателей доказательства этих теорем могли быть отнюдь не простыми, но при современных подходах доказываемые факты выводятся из ряда других фактов, и, таким образом, изначальная сложность оказывается "размытой".)

Кроме того, в математических текстах довольно часто встречаются так называемые *замечания*, про которые студенты спрашивают: "Зачем они нужны и надо ли их учить?" Чаще всего читатели, целью которых является быстрое освоение материала, "не замечают" этих замечаний. Но при более внимательном, повторном чтении пропущенные ранее замечания могут показаться даже более интересными и важными, чем основной текст, ибо в них часто содержатся сведения о дальнейшем развитии темы, о взаимосвязи между, казалось

бы, далекими направлениями и понятиями, информация о возможных приложениях полученных результатов.

Иногда в замечаниях анализируются различия в *терминологии* и *обозначениях*, имеющиеся в учебной и научной литературе. Для начинающих замечания такого рода вряд ли будут представлять серьезный интерес, они рассчитаны на более опытных читателей, которые, возможно, знакомились с излагаемой теорией по другим источникам.

На первом курсе на юных математиков (недавних школьников) обрушивается лавина информации с совершенно непривычным для них уровнем абстракции. В данном пособии также не удается избежать вопросов, трудных для усвоения. Поэтому автор предлагает читателям не огорчаться, если, скажем, при изучении § 2 они "увязнут" в аксиомах, не понимая пока, кому и зачем они нужны. Смело вперед! Осваивайте практические навыки, алгоритмы, но имейте в виду, что к содержанию этого параграфа вам все равно предстоит вернуться, овладев некоторым математическим опытом.

Учебные книги (пособия, курсы лекций и т. п.) всегда содержат больше материала, чем ваши конспекты, поскольку в лекции не удается "втиснуть" все, что хотелось бы донести до слушателей. Так и в настоящем издании отдельные пункты (например, 15.6) отнюдь не являются обязательными для изучения на начальном, минимальном уровне. Но именно они кажутся автору наиболее полезными для заинтересованных, "въедливых" читателей.

Еще раз подчеркнем: данная книга не заменит стандартные руководства.

Учебники тоже надо читать! В них, помимо "важной информации", которой вы будете пользоваться с утилитарной целью — сдать экзамен или зачет, содержится обычно много "интересной информации", необходимой для выработки вашей общей научной культуры.

Например, в книгах [1 — 3, 5] вы найдете много материала, не относящегося напрямую к "чистой" алгебре, но очень ценного для понимания того, где и как применяется эта наука.

Подводя итог нашим вводным наставлениям, подчеркнем, что направление "Математика. Компьютерные науки" имеет целью подготовку *математиков, работающих в области компьютерных наук.* Это — не компьютерные игры! Это — напряженный, требующий значительных временны́х затрат (но благодарный!) труд.

И, наконец, несколько слов о *теории чисел*, дисциплине, не включенной пока в учебный план по направлению "Математика. Ком-

пьютерные науки". На взгляд автора данного пособия, это является досадным пережитком в эпоху все более широкого проникновения теоретикочисловых методов в компьютерную алгебру.

Впрочем, для первичного ознакомления с началами теории чисел вам достаточно (и необходимо!) заглянуть в учебник [2] (или в более раннее его издание [1]), где в первой, вводной главе вы найдете параграф, посвященный арифметике в кольце целых чисел.

Тем не менее несколько определений и простейших фактов из целочисленной арифметики мы приведем здесь.

Сводка информации из теории чисел

1. К о л ь ц о $\mathbb{Z}$. Множество $\mathbb{Z}$ целых чисел с алгебраическими действиями сложения и умножения является коммутативным кольцом (см. определение кольца в п. 14.1).

Кольцо $\mathbb{Z}$ содержит подмножество $\mathbb{N} = \mathbb{Z}_+$, состоящее из целых положительных (натуральных) чисел.

2. Д е л е н и е с о с т а т к о м . Для любого целого числа s и любого натурального числа m существуют однозначно определенные целые числа q (неполное частное) и r (остаток), такие, что

$$s = m \cdot q + r;\ 0 \leqslant r < m.$$

3. О т н о ш е н и е д е л и м о с т и . Говорят, что число $s \in \mathbb{Z}$ *делит* число $t \in \mathbb{Z}$ (или же: t *делится* на s), если существует $q \in \mathbb{Z}$ такое, что

$$t = s \cdot q.$$

Этот факт обозначается:

$$s \mid t.$$

4. С р а в н и м о с т ь ц е л ы х ч и с е л . Пусть m — натуральное число. Два целых числа s и t называются *сравнимыми по модулю* m, если

$$m \mid t - s.$$

Факт сравнимости обозначается следующим образом:

$$s \equiv t \,(\bmod\ m\,).$$

Сравнимость целых чисел по модулю m равносильна их *равноостаточности* по модулю m, т. е. факту совпадения остатков от деления этих чисел на m.

5. Наибольший общий делитель. *Наибольшим общим делителем* (двух или нескольких) целых чисел называется такое *натуральное* число, которое

1) является общим делителем, т. е. *делит* все данные числа;

2) *делится* на любой общий делитель данных чисел.

Для наибольшего общего делителя чисел $a_1, ..., a_n \in \mathbb{Z}$ используются обозначения НОД$(a_1, ..., a_n)$ или $(a_1, ..., a_n)$.

Числа $a_1, ..., a_n$ называются *взаимно простыми*, если их наибольший общий делитель равен единице.

6. Наименьшее общее кратное. *Наименьшим общим кратным* чисел $a_1, ..., a_n \in \mathbb{Z}$ называется такое *натуральное* число, которое

1) является общим кратным, т. е. *делится* на все данные числа;

2) *делит* любое общее кратное данных чисел.

Используются обозначения НОК$(a_1, ..., a_n)$ или $[a_1, ..., a_n]$.

Глава 1

СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ И АЛГЕБРА МАТРИЦ

§ 1. Системы линейных уравнений и их решения. Матрицы и действия над ними

1.1. Развернутая запись системы линейных уравнений. Одно линейное уравнение с одной неизвестной в множестве действительных чисел $\mathbb{R}$ имеет вид

$$a \cdot x = b, \tag{1.1}$$

где $a, b \in \mathbb{R}$ и неизвестная x также ищется в множестве $\mathbb{R}$.

Множеством решений уравнения (1.1) является следующее подмножество в $\mathbb{R}$:

$$L = \begin{cases} \{\frac{b}{a}\}, & \text{если } a \neq 0; \\ \varnothing, & \text{если } a = 0,\ b \neq 0; \\ \mathbb{R}, & \text{если } a = b = 0. \end{cases}$$

В развернутой записи система m линейных уравнений (с коэффициентами из поля $\mathbb{R}$) с n неизвестными $x_1, x_2, ..., x_n$ выглядит следующим образом:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1; \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = b_m, \end{cases} \tag{1.2}$$

где $a_{ij}, b_i, x_j \in \mathbb{R}$; $i = 1, ..., m$; $j = 1, ..., n$.

Развернутая запись (1.2) может быть представлена в более компактной форме с помощью знака суммирования:

$$\sum_{j=1}^{n} a_{ij}x_j = b_i\,;\ i = 1, ..., m. \tag{1.2a}$$

Решением системы линейных уравнений (с.л.у.) (1.2) называется набор чисел $x_1^0, x_2^0, ..., x_n^0$, при подстановке которых вместо соответствующих неизвестных в уравнения системы (1.2) эти уравнения обращаются в истинные равенства.

1.2. Матрицы. В математике очень важны удачные обозначения, такие, которые помогают пониманию сути задачи и облегчают ее решение. Мы постараемся придать записи (1.2) с.л.у. более лаконичную форму, похожую на форму уравнения (1.1). С этой целью вводятся в рассмотрение прямоугольные таблицы, составленные из чисел, называемые *матрицами*.

Общий вид матрицы размера $m \times n$ таков:

$$\underset{m\times n}{A} = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{m1} & a_{m2} & ... & a_{mn} \end{pmatrix}. \tag{1.3}$$

Обратите внимание на то, что в случае необходимости мы указываем под обозначением матрицы ее размеры.

Введем обозначение $\mathrm{Mat}(m, n; \mathbb{R})$ для множества всех $(m \times n)$-матриц с элементами из множества $\mathbb{R}$. Две матрицы считаются *равными*, если одинаковы их размеры и равны все соответствующие элементы.

Сложение двух матриц (одинакового размера) и *умножение матрицы на число* осуществляется *поэлементно*, т. е.

$$\underset{m\times n}{C} = \underset{m\times n}{A} + \underset{m\times n}{B}\,;\ c_{ij} = a_{ij} + b_{ij};\ i = 1, ..., m; j = 1, ..., n; \tag{1.4}$$

$$\underset{m\times n}{B} = \lambda \cdot \underset{m\times n}{A}\,;\ b_{ij} = \lambda \cdot a_{ij};\ i = 1, ..., m; j = 1, ..., n. \tag{1.5}$$

Умножение двух матриц $A \cdot B$ считается возможным, если число столбцов матрицы A равняется числу строк матрицы B, и определяется следующим образом:

1) если

$$\underset{1\times n}{A} = (\, a_1 \quad a_2 \quad ... \quad a_n \,)$$

— матрица-строка, а

$$\underset{n\times 1}{B} = \begin{pmatrix} b_1 \\ b_2 \\ ... \\ b_n \end{pmatrix}$$

— матрица-столбец ("высота" столбца B равна "длине" строки A), то их произведение полагается равным одноэлементной матрице

$$\underset{1\times 1}{C} = (\, c \,);\ c = a_1b_1 + a_2b_2 + ... + a_nb_n = \sum_{j=1}^{n} a_jb_j;$$

2) в общем случае, если A есть $(m \times n)$-матрица, B – $(n \times p)$-матрица, то их произведение $C = A \cdot B$ будет $(m \times p)$-матрицей, элемент c_{ik} которой определяется как произведение i-й строки матрицы A на k-й столбец матрицы B:

$$c_{ik} = (\, a_{i1} \quad a_{i2} \quad ... \quad a_{in} \,) \cdot \begin{pmatrix} b_{1k} \\ b_{2k} \\ ... \\ b_{nk} \end{pmatrix}.$$

Таким образом,

$$\underset{m\times p}{C} = \underset{m\times n}{A} \cdot \underset{n\times p}{B};\ c_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk};\ i = 1, ..., m;\ k = 1, ..., p. \qquad (1.6)$$

Пример 1.1. Для матриц

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & -5 & 6 \end{pmatrix},\ B = \begin{pmatrix} 0 & 1 & -1 \\ 2 & 2 & 1 \end{pmatrix},\ C = \begin{pmatrix} 0 & 1 \\ 2 & 1 \\ 0 & -1 \end{pmatrix}$$

получим

$$A + B = \begin{pmatrix} 1 & 3 & 2 \\ 6 & -3 & 7 \end{pmatrix},\ (-2) \cdot B = \begin{pmatrix} 0 & -2 & 2 \\ -4 & -4 & -2 \end{pmatrix},$$

$$A \cdot C = \begin{pmatrix} 4 & 0 \\ -10 & -7 \end{pmatrix}.$$

Современные компьютерные математические системы (упомянутые в предисловии) давно уже "изучили" алгебру матриц. Например, система MATLAB (Матричная лаборатория) любую вводимую переменную считает по умолчанию матрицей и готова провести с ней какие угодно вычисления. Разумеется, и вам, чтобы квалифицированно общаться с такими программными средствами, необходимо как можно скорее освоить "матричную грамоту".

1.3. Матрицы-столбцы (арифметические векторы). Из геометрии алгебраисты переняли следующую терминологию: матрицы-столбцы (а также матрицы-строки) называются *арифметическими векторами.* (Действительные числа на том же жаргоне именуются *скалярами.*) Такое заимствование не случайно. Дело в том, что *геометрические векторы* (изучавшиеся в школе и изучаемые в курсе аналитической геометрии), будучи заданными своими координатами, складываются и умножаются на скаляры покомпонентно (т. е. по тому же принципу, что и матрицы в алгебре). В геометрии чаще всего координаты векторов записываются в строчку (и отделяются запятыми); мы будем предпочитать запись векторов в столбик, что диктуется удобством оформления вычислений (читатель вскоре должен в этом убедиться). Упомянем еще об одном отличии между векторами в геометрии и в алгебре: в классической аналитической геометрии изучаются векторы на прямой, на плоскости и в пространстве (или, как говорят, векторы одно-, дву- и трехмерные); в алгебре же мы с самого начала работаем с векторами произвольной размерности.

Арифметическим линейным пространством (размерности n) мы будем называть совокупность всех матриц-столбцов (арифметических векторов-столбцов) заданного размера $n \times 1$. Будет использоваться обозначение

$$\mathbb{R}^n = \mathrm{Mat}(n, 1; \mathbb{R}) = \{\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} \,:\, x_i \in \mathbb{R},\, i = 1, ..., n\}. \qquad (1.7)$$

Замечание 1.1. Несколько слов о важнейшем в нашей науке термине "линейный". Если этот термин применяется к множеству (см. выше), то это значит, что в указанном множестве определены две алгебраические операции (также именуемые линейными): сложение и умножение на число.

Если он применяется к подмножеству, то это означает, что указанное подмножество "устойчиво" относительно названных операций (подробнее см. ниже, в п. 3.2).

Если он применяется к отображению, то такое отображение должно "переводить сумму в сумму" и также "сохранять" произведения на скаляр (подробности — далее, см. п. 15.1).

1.4. Матричная запись для с.л.у. Множество решений с.л.у. Сопоставим с.л.у. (1.2) следующие матрицы:

1) матрицу коэффициентов, размера $m \times n$, вида (1.3);

2) матрицу-столбец (вектор), размера $m \times 1$, правых частей

$$\bar{b} = \begin{pmatrix} b_1 \\ b_2 \\ ... \\ b_m \end{pmatrix} \in \mathbb{R}^m; \tag{1.8}$$

3) матрицу-столбец (вектор), размера $n \times 1$, составленный из неизвестных

$$\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{pmatrix} \in \mathbb{R}^n. \tag{1.9}$$

По правилу умножения матриц (1.6), матрицу A можно умножить на столбец $\bar{x}$, при этом получится матрица-столбец размера $m \times 1$:

$$A \cdot \bar{x} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \\ ..... \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \end{pmatrix},$$

т. е. получится столбец левых частей системы (1.2), а т. к. (1.8) — это столбец правых частей (1.2), то с.л.у. (1.2) может быть записана в следующем (матричном) виде:

$$\underset{m\times n}{A} \cdot \underset{n\times 1}{\bar{x}} = \underset{m\times 1}{\bar{b}}\,. \tag{1.10}$$

Соотношение (1.10) представляет из себя матричное уравнение, равносильное системе (скалярных) уравнений (1.2). При таком подходе следует уточнить данное в п. 1.1 определение решения с.л.у.

Определение 1.1. *Решением* с.л.у. (1.10) называется такой вектор-столбец $\bar{x}^0 \in \mathbb{R}^n$, при подстановке которого вместо неизвестного вектора $\bar{x}$ в (1.10) это уравнение обращается в истинное матричное равенство. Множество решений с.л.у. (1.10) определяется как подмножество в $\mathbb{R}^n$, состоящее из всех решений (1.10):

$$L = \{\bar{x} \in \mathbb{R}^n \ : \ A \cdot \bar{x} = \bar{b}\}. \tag{1.11}$$

Решить с.л.у. это значит найти множество L.

Замечание 1.2. В формуле (1.11) мы не ставим верхний нулик в обозначении решения: символ $\bar{x}^0$ будет обозначать одно (какое-либо конкретное) решение системы, а в (1.11) берутся все решения. Далее будет использоваться еще и такая терминология: конкретные решения с.л.у. будут именоваться ее *частными решениями*; формула, описывающая множество L всех решений с.л.у., будет называться *общим решением* этой системы.

Определение 1.2. 1. С.л.у. (1.10) называется *совместной*, если она имеет хотя бы одно решение, т. е. если $L \neq \varnothing$. В противном случае, т. е. если $L = \varnothing$, с.л.у. называется *несовместной*.

2. Совместная с.л.у. называется *определенной*, если она имеет единственное решение, т. е. если множество ее решений L состоит из единственного элемента (вектора). В противном случае, т. е. если L содержит более одного элемента, с.л.у. называется *неопределенной*.

Пример 1.2. Рассмотрим следующую систему из $m = 3$ линейных уравнений с $n = 4$ неизвестными:

$$\begin{cases} -2x_1 + x_2 - x_3 - 2x_4 = 0; \\ 6x_1 - 3x_2 + 2x_3 + 5x_4 = -1; \\ 4x_1 - 2x_2 + x_3 + 3x_4 = -1. \end{cases}$$

Матричная запись этой системы будет иметь вид

$$\begin{pmatrix} -2 & 1 & -1 & -2 \\ 6 & -3 & 2 & 5 \\ 4 & -2 & 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}.$$

Проверьте (подставив вектор в матричное уравнение), что

$$\bar{x}^0 = \begin{pmatrix} -1/2 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

является частным решением данной системы (и таким образом, эта система совместна); проверьте еще один вектор

$$\bar{x}^1 = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

и убедитесь, что он также удовлетворяет системе (и следовательно, данная система является неопределенной).

1.5. Однородные с.л.у.

Определение 1.3. С.л.у. (1.10) называется *однородной*, если вектор-столбец ее правых частей является нулевым: $\bar{b} = \bar{0}$. Однородная с.л.у.

$$\underset{m\times n}{A} \cdot \underset{n\times 1}{\bar{x}} = \underset{m\times 1}{\bar{0}} \tag{1.10h}$$

называется *соответствующей* с.л.у. (1.10).

Очевидно, что любая однородная с.л.у. является совместной, поскольку она всегда имеет нулевое решение $\bar{x} = \bar{0}$. Нулевое решение однородной с.л.у. принято называть *тривиальным*.

Таким образом, для однородных систем возможны лишь две из трех описанных в предыдущем пункте ситуаций: система может быть определенной (и тогда она имеет только тривиальное решение), либо она является неопределенной (и тогда имеет нетривиальное решение).

§ 2. Законы матричной алгебры

2.1. Аксиомы поля. В множестве действительных чисел $\mathbb{R}$ заданы две основные алгебраические операции: сложение и умножение. Эти операции удовлетворяют следующим законам:

$\boxed{1}$ $\forall a, b, c\ [\,(a+b)+c = a+(b+c)\,]$ — ассоциативность сложения;

$\boxed{2}$ $\forall a, b\ [\,a+b = b+a\,]$ — коммутативность сложения;

$\boxed{3}$ $\exists\, 0\, \forall a\ [\,a+0 = a\,]$ — существование нуля (единственность нуля выводима из $\boxed{2}$, $\boxed{3}$);

$\boxed{4}$ $\forall a\, \exists\, b\ [\,a+b = 0\,]$ — существование противоположного элемента (числа) [его единственность выводима из предыдущих законов; однозначно определенный противоположный элемент обозначается $b = -a$; затем определяется операция вычитания: $c - a = c + (-a)$];

$\boxed{5}$ $\forall a, b, c\ [\,(a+b)\cdot c = a\cdot c + b\cdot c\,]$ — дистрибутивность умножения относительно сложения;

$\boxed{6}$ $\forall a, b, c\ [\,(a\cdot b)\cdot c = a\cdot(b\cdot c)\,]$ — ассоциативность умножения;

$\boxed{7}$ $\forall a, b\ [\,a\cdot b = b\cdot a\,]$ — коммутативность умножения;

$\boxed{8}$ $\exists\, 1\, \forall a\ [\,a\cdot 1 = a\,]$ — существование единицы (единственность единицы выводима из $\boxed{7}$, $\boxed{8}$);

$\boxed{9}$ $\forall a \neq 0\, \exists\, b\ [\,a\cdot b = 1\,]$ — существование обратного элемента (числа) [его единственность выводима из $\boxed{6}$ — $\boxed{8}$; однозначно определенный обратный элемент обозначается $b = a^{-1}$; затем определяется операция деления: $c/a = c\cdot a^{-1}$].

Другие свойства операций сложения и умножения (например: $a\cdot 0 = 0$; $(a-b)\cdot c = a\cdot c - b\cdot c$ и т. п.) выводимы из основных законов $\boxed{1}$ — $\boxed{9}$.

Замечание 2.1. Студент-математик должен постепенно "дозреть" до осознания необходимости доказывать свойства алгебраических операций над числами (даже натуральными). Обычно это достигается на старших курсах, где изучается специальная математическая дисциплина "Числовые системы", в которой прослеживается развитие понятия числа, начиная с множества натуральных чисел, с последовательным построением множеств целых, рациональных, действительных чисел и т. д. Но уже в ближайшее время в курсе алгебры нам встретятся другие *алгебраические системы,* составленные не из чисел, а из каких-либо иных математических объектов и также наделенные алгебраическими операциями (типа сложения и умножения). Для этих операций встанет вопрос о выполнимости законов, аналогичных вышеприведенным законам для $\mathbb{R}$.

Если в алгебраической системе выполняются все законы $\boxed{1}$ — $\boxed{9}$ и, кроме того, $1 \neq 0$, то такая система называется *полем.* Сами эти законы именуются *аксиомами поля.*

Помимо поля $\mathbb{R}$, вам уже знакомо поле рациональных чисел $\mathbb{Q}$. А вот множество целых чисел $\mathbb{Z}$ полем не является. (Почему?) В этой книге до конца четвертой главы мы будем работать "над полем $\mathbb{R}$"; рассматриваемые матрицы и векторы будут иметь действительные элементы. В пятой главе вводится в обращение поле $\mathbb{C}$ *комплексных чисел*. Однако поскольку все доказываемые результаты и все проводимые вычисления опираются лишь на аксиомы поля $\boxed{1}$ — $\boxed{9}$, то они остаются в силе "над любым полем P". Если вы готовы к такому абстрагированию, то можете всюду в дальнейшем (мысленно) заменять букву $\mathbb{R}$ на букву P (где P — произвольное поле). В частности, можно рассматривать арифметические линейные пространства P^n, матрицы с элементами из поля P и т. д.

Условие $1 \neq 0$ равносильно требованию наличия в поле как минимум двух различных элементов: нулевого и единичного. Существует поле, в котором, кроме этих элементов, никаких других больше нет. Обозначается это поле $\mathbb{F}_2$, и арифметика в нем полностью задается определением $1 + 1 = 0$.

(Внимание, компьютерщики! С этим, пока необычным для вас, полем вам предстоит активно работать в теории кодирования.)

2.2. Алгебраическая система матриц. Операция транспонирования. В множестве всех матриц с действительными элементами мы уже ввели три алгебраические операции: сложение матриц (1.4), умножение матрицы на скаляр (1.5) и умножение матриц (1.6). Обратим внимание на то, что (в отличие от поля $\mathbb{R}$) операции сложения и умножения матриц *не всюду определены* (т. е. применимы не к любым парам матриц), а также на то, что при умножении матрицы на скаляр сомножители *не равноправны* (принадлежат различным множествам).

Введем еще одну (четвертую) алгебраическую операцию в множестве матриц (у нее тоже будет своя особенность: она будет действовать на *один аргумент*).

Определение 2.1. Матрицей, *транспонированной* к $(m \times n)$-матрице A, называется $(n \times m)$-матрица (обозначаемая A^t), которая получается из матрицы A превращением строк в столбцы (и наоборот) с сохранением их порядка или, иначе говоря, симметричным отражением матрицы относительно так называемой *главной диагонали* (составленной из элементов матрицы, для которых номер строки и номер столбца совпадают).

Формулой операцию транспонирования можно задать так:

$$\underset{n\times m}{B} = \underset{m\times n}{A}^{\,t};\quad b_{ji} = a_{ij};\ i = 1,...,m;\ j = 1,...,n. \tag{2.1}$$

Пример 2.1 иллюстрирует данное выше определение:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & -5 & 6 \end{pmatrix};\quad B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix};\quad C = (\,-3\,);$$

$$A^t = \begin{pmatrix} 1 & 4 \\ 2 & -5 \\ 3 & 6 \end{pmatrix};\quad B^t = B;\quad C^t = C.$$

Замечание 2.2. Вектор-строку можно представить как транспонированный вектор-столбец:

$$(\,a_1 \quad a_2 \quad ... \quad a_n\,) = \begin{pmatrix} a_1 \\ a_2 \\ ... \\ a_n \end{pmatrix}^t = \bar{a}^t. \tag{2.2}$$

Если нам понадобится различать арифметическое линейное пространство векторов-столбцов (1.7) и аналогичное пространство векторов-строк, то для последнего будет использоваться обозначение:

$$\overset{*}{\mathbb{R}}{}^n = \mathrm{Mat}(1,n;\mathbb{R}) = \{\bar{x}^t = (\,x_1\ x_2\ ...\ x_n\,);\ x_i \in \mathbb{R},\ i = 1,...,n\}. \tag{2.3}$$

2.3. Законы для алгебраических операций над матрицами. Ниже будет сформулирована теорема с перечислением основных законов матричной алгебры. Обращайте внимание на указываемые под символами матриц размеры (числа m, n, p, q и т. д.), которые могут принимать произвольные натуральные значения; скалярные величины обозначаются греческими буквами (λ, μ и т. д.) и могут принимать произвольные действительные значения.

Теорема 2.1. *В множестве всех матриц с элементами из поля* $\mathbb{R}$ *определены: 1) частичная алгебраическая операция* (1.4) — *сложение матриц одинакового размера; 2) операция* (1.5) — *умножение матриц на действительные числа; 3) частичная операция* (1.6) — *умножение матриц согласованных размеров; 4) операция транспонирования матриц* (2.1). *Эти алгебраические операции удовлетворяют следующим законам:*

(i) $\forall\underset{m\times n}{A},\ \underset{m\times n}{B},\ \underset{m\times n}{C}\ [(A+B)+C=A+(B+C)]$;

(ii) $\forall\underset{m\times n}{A},\ \underset{m\times n}{B}\ [A+B=B+A]$;

(iii) $\exists\underset{m\times n}{O}\ \forall\underset{m\times n}{A}\ [A+O=O+A=A]$;

(iv) $\forall\underset{m\times n}{A}\ \exists\underset{m\times n}{B}\ [A+B=O]$;

(v) $\forall\lambda,\mu,\underset{m\times n}{A}\ [(\lambda+\mu)\cdot A=\lambda\cdot A+\mu\cdot A]$;

(vi) $\forall\lambda,\underset{m\times n}{A},\ \underset{m\times n}{B}\ [\lambda\cdot(A+B)=\lambda\cdot A+\lambda\cdot B]$;

(vii) $\forall\lambda,\mu,\underset{m\times n}{A}\ [(\lambda\cdot\mu)\cdot A=\lambda\cdot(\mu\cdot A)]$;

(viii) $\forall\underset{m\times n}{A}\ [1\cdot A=A]$;

(ix) $\forall\underset{m\times n}{A},\ \underset{m\times n}{B},\ \underset{n\times p}{C}\ [(A+B)\cdot C=A\cdot C+B\cdot C]$;

(x) $\forall\underset{m\times n}{A},\ \underset{n\times p}{B},\ \underset{n\times p}{C}\ [A\cdot(B+C)=A\cdot B+A\cdot C]$;

(xi) $\forall\lambda,\underset{m\times n}{A},\ \underset{n\times p}{B}\ [(\lambda\cdot A)\cdot B=A\cdot(\lambda\cdot B)=\lambda\cdot(A\cdot B)]$;

(xii) $\forall\underset{m\times n}{A},\ \underset{n\times p}{B},\ \underset{p\times q}{C}\ [(A\cdot B)\cdot C=A\cdot(B\cdot C)]$;

(xiii) $\exists\underset{n\times n}{E}\ \forall\underset{m\times n}{A},\ \underset{n\times p}{B}\ [(A\cdot E=A)\ \wedge\ (E\cdot B=B)]$;

(xiv) $\forall\underset{m\times n}{A}\ [(A^t)^t=A]$;

(xv) $\forall\underset{m\times n}{A},\ \underset{m\times n}{B}\ [(A+B)^t=A^t+B^t]$;

(xvi) $\forall\lambda,\underset{m\times n}{A}\ [(\lambda\cdot A)^t=\lambda\cdot A^t]$;

(xvii) $\forall\underset{m\times n}{A},\ \underset{n\times p}{B}\ [(A\cdot B)^t=B^t\cdot A^t]$.

Доказательство. 1. Прежде всего заметим, что первые восемь законов, (i) — (viii), относящиеся к первым двум алгебраическим операциям (сложению матриц и умножению матрицы на число; напомним, что они называются линейными), фактически не требуют доказательства. Причиной этого является поэлементный характер линейных операций. Они производятся отдельно в каждой матричной позиции над располагающимися в этой позиции действительными числами, поэтому можно просто сослаться на аксиомы поля $\boxed{1}$ — $\boxed{8}$.

В частности, роль нулевого элемента (в каждом размере) играет, очевидно, матрица, составленная только из нулей, а чтобы найти матрицу, противоположную данной, надо поменять знаки перед всеми элементами данной матрицы. Заметим также, что неравноправность сомножителей при умножении матрицы на число приводит к необходимости отдельной формулировки двух дистрибутивных законов (v) и (vi).

2. Обратимся теперь к дистрибутивным законам (ix) — (x). Их тоже два, но по другой причине: как будет показано ниже, умножение матриц не удовлетворяет коммутативному закону. Доказательство мы приведем лишь для первого из этих законов, оставив второе (вполне аналогичное) доказательство в качестве задания для читателей.

Чтобы не вводить новые буквы для промежуточных результатов алгебраических действий, нам будет удобно применять символ *извлечения элемента указанной позиции из матрицы*; для $(m \times n)$-матрицы (1.3) положим:

$$[A]_{ij} = a_{ij};\ i = 1, ..., m;\ j = 1, ..., n. \tag{2.4}$$

Далее заметим, что размеры матриц в правой и левой частях доказываемого равенства совпадают (и равны $m \times p$). Теперь требуется доказать совпадение всех соответствующих элементов двух этих матриц:

$$\begin{aligned}
&[(A+B)\cdot C]_{ik} \overset{(1.6)}{=\!=} \sum_{j=1}^{n} [A+B]_{ij} c_{jk} \overset{(1.4)}{=\!=} \sum_{j=1}^{n} (a_{ij}+b_{ij}) c_{jk} \overset{\boxed{5}}{=\!=} \\
&= \sum_{j=1}^{n} (a_{ij}c_{jk} + b_{ij}c_{jk}) \overset{\boxed{1},\boxed{2}}{=\!=\!=} \sum_{j=1}^{n} a_{ij}c_{jk} + \sum_{j=1}^{n} b_{ij}c_{jk} \overset{(1.6)}{=\!=} \\
&= [A\cdot C]_{ik} + [B\cdot C]_{jk} \overset{(1.4)}{=\!=} [A\cdot C + B\cdot C]_{ik},
\end{aligned}$$

где $i = 1, ..., m;\ k = 1, ..., p$.

Обратите внимание на то, что точки (обозначающие различные умножения) часто опускаются, а также на стиль оформления преобразований, когда над знаками равенства в цепочке вычислений мы указываем номера формул (в качестве ссылок на используемые определения и законы). Например, над четвертым равенством в цепочке мы даем ссылку на ассоциативный и коммутативный законы для

сложения действительных чисел; именно они позволяют *перегруппировать* сумму слева, представив ее как сумму двух сумм справа от этого знака равенства.

Докажем теперь закон (xi), словесно выражаемый следующим правилом: *скалярные множители можно выносить как из первого, так и из второго сомножителя в произведении матриц.*

Это правило в сочетании с двумя предыдущими дистрибутивными законами носит специальное название: *свойство билинейности* для произведения матриц. Само слово "билинейность" понимается как *линейность по каждому из двух аргументов (сомножителей).* А линейность, скажем, по первому сомножителю означает (в соответствии с замечанием 1.1), что при фиксированном втором сомножителе B отображение $A \mapsto A \cdot B$ "сохраняет" суммы и произведения на скаляр.

Очевидно, размеры матриц во всех трех частях формулы (xi) совпадают и равны $m \times n$. Ниже приводится доказательство лишь одного из двух составляющих эту формулу равенств (второе равенство доказывается совершенно аналогично):

$$
\begin{aligned}
&[(\lambda \cdot A) \cdot B]_{ik} \overset{(1.6)}{=\!=} \sum_{j=1}^{n} [\lambda \cdot A]_{ij} b_{jk} \overset{(1.5)}{=\!=} \\
&= \sum_{j=1}^{n} (\lambda a_{ij}) b_{jk} \overset{\boxed{6}}{=\!=} \sum_{j=1}^{n} \lambda (a_{ij} b_{jk}) \overset{\boxed{5}}{=\!=} \lambda \sum_{j=1}^{n} a_{ij} b_{jk} \overset{(1.6)}{=\!=} \\
&= \lambda [A \cdot B]_{ik} \overset{(1.5)}{=\!=} [\lambda \cdot (A \cdot B)]_{ik},
\end{aligned}
$$

где $i = 1, ..., m$; $k = 1, ..., p$. Внимательный читатель, по-видимому, сам разберется в используемых законах и определениях. Поясним только четвертое равенство в цепочке. Здесь используется дистрибутивный закон в $\mathbb{R}$ в форме *правила вынесения постоянного множителя из-под знака суммы*:

$$
\sum_{i=1}^{n} \lambda a_i = \lambda \sum_{i=1}^{n} a_i. \tag{2.5}
$$

Выражение "постоянный множитель" следует понимать как *одно и то же число, присутствующее множителем в каждом из слагаемых суммы.* Признаком "постоянства" служит отсутствие у

этого множителя индекса i, по которому проводится суммирование. Подчеркнем для дальнейшего, что у выносимого из-под знака суммы (вносимого под знак суммы) множителя могут быть другие индексы, отличные от индекса, по которому проводится суммирование.

3. Докажем теперь ассоциативность матричного умножения, т. е. закон (xii). Ключевым звеном в цепочке преобразований будет *перемена порядка суммирования* в двойной сумме, которая производится в соответствии со следующим "бухгалтерским" правилом перегруппировки слагаемых в сумме по двум индексам:

$$\sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}\right) = \sum_{j=1}^{n}\left(\sum_{i=1}^{m} a_{ij}\right). \tag{2.6}$$

Сумма сумм по строкам (в заполняемой бухгалтером таблице) должна сходится с суммой сумм по столбцам. Это правило следует из коммутативности и ассоциативности сложения чисел. Его справедливость позволяет в принципе не ставить в двойных суммах скобки.

Понадобится также правило (2.5) и следующее за ним пояснение.

Как обычно, рассуждение начинается с того, что мы убеждаемся в совпадении размеров матриц в левой и правой частях (xii): матрица $A \cdot B$ имеет размеры $m \times p$, матрица $B \cdot C$ — размеры $n \times q$, обе части равенства являются $(m \times q)$-матрицами. Проверяем совпадение соответствующих элементов (индексы i, j, k, l будут изменяться в пределах от 1 до m, n, p, q соответственно):

$$\begin{aligned}
[(A \cdot B) \cdot C]_{il} &\overset{(1.6)}{=\!=} \sum_{k=1}^{p} [A \cdot B]_{ik} c_{kl} \overset{(1.6)}{=\!=} \\
&= \sum_{k=1}^{p}\left(\sum_{j=1}^{n} a_{ij} b_{jk}\right) c_{kl} \overset{(2.5)}{=\!=} \sum_{k=1}^{p}\left(\sum_{j=1}^{n} a_{ij} b_{jk} c_{kl}\right) \overset{(2.6)}{=\!=} \\
&= \sum_{j=1}^{n}\left(\sum_{k=1}^{p} a_{ij} b_{jk} c_{kl}\right) \overset{(2.5)}{=\!=} \sum_{j=1}^{n} a_{ij}\left(\sum_{k=1}^{p} b_{jk} c_{kl}\right) \overset{(1.6)}{=\!=} \\
&= \sum_{j=1}^{n} a_{ij} [B \cdot C]_{jl} \overset{(1.6)}{=\!=} [A \cdot (B \cdot C)]_{il}.
\end{aligned}$$

4. Докажем закон (xiii), указав *единичные матрицы* (обладающие свойствами единиц в матричной алгебре). Такими матрицами

являются квадратные матрицы вида

$$\underset{n\times n}{E} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}. \tag{2.7}$$

Элементы матрицы E имеют специальное обозначение и имя — они образуют так называемый *символ Кронекера*:

$$[E]_{ij} = \delta_{ij}; \quad \delta_{ij} = \begin{cases} 0, & \text{если } i \neq j; \\ 1, & \text{если } i = j, \end{cases} \tag{2.8}$$

где $i, j = 1, \dots, n$.

Снова ограничимся доказательством одного из двух равенств, входящих в (xiii). Размеры матриц слева и справа от знака равенства одинаковы; проверим равенство соответствующих элементов:

$$[A \cdot E]_{ik} \overset{(1.6)}{=\!=} \sum_{j=1}^{n} a_{ij}\delta_{jk} \overset{(2.8)}{=\!=} a_{ik} \overset{(2.4)}{=\!=} [A]_{ik},$$

где $i = 1, \dots, m$; $k = 1, \dots, n$ (поясним, что в сумме, выписанной выше, лишь одно из значений символа Кронекера отлично от нуля, вот почему от этой суммы осталось одно слагаемое).

5. В заключение нам предстоит доказать четыре закона для операции транспонирования. Но закон (xiv) совершенно очевиден, а доказательство законов (xv) и (xvi) должно составить простейшее упражнение для читателей (после целого ряда разобранных выше более сложных доказательств).

Кстати, последние законы выражают факт линейности отображения транспонирования $A \mapsto A^t$; см. замечание 1.1.

Обратимся к проверке равенства (xvii). Обе части этого равенства являются матрицами размера $p \times n$; проверим равенство соответствующих элементов:

$$\begin{aligned} &[(A \cdot B)^t]_{ki} \overset{(2.1)}{=\!=} [A \cdot B]_{ik} \overset{(1.6)}{=\!=} \\ &= \sum_{j=1}^{n} a_{ij}b_{jk} \overset{(2.1)}{=\!=} \sum_{j=1}^{n} [A^t]_{ji}[B^t]_{kj} \overset{\boxed{7}}{=\!=} \\ &= \sum_{j=1}^{n} [B^t]_{kj}[A^t]_{ji} \overset{(1.6)}{=\!=} [B^t \cdot A^t]_{ki}, \end{aligned}$$

где $i = 1, ..., m$; $k = 1, ..., p$ и в цепочке равенств использованы лишь определения умножения и транспонирования матриц, а также (на предпоследнем шаге) коммутативный закон для умножения в поле $\mathbb{R}$ (мы переставили числовые множители под знаком суммы). $\square$

Замечание 2.3. Как и в случае аксиом поля, основные законы алгебры матриц (i) — (xvii) влекут множество других, например: $\lambda \cdot O = O$; $0 \cdot A = O$ (где во втором равенстве "два разных нуля": 0 — число и O — матрица) и т. п.

Замечание 2.4. Если рассматривать множество $\mathrm{Mat}(m, n, \mathbb{R})$ матриц фиксированного размера, то на нем (при $m \neq n$) будут определены лишь две операции (сложение и умножение на скаляр). В частности, так будет обстоять дело в случае арифметического линейного пространства $\mathbb{R}^n$. Совокупность законов, действующих в этой ситуации, сводится к законам (i) — (viii).

Ассоциативность сложения обеспечивает корректность определения сумм более чем двух слагаемых, а вся совокупность законов (i) — (viii) обеспечивает возможность образования так называемых *линейных комбинаций* матриц (векторов), т. е. сумм нескольких матриц (векторов) с числовыми коэффициентами вида

$$B = \sum_{i=1}^{s} \lambda_i \cdot A_i; \ \bar{b} = \sum_{i=1}^{s} \lambda_i \cdot \bar{a}_i \,. \tag{2.9}$$

Замечание 2.5. Остановимся на вопросе о выполнимости для матричного умножения коммутативного закона. Если матрица A имеет размеры $m \times n$, а матрица B — размеры $n \times p$, то определено произведение $A \cdot B$, в то время как произведение $B \cdot A$ будет определено лишь при дополнительном условии $p = m$, причем и в этом случае размеры матриц $A \cdot B$ и $B \cdot A$ будут, вообще говоря, различными: первая из них будет квадратной $(n \times n)$-матрицей, а вторая — квадратной $(m \times m)$-матрицей. И даже при дополнительном предположении $m = n$, т. е. при условии, что данные матрицы являются квадратными матрицами одинакового размера, коммутативность умножения не обязана иметь место. Попробуйте сами вычислить и сравнить два произведения $A \cdot B$ и $B \cdot A$ для следующих (2×2)-матриц:

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

Разумеется, для (1×1)-матриц (которые отождествляются с действительными числами) справедливы все аксиомы поля $\boxed{1}$ — $\boxed{9}$.

Замечание 2.6. Приведем небольшую информацию о языке математической компьютерной системы Maple. Подробнее вы сможете ознакомиться с синтаксисом этой системы по обширной, непрерывно обновляющейся литературе.

Здесь мы будем считать, что основные понятия этой системы вы уже освоили, и расскажем о некоторых командах пакета linalg, позволяющих вводить матрицы и выполнять над ними простейшие операции.

Прежде всего, в начале *сессии* упомянутый пакет надо "подгрузить":

```
> with ( linalg ) ;
```

Maple "расскажет" вам, что он "умеет": выведет имена доступных в используемом пакете команд, всего их порядка сотни; по поводу каждой из них можно будет использовать help.

Теперь введем матрицы (из примера 2.1):

```
> A := matrix ( [ [ 1, 2, 3 ] , [ 4, −5, 6 ] ] ) ;
```

$$A := \begin{bmatrix} 1 & 2 & 3 \\ 4 & -5 & 6 \end{bmatrix}$$

```
> B := matrix ( [ [ 0, 1, −1 ] ,[ 2, 2, 1 ] ] ) ;
```

$$B := \begin{bmatrix} 0 & -1 & 1 \\ 2 & 2 & 1 \end{bmatrix}$$

```
> C := matrix ( [ [ 0, 1 ] , [ 2, 1 ] , [ 0, −1 ] ] ) ;
```

$$C := \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 0 & -1 \end{bmatrix}$$

Аргументом команды matrix служит *массив*, вводимый (в круглых скобках) как *список списков*. Подробнее: (двумерный) массив — это список строк, каждая из которых сама является списком. Матрица — это особый вид двумерного массива, в котором нумерация строк и столбцов начинается с 1; вместо круглых скобок Maple заключает матрицы и векторы в квадратные.

Вычислим линейную комбинацию матриц A и B:

> M := evalm (2 * A + 3 * B) ;

$$M := \begin{bmatrix} 2 & 7 & 3 \\ 14 & 4 & 15 \end{bmatrix}$$

Обратите внимание на важнейшую команду evalm (вычислить матрицу). Если A — (ранее введенная) матрица, то простое *присваивание*

> A1 := A ;

лишь "резервирует имя"; по-настоящему матрица $A1$ заполняется, если дана команда:

>A1 := evalm (A) ;

А теперь перемножим матрицы A и C двумя способами (размеры матриц позволяют в данном случае вычислить как $A \cdot C$, так и $C \cdot A$).

> P := evalm (A &* C) ; Q := evalm (C &* A) ;

$$P := \begin{bmatrix} 4 & 0 \\ -10 & -7 \end{bmatrix}$$

$$Q := \begin{bmatrix} 4 & -5 & 6 \\ 6 & -1 & 12 \\ -4 & 5 & -6 \end{bmatrix}$$

Здесь обратите внимание на особый знак *некоммутативного умножения* &* .

Покажем еще операцию транспонирования:

> S := transpose (A) ;

$$S := \begin{bmatrix} 1 & 4 \\ 2 & -5 \\ 3 & 6 \end{bmatrix}$$

§ 3. Свойства решений систем линейных уравнений

3.1. Свойства решений однородных и неоднородных с.л.у. Применим законы алгебры матриц, доказанные в теореме 2.1, для установления свойств решений (однородных и неоднородных) с.л.у.

Рассмотрим с.л.у. (1.10) и соответствующую однородную с.л.у. (1.10h).

Предложение 3.1. *1. Сумма двух решений однородной системы* (1.10h) *снова является решением этой системы.*

2. При умножении решения однородной системы на любой скаляр снова получается решение этой системы.

3. Сумма решения неоднородной системы (1.10) *и решения соответствующей однородной системы* (1.10h) *является решением* (1.10).

4. Разность двух решений системы (1.10) *является решением соответствующей однородной системы* (1.10h).

Доказательство. 1. Пусть векторы $\bar{u}$ и $\bar{v}$ являются решениями системы (1.10h), т. е. $A \cdot \bar{u} = \bar{0}$ и $A \cdot \bar{v} = \bar{0}$. Тогда

$$A \cdot (\bar{u} + \bar{v}) \overset{\text{(x)}}{=} A \cdot \bar{u} + A \cdot \bar{v} = \bar{0} + \bar{0} \overset{\text{(iii)}}{=\!=} \bar{0},$$

т. е. вектор $\bar{u} + \bar{v}$ также является решением (1.10h).

2. Аналогично, если $A \cdot \bar{u} = \bar{0}$ и $\lambda \in \mathbb{R}$, то

$$A \cdot (\lambda \cdot \bar{u}) \overset{\text{(xi)}}{=} \lambda \cdot (A \cdot \bar{u}) = \lambda \cdot \bar{0} = \bar{0},$$

т. е. вектор $\lambda \cdot \bar{u}$ является решением (1.10h).

3. Пусть вектор $\bar{u}$ является решением (1.10), а вектор $\bar{v}$ — решением (1.10h), т. е. $A \cdot \bar{u} = \bar{b}$ и $A \cdot \bar{v} = \bar{0}$. Тогда

$$A \cdot (\bar{u} + \bar{v}) \overset{\text{(x)}}{=} A \cdot \bar{u} + A \cdot \bar{v} = \bar{b} + \bar{0} \overset{\text{(iii)}}{=\!=} \bar{b},$$

т. е. вектор $\bar{u} + \bar{v}$ является решением (1.10).

4. Аналогично, если $A \cdot \bar{u} = \bar{b}$ и $A \cdot \bar{v} = \bar{b}$, то

$$A \cdot (\bar{u} - \bar{v}) \overset{\text{(x)},\text{(xi)}}{=\!=\!=} A \cdot \bar{u} + (-1) \cdot A \cdot \bar{v} = \bar{b} - \bar{b} = \bar{0},$$

т. е. вектор $\bar{u} - \bar{v}$ является решением (1.10h). □

Замечание 3.1. Из доказанного выше предложения вытекает такой факт: если с.л.у. имеет два различных решения, то она имеет бесконечно много решений.

В самом деле, если $\bar{x}^0$ и $\bar{x}^1$ — два различных решения системы (1.10), то их разность $\bar{x}^1 - \bar{x}^0$ будет (см. п. 4 предложения 3.1) решением (1.10h), а значит (см. п. 2), при любом $\lambda \in \mathbb{R}$ вектор $\lambda(\bar{x}^1 - \bar{x}^0)$ будет решением (1.10h). Так как $\bar{x}^1 - \bar{x}^0 \neq \bar{0}$, то все такие векторы попарно различны. Таким образом, однородная система уже имеет

бесконечно много решений. Но тогда (см. п. 3) и исходная неоднородная система будет иметь бесконечно много попарно различных решений вида

$$\bar{x}^0 + \lambda(\bar{x}^1 - \bar{x}^0) = (1-\lambda)\bar{x}^0 + \lambda\bar{x}^1.$$

Замечание 3.2. Сделаем (на будущее) оговорку: все сказанное в данном замечании существенно опирается на *бесконечность* нашего основного поля $\mathbb{R}$. В дальнейшем студентам-компьютерщикам, которые будут изучать теорию кодирования, предстоит знакомство с *конечными полями* (см. замечание 2.1). В случае конечного поля коэффициентов утверждения замечания 3.1 потеряют силу, поскольку сами арифметические линейные пространства, а следовательно, и их подмножества будут конечными множествами.

Далее результатам этого пункта мы придадим иную форму с помощью важного понятия линейного подпространства в арифметическом линейном пространстве.

3.2. Линейные подпространства пространства $\mathbb{R}^n$. Рассмотрим арифметическое линейное пространство $\mathbb{R}^n$.

Определение 3.1. Непустое подмножество $V \subseteq \mathbb{R}^n$ называется линейным подпространством в $\mathbb{R}^n$, если V *устойчиво* (или, как иначе выражаются, *замкнуто*) относительно линейных операций (сложения и умножения на скаляр), т. е. справедливы следующие два свойства:

$$(\,\bar{x},\, \bar{y} \in V\,) \;\Rightarrow\; (\,x + y \in V\,); \tag{3.1}$$

$$(\forall \lambda \in \mathbb{R})\;[\,(\,\bar{x} \in V\,) \;\Rightarrow\; (\,\lambda\bar{x} \in V\,)\,]. \tag{3.2}$$

Из определения линейного подпространства следует, что вместе с любым набором векторов оно должно содержать произвольную линейную комбинацию векторов, входящих в этот набор. Кроме того, очевидно, что всякое подпространство должно содержать нулевой вектор пространства $\mathbb{R}^n$ (в самом деле, взяв любой вектор $\bar{a} \in V$, получим $\bar{0} = 0 \cdot \bar{a} \in V$).

Подмножество $O = \{\bar{0}\}$, состоящее из одного нулевого вектора, является наименьшим из подпространств пространства $\mathbb{R}^n$; само это пространство является своим подпространством, причем наибольшим из всех подпространств.

Замечание 3.3. При изучении с.л.у. нам встретятся также "сдвинутые подпространства", т. е. подмножества в $\mathbb{R}^n$ следующего вида:

$$M = V + \bar{x}^0,$$

где V — подпространство в $\mathbb{R}^n$, а $\bar{x}^0 \in \mathbb{R}^n$.

(Сумма подпространства V и вектора $\bar{x}^0$, стоящая в правой части равенства, определяющего M, понимается как подмножество в $\mathbb{R}^n$, состоящее из всевозможных векторов вида $\bar{y} + \bar{x}^0$, где вектор $\bar{y}$ пробегает все V.)

При $\bar{x}^0 \in V$ оказывается, что $V + \bar{x}^0 = V$. (Докажите!)

А при $\bar{x}^0 \notin V$ подмножество M линейным подпространством уже не является. (Почему?)

У математиков для "сдвинутых подпространств" имеется особый термин — *аффинные подпространства*. Так называются подмножества в $\mathbb{R}^n$, содержащие вместе с любыми двумя векторами $\bar{x}^0$ и $\bar{x}^1$ произвольную их *аффинную* комбинацию $(1-\lambda)\bar{x}^0 + \lambda\bar{x}^1$ (ср. с замечанием 3.1).

3.3. Подмножества решений однородных и неоднородных с.л.у. Вернемся к с.л.у. (1.10) и (1.10h).

Предложение 3.2. *1. Подмножество L^0 решений однородной системы* (1.10h) *является линейным подпространством в $\mathbb{R}^n$.*

2. Подмножество L решений совместной неоднородной системы (1.10) *имеет вид*

$$L = L^0 + \bar{x}^0, \tag{3.3}$$

где $\bar{x}^0$ — произвольное решение системы (1.10).

Доказательство. 1. Первое утверждение с очевидностью вытекает из пп. 1, 2 предложения 3.1, которые в обозначениях настоящего предложения могут быть переписаны в следующем виде:

$$(\,\bar{u}, \bar{v} \in L^0\,) \;\Rightarrow\; (\,\bar{u} + \bar{v} \in L^0\,); \tag{3.4}$$

$$(\forall \lambda \in \mathbb{R})\;[\,(\,\bar{u} \in L^0\,) \;\Rightarrow\; (\,\lambda\bar{u} \in L^0\,)\,]. \tag{3.5}$$

2. Равенство (3.3) — это равенство множеств. Всякое равенство множеств $M_1 = M_2$ равносильно одновременному выполнению двух включений: $M_1 \subseteq M_2$ и $M_2 \subseteq M_1$.

Каждое из включений можно проверять "на элементах", доказывая утверждения: $(x \in M_1) \Rightarrow (x \in M_2)$ и $(x \in M_2) \Rightarrow (x \in M_1)$.

Так мы и поступим при доказательстве (3.3).

Пусть вектор $\bar{x}$ принадлежит левой части (3.3), т. е. является решением (1.10). Пусть $\bar{x}^0$ — еще одно (произвольное) решение этой системы. [Заметьте, что именно в этом месте доказательства используется непустота L, т. е. совместность системы (1.10).] Тогда, в силу п. 4 предложения 3.1, разность $\bar{x} - \bar{x}^0$ будет решением (1.10h), т. е. $\bar{x} - \bar{x}^0 \in L^0$, или $\bar{x} = (\bar{x} - \bar{x}^0) + \bar{x}^0 \in L^0 + \bar{x}^0$, т. е. $\bar{x}$ принадлежит правой части (3.3).

Пусть теперь $\bar{x}$ принадлежит правой части (3.3), т. е. представляется в виде $\bar{x} = \bar{y} + \bar{x}^0$, где $\bar{y}$ — некоторый вектор, принадлежащий подпространству L^0. Тогда, в силу п. 3 предложения 3.1, вектор $\bar{x}$ принадлежит L, т. е. левой части (3.3).

Равенство (3.3) доказано. □

Замечание 3.4. Вспоминая о понятиях частных и общего решений с.л.у. (см. замечание 1.1), мы можем трактовать подпространство L^0 (или формулу, описывающую это подпространство) как *общее решение однородной системы* (1.10h) (о.р.о.), а сдвинутое подпространство L (или формулу, его описывающую) — как *общее решение неоднородной системы* (1.10) (о.р.н.); вектор $\bar{x}^0$ трактуется как некоторое частное решение неоднородной системы (ч.р.н.).

Получается следующая словесная формулировка: *общее решение (совместной) неоднородной с.л.у. складывается из общего решения соответствующей однородной с.л.у. и (произвольного) частного решения неоднородной с.л.у.* Краткая запись этой формулировки:

$$\text{о. р. н.} \;=\; \text{о. р. о.} \;+\; \text{ч. р. н.} \tag{3.6}$$

Символическая формула (3.6) выражает одно из самых знаменитых правил математики. Она многократно должна встретиться вам в "продвинутых" разделах математики (например, в теории линейных дифференциальных уравнений; обратите внимание на присутствующее и здесь слово "линейных": именно с понятием линейности — одним из самых важных во всей математике — связано это правило).

Замечание 3.5. В случае определенности системы (1.10) множество ее решений состоит из единственного вектора: $L = \{\bar{x}^0\}$. Из формулы (3.3) немедленно следует в этом случае (см. п. 1.5), что однородная система (1.10h) будет иметь лишь тривиальное решение, т. е. подпространство L^0 будет состоять из одного нулевого вектора.

§ 4. Равносильные системы линейных уравнений. Элементарные преобразования. Понятие о методе Гаусса

4.1. Равносильные с.л.у. Рассмотрим две с.л.у. (с одним и тем же неизвестным вектором $\bar{x}$):

$$\underset{m\times n}{A}\cdot\underset{n\times 1}{\bar{x}}=\underset{m\times 1}{\bar{b}} \tag{4.1}$$

и

$$\underset{m'\times n}{A'}\cdot\underset{n\times 1}{\bar{x}}=\underset{m'\times 1}{\bar{b}'}\,. \tag{4.2}$$

Пусть L и L' — подмножества их решений (оба они содержатся в арифметическом линейном пространстве $\mathbb{R}^n$).

Определение 4.1. Говорят, что с.л.у. (4.2) является *следствием* с.л.у. (4.1), если всякое решение (4.1) является также решением (4.2), т. е. если $L \subseteq L'$. Две с.л.у. (4.1) и (4.2) называются *равносильными* (или *эквивалентными*), если каждая из них является следствием другой, т. е. если эти системы имеют одинаковые множества решений: $L = L'$.

Подчеркнем, что равносильные с.л.у. могут иметь разное количество уравнений, но обязаны иметь один и тот же набор неизвестных.

Целью нашей дальнейшей работы будет изучение так называемого метода Гаусса и его модификации — метода Жордана — Гаусса, позволяющих с помощью простых преобразований над произвольной с.л.у. привести эту систему к равносильной системе, которую уже можно легко решить.

4.2. Элементарные преобразования с.л.у. Рассмотрим следующие четыре типа преобразований над с.л.у. вида (4.1).

I. Перестановка двух уравнений системы. Символически это действие выражаться записью: $i^{\text{ур}} \leftrightarrow k^{\text{ур}}$, где i, k — номера уравнений, $i \neq k$.

II. Прибавление к одному из уравнений системы другого уравнения, домноженного на некоторое число. Символическое выражение: $i^{\text{ур}} + k^{\text{ур}} \cdot c$. При этом действии к уравнению с номером i (почленно) прибавляется уравнение с номером k (где $i \neq k$), домноженное (почленно) на число c, остальные же уравнения (в том числе и уравнение с номером k) остаются неизменными.

III. Почленное домножение одного из уравнений системы на ненулевое число. Символически: $i^{\text{ур}} \cdot c$, где $c \in \mathbb{R}$, $c \neq 0$.

IV. Удаление из системы (включение в систему) тривиального уравнения, все коэффициенты которого, а также его правая часть равны нулю.

Определение 4.2. Описанные выше преобразования называются *элементарными преобразованиями* типов I — IV над с.л.у.

Предложение 4.1. *Под действием элементарных преобразований с.л.у. переходит в равносильную с.л.у.*

Доказательство. Характерной особенностью преобразований I — IV типов является их *обратимость*. В самом деле, преобразование типа I является "самообратным", т. е. его повторное применение возвращает систему к исходному виду. Обратимость преобразования типа IV заложена в самом его определении. Преобразованием, обратным к преобразованию типа II, является преобразование (такого же типа) $i^{\text{ур}} + k^{\text{ур}} \cdot (-c)$. Аналогично, обратным к III будет преобразование типа III: $i^{\text{ур}} \cdot c^{-1}$.

Далее заметим, что всякий арифметический вектор $\bar{x}$, являющийся решением исходной системы, останется решением и для преобразованной системы. В самом деле, вектор $\bar{x}$ обращает все уравнения системы в истинные равенства, а преобразования типов I — III, будучи произведены над истинными равенствами, снова приводят, очевидно, к истинным равенствам. Что касается преобразования типа IV, то оно удаляет из системы (добавляет к системе) "уравнение" вида $0 \cdot x_1 + \cdots + 0 \cdot x_n = 0$, которое является "абсолютным тождеством" (т. е. справедливо для любых значений неизвестных) и поэтому "содержит нуль информации": его исключение (добавление) никак не отражается на множестве решений с.л.у.

Теперь можно утверждать, что преобразованная с помощью элементарных преобразований с.л.у. является следствием исходной системы. Отмеченная выше обратимость этих преобразований позволяет заключить, что верно и обратное: исходная с.л.у. будет следствием преобразованной. А значит, эти с.л.у. будут равносильны. □

4.3. Расширенная матрица с.л.у. Матричное выражение элементарных преобразований над с.л.у. Производить элементарные преобразования над с.л.у. (4.2) удобнее используя вспомогательную матрицу B размера $m \times (n+1)$, получающуюся слиянием матрицы коэффициентов (1.3) и столбца правых частей (1.8).

Определение 4.3. Матрица

$$B = (A|\,\bar{b}) \tag{4.3}$$

называется *расширенной матрицей* с.л.у. (4.1).

Элементарные преобразования над с.л.у. будут описываться с помощью *элементарных преобразований над строками* расширенной матрицы этой системы. Эти последние преобразования будут оформляться в стиле, аналогичном описанному в предыдущем пункте: преобразование типа II, например, будет выражаться записью: $i^{\text{стр}} + k^{\text{стр}} \cdot c$ и т. п.

Далее мы убедимся, что при практическом решении с.л.у. часто бывает удобным переставлять неизвестные. Поэтому, чтобы "не потерять" истинные имена неизвестных, нам придется приписывать их над столбцами матрицы A. Матрица (4.3) будет в этом случае выглядеть следующим образом:

$$\underset{m\times(n+1)}{B} = \left(\begin{array}{cccc|c} \boxed{x_1} & \boxed{x_2} & \dots & \boxed{x_n} & \\ a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right). \tag{4.4}$$

Подчеркнем, что имена неизвестных (в квадратиках) не являются элементами матрицы (4.4) и служат лишь метками столбцов. Если вы не собираетесь переставлять столбцы, то и метки вам не понадобятся.

Замечание 4.1. Ясно, что однородная с.л.у. под действием элементарных преобразований снова переходит в однородную с.л.у. Поэтому для "слежения" за преобразованиями не нужна расширенная матрица системы B: достаточно преобразовывать матрицу A.

4.4. Идея метода Жордана — Гаусса (на примере). Сейчас мы вернемся к с.л.у. из примера 1.2, выпишем для нее расширенную матрицу системы, подвергнем эту матрицу элементарным преобразованиям и приведем ее сначала к так называемому *ступенчатому виду*, а затем к *виду Жордана — Гаусса*. Восстановим по полученной матрице с.л.у. (она, в силу предложения 4.1, будет равносильна исходной с.л.у.) и решим вновь полученную систему.

Пример 4.1. Решим следующую систему из трех линейных уравнений с четырьмя неизвестными.

$$\begin{cases} -2x_1 + x_2 - x_3 - 2x_4 = 0; \\ 6x_1 - 3x_2 + 2x_3 + 5x_4 = -1; \\ 4x_1 - 2x_2 + x_3 + 3x_4 = -1. \end{cases}$$

$$B = \left(\begin{array}{cccc|c} \boxed{x_1} & \boxed{x_2} & \boxed{x_3} & \boxed{x_4} & \\ \boxed{-2} & 1 & -1 & -2 & 0 \\ 6 & -3 & 2 & 5 & -1 \\ 4 & -2 & 1 & 3 & -1 \end{array}\right) \xrightarrow[3^{\text{стр}}+1^{\text{стр}}\cdot 2]{2^{\text{стр}}+1^{\text{стр}}\cdot 3}$$

$$\longrightarrow \left(\begin{array}{cccc|c} \boxed{x_1} & \boxed{x_2} & \boxed{x_3} & \boxed{x_4} & \\ \boxed{-2} & 1 & -1 & -2 & 0 \\ 0 & 0 & \boxed{-1} & -1 & -1 \\ 0 & 0 & -1 & -1 & -1 \end{array}\right) \xrightarrow{3^{\text{стр}}+2^{\text{стр}}\cdot(-1)}$$

$$\longrightarrow \left(\begin{array}{cccc|c} \boxed{x_1} & \boxed{x_2} & \boxed{x_3} & \boxed{x_4} & \\ \boxed{-2} & 1 & -1 & -2 & 0 \\ 0 & 0 & \boxed{-1} & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \longrightarrow \ldots$$

Достигнут ступенчатый вид. "Ступеньки" начинаются с обведенных в квадратики (так называемых *ключевых*) элементов, которые обязаны быть ненулевыми.

Как только получен ступенчатый вид, все неизвестные могут быть разбиты на два класса: *главные* и *свободные*. В качестве главных выбираются те неизвестные, которые соответствуют *ключевым столбцам* (тем, в которых располагаются ключевые элементы). Остальные неизвестные объявляются свободными, они могут принимать произвольные значения.

Запишем с.л.у., отвечающую полученной матрице:

$$\begin{cases} -2x_1 + x_2 - x_3 - 2x_4 = 0; \\ - x_3 - x_4 = -1, \end{cases}$$

Эту систему уже можно решать "снизу вверх", выражая главные неизвестные (в данном случае — это x_1 и x_3) через свободные неизвестные (x_2 и x_4):

$$x_3 = -x_4 + 1;$$

$$x_1 = \frac{1}{2}(x_2 - x_3 - 2x_4) = \frac{1}{2}(x_2 - x_4 - 1)$$

и т. д.

В этом, собственно, и состоит идея простейшей версии метода Гаусса: выбираются ключевые элементы и под ними "делаются нули"; полученная ступенчатая система решается относительно главных неизвестных.

Но мы продолжим вычисления: исключим нулевую строку, получим нули не только *под*, но и *над* ключевыми элементами и сделаем ключевые элементы единичными (что будет характерно для так называемого метода Жордана — Гаусса). Кроме того, мы модифицируем метод, производя группировку главных неизвестных в начале списка, после чего в начале матрицы должна сформироваться *единичная подматрица* E. (Именно здесь нам необходимы метки над столбцами, соответствующие именам неизвестных.)

$$\dots \xrightarrow[1^{\text{стр}}\cdot(-1/2);\ 2^{\text{стр}}\cdot(-1)]{1^{\text{стр}}+2^{\text{стр}}\cdot(-1)} \left(\begin{array}{cccc|c} \boxed{x_1} & \boxed{x_2} & \boxed{x_3} & \boxed{x_4} & \\ \boxed{1} & -1/2 & 0 & 1/2 & -1/2 \\ 0 & 0 & \boxed{1} & 1 & 1 \end{array}\right) \longrightarrow$$

$$\longrightarrow \left(\begin{array}{cccc|c} \boxed{x_1} & \boxed{x_3} & \boxed{x_2} & \boxed{x_4} & \\ \boxed{1} & 0 & -1/2 & 1/2 & -1/2 \\ 0 & \boxed{1} & 0 & 1 & 1 \end{array}\right).$$

Запишем теперь с.л.у. по последней полученной матрице:

$$\begin{cases} x_1 & - & 1/2\,x_2 & + & 1/2\,x_4 & = & -1/2\,; \\ x_3 & & & + & x_4 & = & 1\,. \end{cases}$$

Выразим из этой системы главные неизвестные через свободные. Кроме того, поскольку (по определению 1.1) решение с.л.у. представляет собой вектор-столбец, включающий выражения для всех неизвестных, учтем в записи решения свободные неизвестные, включая в ответ "тавтологические" равенства $x_2 = x_2$; $x_4 = x_4$.

$$\begin{cases} x_1 & = & 1/2\,x_2 & - & 1/2\,x_4 & - & 1/2\,; \\ x_2 & = & x_2 & & & & ; \\ x_3 & = & & & x_4 & + & 1\,; \\ x_4 & = & & & x_4 & & . \end{cases}$$

Полученные формулы представляют собой не что иное, как общее решение исходной с.л.у. (см. замечание 1.1). Придадим этим

формулам векторный вид:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_2 \begin{pmatrix} 1/2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1/2 \\ 0 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -1/2 \\ 0 \\ 1 \\ 0 \end{pmatrix}.$$

Обозначив три вектора-столбца в правой части последнего равенства $\bar{f}_2$, $\bar{f}_4$ и $\bar{x}^0$, получим краткую запись ответа:

$$\bar{x} = x_2\bar{f}_2 + x_4\bar{f}_4 + \bar{x}^0.$$

При любых конкретных значениях свободных неизвестных (фигурирующих в качестве скалярных множителей перед векторами $\bar{f}_2$ и $\bar{f}_4$) мы получаем конкретное частное решение с.л.у. Таким образом, эта система (как мы уже знаем из примера 1.2) является неопределенной. Из всего (бесконечного) множества L решений с.л.у. можно выделить одно особое, которое получается, если положить равными нулю все свободные неизвестные: $x_2 = x_4 = 0$. Это решение $\bar{x} = \bar{x}^0$ принято называть *опорным* (оно также уже встречалось нам в примере 1.2).

Проясним теперь смысл векторов-столбцов $\bar{f}_2$ и $\bar{f}_4$. Представьте себе, что в правых частях исходной системы стоят нули, т. е. с.л.у. является однородной. Она останется таковой и при всех элементарных преобразованиях (см. замечание 4.1). Ответом для нее будет выражение

$$\bar{x} = x_2\bar{f}_2 + x_4\bar{f}_4,$$

которое является, таким образом, *общим решением для однородной системы, соответствующей исходной неоднородной системе.* При конкретных значениях свободных неизвестных получаются конкретные *частные решения однородной системы.* Среди них выделяются так называемые *базисные* частные решения однородной системы: если положить $x_2 = 1$, а $x_4 = 0$, то останется $\bar{x} = \bar{f}_2$, при противоположном выборе значений свободных неизвестных получится $\bar{x} = \bar{f}_4$.

Таким образом, векторы, стоящие при свободных неизвестных в общем решении исходной системы, являются решениями соответствующей однородной системы. То, почему эти решения заслуживают названия базисных, будет строго объяснено в § 10. Пока вы можете заметить, что всякое решение однородной системы является линейной комбинацией (см. замечание 2.4) базисных решений. Еще

раз возвратившись к примеру 1.2, заметьте, что указанное там решение $\bar{x}^1$ получается из найденного выше общего решения при $x_2 = 0$, $x_4 = -1$, т. е. $\bar{x}^1 = 0 \cdot \bar{f}_2 + (-1) \cdot \bar{f}_4 + \bar{x}^0$.

При решении практических примеров настоятельно рекомендуется делать *проверку:* опорное решение следует проверять, подставляя во все уравнения исходной системы, а каждое из базисных решений — подставляя во все уравнения соответствующей однородной системы.

Замечание 4.2. Проведенный анализ подтверждает на конкретном примере представление общего решения совместной неоднородной с.л.у. в виде суммы общего решения соответствующей однородной системы и частного решения неоднородной системы [см. правило (3.6)]. В примере это правило дополнено новой информацией: 1) в качестве ч.р.н. можно брать опорное частное решение неоднородной системы; 2) о.р.о. само имеет представление в виде линейной комбинации базисных частных решений однородной системы (ч.р.о.). Ниже, в § 6, мы убедимся, что такое выражение для общего решения получается и для произвольной совместной неоднородной системы.

А пока задумаемся над тем, как в ходе применения алгоритма Жордана — Гаусса мы можем обнаружить, что исходная система является несовместной. Представьте себе, что в системе, рассмотренной в примере 4.1, в правой части третьего уравнения, вместо числа -1 стоит, скажем, число 5. Тогда в момент достижения ступенчатого вида (т. е. на втором шаге преобразований матрицы B) мы вместо последней нулевой строки получили бы строку с четырьмя нулями и с числом 6 в последнем столбце (это число было бы последней ступенькой в ступенчатом виде матрицы B). Такая строка соответствовала бы противоречивому уравнению $0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 + 0 \cdot x_4 = 6$, что свидетельствовало бы о несовместности исходной с.л.у.

И в общем случае (см. § 6) с.л.у. будет несовместной, если последняя ступенька в ступенчатом виде расширенной матрицы этой системы будет приходиться на последний столбец.

Для совместной с.л.у. возможны две ситуации:

1) либо число ступенек меньше числа неизвестных, и тогда в решении будут присутствовать свободные неизвестные и, следовательно, с.л.у. будет неопределенной;

2) либо число ступенек равно числу неизвестных, и тогда свободных неизвестных не будет, решение будет единственным, система — определенной.

Замечание 4.3. Говоря о выборе главных (свободных) неизвестных, необходимо отметить, что такой выбор далеко не однозначен. Так, в разобранном примере в качестве главных могли бы фигурировать неизвестные x_1, x_4 или x_2, x_3, но не x_1, x_2. (При изучении следующих глав вы должны будете разобраться в причинах этого явления, а также понять, почему, несмотря на неоднозначность в определении ступенчатого вида для данной матрицы, "количество ступенек" в ступенчатом виде определено однозначно; см. понятие *ранга матрицы*.)

И еще: может быть, это покажется вам непривычным, но некоторые неизвестные могут в системе вообще отсутствовать. Такие неизвестные неизбежно попадают в разряд свободных: если о них ничего не говорится ("нуль информации"), то их значения совершенно произвольны.

Например, система из одного уравнения

$$x_2 - x_3 = 0,$$

будучи рассмотрена в пространстве $\mathbb{R}^4$, дает такой результат: неизвестные x_1, x_4 — свободны в силу невхождения в уравнение; в качестве еще одной свободной неизвестной можно (по произволу) выбрать либо x_2, либо x_3. Если выбрано x_2, то ответ будет таким:

$$\begin{cases} x_1 = x_1 \quad ; \\ x_2 = \quad x_2 \quad ; \\ x_3 = \quad x_2 \quad ; \\ x_4 = \quad\quad x_4, \end{cases}$$

или

$$\bar{x} = x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Замечание 4.4. (Студентам не следует обращать внимание на это замечание: оно — для преподавателей, знакомых с различными методиками изложения основ линейной алгебры.)

В большинстве учебников и задачников по алгебре в качестве ответов для с.л.у. приводятся выражения главных неизвестных через свободные, что, по мнению автора данного пособия, может считаться лишь "полуфабрикатом" для представления общего решения системы. В соответствии с определением 1.1, решения с.л.у. являются

векторами, следовательно, и ответ должен быть представлен в векторном виде, причем очень важно (для учебных и для прикладных целей), чтобы была выявлена *структура решения* (опорное решение плюс линейная комбинация базисных решений однородной системы со свободными неизвестными в качестве неопределенных коэффициентов этой линейной комбинации). Столь же неудачным представляется автору и переобозначение (в ответах в задачниках) свободных неизвестных (как параметров) другими буквами (с последующим приданием этим параметрам конкретных значений), что затемняет суть явного выражения. Ключевыми моментами пропагандируемого в настоящем издании подхода к оформлению вычислений являются: запись векторов в столбик, включение в ответ тавтологических равенств (типа $x_j = x_j$) для каждой свободной неизвестной и аккуратное расположение (строго в столбик) одноименных неизвестных.

Замечание 4.5. Немного информации о решении с.л.у. в системе Maple.

Поставив компьютеру задачу решить неоднородную с.л.у., мы (в случае неопределенности этой с.л.у.) в качестве ответа вправе были бы ожидать на выходе: вектор опорного решения и так называемую *фундаментальную матрицу* системы, составленную из векторов-столбцов базисных решений соответствующей однородной с.л.у. И действительно, вычисления можно организовать так, чтобы получить ответ в желаемой форме. Однако, по умолчанию, системы настроены на иные способы выдачи результатов.

Приведем ниже протокол небольшой сессии с системой Maple, посвященной решению рассмотренного выше примера 4.1.

Разумеется, мы не имеем здесь возможности углубляться в подробности синтаксиса системы Maple, однако надеемся, что будущие компьютерщики сами прочитают нужные "help'ы" или изучат руководства.

```
> with ( linalg ) :
> A := matrix ([ [ −2, 1, −1, −2 ] , [ 6, −3, 2, 5 ] , [ 4, −2, 1, 3 ] ]);
```

$$\mathrm{A} := \begin{bmatrix} -2 & 1 & -1 & -2 \\ 6 & -3 & 2 & 5 \\ 4 & -2 & 1 & 3 \end{bmatrix}$$

```
> b := vector ( [ 0, −1, −1 ] ) ;
```

$$b := [0, -1, -1]$$

Вектор — это одномерный массив с нумерацией элементов, начинающейся с 1. Maple воспринимает векторы как столбцы, хотя вводятся и записываются они как строки.

Можно, например, "слить" матрицу A и вектор b, составив расширенную матрицу системы $B = (A|b)$ с помощью команды

> B := concat (A, b) ;

$$B := \begin{bmatrix} -2 & 1 & -1 & -2 & 0 \\ 6 & -3 & 2 & 5 & -1 \\ 4 & -2 & 1 & 3 & -1 \end{bmatrix}$$

Решение системы $A \cdot x = b$ находится с помощью команды

> x := linsolve (A, b, 'r', t) ;

$$x := [t_1,\ 2t_1 + 1 + t_2,\ 1 - t_2,\ t_2]$$

Поясним, что означают два последних (необязательных) аргумента функции linsolve. Третий аргумент 'r' резервирует имя для переменной "ранг матрицы A" (количество ступенек в ступенчатом виде этой матрицы). По дополнительному запросу значение ранга выводится:

> r;

2

Четвертый аргумент определяет обозначения для свободных неизвестных. Судя по ответу, свободными являются неизвестные x_1 и x_4. Maple обозначил их (предписанной) буквой t и занумеровал по-своему. (По умолчанию свободные неизвестные обозначаются именно буквой t, но — с "системным префиксом": $\_t_1, \_t_2$.)

Решение, выведенное системой в строчку, можно при желании конвертировать в столбец:

> x := convert (x, matrix) ;

$$x := \begin{bmatrix} t_1 \\ 2t_1 + 1 + t_2 \\ 1 - t_2 \\ t_2 \end{bmatrix}$$

Можно попросить Maple проверить результат:

> evalm (A & ∗ x – b) ;

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Подсчитанная "невязка" оказалась нулевым вектором, значит решение найдено верно. Может понадобиться не сразу получить ответ, а остановиться на этапе достижения вида Жордана — Гаусса. Тогда применяем к расширенной матрице B команду gaussjord:

> BJ := gaussjord (B) ;

$$BJ := \begin{bmatrix} 1 & -1/2 & 0 & 1/2 & -1/2 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Именно такой вид Жордана — Гаусса у нас получился при решении вручную; свободными неизвестными здесь будут x_2 и x_4. Можно догадаться, что алгоритм, реализующий команду linsolve, не столь "прямолинеен" (поскольку приводит к другому, равносильному ответу).

§ 5. Метод Жордана — Гаусса для матриц

5.1. Матрицы ступенчатого вида, вида Жордана — Гаусса, модифицированного вида Жордана — Гаусса, скелетного вида

Определение 5.1. 1. Матрицу $A \in \mathrm{Mat}(m, n; \mathbb{R})$ будем называть матрицей *ступенчатого вида*, если

$$A = \begin{pmatrix} \boxed{*} & * & \dots & * & * & \dots & * & * & \dots & * & * & \dots \\ & & & \boxed{*} & * & \dots & * & * & \dots & * & * & \dots \\ & & & & & & \boxed{*} & * & \dots & * & * & \dots \\ & & & & & & & & \ddots & \dots & \dots & \dots \\ & & & & & & & & & \boxed{*} & * & \dots \\ & & & & & & & & & & & \end{pmatrix}, \quad (5.1)$$

где символы $\boxed{*}$ обозначают ненулевые элементы (именуемые *ключевыми*), $*$ обозначают произвольные элементы, а на пустых местах подразумеваются нули.

2. Будем говорить, что ступенчатая матрица (5.1) имеет *вид Жордана — Гаусса (Ж.—Г.)*, если все ключевые элементы равны единице и нули в ключевых столбцах стоят не только ниже, но и выше ключевых элементов, т. е. если матрица имеет вид

$$A = \begin{pmatrix} \boxed{1} & * & \dots & 0 & * & \dots & 0 & * & \dots & 0 & * & \dots \\ & & & \boxed{1} & * & \dots & 0 & * & \dots & 0 & * & \dots \\ & & & & & & \boxed{1} & * & \dots & 0 & * & \dots \\ & & & & & & & & \ddots & \dots & \dots & \dots \\ & & & & & & & & & \boxed{1} & * & \dots \\ & & & & & & & & & & & \end{pmatrix}. \quad (5.2)$$

3. Если все ключевые столбцы матрицы (5.2) сгруппированы в начале этой матрицы, т. е.

$$A = \begin{pmatrix} \boxed{1} & 0 & 0 & \dots & 0 & * & \dots & * \\ 0 & \boxed{1} & 0 & \dots & 0 & * & \dots & * \\ 0 & 0 & \boxed{1} & \dots & 0 & * & \dots & * \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \boxed{1} & * & \dots & * \\ & & & & & & & \end{pmatrix}, \quad (5.3)$$

то будем говорить, что матрица A имеет *модифицированный* вид Жордана — Гаусса.

4. Если в матрице (5.3), кроме нескольких единиц, расположенных в начале главной диагонали, нет других ненулевых элементов, т. е. если

$$A = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & \end{pmatrix}, \tag{5.4}$$

то говорят, что матрица A имеет *скелетный* вид.

Замечание 5.1. Модифицированный вид Ж.—Г. и скелетный вид матрицы могут быть описаны с помощью так называемых *блочных* матриц. Простейшим вариантом блочной матрицы является матрица, разбитая на четыре прямоугольных блока следующим образом:

$$\underset{(k_1+k_2)\times(l_1+l_2)}{M} = \left(\begin{array}{c|c} \underset{k_1\times l_1}{A} & \underset{k_1\times l_2}{B} \\ \hline \underset{k_2\times l_1}{C} & \underset{k_2\times l_2}{D} \end{array}\right). \tag{5.5}$$

Обозначив буквой r количество ступенек в матрице (5.3), мы можем представить эту матрицу в блочном виде:

$$\underset{m\times n}{A} = \left(\begin{array}{c|c} \underset{r\times r}{E} & \underset{r\times(n-r)}{H} \\ \hline \underset{(m-r)\times r}{O} & \underset{(m-r)\times(n-r)}{O} \end{array}\right), \tag{5.3a}$$

где E — единичная матрица (см. п. 2.3), H — произвольная матрица.

Скелетная матрица (5.4) имеет блочное представление

$$\underset{m\times n}{A} = \left(\begin{array}{c|c} \underset{r\times r}{E} & \underset{r\times(n-r)}{O} \\ \hline \underset{(m-r)\times r}{O} & \underset{(m-r)\times(n-r)}{O} \end{array}\right). \tag{5.4a}$$

5.2. Теорема Жордана – Гаусса для матриц. В этом пункте будет сформулирована и доказана основная теорема метода Ж.—Г. для матриц. В § 4 были определены элементарные преобразования для с.л.у. и соответствующие элементарные преобразования над строками расширенных матриц этих систем. Однако при изучении элементарных преобразований для матриц (безотносительно к системам линейных уравнений) принято рассматривать преобразования,

сохраняющие размеры матриц, т. е. преобразования типа IV исключаются из рассмотрения. С другой стороны, при изучении "матриц вообще" строки и столбцы совершенно равноправны, поэтому наряду с элементарными преобразованиями (типов I — III) над строками следует рассматривать аналогичные преобразования над столбцами матриц. (Ниже, в § 6, мы снова вернемся к матрицам, происходящим из с.л.у., и докажем соответствующую теорему, учитывающую специфику таких матриц.)

Теорема 5.1. *Рассмотрим произвольную матрицу A с действительными элементами.*

1. С помощью элементарных преобразований типов I — II *над строками матрица A может быть приведена к ступенчатому виду.*

2. Если дополнительно разрешить элементарные преобразования типа III *(над строками), то матрицу можно привести к виду Жордана — Гаусса.*

3. Если дополнительно разрешить элементарные преобразования типа I *над столбцами, то матрицу можно привести к модифицированному виду Жордана — Гаусса.*

4. Если дополнительно разрешить элементарные преобразования типа II *над столбцами, то матрицу можно привести к скелетному виду.*

Доказательство. 1. Пусть дана $(m \times n)$-матрица (1.3). Если $A = O$, то нечего доказывать: нулевую матрицу можно рассматривать как ступенчатую (с числом ступенек, равным нулю). Если матрица ненулевая, то будем просматривать столбцы этой матрицы, начиная с первого, до тех пор, пока нам не встретится ненулевой столбец. Этот столбец мы примем за первый ключевой. С помощью перестановки строк (т. е. элементарных преобразований типа I) добьемся того, чтобы верхний элемент ключевого столбца стал ненулевым, и примем этот элемент за первый ключевой элемент. Допустим, он равен a.

С помощью элементарных преобразований типа II (над строками) сделаем нули под ключевым элементом. Если, скажем, в i-й строке под ключевым элементом стоит число c, то оно "обнуляется" преобразованием $i^{\text{стр}} + 1^{\text{стр}} \cdot (-\frac{c}{a})$.

После завершения этого этапа мы получим:

$$A \longrightarrow \cdots \longrightarrow \begin{pmatrix} \boxed{a} & * & \ldots & * \\ 0 & * & \ldots & * \\ \ldots & \ldots & \ldots & \ldots \\ 0 & * & \ldots & * \end{pmatrix}.$$

Обозначим A' блок полученной матрицы, расположенный справа-снизу от ключевого элемента. Повторим описанный выше шаг алгоритма применительно к строкам, начиная со второй, т. е. фактически — применительно к матрице A' (поскольку очевидно, что при преобразованиях над этими строками ранее полученные нули "не портятся"). Далее работа алгоритма продолжается и следует оговорить условия "останова". Алгоритм заканчивает работу, если

— либо достигнута последняя строка матрицы (и тогда получается ступенчатый вид (5.1) для A с числом ступенек r, равным числу строк m);

— либо на некотором шаге (с номером r) достигнут последний столбец матрицы (и тогда, после обнуления элементов ниже последнего ключевого, строки, начиная с $(r+1)$-й, окажутся нулевыми);

— либо после получения r-го ключевого элемента снизу-справа от него образуется нулевая подматрица (и тогда получается ступенчатый вид (5.1) с числом ступенек $r < m$).

2. Получив ступенчатый вид (5.1), мы продолжим вычисления с применением элементарных преобразований над строками типа III. Можно сначала превратить все ключевые элементы в единицы. Если, скажем, в i-й строке ключевым является элемент a, то, применяя преобразование $i^{\text{стр}} \cdot \frac{1}{a}$, мы сделаем этот элемент единичным. Далее, если в столбце над i-й ключевой единицей (в строке с номером k) находится ненулевой элемент c, то мы его можем обнулить с помощью преобразования типа II: $k^{\text{стр}} + i^{\text{стр}} \cdot (-c)$. Так мы придем к виду Ж.—Г. (5.2).

3. Если разрешается переставлять столбцы матрицы, то мы можем все ключевые столбцы матрицы (5.2) собрать в начале матрицы и получить модифицированный вид Ж.—Г. (5.3). При этом в левом верхнем углу матрицы сформируется единичная подматрица (блок) E размера $r \times r$ [см. (5.3а)], где r — количество ступенек.

4. Если разрешены элементарные преобразования над столбцами типа II, то мы можем обнулить оставшиеся ненулевыми столбцы (в зоне блока H). Скажем, j-й столбец матрицы (5.3), "пересекающий" блок H

$$A \to \cdots \to \begin{pmatrix} \boxed{1} & 0 & \ldots & 0 & * & \ldots & h_{1j} & \ldots \\ 0 & \boxed{1} & \ldots & 0 & * & \ldots & h_{2j} & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \boxed{1} & * & \ldots & h_{rj} & \ldots \\ & & & & & & & \\ & & & & & & & \end{pmatrix},$$

обнуляется одним "комбинированным" действием

$$j^{\text{стб}} + 1^{\text{стб}} \cdot (-h_{1j}) + 2^{\text{стб}} \cdot (-h_{2j}) + \cdots + r^{\text{стб}} \cdot (-h_{rj}).$$

Таким образом мы получим скелетный вид (5.4). □

Замечание 5.2. К ступенчатому виду матрицы можной прийти "разными путями" (с помощью различных последовательностей элементарных преобразований). И сам этот вид определен не однозначно. В последствии, однако, в п. 11.4, будет установлено, что "количество ступенек" r *инвариантно* (не зависит от способа приведения матрицы к ступенчатому виду). Это количество будет важнейшей числовой характеристикой матрицы и получит имя *ранг матрицы*. Скелетный вид матрицы, поскольку он полностью определяется числом r, также будет инвариантным.

§ 6. Метод Жордана — Гаусса для систем линейных уравнений

6.1. Теорема Жордана — Гаусса для с.л.у. В этом пункте мы применим результаты теоремы 5.1 к матрице, являющейся расширенной матрицей с.л.у. Данный случай имеет следующие особенности:

1) элементарные преобразования типов I — III, безусловно, можно производить лишь над строками матрицы;

2) над столбцами (кроме последнего, состоящего из правых частей уравнений) можно производить лишь преобразования типа I, причем для того чтобы не забыть, какой неизвестной отвечает тот или иной столбец, приходится ставить над столбцами *метки* (знаки неизвестных); последний столбец никуда переставлять нельзя;

3) преобразования над столбцами типов II — III в этой ситуции применять нельзя, поскольку они "смешивают" неизвестные;

4) можно использовать преобразования над строками типа IV, выбрасывая нулевые строки.

Теорема 6.1. *Пусть B — расширенная матрица* (4.3), *соответствующая с.л.у.* (1.10); $B' = (A'|\bar{b}')$ *— ступенчатый вид матрицы* B; r *— количество ступенек в матрице* A'. *Тогда в матрице* B' *ступенек либо столько же, либо на одну больше.*

1. Если количество ступенек для B' *равно* $r+1$, *то последняя ступенька приходится на последний,* $(n+1)$*-й столбец* B' *и с.л.у.* (1.10) *несовместна.*

2. Если количество ступенек для B' *равно* r, *то с.л.у.* (1.10) *совместна и возможны следующие два случая.*

2а. Если количество ступенек r *равняется числу неизвестных* n, *то с.л.у.* (1.10) *является определенной. Приводя в этом случае матрицу* B' *к виду Ж.—Г. (и вычеркивая нулевые строки), мы получим матрицу*

$$\underset{n\times(n+1)}{B''} = \left(\begin{array}{c|c} \underset{n\times n}{E} & \underset{n\times 1}{\bar{\beta}} \end{array}\right), \tag{6.1}$$

последний столбец которой определяет единственное решение системы (1.10)*:* $\bar{x}^0 = \bar{\beta}$.

2б. Если $r < n$, *то с.л.у.* (1.10) *является неопределенной и ее общее решение содержит* $n-r$ *свободных неизвестных, принимающих произвольные значения. Для отыскания общего решения приведем матрицу* B' *к модифицированному виду Ж.—Г. (и вычеркнем нулевые строки), получим матрицу вида*

$$\underset{r\times(n+1)}{B''} = \left(\begin{array}{c|c|c} \underset{r\times r}{E} & \underset{r\times(n-r)}{H} & \underset{r\times 1}{\bar{\beta}} \end{array}\right). \tag{6.2}$$

В качестве свободных неизвестных можно выбрать неизвестные, приходящиеся на зону матрицы H. *Если предположить, что порядок неизвестных в ходе преобразований не был изменен, то в качестве свободных неизвестных будут фигурировать*

$$x_{r+1}, x_{r+2}, \ldots, x_n\,,$$

а общее решение системы (1.10) *будет иметь следующую структуру:*

$$\bar{x} = x_{r+1}\bar{f}_{r+1} + x_{r+2}\bar{f}_{r+2} + \cdots + x_n\bar{f}_n + \bar{x}^0, \tag{6.3}$$

где вектор $\bar{x}^0$ является частным решением системы (1.10), *которое получается, если положить все свободные неизвестные равными нулю, а векторы $\bar{f}_j$ ($j = r+1, r+2, \ldots, n$) являются частными решениями однородной системы* (1.10h), *отвечающей данной системе* (1.10), *причем вектор $\bar{f}_j$ получается, если положить $x_j = 1$, а остальные свободные неизвестные приравнять нулю.*

Доказательство. Может быть, у читателей сложилось впечатление, что сформулированная выше теорема чрезвычайно сложная и что вообще таких теорем "не бывает". Действительно, математики обычно стремятся к лаконичным формулировкам. И теорему 6.1 тоже можно было бы сформулировать гораздо короче, указав только условия, когда существует хотя бы одно решение и когда — ровно одно. Мы же, ориентируясь на практические вычисления, предпочли в каждом из случаев указать явный способ отыскания и структуру решения. К тому же при таком подходе доказательство будет не намного длиннее формулировки.

Итак, в соответствии с теоремой 5.1, матрицу B можно с помощью элементарных преобразований типов I — II над строками привести к ступенчатому виду B'. Если последняя, $(r+1)$-я ступенька в B' (с ключевым элементом $a \neq 0$) приходится на последний столбец, то $(r+1)$-я строка этой матрицы будет соответствовать уравнению $0 \cdot x_1 + \cdots + 0 \cdot x_n = a$, что будет свидетельствовать о несовместности с.л.у., отвечающей матрице B', а значит, и исходная система несовместна.

В противном случае число ступенек в матрицах A' и B' одинаково (и равно r).

Если $r = n$, то ступеньки в A' идут подряд по главной диагонали, а все строки матрицы B', начиная с $(n+1)$-й (если они есть), являются нулевыми и их можно вычеркнуть. Продолжая преобразования до достижения вида Ж.—Г., мы приведем эту матрицу к виду (6.1). Теперь нам остается "считать" из последнего столбца ответ — единственное решение с.л.у. (1.10).

Если $r < n$, то ступеньки в матрице B' не будут идти все подряд и неизвестные разделятся на два класса: главные (те, которые приходятся на ключевые столбцы) и свободные (все остальные). Нумерация неизвестных есть фактор не слишком важный, и, хотя в практических примерах это будет не всегда так, ничто не мешает нам при теоретическом анализе считать, что главными являются первые r неизвестных. (Мы их занумеровали — мы и перенумеровать можем!)

Удаляя нулевые строки в матрице, представляющей собой вид Ж.—Г. для матрицы B, мы получим матрицу B'' вида (6.2). Запишем систему, соответствующую этой матрице:

$$\begin{cases} x_1 \quad\quad\quad\quad +h_{1(r+1)}x_{r+1}+h_{1(r+2)}x_{r+2}+\ldots+h_{1n}x_n=\beta_1; \\ \quad x_2 \quad\quad\quad +h_{2(r+1)}x_{r+1}+h_{2(r+2)}x_{r+2}+\ldots+h_{2n}x_n=\beta_2; \\ \quad\quad \ddots \quad\quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \quad\quad\quad x_r \quad +h_{r(r+1)}x_{r+1}+h_{r(r+2)}x_{r+2}+\ldots+h_{rn}x_n=\beta_r. \end{cases}$$

Решим полученную систему относительно главных неизвестных и добавим к ответу *тавтологические* уравнения $x_j = x_j$ для каждой свободной неизвестной. (Эти тавтологии выражают тот факт, что свободные неизвестные совершенно произвольны.) Будем иметь:

$$\begin{cases} x_1 = -h_{1(r+1)}x_{r+1} - h_{1(r+2)}x_{r+2} - \ldots - h_{1n}x_n + \beta_1; \\ x_2 = -h_{2(r+1)}x_{r+1} - h_{2(r+2)}x_{r+2} - \ldots - h_{2n}x_n + \beta_2; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_r = -h_{r(r+1)}x_{r+1} - h_{r(r+2)}x_{r+2} - \ldots - h_{rn}x_n + \beta_r; \\ x_{r+1} = x_{r+1}; \\ x_{r+2} = x_{r+2}; \\ \ldots\ldots \ldots \quad \ldots \\ x_n = x_n, \end{cases}$$

или, в векторной форме:

$$\bar{x} = x_{r+1}\begin{pmatrix} -h_{1(r+1)} \\ -h_{2(r+1)} \\ \ldots \\ -h_{r(r+1)} \\ 1 \\ 0 \\ \ldots \\ 0 \end{pmatrix} + x_{r+2}\begin{pmatrix} -h_{1(r+2)} \\ -h_{2(r+2)} \\ \ldots \\ -h_{r(r+2)} \\ 0 \\ 1 \\ \ldots \\ 0 \end{pmatrix} + \ldots + x_n\begin{pmatrix} -h_{1n} \\ -h_{2n} \\ \ldots \\ -h_{rn} \\ 0 \\ 0 \\ \ldots \\ 1 \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_r \\ 0 \\ 0 \\ \ldots \\ 0 \end{pmatrix}.$$

Обозначая $n-r+1$ векторов в правой части последнего равенства символами $\bar{f}_{r+1}, \bar{f}_{r+2}, \ldots, \bar{f}_n, \bar{x}^0$, получим формулу (6.3).

Проанализируем смысл введенных векторов. Если приравнять нулю все свободные неизвестные, то в ответе останется только последний вектор $\bar{x}^0$. Следовательно, этот вектор представляет собой частное решение неоднородной системы. (В примерах § 4 этому вектору уже присвоено было наименование *опорное* ч.р.н.)

Далее, как уже отмечалось в замечании 4.1, если бы данная система была однородной, то она осталась бы таковой и под действием элементарных преобразований. Для однородной с.л.у. мы получили бы $\bar{x}^0 = \bar{0}$, и следовательно, первые $n - r$ слагаемых в (6.3) дают не что иное, как общее решение для однородной системы (1.10h), соответствующей исходной неоднородной системе (1.10).

Каждое слагаемое в о.р.о.

$$\bar{x} = x_{r+1}\bar{f}_{r+1} + x_{r+2}\bar{f}_{r+2} + \cdots + x_n\bar{f}_n \tag{6.3h}$$

разъясняется следующим образом: если в (6.3h) положить одну из свободных неизвестных $x_j = 1$, а все остальные приравнять нулю, то в правой части останется только вектор $\bar{f}_j$. Таким образом, векторы $\bar{f}_j$ $(j = r+1, \ldots, n)$ являются частными решениями (1.10h). (Ранее, в примерах, они тоже получили специальное наименование — *базисные* ч.р.о.)

Этим и завершается доказательство теоремы. □

Здесь читателю можно посоветовать еще раз просмотреть "опередивший теорию" пример 4.1 и последующие замечания. Изложенная в § 4 идея метода Ж.—Г. получила в теореме 6.2 детальное обоснование. Кроме того, получило подтверждение и дополнительное уточнение правило (3.6): в качестве ч.р.н. в этой формуле может быть выбрано опорное частное решение неоднородной системы, а о.р.о. может быть расписано как линейная комбинация базисных ч.р.о.

6.2. Случай однородной с.л.у.

Предложение 6.1. *Однородная с.л.у.* (1.10h) *имеет ненулевое решение тогда и только тогда, когда количество ступенек в ступенчатом виде матрицы A меньше числа неизвестных n. В частности, если число уравнений $m < n$, то однородная система имеет ненулевое решение.*

Доказательство. Первое утверждение немедленно следует из теоремы 6.1, а второе — из того факта, что количество ступенек r не превосходит количества уравнений m. □

Заметьте, что у нас уже есть формула (6.3h) для общего решения системы (1.10h). А только что доказанное (практически тривиальное) предложение 6.1 выделено особо, т. к. в последующих главах на него придется неоднократно ссылаться.

6.3. Случай квадратной с.л.у. Альтернатива Фредгольма

Определение 6.1. С.л.у. (1.10) называется *квадратной*, если является квадратной матрица A, т. е. если число уравнений m равняется числу неизвестных n.

Следующее предложение обычно называют *альтернативой Фредгольма*. Автору хотелось бы, чтобы это словосочетание сохранилось в памяти повзрослевших (и "доучившихся" до *функционального анализа и интегральных уравнений*) читателей и чтобы формулируемое ниже простое (*конечномерное*) утверждение облегчило им понимание новой, значительно более сложной (*бесконечномерной*) задачи.

Предложение 6.2. *Для квадратной с.л.у.* (1.10) *и соответствующей однородной с.л.у.* (1.10h) *выполняется одно (и только одно) из следующих утверждений:*

1) либо система (1.10h) *имеет только нулевое решение, и тогда система* (1.10) *является определенной при любой правой части* $\bar{b}$;

2) либо (1.10h) *имеет нетривиальное решение, и тогда, в зависимости от* $\bar{b}$, *система* (1.10) *может быть либо несовместной, либо неопределенной.*

Доказательство. Пусть система (1.10h) имеет только нулевое решение. Тогда, по предложению 6.1, число ступенек r в ступенчатом виде A' матрицы A равняется числу неизвестных n. А поскольку система предполагается квадратной ($m = n$), то эти ступеньки идут подряд и имеются в каждой строке матрицы A'. При переходе к расширенной матрице (4.3), отвечающей неоднородной с.л.у. (1.10), мы замечаем, что вне зависимости от правой части $\bar{b}$ дополнительная ступенька в последнем столбце появиться не может (на нее просто "не хватит строк"). Значит, по теореме 6.1, система (1.10) будет совместной и определенной.

Если же система (1.10h) имеет ненулевое решение, то, по предложению 6.1, $r < n$. В силу теоремы 6.1, система (1.10) будет неопределенной, если окажется совместной. Но она может оказаться и несовместной, ибо для последней, $(r+1)$-й ступеньки в последнем столбце теперь "есть место" и она может появиться. □

§ 7. Некоторые типовые задачи: системы линейных уравнений с параметром, линейные матричные уравнения

7.1. С.л.у. с параметром. Рассмотрим с.л.у. вида (1.10), но в которой матрица A и вектор $\bar{b}$ зависят от некоторого параметра (действительного числа) λ:

$$A(\lambda) \cdot \bar{x} = \bar{b}(\lambda)\,. \tag{7.1}$$

Задача с параметром — это всегда "исследовательская" задача. Ход ее решения и ответ зависят от значения параметра. В различных случаях может быть различным *характер* системы: при одних значениях λ она окажется определенной, при других — неопределенной или несовместной. Разберем некоторые особенности задач с параметром на простом примере.

Пример 7.1. Решим при всех значениях параметра λ с.л.у.

$$\begin{cases} \lambda x_1 + x_2 + x_3 = 1; \\ x_1 + \lambda x_2 + x_3 = 1; \\ x_1 + x_2 + \lambda x_3 = 1. \end{cases}$$

Как и в примере 4.1, составим расширенную матрицу системы B и подвергнем ее элементарным преобразованиям. Очень частой ошибкой студентов при решении систем с параметром является поспешное рассмотрение случаев, кажущихся им "особыми". Например, очень часто первой фразой решения становится: "Рассмотрим случай $\lambda = 0$". (Аналогичную ошибку совершают многие абитуриенты в школьных задачах с параметрами.) Однако случай $\lambda = 0$ совсем не обязан быть особым. Как правило, анализ случаев (особых и неособых) следует начинать лишь после того, как "наметился" ступенчатый вид матрицы B, и различные случаи будут как раз отличаться различным количеством ступенек (для всей матрицы B и в зоне матрицы A).

$$B = \left(\begin{array}{ccc|c} \boxed{x_1} & \boxed{x_2} & \boxed{x_3} & \\ \lambda & 1 & 1 & 1 \\ 1 & \lambda & 1 & 1 \\ 1 & 1 & \lambda & 1 \end{array}\right) \xrightarrow{3^{\text{стр}} \leftrightarrow 1^{\text{стр}}}$$

$$\longrightarrow \left(\begin{array}{ccc|c} \boxed{x_1} & \boxed{x_2} & \boxed{x_3} & \\ \boxed{1} & 1 & \lambda & 1 \\ 1 & \lambda & 1 & 1 \\ \lambda & 1 & 1 & 1 \end{array}\right) \xrightarrow[3^{\text{стр}} + 1^{\text{стр}} \cdot (-\lambda)]{2^{\text{стр}} + 1^{\text{стр}} \cdot (-1)}$$

$$\longrightarrow \left(\begin{array}{ccc|c} \boxed{x_1} & \boxed{x_2} & \boxed{x_3} & \\ \boxed{1} & 1 & \lambda & 1 \\ 0 & \lambda-1 & 1-\lambda & 0 \\ 0 & 1-\lambda & 1-\lambda^2 & 1-\lambda \end{array}\right) \longrightarrow \ldots$$

Хотя ступенчатый вид наметился пока не полностью, тем не менее мы можем заметить первый особый случай:

$$1) \ \boxed{\lambda = 1}.$$

При таком значении параметра две последние строки полностью обнуляются и могут быть вычеркнуты, количество ступенек равняется 1 и остается решить систему из одного уравнения:

$$x_1 + x_2 + x_3 = 1.$$

Ответ для первого случая:

$$\boxed{\bar{x} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}}.$$

Система в этом случае является неопределенной.

Рассмотрим теперь противоположный случай:

$$2) \ \boxed{\lambda \neq 1}.$$

Продолжим преобразования:

$$\ldots \xrightarrow[3^{\text{стр}} \cdot \frac{1}{1-\lambda}]{2^{\text{стр}} \cdot \frac{1}{\lambda-1}} \left(\begin{array}{ccc|c} \boxed{x_1} & \boxed{x_2} & \boxed{x_3} & \\ \boxed{1} & 1 & \lambda & 1 \\ 0 & \boxed{1} & -1 & 0 \\ 0 & 1 & 1+\lambda & 1 \end{array}\right) \longrightarrow$$

$$\xrightarrow[3^{\text{стр}}+2^{\text{стр}} \cdot (-1)]{1^{\text{стр}}+2^{\text{стр}} \cdot (-1)} \left(\begin{array}{ccc|c} \boxed{x_1} & \boxed{x_2} & \boxed{x_3} & \\ \boxed{1} & 0 & \lambda+1 & 1 \\ 0 & \boxed{1} & -1 & 0 \\ 0 & 0 & \lambda+2 & 1 \end{array}\right) \longrightarrow \ldots$$

Теперь ступенчатый вид намечен полностью. Но именно "намечен", поскольку расположение ступенек зависит от λ. Второй случай

разбивается на два подслучая. Разберем сначала первый подслучай (особый):

$$2.1)\ \boxed{\lambda = -2}.$$

В этом подслучае последняя строка дает противоречие и, следовательно, ответ таков: $\boxed{\bar{x} \in \varnothing}$; система является несовместной.

Остается рассмотреть подслучай, который можно назвать неособым:

$$2.2)\ \boxed{\lambda \notin \{1, -2\}}.$$

Продолжим преобразования:

$$\ldots \xrightarrow{3^{\text{стр}} \cdot \frac{1}{\lambda+2}} \left(\begin{array}{ccc|c} \boxed{x_1} & \boxed{x_2} & \boxed{x_3} & \\ \boxed{1} & 1 & \lambda+1 & 1 \\ 0 & \boxed{1} & -1 & 0 \\ 0 & 1 & \boxed{1} & \frac{1}{\lambda+2} \end{array}\right) \longrightarrow$$

$$\xrightarrow[2^{\text{стр}}+3^{\text{стр}}]{1^{\text{стр}}+3^{\text{стр}} \cdot (-\lambda-1)} \left(\begin{array}{ccc|c} \boxed{x_1} & \boxed{x_2} & \boxed{x_3} & \\ \boxed{1} & 0 & 0 & \frac{1}{\lambda+2} \\ 0 & \boxed{1} & 0 & \frac{1}{\lambda+2} \\ 0 & 0 & \boxed{1} & \frac{1}{\lambda+2} \end{array}\right).$$

Ответ в этом подслучае:

$$\boxed{\bar{x} = \frac{1}{\lambda+2}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}};$$

система является определенной.

В конце решения задачи с параметром полагается приводить обычно довольно длинный ответ, содержащий все случаи, причем эти случаи должны быть, во-первых, *исчерпывающими*, т. е. их объединение должно покрывать всю числовую ось параметра λ, а во-вторых, взаимоисключающими, т. е. попарно не пересекаться. По ходу решения мы специально выделяли (и вам советуем) рамочками описания всех случаев и ответы в каждом из них. Теперь нам легко сформировать полный

О т в е т:

1. При $\lambda = 1$ система является неопределенной, общее решение:

$$\bar{x} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

2. При $\lambda = -2$ система несовместна, $\bar{x} \in \varnothing$.

3. При $\lambda \notin \{1, -2\}$ система является определенной, единственным решением является вектор

$$\bar{x} = \frac{1}{\lambda + 2} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Замечание 7.1. В дальнейшем (см. п. 29.2) для квадратных систем линейных уравнений мы (в случае их определенности) выведем явные формулы для решения — так называемые *формулы Крамера*, использующие понятие *определителя* квадратной матрицы. С помощью этих формул решение рассмотренной выше задачи (а мы собираемся к ней еще вернуться) получится более компактным.

Замечание 7.2. Любопытно, что Maple легко справляется также и с системами, содержащими параметр, однако анализ случаев не проводится. Так, в условиях примера 7.1 выдается лишь ответ для неособого случая.

```
> restart; with ( linalg ) :
> A := matrix ([[ lambda, 1, 1 ], [ 1, lambda, 1 ], [ 1, 1, lambda ]]);
```

$$A := \begin{bmatrix} \lambda & 1 & 1 \\ 1 & \lambda & 1 \\ 1 & 1 & \lambda \end{bmatrix}$$

```
> b := vector ( [ 1, 1, 1 ] ) ;
```

$$b := [1, 1, 1]$$

```
> x := linsolve ( A, b, 'r', t ) ;
```

$$x := \left[\frac{1}{\lambda + 2}, \frac{1}{\lambda + 2}, \frac{1}{\lambda + 2} \right]$$

По этому ответу, разумеется, видно, что при $\lambda = -2$ он теряет силу, но другой особый случай ($\lambda = 1$) по ответу "не чувствуется" (хотя, внимательно присмотревшись к системе, легко обнаружить, что при таком значении параметра все три уравнения становятся одинаковыми).

Освоив систему Maple с достаточной глубиной, вы сможете заставлять ее "тоньше" подходить к задачам с параметром, проводить необходимые исследования.

Составим расширенную матрицу системы.

> B := concat (A, b) ;

$$B := \begin{bmatrix} \lambda & 1 & 1 & 1 \\ 1 & \lambda & 1 & 1 \\ 1 & 1 & \lambda & 1 \end{bmatrix}$$

Приведем матрицу B к ступенчатому виду, не используя операцию деления на выражения, содержащие параметр.

Для этого предусмотрена особая команда ffgausselim — свободное от дробей (= fraction free) гауссово исключение. (Более простая команда gausselim — это обычное гауссово исключение, приведение матрицы к ступенчатому виду.)

> BG := ffgausselim (B) ;

$$BG := \begin{bmatrix} 1 & \lambda & 1 & 1 \\ 0 & 1 & -\lambda & -1+\lambda \\ 0 & 0 & 2-3\lambda+\lambda^3 & 1-2\lambda+\lambda^2 \end{bmatrix}$$

Разложим на множители многочлены от переменной λ, фигурирующие в ступенчатой матрице BG. Для отдельного многочлена его разложение на множители осуществляется командой factor. Но нам надо применить эту команду к каждому элементу матрицы. Здесь на помощь приходит еще одна команда — map (применить).

Не забудем также о команде evalm, о значении которой говорилось в замечании 2.6.

> BG := evalm (map (factor, BG)) ;

$$BG := \begin{bmatrix} 1 & \lambda & 1 & 1 \\ 0 & 1-\lambda & -1+\lambda & 0 \\ 0 & 0 & (\lambda+2)(-1+\lambda)^2 & (-1+\lambda)^2 \end{bmatrix}$$

Теперь видны "особые" случаи, влияющие на количество и расположение ступенек в ступенчатом виде матрицы: 1) $\lambda = 1$ и 2) $\lambda = -2$.

Подставим значение $\lambda = 1$ в матрицу BG.

В "скалярное" выражение подстановку можно осуществить универсальной командой eval. (Это одна из базовых команд системы Maple, обязательно изучите ее синтаксис самостоятельно.) Но в данном случае команду eval нужно применить (map) к каждому элементу матрицы BG.

Чтобы сделать это, можно сначала определить вспомогательную функцию:

```
> f1 := ex -> eval ( ex , lambda = 1 ) ;
```

$$f1 := ex \to ex\Big|_{\lambda=1}$$

Функция $f1$ сопоставляет (произвольному) выражению ex результат подстановки в это выражение числа 1 вместо переменной λ.

```
> BG1 := evalm ( map ( f1, BG ) ) ;
```

$$BG1 := \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Теперь решим систему линейных уравнений с расширенной матрицей $BG1$. Чтобы применить команду linsolve, надо выделить из $BG1$ подматрицу $AG1$ коэффициентов и вектор $bG1$ правых частей.

```
> AG1 := submatrix ( BG1, 1..3, 1..3 ) ;
  bG1 := convert ( submatrix ( BG1, 1..3, 4..4 ) , vector ) ;
```

$$AG1 := \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$bG1 := [1, 0, 0]$$

```
> x := linsolve ( AG1, bG1, 'r1', t ) ;
```

$$x := [1 - t_1 - t_2, t_1, t_2]$$

Переходим ко второму особому случаю.

> f2 := ex – > eval (ex , lambda = −2) ;

$$f2 := ex \rightarrow ex\Big|_{\lambda=-2}$$

> BG2 := evalm (map (f2, BG)) ;

$$BG2 := \begin{bmatrix} 1 & -2 & 1 & 1 \\ 0 & 3 & -3 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}$$

> AG2 := submatrix (BG2, 1..3, 1..3) ;
bG2 := convert (submatrix (BG2, 1..3, 4..4) , vector) ;

$$AG2 := \begin{bmatrix} 1 & -2 & 1 \\ 0 & 3 & -3 \\ 0 & 0 & 0 \end{bmatrix}$$

$$bG2 := [1, 0, 9]$$

> x := linsolve (AG2, bG2, 'r2', t) ;

$$x :=$$

Во втором случае получили "пустой ответ", что свидетельствует о несовместности системы.

Замечание 7.3. Разумеется, бывают задачи и с более чем одним параметром. Попробуйте, например, проанализировать следующую однородную систему:

$$\begin{cases} x_1 & + & x_2 & + & x_3 & = & 0; \\ \alpha x_1 & + & \beta x_2 & + & \gamma x_3 & = & 0; \\ \alpha^2 x_1 & + & \beta^2 x_2 & + & \gamma^2 x_3 & = & 0. \end{cases}$$

Матрица этой системы имеет название *матрица Вандермонда.* (Аналогичная матрица может быть записана в произвольном квадратном размере $n \times n$. Как?)

7.2. Линейные матричные уравнения. Рассмотрим уравнение вида

$$\underset{m\times n}{A}\cdot X=\underset{m\times p}{B} \tag{7.2}$$

с неизвестной матрицей X. Размеры этой матрицы однозначно восстанавливаются по заданным размерам матриц A и B: для того чтобы умножение в левой части было возможным и результат имел размеры, указанные в правой части, необходимо, чтобы X была $(n\times p)$-матрицей. Решение таких (и более сложных) линейных уравнений является стандартной задачей на практических занятиях по курсу алгебры.

Покажем, как уравнение (7.2) с неизвестной матрицей X заменить на равносильное уравнение вида (1.10) с неизвестным вектором-столбцом $\bar{x}$. Можно (как это фактически и делает компьютер) переформатировать матрицу X в столбец

$$\underset{(n\cdot p)\times 1}{\bar{\bar{x}}}=\begin{pmatrix}\bar{x}_1\\ \bar{x}_2\\ \ldots\ldots\ldots\\ \bar{x}_p\end{pmatrix}\in\mathbb{R}^{n\cdot p},$$

где

$$\bar{x}_k=\begin{pmatrix}x_{1k}\\ x_{2k}\\ \ldots\ldots\ldots\\ x_{nk}\end{pmatrix}\in\mathbb{R}^n,\ k=1,\ldots,p$$

— столбцы матрицы X.

Аналогичным образом переформатируем матрицу B:

$$\underset{(m\cdot p)\times 1}{\bar{\bar{b}}}=\begin{pmatrix}\bar{b}_1\\ \bar{b}_2\\ \ldots\ldots\ldots\\ \bar{b}_p\end{pmatrix}\in\mathbb{R}^{m\cdot p},$$

где

$$\bar{b}_k=\begin{pmatrix}b_{1k}\\ b_{2k}\\ \ldots\ldots\ldots\\ b_{mk}\end{pmatrix}\in\mathbb{R}^m,\ k=1,\ldots,p.$$

Вспоминая определение умножения матриц [см. формулу (1.6)], мы можем придать теперь левой части матричного уравнений (7.2) следующий вид:

$$A\cdot X=(A\cdot\bar{x}_1|\ A\cdot\bar{x}_2|\ \ldots|\ A\cdot\bar{x}_p)\ .$$

Правую часть уравнения также запишем по столбцам:

$$B = \left(\bar{b}_1 \middle| \bar{b}_2 \middle| \ldots \middle| \bar{b}_p\right) .$$

Видим, что матричное уравнение (7.2) оказывается равносильным системе векторных уравнений

$$A \cdot \bar{x}_k = \bar{b}_k;\ k = 1, ..., p\,,$$

которую, с использованием вектора $\bar{x}$, можно переписать в виде одного векторного уравнения

$$\mathbf{A} \cdot \bar{x} = \bar{b} \tag{7.3}$$

с новой матрицей, которая приводится ниже в блочном виде:

$$\underset{(m\cdot p)\times(n\cdot p)}{\mathbf{A}} = \begin{pmatrix} A & O & ... & O \\ O & A & ... & O \\ ... & ... & ... & ... \\ O & O & ... & A \end{pmatrix} .$$

[Попробуйте сами перемножить матрицу $\mathbf{A}$ и вектор $\bar{x}$. Хотя теперь в качестве "элементов" фигурируют не числа, а целые блоки, правило умножения (1.6) сохраняется. Вам предлагается самим осознать, почему это так.]

Пример 7.2. Решим матричное уравнение вида (7.2) с данными матрицами:

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

Элементы неизвестной (3×2)-матрицы X сразу занумеруем "под столбец" $\bar{x}$:

$$X = \begin{pmatrix} x_1 & x_4 \\ x_2 & x_5 \\ x_3 & x_6 \end{pmatrix} .$$

Выпишем векторное уравнение вида (7.3), соответствующее исходному матричному уравнению:

$$\begin{pmatrix} 1 & 0 & 2 & 0 & 0 & 0 \\ 1 & -1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & -1 & 3 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} .$$

Решая это уравнение методом Ж.—Г., получим:

$$\bar{x} = x_3 \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_6 \begin{pmatrix} 0 \\ 0 \\ 0 \\ -2 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

или, после обратного переформатирования векторов в матрицы:

$$X = x_3 \begin{pmatrix} -2 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} + x_6 \begin{pmatrix} 0 & -2 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 0 & 0 \end{pmatrix} .$$

Стандартная команда linsolve предусматривает возможность ее применения к линейным матричным уравнениям типа (7.2).

```
> restart; with( linalg ) ;

> A := matrix ( [ [ 1, 0, 2 ] , [ 1, −1, 3 ] ] ) ;
  B := matrix ( [ [ 1, 1 ] , [ −1, 0 ] ] ) ;
```

$$A := \begin{bmatrix} 1 & 0 & 2 \\ 1 & -1 & 3 \end{bmatrix}$$

$$B := \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}$$

```
> X := linsolve ( A, B, 'r', t ) ;
```

$$X := \begin{bmatrix} 1 - 2t_{1_1} & 1 - 2t_{2_1} \\ 2 + t_{1_1} & 1 + t_{2_1} \\ t_{1_1} & t_{2_1} \end{bmatrix}$$

Maple при решении матричных уравнений свободные неизвестные нумерует двойными индексами. Например, t_{2_1} — это первая свободная неизвестная во втором столбце и т. п.

Аналогично уравнению (7.2) можно рассмотреть линейное матричное уравнение

$$\underset{n\times p}{X} \cdot \underset{p\times q}{C} = \underset{n\times q}{B} . \tag{7.4}$$

В этом случае матрицу X надо разбить на n строк:

$$(\bar{x}^j)^t = (x_{j1}\ x_{j2}\ ...\ x_{jp}) \in \overset{*}{\mathbb{R}}{}^p\,;\ j = 1, ..., n,$$

каждая из которых по отдельности умножается справа на матрицу C; произведение приравнивается соответствующей строке матрицы B с последующим переходом к векторному уравнению для "длинной" (составной) строки.

При желании можно свести уравнение вида (7.4) к уравнению вида (7.2) с помощью транспонирования:

$$\underset{q\times p}{C^t} \cdot \underset{p\times n}{Y} = \underset{q\times n}{B^t},$$

где $Y = X^t$. Решив это уравнение относительно Y, нужно затем вернуться к X (обратным транспонированием).

Наиболее общим видом линейного матричного уравнения с одной неизвестной матрицей X является следующее уравнение:

$$A_1 \cdot X \cdot C_1 + A_2 \cdot X \cdot C_2 + \cdots + A_n \cdot X \cdot C_n = B\,,$$

где размеры всех матриц должны быть согласованы так, чтобы все умножения и все сложения были осуществимы и равенство было возможным.

На практике решение таких уравнений ничем принципиально не отличается от разобранного случая: как угодно обозначив элементы неизвестной матрицы, надо свести дело к решению обычной с.л.у.

Замечание 7.4. Можно (вручную или с помощью компьютерных систем) решать и нелинейные матричные уравнения.

Решим, например, средствами Maple простейшее квадратичное матричное уравнение

$$X^2 = E,$$

где единичная матрица E и искомая матрица X имеют размеры 2×2:

$$E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix};\ X = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix}.$$

Введем матрицы E и X:

> E := matrix ([[1, 0] , [0, 1]]) ;

x := array (1..4) ; X := matrix (2, 2, x) ;

$$E := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$x := array(1..4, [\,])$$

$$X := \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$$

Вычислим квадрат матрицы X:

> Y := evalm (X ^ 2) ;

$$Y := \begin{bmatrix} x_1^2 + x_2x_3 & x_1x_2 + x_2x_4 \\ x_3x_1 + x_4x_3 & x_3x_3 + x_4^2 \end{bmatrix}$$

Приступим к составлению системы уравнений. Сначала введем пустое множество уравнений:

> sys := { };

$$sys := \{\}$$

Затем (с помощью простой *процедуры*, используя команду объединения множеств union) добавим к множеству sys четыре уравнения, в которых приравниваются соответствующие элементы матриц Y и E:

```
> for i from 1 to 2 do
    for j from 1 to 2 do
      sys := sys union { Y[i,j] = E[i,j] } ;
    end do;
  end do;
  sys:=eval ( sys ) ;
```

Получим множество:

$$sys := \{x_1^2 + x_2x_3 = 1,\ x_1x_2 + x_2x_4 = 0,$$

$$x_3x_1 + x_4x_3 = 0,\ x_2x_3 + x_4^2 = 1\}$$

Решим теперь полученную систему из четырех квадратичных уравнений с четырьмя неизвестными. Делается это с помощью еще одной важнейшей команды solve, входящей в ядро системы Maple:

```
> solve(sys);
```

Ответ выдается как последовательность множеств, каждое из которых либо определяет "одиночное" решение системы, либо — целую "серию решений", содержащую произвольные (свободные) неизвестные.

$$\{x_2 = 0, x_3 = 0, x_4 = 1, x_1 = 1\},\ \{x_2 = 0, x_3 = 0, x_4 = -1, x_1 = -1\},$$

$$\{x_2 = 0, x_4 = -1, x_1 = 1, x_3 = x_3\},\ \{x_2 = 0, x_4 = 1, x_1 = -1, x_3 = x_3\},$$

$$\{x_1 = -x_4, x_3 = -\frac{-1 + x_4^2}{x_2}, x_2 = x_2, x_4 = x_4\}$$

Глава 2

АРИФМЕТИЧЕСКИЕ ЛИНЕЙНЫЕ ПРОСТРАНСТВА И ИХ ПОДПРОСТРАНСТВА. БАЗИСЫ И РАЗМЕРНОСТИ

§ 8. Системы векторов в пространстве $\mathbb{R}^n$ и их линейные оболочки

8.1. Конечные системы арифметических векторов и соответствующие матрицы

Определение 8.1. Конечной *системой векторов (с.в.)* в арифметическом линейном пространстве $\mathbb{R}^n$ называется *упорядоченный* конечный набор (конечная последовательность) арифметических векторов (не обязательно различных)

$$\mathcal{A} = [\,\bar{a}_1, \bar{a}_2, ..., \bar{a}_k\,], \tag{8.1}$$

где $\bar{a}_i \in \mathbb{R}^n$; $i = 1, ..., k$; $k \in \mathbb{N}$.

Замечание 8.1. (Это замечание, как и некоторые другие в дальнейшем, относится к типу "для служебного пользования", т. е. ориентировано на преподавателей. Студенты могут пока пропускать такие замечания.)

По причине упорядоченности набора, а также с учетом возможного наличия повторяющихся элементов в нем было бы некорректным называть $\mathcal{A}$ подмножеством пространства $\mathbb{R}^n$. Далее, однако, при изучении линейных оболочек конечных с.в., а также их свойств (типа линейной зависимости) будут неоднократно делаться замечания о независимости указанных понятий и свойств от порядка векторов в системе. Это провоцирует некоторых авторов на незаконное употребление включений типа $\mathcal{A} \subseteq \mathbb{R}^n$. Беда здесь не очень велика: в

конце концов, можно как-то условиться вместо упорядоченных наборов рассматривать неупорядоченные (и не содержащие повторений) подмножества. Но для "воспитания программистов" такой подход, несомненно, является вредным. Программист четко должен определять в задаче *тип переменной* (множество, список, последовательность, массив и т. п.).

Надо еще оговорить, в каком смысле мы будем употреблять термин *подсистема* конечной с.в.: в том же самом смысле, в котором понимается термин *подпоследовательность* в курсе математического анализа, с поправкой на конечность последовательности и подпоследовательностей.

Определение 8.2. Матрицей, соответствующей с.в. (8.1), называется матрица, составленная из векторов-столбцов, входящих в эту систему:

$$\underset{n\times k}{A} = (\bar{a}_1|\ \bar{a}_2|\ \dots\ |\bar{a}_k)\ . \tag{8.2}$$

Замечание 8.2. (Это замечание — для студентов.) Привыкайте к "математическим обычаям". Близкие, но различные математические понятия часто обозначают одинаковыми буквами разного начертания. Ниже мы будем обозначать системы векторов *рукописными* большими латинскими буквами, а соответствующие матрицы — соответствующими *печатными* буквами.

Определение 8.3. *Системой единичных векторов* в пространстве $\mathbb{R}^n$ называется с.в.

$$\mathcal{E}_n = [\,\bar{e}_1, \bar{e}_2, ..., \bar{e}_n\,], \tag{8.3}$$

где единичные векторы $\bar{e}_j$ $(j = 1, ..., n)$ имеют все нулевые элементы, кроме j-го элемента, равного 1:

$$\bar{e}_1 = \begin{pmatrix} 1 \\ 0 \\ ... \\ 0 \end{pmatrix};\ \bar{e}_2 = \begin{pmatrix} 0 \\ 1 \\ ... \\ 0 \end{pmatrix};\ \ \dots\ ;\ \bar{e}_n = \begin{pmatrix} 0 \\ 0 \\ ... \\ 1 \end{pmatrix}\ .$$

Замечание 8.3. Ясно, что системе векторов $\mathcal{E}_n$ соответствует единичная $(n \times n)$-матрица E_n. Загляните в первую главу: матрица E определялась формулой (2.7). Там мы индекс n не ставили, а теперь будем ставить в случае необходимости явно уточнить размер матрицы. Учтите, что вектор $\bar{e}_1$ в пространстве $\mathbb{R}^2$ и вектор $\bar{e}_1$ в пространстве $\mathbb{R}^3$ — это разные векторы.

8.2. Линейная оболочка конечной с.в. Линейной комбинацией с.в. (8.1) мы будем называть *выражение* вида $\sum_{i=1}^{k} \lambda_i \bar{a}_i$, где $\lambda_i \in \mathbb{R}$ $(i = 1, ..., k)$.

Если в этом выражении произвести действия (умножения на скаляры и сложение), то получится вектор

$$\bar{a} = \sum_{i=1}^{k} \lambda_i \bar{a}_i \,,$$

который называется *значением линейной комбинации.* (Сравните с обсуждением понятия линейной комбинации в замечании 2.4.)

Следует иметь в виду, что различные линейные кмбинации (одной и той же с.в.) могут иметь одинаковые значения; простейший пример: $\mathcal{A} = [\bar{a}, \bar{a}]$; $\bar{a} \neq \bar{0}$; $1 \cdot \bar{a} + (-1) \cdot \bar{a} = 2 \cdot \bar{a} + (-2) \cdot \bar{a} = \bar{0}$.

Из коммутативного закона для сложения векторов вытекает, что значение линейной комбинации не зависит от порядка слагаемых, т. е. от порядка векторов в системе $\mathcal{A}$.

Определение 8.4. *Линейной оболочкой* конечной с.в. (8.1) называется подмножество $\langle \mathcal{A} \rangle$ пространства $\mathbb{R}^n$, состоящее из значений всевозможных линейных комбинаций векторов этой системы.

Линейную оболочку $\langle \mathcal{A} \rangle$ можно описать одной из следующих формул:

$$\langle \mathcal{A} \rangle = \left\{ \sum_{i=1}^{k} \lambda_i \bar{a}_i \ : \ \lambda_i \in \mathbb{R};\ i = 1, ..., k \right\}; \tag{8.4}$$

$$\langle \mathcal{A} \rangle = \left\{ \bar{a} \in \mathbb{R}^n \ : \ [\, \exists \lambda_i \in \mathbb{R} \ (i = 1, ..., k)\,] \ [\bar{a} = \sum_{i=1}^{k} \lambda_i \bar{a}_i \,] \right\}. \tag{8.4a}$$

Формулу (8.4а) можно словесно выразить следующим образом: линейная оболочка $\langle \mathcal{A} \rangle$ состоит из всех векторов $\bar{a} \in \mathbb{R}^n$, представимых в виде значения какой-либо линейной комбинации векторов, входящих в с.в. $\mathcal{A}$.

Замечание 8.4. Линейная оболочка (в силу коммутативности сложения) не зависит от порядка векторов в с.в. $\mathcal{A}$, т. е. она не изменится при произвольной перестановке векторов в этой системе.

Удобно распространить определение линейной оболочки на случай пустой системы векторов; считается, что оболочка пустой с.в. состоит из одного нулевого вектора: $\langle \varnothing \rangle = \{\bar{0}\}$.

И несколько слов об обозначении линейной оболочки. В нашем основном учебнике [2] и в данном пособии она обозначается угловыми скобками: $\langle \mathcal{A} \rangle$. Используются, однако, и другие обозначения, например $\mathrm{lin}(\mathcal{A})$.

Далее нам необходимо вспомнить определение линейного подпространства в арифметическом линейном пространстве (см. определение 3.1): так называлось подмножество пространства $\mathbb{R}^n$, устойчивое относительно линейных операций над векторами.

Здесь мы введем специальное обозначение для свойства подмножества $V \subseteq \mathbb{R}^n$ *быть линейным подпространством* в $\mathbb{R}^n$. Мы будем записывать этот факт следующим образом: $V \leqslant \mathbb{R}^n$. Не упустите из внимания то, что хорошо знакомый вам знак числового неравенства $\leqslant$ будет в дальнейшем нести двойную нагрузку.

Предложение 8.1. *1. Для любой с.в. $\mathcal{A}$ ее линейная оболочка $\langle \mathcal{A} \rangle$ является линейным подпространством пространства $\mathbb{R}^n$.*

2. Это линейное подпространство является наименьшим из подпространств в $\mathbb{R}^n$, содержащих все векторы, входящие в $\mathcal{A}$, т. е. если для какого-либо линейного подпространства $V \leqslant \mathbb{R}^n$ справедливо $\bar{a}_i \in V$ $(\forall i = 1, ..., k)$, то $\langle \mathcal{A} \rangle \subseteq V$.

Доказательство. 1. Если

$$\bar{x} = \sum_{i=1}^{k} \lambda_i \bar{a}_i;\ \bar{y} = \sum_{i=1}^{k} \mu_i \bar{a}_i;\ (\lambda_i, \mu_i \in \mathbb{R};\ i = 1, ..., k)$$

— два вектора, принадлежащие $\langle \mathcal{A} \rangle$, то их сумма может быть с помощью законов (i) — (viii) векторной алгебры (см. теорему 2.1) приведена к виду

$$\bar{x} + \bar{y} = \sum_{i=1}^{k} (\lambda_i + \mu_i) \bar{a}_i$$

и, следовательно, тоже принадлежит $\langle \mathcal{A} \rangle$.

Столь же легко убедиться в том, что вместе с вектором $\bar{x}$ линейная оболочка содержит все пропорциональные векторы. Таким образом, доказано, что $\langle \mathcal{A} \rangle \leqslant \mathbb{R}^n$.

2. Пусть теперь $V \leqslant \mathbb{R}^n$ и для любого $i = 1, ..., k$ вектор $\bar{a}_i \in V$. Тогда, по определению подпространства, любая линейная комбинация векторов $\bar{a}_i$ принадлежит V. А это означает, что $\langle \mathcal{A} \rangle$ содержится в V. $\square$

В связи с только что доказанным предложением введем еще один термин.

Определение 8.5. Если линейное подпространство $V \leqslant \mathbb{R}^n$ представляется как линейная оболочка с.в. $\mathcal{A}$, то говорят, что векторы этой системы *порождают* V, а также что с.в. $\mathcal{A}$ является *порождающей* для подпространства V.

Рекомендуем вам подумать и понять следующие свойства порождающих с.в: 1) всякая система, содержащая порождающую с.в. для некоторого подпространства, сама является порождающей для этого подпространства; 2) из порождающей системы можно выбросить повторно встречающиеся, а также нулевые векторы, и она при этом останется порождающей (для того же самого подпространства).

Замечание 8.5. В алгебре используется также несколько иное (отличное от данного в определении 8.4) понятие линейной оболочки. Оно относится не к системам векторов, а к подмножествам, причем не обязательно конечным, пространства $\mathbb{R}^n$. Для $M \subseteq \mathbb{R}^n$ линейная оболочка $\langle M \rangle$, по определению, состоит из всевозможных конечных линейных комбинаций векторов из M. Это будет наименьшее из подпространств в $\mathbb{R}^n$, содержащих M. При таком подходе подмножество совпадает со своей линейной оболочкой тогда и только тогда, когда оно само является подпространством, и для любого подмножества M справедливо равенство $\langle\langle M \rangle\rangle = \langle M \rangle$.

8.3. Критерий принадлежности вектора линейной оболочке системы векторов. Теперь нам необходимо вспомнить из первой главы различные формы записи систем линейных уравнений. Прежде всего это развернутая запись (1.2):

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

Далее, изучив алгебраические операции над матрицами и векторами, мы получили матричную запись (1.10) этой же самой системы (здесь мы присвоим ей новый номер):

$$\underset{m\times n}{A} \cdot \underset{n\times 1}{\bar{x}} = \underset{m\times 1}{\bar{b}}\,. \tag{8.5}$$

Но можно также, пользуясь линейными алгебраическими операциями над векторами, получить еще одну форму записи с.л.у., кото-

рую мы будем называть *векторной*:

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ ... \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ ... \\ a_{m2} \end{pmatrix} + ... + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ ... \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ ... \\ b_m \end{pmatrix}. \qquad (8.6)$$

Формула (8.6) представляет собой векторное равенство в арифметическом линейном пространстве $\mathbb{R}^m$; неизвестные из с.л.у. (8.5) предстают здесь в облике неизвестных коэффициентов линейной комбинации

$$x_1\bar{a}_1 + x_2\bar{a}_2 + ... + x_n\bar{a}_n = \bar{b}\,, \qquad (8.7)$$

где символами $\bar{a}_j$ $(j = 1, 2, ..., n)$ обозначены столбцы матрицы A.

Замечание 8.6. Снова деликатное замечание "для служебного пользования". Многие учебные курсы по алгебре хранят верность традиции записывать векторы в строчку. Отчасти это мотивируется экономией бумаги и более легким набором. В нашу эпоху эти аргументы, кажется, потеряли силу: бумага стала такой дорогой, что ее уже никто не экономит, а новые компьютерные издательские системы позволяют легко набирать математические тексты любой степени сложности. (Здесь, прежде всего, следует упомянуть выдающееся изобретение гения программирования Д. Кнута — математическую издательскую систему TEX, используемую сейчас при верстке всех математических книг и журналов; применялась она и при подготовке к изданию настоящего пособия.)

Автору этого пособия преимущества использования в линейной алгебре векторов-столбцов представляются неоспоримыми. Причину существования этих преимуществ также можно усмотреть в традициях, причем значительно более глубоких. В мире, ориентированном на Европу, тексты пишут в строчку, слева направо, а сами строки идут сверху вниз. Именно поэтому развернутая запись с.л.у. (1.10) приводит к равенству (8.7) для векторов-столбцов. (Можно пофантазировать, какой была бы естественная векторная запись с.л.у., если бы наша письменность была организована иначе, например при письме в столбик, сверху вниз, с расположением столбцов слева направо.)

А теперь мы вернемся к теме настоящего параграфа и зададимся вопросом: как узнать принадлежит ли заданный вектор $\bar{b} \in \mathbb{R}^n$ линейной оболочке заданной с.в. (8.1)? Ответом служит следующее

Предложение 8.2. *Вектор $\bar{b}$ принадлежит линейной оболочке $\langle\mathcal{A}\rangle$ системы векторов (8.1) тогда и только тогда, когда совместна система линейных уравнений*

$$\underset{n\times k}{A}\cdot\underset{k\times 1}{\bar{\lambda}}=\underset{n\times 1}{\bar{b}}\,, \tag{8.8}$$

где A — матрица, соответствующая с.в. $\mathcal{A}$, $\bar{\lambda}$ — неизвестный вектор, принадлежащий пространству $\mathbb{R}^k$.

Доказательство. В соответствии с определением 8.4, вектор $\bar{b}$ принадлежит $\langle\mathcal{A}\rangle$ тогда и только тогда, когда он представляется в виде линейной комбинации

$$\bar{b}=\lambda_1\bar{a}_1+\lambda_2\bar{a}_2+...+\lambda_k\bar{a}_k\,, \tag{8.9}$$

т. е. должны найтись скаляры λ_j $(j=1,2,...,k)$, превращающие (8.9) в истинное векторное равенство.

Как только что объяснено, векторному соотношению (8.9) можно придать вид с.л.у., причем именно с такими размерами матриц и столбцов, как это указано в формуле (8.8). Вопрос, таким образом, сводится к выяснению характера этой с.л.у.: будет ли она совместной или же несовместной. Что и требовалось доказать. □

Замечание 8.7. Теперь у вас есть прекрасная возможность "запутаться в системах". Вообще слово "система" в математике (и не только в математике) весьма многозначно. Пока в алгебре нам встретились два термина, включающих это слово: "система линейных уравнений" и "система векторов". Разумеется, это совершенно различные, хотя и связанные, понятия. Наши аббревиатуры (с.л.у. и с.в.), помимо цели сокращения записей, имеют еще одну цель — облегчить вашему подсознанию разграничение соответствующих понятий.

8.4. Примеры линейных оболочек. Приведем несколько примеров "вычисления" линейной оболочки с.в. Каждый из этих примеров найдет применения в дальнейшей теории.

Пример 8.1. Рассмотрим с.в. $\mathcal{A}=[\bar{a}]$, состоящую из единственного вектора $\bar{a}\in\mathbb{R}^n$. Ясно, что если этот вектор является нулевым, то и линейная оболочка $\langle\mathcal{A}\rangle$ является нулевым подпространством в $\mathbb{R}^n$. В противном случае $\langle\mathcal{A}\rangle$ будет состоять из всех векторов, пропорциональных вектору $\bar{a}$ (такое подпространство в $\mathbb{R}^n$ называется *прямой*, порожденной вектором $\bar{a}$).

Пример 8.2. Система единичных векторов $\mathcal{E}_n$ является порождающей для пространства $\mathbb{R}^n$. В самом деле, любой вектор $\bar{x} \in \mathbb{R}^n$ можно представить в виде линейной комбинации:

$$\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ ... \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ ... \\ 0 \end{pmatrix} + ... + x_n \begin{pmatrix} 0 \\ 0 \\ ... \\ 1 \end{pmatrix} = \sum_{i=1}^{n} x_i \bar{e}_i \, .$$

Перед тем как рассмотреть следующий важный (теоретический) пример, вспомним формулу (6.3h), представляющую общее решение однородной системы линейных уравнений (1.10h). (Ниже мы снабдим эту с.л.у. новым номером, согласованным с общей нумерацией формул настоящего параграфа.]

Пример 8.3. Рассмотрим однородную с.л.у.

$$\underset{m\times n}{A} \cdot \underset{n\times 1}{\bar{x}} = \underset{m\times 1}{\bar{0}} \, . \tag{8.5h}$$

Пусть количество ступенек в ступенчатом виде матрицы A равняется r, причем главными являются первые r неизвестных. Рассмотрим систему векторов

$$\mathcal{F} = \left[\bar{f}_{r+1}, \bar{f}_{r+2}, ..., \bar{f}_n \right] , \tag{8.10}$$

состоящую из базисных частных решений системы (8.5h).

Согласно теореме 6.1, множество L^0 решений однородной с.л.у. (8.5h) исчерпывается линейными комбинациями

$$\bar{x} = x_{r+1}\bar{f}_{r+1} + x_{r+2}\bar{f}_{r+2} + ... + x_n\bar{f}_n \, , \tag{8.11}$$

а это означает, что подпространство L^0 решений с.л.у. (8.5h) является линейной оболочкой с.в. (8.10).

§ 9. Линейно зависимые и линейно независимые системы векторов

9.1. Понятие линейно зависимой (линейно независимой) с.в. Снова рассмотрим конечную с.в. (8.1) в пространстве $\mathbb{R}^n$.

Определение 9.1. 1. С.в. $\mathcal{A}$ называется *линейно зависимой*, если существует такая линейная комбинация векторов системы $\mathcal{A}$, значение которой равно нулю, хотя не все коэффициенты этой линейной комбинации равны нулю.

2. В противном случае, т. е. если из равенства нулю значения линейной комбинации вытекает, что все ее коэффициенты равны нулю, с.в. называется *линейно независимой*.

Замечание 9.1. Данное выше определение удобно дополнить следующим образом: пустая с.в. считается линейно независимой. Если вы склонны к логическим упражнениям, то попробуйте найти чисто логическое обоснование такой условности.

Свойство с.в. быть линейно зависимой (независимой) не зависит от порядка векторов в этой системе (ср. с замечанием 8.4).

Замечание 9.2. Понятие линейно зависимой (линейно независимой) с.в., по-видимому, уже встречалось вам в курсе аналитической геометрии. Будет оно встречаться и в других математических дисциплинах.

Однако понятие это является в достаточной степени нетривиальным и обычно вызывает серьезные затруднения у студентов. В виде исключения преподаватели даже рекомендуют прием "зазубривания" определения 9.1 (в надежде, что после очередного повторения вслух вдруг прояснится ускользавший от вас ранее смысл этого определения). В частности, для вас должны стать совершенно очевидными следующие простейшие свойства и примеры линейно зависимых (линейно независимых) с.в.

Любая с.в., содержащая нулевой вектор, линейно зависима. (В самом деле, линейная комбинация векторов этой системы, в которой все коэффициенты равны 0, кроме одного, стоящего при нулевом векторе, имеет значение $\bar{0}$.) Столь же легко можно убедиться в том, что с.в., содержащая повторяющиеся векторы, линейно зависима.

Система, состоящая из одного вектора, является линейно независимой тогда и только тогда, когда этот вектор ненулевой. Система, состоящая из двух векторов, линейно независима тогда и только тогда, когда эти векторы не пропорциональны.

Более серьезные свойства и примеры будут приведены далее.

9.2. Критерий линейной зависимости (линейной независимости) с.в.

Предложение 9.1. *Пусть $\mathcal{A}$ — конечная с.в. (8.1), A — соответствующая ей матрица (8.2). Система $\mathcal{A}$ является линейно зависимой*

(линейно независимой) тогда и только тогда, когда однородная с.л.у.

$$\underset{n\times k}{A}\cdot\underset{k\times 1}{\bar{\lambda}}=\underset{n\times 1}{\bar{0}} \tag{8.8h}$$

имеет нетривиальное решение (имеет только тривиальное решение).

Доказательство. С.в. $\mathcal{A}$ будет линейно зависимой, если векторное равенство

$$\lambda_1\bar{a}_1+\lambda_2\bar{a}_2+...+\lambda_k\bar{a}_k=\bar{0} \tag{8.9h}$$

может быть истинным для некоторых (не всех нулевых) значений коэффициентов λ_j $(j=1,2,...,k)$. Если же равенство (8.9h) влечет обращение в нуль всех λ_j, то с.в. является линейно независимой.

При доказательстве предложения 8.2 объяснялось, что векторное равенство (8.9) равносильно матричному равенству (8.8). Сейчас мы рассматриваем "однородный вариант" (8.9h) равенства (8.9). Он будет равносилен однородной с.л.у. (8.8h), и следовательно, факт линейной зависимости с.в. будет равносилен факту наличия у с.л.у. (8.8h) ненулевого решения. □

Следствие. *Всякая с.в. в пространстве $\mathbb{R}^n$, содержащая более чем n векторов, является линейно зависимой.*

Доказательство. Если количество векторов в данной с.в. $k>n$, то число неизвестных в однородной с.л.у. (8.8h) будет больше числа уравнений. По предложению 6.1, эта с.л.у. будет иметь нетривиальное решение. Следовательно, с.в. будет линейно зависимой. □

9.3. Свойства линейно зависимых (линейно независимых) с.в. В следующем предложении собраны важнейшие свойства линейно зависимых (независимых) систем векторов. Некоторые из упомянутых ранее свойств могут быть выведены иначе, с использованием этого предложения.

Предложение 9.2. *1. Пусть $\mathcal{B}$ — конечная с.в., $\mathcal{A}$ — ее подсистема. Тогда если система $\mathcal{A}$ является линейно зависимой, то и система $\mathcal{B}$ является линейно зависимой.*

2. С.в. $\mathcal{A}$ является линейно зависимой тогда и только тогда, когда некоторый вектор, входящий в систему $\mathcal{A}$, линейно выражается через остальные векторы этой системы.

3. Пусть с.в. векторов $\mathcal{A}$ линейно независима, а с.в. $\mathcal{B}$ получается из системы $\mathcal{A}$ присоединением к ней одного вектора $\bar{b}$. Тогда с.в. $\mathcal{B}$

будет линейно зависимой в том и только том случае, когда вектор $\bar{b}$ *принадлежит линейной оболочке* $\langle A \rangle$.

Доказательство. 1. Свойство с.в. быть линейно зависимой (линейно независимой) не зависит от порядка векторов в этой системе (см. замечание 9.1). Поэтому можно считать, что с.в. $\mathcal{B}$ занумерована следующим образом:

$$\mathcal{B} = [\,\bar{a}_1, \bar{a}_2, ..., \bar{a}_k\,, \bar{a}_{k+1}, ..., \bar{a}_l]\,,$$

где первые k векторов составляют с.в. $\mathcal{A}$.

Если система $\mathcal{A}$ линейно зависима, т. е. найдутся скаляры λ_i $(i = 1, ..., k)$, не все равные нулю и такие, что $\sum_{i=1}^{k} \lambda_i \bar{a}_i = \bar{0}$, то, полагая $\lambda_i = 0$ для $i = k+1, ..., l$, мы получим $\sum_{i=1}^{l} \lambda_i \bar{a}_i = \bar{0}$, что (в силу наличия ненулевых коэффициентов в последней линейной комбинации) свидетельствует о линейной зависимости с.в. $\mathcal{B}$.

Добавим, что доказанное выше утверждение мы будем использовать еще и в такой равносильной форме: всякая подсистема линейно независимой с.в. является линейно независимой.

2. Если с.в. $\mathcal{A}$, заданная формулой (8.1), линейно зависима, то для некоторого набора коэффициентов λ_i $(i = 1, ..., k)$, среди которых имеются ненулевые, будет выполняться векторное равенство (8.9h). Снова, пользуясь несущественностью порядка векторов, можно считать, что первый коэффициент $\lambda_1 \neq 0$. Тогда первый вектор $\bar{a}_1$ можно линейно выразить через остальные:

$$\bar{a}_1 = -\frac{\lambda_2}{\lambda_1}\bar{a}_2 - \cdots - \frac{\lambda_k}{\lambda_1}\bar{a}_k.$$

Обратно, если какой-то (можно считать, что — первый) вектор системы $\mathcal{A}$ линейно выражается через остальные:

$$\bar{a}_1 = \mu_2 \bar{a}_2 + \cdots + \mu_k \bar{a}_k\,,$$

то

$$1 \cdot \bar{a}_1 + (-\mu_2) \cdot \bar{a}_2 + \cdots + (-\mu_k) \cdot \bar{a}_k = \bar{0},$$

что (в силу присутствия ненулевого коэффициента в последней линейной комбинации) свидетельствует о линейной зависимости с.в. $\mathcal{A}$.

3. Пусть с.в. (8.1) линейно независима, а после добавления к ней вектора $\bar{b}$ получается линейно зависимая с.в. $\mathcal{B} = [\mathcal{A}, \bar{b}]$. Это означает,

что для некоторых (не всех нулевых) скаляров $\lambda_1, ..., \lambda_k, \mu$ выполняется равенство

$$\lambda_1 \bar{a}_1 + ... + \lambda_k \bar{a}_k + \mu \bar{b} = \bar{0}.$$

Ясно, что коэффициент $\mu \neq 0$ (в противном случае последнее равенство свидетельствовало бы о линейной зависимости с.в. $\mathcal{A}$, что противоречит предположению). А это позволяет (см. доказательство предыдущего пункта) линейно выразить вектор $\bar{b}$ через векторы системы $\mathcal{A}$. Последнее же означает, что вектор $\bar{b}$ принадлежит линейной оболочке $\langle \mathcal{A} \rangle$.

Обратно, если вектор $\bar{b} \in \langle \mathcal{A} \rangle$, то, в силу п. 2 настоящего предложения, система $\mathcal{B}$ будет линейно зависимой. □

Замечание 9.3. Рассмотрим конечную с.в. (8.1) в пространстве $\mathbb{R}^n$. Выберем какие-либо r компонент из общего набора $x_1, ..., x_n$ и выбросим эти компоненты у всех векторов системы $\mathcal{A}$, перейдя тем самым к системе векторов в пространстве $\mathbb{R}^{n-r}$. Эту операцию назовем "укорачиванием" векторов и обозначим символом $\mathcal{A}'$ систему укороченных векторов.

Из предложения 9.1 легко усматривается, что линейная зависимость с.в. $\mathcal{A}$ влечет линейную зависимость с.в. $\mathcal{A}'$ (если в истинном векторном равенстве выбросить у всех векторов некоторые компоненты, то равенство останется истинным).

(Вместо несколько жаргонного термина "укорачивание" хотелось бы употребить более адекватный термин "проектирование", но, наверное, это потребовало бы дополнительных объяснений, здесь не очень своевременных.)

9.4. Примеры линейно независимых с.в. Приводимые ниже примеры являются продолжением примеров 8.2 и 8.3.

Пример 9.1. Система единичных векторов $\mathcal{E}_n$ в пространстве $\mathbb{R}^n$ является линейно независимой. В самом деле, если $\sum_{i=1}^{n} \lambda_i \bar{e}_i = \bar{0}$, то нулевым будет вектор $\bar{\lambda}$, составленный из коэффициентов этой линейной комбинации, а значит, и сами эти коэффициенты будут все равны нулю.

Пример 9.2. Рассмотрим с.в. (8.10), составленную из базисных решений однородной с.л.у. (8.5h), т. е. систему

$$\bar{f}_{r+1} = \begin{pmatrix} -h_{1(r+1)} \\ -h_{2(r+1)} \\ \dots \\ -h_{r(r+1)} \\ 1 \\ 0 \\ \dots \\ 0 \end{pmatrix}, \bar{f}_{r+2} = \begin{pmatrix} -h_{1(r+2)} \\ -h_{2(r+2)} \\ \dots \\ -h_{r(r+2)} \\ 0 \\ 1 \\ \dots \\ 0 \end{pmatrix}, \dots, \bar{f}_n = \begin{pmatrix} -h_{1n} \\ -h_{2n} \\ \dots \\ -h_{rn} \\ 0 \\ 0 \\ \dots \\ 1 \end{pmatrix}.$$

Если "укоротить" эти векторы, выбросив первые r компонент (см. замечание 9.3), то останется система $\mathcal{E}_{n-r}$ единичных векторов пространства $\mathbb{R}^{n-r}$, которая линейно независима (см. предыдущий пример). А значит, в силу замечания 9.3, является линейно независимой и исходная с.в. (8.10).

§ 10. Базисы в линейных подпространствах пространства $\mathbb{R}^n$

10.1. Понятие базиса в линейном подпространстве арифметического линейного пространства. Пусть V является линейным подпространством в пространстве $\mathbb{R}^n$.

Определение 10.1. Конечная с.в. $\mathcal{A}$ называется *базисом* подпространства V, если

1) система $\mathcal{A}$ порождает V, т. е. $\langle A \rangle = V$;

2) система $\mathcal{A}$ линейно независима.

Пример 10.1. В соответствии с замечаниями 8.4 и 9.1, пустую с.в. можно считать базисом нулевого подпространства $V = O$.

Система из одного ненулевого вектора $\bar{a} \in \mathbb{R}^n$ является базисом линейной оболочки $V = \langle \bar{a} \rangle$. Базисом V будет также и любая одноэлементная система $[\bar{b}]$, где $\bar{b} = \alpha\bar{a}$ ($\alpha \in \mathbb{R}$, $\alpha \neq 0$). (Почему?)

Пример 10.2. Система единичных векторов $\mathcal{E}_n$ (см. примеры 8.2 и 9.1) является базисом во всем пространстве ($V = \mathbb{R}^n$).

Определение 10.2. Базис $\mathcal{E}_n$ называется *естественным базисом* пространства $\mathbb{R}^n$.

Пример 10.3. Рассмотрим однородную с.л.у. (8.5h). Обозначим (как уже делалось в первой главе) множество решений этой системы символом L^0. В предложении 3.2 было доказано, что L^0 является линейным подпространством в пространстве $\mathbb{R}^n$. Применяя результаты примеров 8.3 и 9.2, мы можем утверждать, что с.в. $\mathcal{F}$ [см. формулу (8.10)], составленная из базисных решений с.л.у. (8.5h), является базисом подпространства L^0. (Теперь мы можем считать оправданным сам термин "базисное решение".)

10.2. Свойство единственности для разложения по базису. Если с.в. $\mathcal{A}$ порождает подпространство $V \leqslant \mathbb{R}^n$, то любой вектор $\bar{b} \in V$ представляется в виде линейной комбинации векторов системы $\mathcal{A}$ [см. формулу (8.9)]. Коэффициенты разложения (8.9) определены, вообще говоря, неоднозначно. Однозначность определения этих коэффициентов имеет место в том и только том случае, когда с.в. является базисом подпространства V. Точнее, справедливо следующее

Предложение 10.1. *С.в. $\mathcal{A}$, порождающая линейное подпространство V, является базисом этого подпространства тогда и только тогда, когда для любого вектора $\bar{b} \in V$ коэффициенты λ_i $(i = 1, ..., k)$ разложения $\bar{b} = \sum_{i=1}^{k} \lambda_i \bar{a}_i$ определены однозначно.*

Доказательство. 1. Пусть с.в. $\mathcal{A}$ является базисом подпространства V. Возьмем любой вектор $\bar{b} \in V$. Допустим, имеются два разложения этого вектора по базису $\mathcal{A}$:

$$\bar{b} = \sum_{i=1}^{k} \lambda_i \bar{a}_i = \sum_{i=1}^{k} \mu_i \bar{a}_i; \;\; \lambda_i, \mu_i \in \mathbb{R}; \;\; i = 1, ..., k.$$

Тогда, пользуясь законами алгебры векторов, мы получим равенство

$$\sum_{i=1}^{k} (\lambda_i - \mu_i) \bar{a}_i = \bar{0},$$

из которого, в силу линейной независимости с.в. $\mathcal{A}$, вытекает, что $\lambda_i - \mu_i = 0$ для любого номера i, т. е. два разложения совпадают.

2. Обратное утверждение мы докажем даже в более сильной форме: если единственность разложения по порождающей с.в. $\mathcal{A}$ имеет место *для некоторого* вектора $\bar{a} \in V$, то эта с.в. является базисом V

(и следовательно, по уже доказанному, единственность разложения будет иметь место и *для любого* вектора $\bar{b}$ из подпространства V).

В самом деле, если

$$\bar{0} = \sum_{i=1}^{k} \lambda_i \bar{a}_i,$$

а

$$\bar{a} = \sum_{i=1}^{k} \alpha_i \bar{a}_i$$

есть разложение вектора $\bar{a}$ по системе $\mathcal{A}$, то, складывая (с использованием законов векторной алгебры) эти формулы почленно, мы получим

$$\bar{a} = \sum_{i=1}^{k} (\lambda_i + \alpha_i) \bar{a}_i.$$

А это есть еще одно разложение вектора $\bar{a}$ по системе $\mathcal{A}$. В силу предположения о единственности разложения для вектора $\bar{a}$, мы получим $\lambda_i + \alpha_i = \alpha_i$ для любого номера i, откуда вытекает обращение в нуль всех λ_i, что свидетельствует о линейной независимости с.в. $\mathcal{A}$. □

10.3. Теорема существования базиса. В этом пункте нам предстоит доказать факт первостепенной важности — теорему, утверждающую, что в любом линейном подпространстве арифметического линейного пространства существует базис.

Теорема 10.1. *1. В любом линейном подпространстве $V \leqslant \mathbb{R}^n$ существует базис $\mathcal{B}$. Количество векторов в базисе $\mathcal{B}$ не превосходит числа n.*

2. Если подпространство V представлено как линейная оболочка $V = \langle \mathcal{A} \rangle$ некоторой системы векторов $\mathcal{A}$, то базисные векторы можно выбрать из числа порождающих векторов.

Доказательство. 1. Если подпространство $V = O$, то (см. пример 10.1) в качестве базиса в нем можно рассматривать пустую с.в. $\mathcal{B} = [\,]$. Пусть далее подпространство V является ненулевым.

Выберем произвольный ненулевой вектор $\bar{b}_1 \in V$ и начнем индуктивный процесс построения базиса в V с одноэлементной с.в. $\mathcal{B}_1 = [\bar{b}_1]$. Эта с.в. является линейно независимой (см. замечание 9.2).

На каждом шаге последующего процесса у нас будет получаться линейно независимая с.в. $\mathcal{B}_1, \mathcal{B}_2, \dots$, и мы остановим этот процесс тогда, когда сможем утверждать, что очередная с.в. является базисом в V.

Пусть уже построена линейно независимая система

$$\mathcal{B}_k = \left[\bar{b}_1, ..., \bar{b}_k\right],$$

составленная из векторов подпространства V. Имеются две логические возможности:

— либо для любого вектора $\bar{b} \in V$ система $[\bar{b}_1, ..., \bar{b}_k, \bar{b}]$ линейно зависима;

— либо существует вектор $\bar{b} \in V$ такой, что с.в. $[\bar{b}_1, ..., \bar{b}_k, \bar{b}]$ линейно независима.

В первом случае п. 3 предложения 9.2 позволит нам утверждать, что любой вектор $\bar{b} \in V$ принадлежит линейной оболочке $\langle \mathcal{B}_k \rangle$, т. е. эта оболочка совпадает с подпространством V, т. е. система $\mathcal{B}_k$ является порождающей для V и, следовательно,— базисом V.

Во втором случае мы переобозначим "вновь обнаруженный" вектор ($\bar{b} = \bar{b}_{k+1}$) и получим более широкую линейно независимую с.в.

$$\mathcal{B}_{k+1} = \left[\bar{b}_1, ..., \bar{b}_k, \bar{b}_{k+1}\right];$$

процесс продолжается.

Выстраиваемая последовательность линейно независимых конечных систем векторов является строго расширяющейся; она не может "продолжаться до бесконечности", поскольку, в силу следствия из предложения 9.1, всякая линейно независимая с.в. в пространстве $\mathbb{R}^n$ содержит не более n векторов. Значит, на некотором шаге (с номером $r \leqslant n$) обязательно осуществится первая возможность и полученная с.в. $\mathcal{B}_r$ окажется базисом подпространства V. Первое утверждение теоремы доказано.

2. Пусть теперь данное (ненулевое) подпространство $V \leqslant \mathbb{R}^n$ представлено в виде $V = \langle \mathcal{A} \rangle$, где $\mathcal{A} = [\bar{a}_1, ..., \bar{a}_s]$ — конечная с.в. Модифицируем индуктивный процесс построения базиса следующим образом: на каждом шаге будем выбирать векторы *из числа порождающих,* т. е. из системы $\mathcal{A}$. Точно так же, как и в предыдущем пункте, этот процесс оборвется на некотором шаге [с номером $r \leqslant \min(n, s)$]. В момент "останова" будет получена линейно независимая с.в.

$$\mathcal{B}_r = \left[\bar{b}_1, ..., \bar{b}_r\right], \tag{10.1}$$

такая, что всякий вектор системы $\mathcal{A}$ линейно выражается через векторы системы $\mathcal{B}_r$. По предположению, с.в. $\mathcal{A}$ является порождающей для V, т. е. всякий вектор $\bar{b} \in V$ линейно выражается через векторы этой системы. Подставляя в полученное выражение для $\bar{b}$ через $\bar{a}_i$ $(i = 1, ..., s)$ выражения каждого из $\bar{a}_i$ через векторы системы $\mathcal{B}_r$ и пользуясь законами векторной алгебры, мы получим (после перегруппировки) выражение $\bar{b}$ через $\mathcal{B}_r$. Этим будет установлено, что с.в. (10.1) является базисом подпространства V. □

Замечание 10.1. Внимательно проанализировав ход доказательства первого утверждения теоремы 10.1, мы можем заметить, что фактически установлено следующее: *максимальная линейно независимая система векторов* подпространства V (т. е. такая линейно независимая с.в., которая не содержится ни в какой строго более широкой линейно независимой системе) является базисом в V.

Обратное утверждение (о том, что всякий базис является максимальной линейно независимой с.в.) также легко усматривается из п. 3 предложения 9.2.

Таким образом, определение базиса (см. определение 10.1) равносильно другому возможному определению: базисом подпространства можно назвать максимальную линейно независимую систему векторов из этого подпространства.

В качестве логического упражнения можно предложить читателям доказать, что имеется еще и третий равносильный способ характеризации базисов: система векторов является базисом в подпространстве V тогда и только тогда, когда она является *минимальной порождающей с.в. для* V. (Вернитесь в связи с этим к определению 8.4 и последующим рекомендациям.)

Замечание 10.2. Обратим внимание юных компьютерщиков на радикальное различие рассуждений в первой и во второй частях доказательства теоремы 10.1.

В своей первой части эта теорема является, как говорят, "теоремой существования". На каждом шаге построения базиса мы довольствуемся тем, что вектор, подлежащий добавлению к ранее построенной системе, *существует.* Не указывается (и не может быть указано) никакого явного способа выбора этого вектора из (бесконечного) множества векторов, которые могут быть (на данном шаге) добавлены.

Вторая часть доказательства является, по сути, описанием работы *алгоритма* выбора базиса в подпространстве. Изначально это

подпространство может быть задано $(n \times s)$-матрицей A, составленной из порождающих векторов-столбцов. На выходе остается подматрица B_r матрицы A, содержащая те из столбцов этой матрицы, которые входят в некоторый базис данного подпространства.

Алгоритм выбора базиса нетрудно записать на каком-либо из алгоритмических языков; различные его варианты используются компьютерными математическими системами Maple и MATLAB. В следующих пунктах мы (хотя и бегло) познакомимся с работой Maple в задачах построения базисов.

10.4. Свойство продолжения базисов. В дальнейшем нам понадобится так называемое *свойство продолжения базисов*: базис в некотором подпространстве можно "продолжить", т. е. включить его в базис в каком-либо более широком подпространстве (в частности, базис в подпространстве можно продолжить до базиса во всем пространстве $\mathbb{R}^n$). Точнее, справедливо следующее

Предложение 10.2. *1. Всякая линейно независимая система векторов в линейном подпространстве $V \leqslant \mathbb{R}^n$ может быть включена в некоторый базис этого подпространства.*

2. Пусть V и W два линейных подпространства в $\mathbb{R}^n$, причем V содержится в W (т. е. $V \subseteq W$). Тогда всякий базис подпространства V может быть дополнен до базиса в подпространстве W.

Доказательство. 1. Пусть $\mathcal{A} = [\bar{a}_1, ..., \bar{a}_k]$ — линейно независимая с.в. в подпространстве V. Мы можем считать, что первые k шагов в процессе построения базиса, описанном в первой части доказательства теоремы 10.1, уже сделаны. Доведя этот процесс до конца, мы получим базис в V, содержащий данную линейно независимую с.в. $\mathcal{A}$.

2. Если имеется пара вложенных одно в другое подпространств $V \subseteq W$, то всякий базис в V, будучи линейно независимой с.в. в W, допускает (по доказанному в п. 1) продолжение до базиса в W.

(Дополнительно заметим, что если W строго содержит V, то и полученный базис в W будет строго содержать исходный базис в V. Почему?) $\square$

§ 11. Равномощность базисов в подпространстве. Понятие размерности подпространства. Ступенчатый ранг матрицы

11.1. Теорема о равномощности всех базисов в подпространстве. Теперь пришла очередь доказать второй, после теоремы существования базиса, фундаментальный факт линейной алгебры — теорему о том, что все базисы в одном и том же линейном подпространстве $V \leqslant \mathbb{R}^n$ имеют одинаковое количество векторов.

Напомним из курса "Введение в математику", что два множества называются *равномощными*, если между ними существует взаимно однозначное (биективное) соответствие. Гораздо сложнее определяется понятие *мощности* множества. Эти вопросы вы будете изучать позже, в курсе математической логики. Но для конечных множеств понятие мощности вполне элементарно: мощностью конечного множества называется количество элементов в этом множестве.

Для *списков* (которые отличаются от конечных множеств наличием порядка и, возможно, повторяющихся элементов) вместо понятия мощности можно рассматривать понятие *длины списка*. Но поскольку базисы не могут содержать повторяющихся векторов и остаются базисами после произвольной перестановки своих элементов, то (вообще говоря, не очень корректный) термин "мощность базиса" имеет право на существование. И мы будем его применять, что позволит ниже весьма лаконично сформулировать основной результат параграфа — теорему 11.1.

Но сначала докажем следующее вспомогательное

Предложение 11.1. *Пусть V — ненулевое линейное подпространство в $\mathbb{R}^n$;*

$$\mathcal{A} = [\bar{a}_1, ..., \bar{a}_s] \tag{11.1}$$

и

$$\mathcal{B} = \left[\bar{b}_1, ..., \bar{b}_r\right] \tag{11.2}$$

— две системы векторов в подпространстве V, причем система $\mathcal{B}$ является базисом V, а система $\mathcal{A}$ линейно независима. Тогда $s \leqslant r$.

Доказательство. Достаточно доказать, что если справедливо противоположное неравенство $s > r$, то с.в. (11.1) является линейно зависимой, т. е. найдутся скаляры λ_j $(j = 1, ..., s)$, не все равные

нулю и такие, что имеет место равенство

$$\sum_{j=1}^{s} \lambda_j \bar{a}_j = \bar{0}\,. \tag{11.3}$$

Каждый из векторов системы (11.1) можно разложить по базису (11.2):

$$\bar{a}_j = \sum_{i=1}^{r} a_{ij} \bar{b}_i;\ j = 1, ..., s. \tag{11.4}$$

Обратите внимание на принцип нумерации скаляров a_{ij}: второй номер j — это номер вектора в с.в. (11.1), а первый номер i — это номер коэффициента в разложении этого вектора по базису (11.2).

Подставив выражения (11.4) в равенство (11.3), мы придадим этому равенству равносильный вид

$$\sum_{j=1}^{s} \lambda_j \sum_{i=1}^{r} a_{ij} \bar{b}_i = \bar{0}\,. \tag{11.5}$$

Теперь нам надо вспомнить правила обращения с суммами:

— внесение постоянного (не зависящего от i) множителя λ_j под знак суммы (по индексу i) [см. формулу (2.5) и последующие комментарии в первой главе];

— перемена порядка суммирования [см. формулу (2.6)];

— вынесение постоянного (не зависящего от i) векторного множителя $\bar{b}_i$ вправо из-под знака суммы (по индексу i);

— перестановка скалярных множителей λ_j и a_{ij} под знаком суммы (по индексу j).

Таким образом, получается следующая цепочка преобразований левой части равенства (11.5):

$$\sum_{j=1}^{s} \lambda_j \bar{a}_j = \sum_{j=1}^{s} \lambda_j \left(\sum_{i=1}^{r} a_{ij} \bar{b}_i \right) =$$

$$= \sum_{j=1}^{s} \left(\sum_{i=1}^{r} \lambda_j a_{ij} \bar{b}_i \right) = \sum_{i=1}^{r} \left(\sum_{j=1}^{s} \lambda_j a_{ij} \bar{b}_i \right) =$$

$$= \sum_{i=1}^{r} \left(\sum_{j=1}^{s} a_{ij} \lambda_j \right) \bar{b}_i = \sum_{i=1}^{r} \mu_i \bar{b}_i\,,$$

где обозначено:

$$\mu_i = \sum_{j=1}^{s} a_{ij}\lambda_j \,. \tag{11.6}$$

После этих преобразований равенство (11.5) приобретет равносильный вид

$$\sum_{i=1}^{r} \mu_i \bar{b}_i = \bar{0} \,. \tag{11.7}$$

Но с.в. (11.2) является базисом и, следовательно, линейно независима. Поэтому равенство (11.7) равносильно утверждению

$$(\,\forall i = 1, ..., r\,)\ [\,\mu_i = 0\,]\,. \tag{11.8}$$

С учетом (11.6) утверждение (11.8) равносильно однородной системе линейных уравнений [здесь она получилась в компактной записи, см. формулу (1.2а)]:

$$\sum_{j=1}^{s} a_{ij}\lambda_j = 0;\ \ i = 1, ..., r\,. \tag{11.9}$$

Количество уравнений в этой системе равно r, неизвестными в ней служат λ_j, их количество равно s. В соответствии со сделанным в начале доказательства предположением, получаем, что количество неизвестных больше количества уравнений, и следовательно, по предложению 6.1, однородная с.л.у. (11.9) имеет ненулевое решение.

С другой стороны, эта с.л.у. равносильна равенству (11.3), которое, таким образом, окажется истинным для некоторого набора λ_j $(j = 1, ..., s)$, не все элементы которого нулевые, что и доказывает линейную зависимость с.в. (11.1). □

Теперь сформулируем и докажем теорему о равномощности всех базисов в одном и том же подпространстве.

Теорема 11.1. *Любые два базиса в линейном подпространстве $V \leqslant \mathbb{R}^n$ равномощны.*

Доказательство. После доказательства предложения 11.1 доказательство теоремы становится совершенно очевидным. Предположим, что (11.1) и (11.2) являются базисами подпространства V. Тогда, в частности, обе эти с.в. линейно независимы. Применяя

предложение 11.1 дважды [сначала рассматривая (11.2) как базис, а (11.1) — как линейно независимую с.в., а потом наоборот], мы получим два неравенства: $s \leqslant r$ и $r \leqslant s$, т. е. получим равенство $s = r$, утверждающее равномощность двух базисов. □

11.2. Понятие размерности для линейного подпространства в пространстве $\mathbb{R}^n$. Согласно теоремам 10.1 и 11.1, в любом подпространстве $V \leqslant \mathbb{R}^n$ существует базис и все базисы в одном и том же подпространстве имееют одинаковую мощность. Поэтому корректно следующее

Определение 11.1. Размерностью линейного подпространства $V \leqslant \mathbb{R}^n$ называется мощность некоторого базиса в этом подпространстве. Размерность подпространства V обозначается $\dim(V)$.

В следующем предложении собраны основные свойства размерности.

Предложение 11.2. *1. Размерность любого линейного подпространства $V \leqslant \mathbb{R}^n$ является целым неотрицательным числом, не превосходящим n, причем $\dim(V) = 0$ тогда и только тогда, когда $V = O$, и $\dim(V) = n$ тогда и только тогда, когда $V = \mathbb{R}^n$.*

2. Справедливо следующее свойство строгой монотонности размерности: если V и W — два подпространства в пространстве $\mathbb{R}^n$, одно из которых строго содержится в другом ($V \subset W$), то их размерности связаны строгим неравенством $\dim(V) < \dim(W)$.

3. Если два линейных подпространства V и W связаны включением $V \subseteq W$, а их размерности одинаковы [$\dim(V) = \dim(W)$], то и сами подпространства совпадают: $V = W$.

Доказательство немедленно следует из определения 11.1, примера 10.2 (в котором описывался естественный базис в пространстве $\mathbb{R}^n$) и предложения 10.2 (свойства продолжения базисов); подробности остаются для обдумывания читателям. □

Замечание 11.1. Следует предостеречь читателя: из одного только совпадения размерностей $\dim(V) = \dim(W)$ (без предположения о наличии включения между подпространствами) отнюдь не следует, что совпадают сами подпространства.

Замечание 11.2. Из изложенного в замечании 10.1 легко усматривается тот факт, что величина $\dim(V)$ доставляет максимум для мощностей линейно независивых с.в. в подпространстве V. Если вы

разобрались с упражнением, сформулированным в конце упомянутого замечания, то вам будет понятно, что эта же величина доставляет минимум для мощностей порождающих с.в. в подпространстве V.

11.3. Размерность подпространства решений однородной с.л.у. В этом пункте мы подведем итог серии примеров, посвященных подпространству L^0 решений однородной с.л.у. $A \cdot \bar{x} = \bar{0}$.

Предложение 11.3. *Рассмотрим однородную с.л.у. (8.5h) с матрицей A размера $m \times n$ и подпространство $L^0 \leqslant \mathbb{R}^n$ решений этой системы. Пусть r — количество ступенек в ступенчатом виде матрицы A. Тогда размерность подпространства L^0 может быть определена по формуле*

$$\dim(L^0) = n - r. \tag{11.10}$$

Доказательство следует из утверждения примера 10.3. □

Замечание 11.3. В линейной алгебре используется еще одна числовая характеристика линейного подпространства — его *коразмерность*. Коразмерностью подпространства $V \leqslant \mathbb{R}^n$ называется число

$$\operatorname{codim}(V) = n - \dim(V).$$

Утверждение предложения 11.3 с помощью понятия коразмерности выражается следующим образом: $\operatorname{codim}(L^0) = r$.

11.4. Ступенчатый ранг матрицы. В первой главе мы неоднократно обещали доказать *инвариантность* количества ступенек в ступенчатом виде матрицы (т. е. независимость этой величины от способа приведения матрицы к ступенчатому виду; см. замечание 5.2). Теперь мы можем выполнить это обещание. А именно: справедливо следующее

Предложение 11.4. *Пусть A — матрица размера $m \times n$. При любом способе приведения матрицы A к ступенчатому виду (с помощью элементарных преобразований над строками матрицы) количество ступенек r получается одним и тем же: оно равняется разности $n - \dim(L^0)$ числа столбцов n матрицы A и размерности подпространства L^0 решений однородной с.л.у. $A \cdot \bar{x} = \bar{0}$.*

Доказательство. Данное предложение является, очевидно, лишь переформулировкой предыдущего. Подчеркнем только, что величина $\dim(L^0)$ определена по матрице A абсолютно инвариантно (не

зависит ни от каких "случайных факторов" типа "способа приведения к ступенчатому виду"): матрице однозначно сопоставляется однородная с.л.у., для которой однозначно определено множество (подпространство) ее решений, которому однозначно сопоставляется число — его размерность. □

Доказанное предложение делает корректным следующее

Определение 11.2. Количество ступенек в ступенчатом виде матрицы A называется (*ступенчатым*) *рангом* этой матрицы и обозначается $\operatorname{rank}(A)$.

Замечание 11.4. Ранг матрицы является важнейшей числовой характеристикой этой матрицы.

Число $\operatorname{rank}(A)$ является целым неотрицательным и не может превышать наименьшего из размеров матрицы.

Существует еще три (равносильных) способа определения ранга, обладающих бóльшей естественностью и наглядностью. Два из них (*столбцовый* и *строчный* ранги) будут изучены уже в следующем параграфе. Третий способ (понятие *минорного* ранга) будет рассматриваться в § 30.

К концу четвертой главы вы получите доказательство совпадения всех четырех рангов.

Описанное выше понятие ступенчатого ранга обладает лишь одним (но весомым) преимуществом: легче всего ранг вычисляется именно с помощью приведения матрицы к ступенчатому виду.

§ 12. Столбцовый и строчный ранги матрицы

12.1. Ранг системы векторов и его свойства. Рассмотрим с.в. $\mathcal{A}$ [см. (8.1)] в арифметическом линейном пространстве $\mathbb{R}^n$ и ее линейную оболочку $\langle\mathcal{A}\rangle$ [см. (8.4)], являющуюся линейным подпространством в $\mathbb{R}^n$.

Определение 12.1. Рангом системы векторов называется размерность ее линейной оболочки. Обозначение:

$$\operatorname{rank}(\mathcal{A}) = \dim(\langle\mathcal{A}\rangle). \tag{12.1}$$

По теореме 10.1, в $\langle\mathcal{A}\rangle$ можно выбрать базис из числа порождающих векторов. Поэтому $0 \leqslant \operatorname{rank}(\mathcal{A}) \leqslant \min(n, k)$, где k — количество векторов в с.в. $\mathcal{A}$.

В пункте 5.1 определялись элементарные преобразования над строками и столбцами матрицы (I — III типов). Лишь слегка сменив терминологию, мы можем говорить об *элементарных преобразованиях над системами векторов*, понимая под этим элементарные преобразования над столбцами соответствующих матриц.

Предложение 12.1. *Элементарные преобразования над с.в. не изменяют линейную оболочку этой ситемы и, следовательно, не изменяют ее ранг.*

Доказательство. Преобразования типа I (перестановка двух векторов в системе) не отражаются на линейной оболочке в силу замечания 8.4.

Рассмотрим преобразование типа II, при котором все векторы системы (8.1) остаются неизменными, кроме $\bar{a}_i$, который заменяется на $\bar{a}'_i = \bar{a}_i + c\bar{a}_j$ $(i, j = 1, ..., k;\ i \neq j;\ c \in \mathbb{R})$. Рассмотрим новую с.в.

$$\mathcal{A}' = [\bar{a}_1, ..., \bar{a}'_i, ..., \bar{a}_k]$$

и ее линейную оболочку $\langle \mathcal{A}' \rangle$.

Поскольку все векторы с.в. $\mathcal{A}'$ принадлежат $\langle \mathcal{A} \rangle$, то и всякий вектор, принадлежащий $\langle \mathcal{A}' \rangle$, принадлежит $\langle \mathcal{A} \rangle$. (Почему? Похожее рассуждение уже было.) Следовательно, $\langle \mathcal{A}' \rangle \subseteq \langle \mathcal{A} \rangle$.

Обратимость элементарных преобразований приводит к противоположному включению. Значит, $\langle \mathcal{A}' \rangle = \langle \mathcal{A} \rangle$.

Для преобразований типа III (умножение одного из векторов на ненулевое число) доказательство совершенно аналогично. □

12.2. Столбцовый и строчный ранги матрицы. В предыдущем пункте мы начинали с рассмотрения системы векторов в пространстве $\mathbb{R}^n$, затем эти векторы записывались в матрицу. В данном пункте мы начнем с произвольной $(m \times n)$-матрицы A и сопоставим ей две системы векторов в двух различных арифметических линейных пространствах.

Первая система будет составлена из столбцов данной матрицы:

$$\mathcal{A}_{\text{стб}} = [\bar{a}_1, \bar{a}_2, ..., \bar{a}_n]\,; \tag{12.2}$$

элементы этой с.в. принадлежат арифметическому линейному пространству векторов-столбцов $\mathbb{R}^m$.

Вторая система будет составлена из строк матрицы A:

$$\mathcal{A}_{\text{стр}} = \left[(\bar{a}^1)^t, (\bar{a}^2)^t, ..., (\bar{a}^m)^t\right]; \tag{12.3}$$

ее элементы принадлежат арифметическому линейному пространству векторов-строк $\overset{*}{\mathbb{R}}{}^n$ [см. обозначение (2.3) в первой главе].

Определение 12.2. *Столбцовым* (*строчным*) *рангом* матрицы A называется ранг с.в. (12.2) [соответственно (12.3)]. Обозначения:

$$\text{rank}_{\text{стб}}(A) = \text{rank}(\mathcal{A}_{\text{стб}});\ \text{rank}_{\text{стр}}(A) = \text{rank}(\mathcal{A}_{\text{стр}}). \tag{12.4}$$

Ясно, что оба ранга являются целыми числами, заключенными в пределах от 0 до $\min(m, n)$, причем значение 0 достигается (в обоих случаях) лишь для нулевой матрицы. Кроме того, поскольку при транспонировании столбцы и строки матрицы меняются ролями, то, очевидно,

$$\text{rank}_{\text{стб}}(A^t) = \text{rank}_{\text{стр}}(A);\ \text{rank}_{\text{стр}}(A^t) = \text{rank}_{\text{стб}}(A). \tag{12.5}$$

Предложение 12.1 утверждает, что элементарные преобразования над столбцами матрицы A не меняют $\langle\mathcal{A}_{\text{стб}}\rangle$ и, следовательно, сохраняют $\text{rank}_{\text{стб}}(A)$, а поскольку векторы-столбцы и векторы-строки алгебраически совершенно равноправны (действиям над строками данной матрицы отвечают аналогичные действия над столбцами транспонированной матрицы), то справедлива версия предложения 12.1, относящаяся к строкам. Поэтому элементарные преобразования над строками матрицы не меняют $\langle\mathcal{A}_{\text{стр}}\rangle$ и, соответственно, сохраняют $\text{rank}_{\text{стр}}(A)$.

В следующем пункте мы убедимся в том, что столбцовый ранг сохраняется и при элементарных преобразованиях над строками, а строчный ранг — при преобразованиях над столбцами.

12.3. Инвариантность столбцового (строчного) ранга при элементарных преобразованиях над строками (столбцами).

Докажем сначала следующее вспомогательное

Предложение 12.2. *1. Линейные соотношения между столбцами матрицы сохраняются при элементарных преобразованиях над ее строками.*

2. При элементарных преобразованиях над строками матрицы линейно зависимые (линейно независимые) столбцы остаются линейно зависимыми (линейно независимыми).

Доказательство. 1. Нам надо доказать, что если какой-либо столбец матрицы A линейно выражается через несколько каких-то других ее столбцов, то это линейное соотношение (с теми же скалярными коэффициентами) сохранится и после применения элементарных преобразований над строками для соответствующих столбцов преобразованной матрицы A'.

Пусть, скажем, столбец $\bar{a}_j$ с номером $j \in \{k+1, ..., n\}$ выражается через первые k столбцов (хотя в принципе номера столбцов могут идти не подряд и не по порядку):

$$\bar{a}_j = \lambda_1 \bar{a}_1 + \cdots + \lambda_k \bar{a}_k. \tag{12.6}$$

Соотношение (12.6) означает, что вектор $\bar{\lambda} \in \mathbb{R}^k$, составленный из коэффициентов соотношения (12.6), является решением с.л.у.

$$\underset{m\times k}{A_k} \cdot \underset{k\times 1}{\bar{\mu}} = \underset{m\times 1}{\bar{a}_j}\ , \tag{12.7}$$

где $A_k = (\bar{a}_1|...|\,\bar{a}_k)$ — подматрица данной матрицы A [т. е. при подстановке в (12.7) вектора $\bar{\lambda}$ вместо неизвестного вектора $\bar{\mu}$ получится истинное равенство; вспомните векторную форму записи с.л.у. из § 8].

Расширенной матрицей для с.л.у. (12.7) будет служить подматрица $(A_k|\,\bar{a}_j)$ матрицы A. Элементарные преобразования над строками A будут порождать элементарные преобразования над строками этой подматрицы.

Им, в свою очередь, будут соответствовать элементарные преобразования над уравнениями системы (12.7). С помощью этих преобразований с.л.у. (12.7) будет сведена к равносильной с.л.у.

$$\underset{m\times k}{A'_k} \cdot \underset{k\times 1}{\bar{\mu}} = \underset{m\times 1}{\bar{a}'_j}\ . \tag{12.8}$$

А значит, вектор $\bar{\lambda}$ будет удовлетворять и этой системе, и следовательно, получится соотношение

$$\bar{a}'_j = \lambda_1 \bar{a}'_1 + \cdots + \lambda_k \bar{a}'_k$$

для векторов преобразованной матрицы. Первое утверждение предложения доказано.

2. Второе утверждение немедленно следует из первого. В самом деле, в силу обратимости элементарных преобразований, это утверждение достаточно доказать для свойства линейной зависимости (и тогда для свойства линейной независимости оно будет выполняться автоматически).

Но линейная зависимость системы векторов равносильна существованию линейного выражения одного из векторов системы через остальные (см. п. 2 предложения 9.2).

Линейные соотношения между столбцами, как доказано выше, сохраняются при элементарных преобразованиях над строками, а значит, сохранится и факт линейной зависимости столбцов. □

Замечание 12.1. Разумеется, справедлив также и "транспонированный факт" для линейных соотношений над строками и преобразований над столбцами.

Теперь мы можем доказать основное предложение данного пункта.

Предложение 12.3. *Элементарные преобразования над строками матрицы сохраняют ее столбцовый ранг, а преобразования над столбцами сохраняют строчный ранг.*

Доказательство (для столбцового ранга). Пусть столбцовый ранг $\operatorname{rank}_{\text{стб}}(A) = r$. Это означает, что в матрице A существует r линейно независимых столбцов, а остальные столбцы через них линейно выражаются. Пусть A' — матрица, полученная из A с помощью элементарных преобразований над строками. Из предыдущего предложения вытекает, что в матрице A' столбцы (с теми же номерами, что и отмеченные выше r столбцов в матрице A) также будут линейно независимыми, а остальные столбцы A' будут через них линейно выражаться. Следовательно, столбцовый ранг матрицы A' также будет равняться r. □

12.4. Первая теорема о ранге матрицы. В замечании 11.4 говорилось о четырех подходах к понятию ранга матрицы. На настоящий момент нам знакомы три из них: ступенчатый, столбцовый и строчный ранги. Ниже будет доказано совпадение этих числовых характеристик для матриц.

Теорема 12.1 (первая теорема о ранге матрицы). *Столбцовый и строчный ранги матрицы совпадают между собой и совпадают со ступенчатым рангом.*

Доказательство. Пользуясь теоремой 5.1, приведем данную матрицу A к скелетному виду $\hat{A}$. Для этого придется использовать элементарные преобразования над строками и над столбцами матрицы, которые, как доказано в предложении 12.3, сохраняют и столбцовый, и строчный ранги. Поэтому для доказательства равенства столбцового и строчного рангов для матрицы A достаточно доказать их совпадение для матрицы $\hat{A}$.

Матрица $\hat{A}$ имеет в начале главной диагонали r единиц (где r — количество ступенек в ступенчатом виде матрицы A); остальные элементы матрицы $\hat{A}$ равны нулю. Таким образом, система векторов-столбцов для этой матрицы состоит из единичных векторов $\bar{e}_1, ..., \bar{e}_r$ пространства $\mathbb{R}^m$ и нулевых векторов этого пространства в количестве $n-r$ штук. Поэтому столбцовый ранг матрицы $\hat{A}$ равен r. Аналогично система векторов-строк матрицы $\hat{A}$ содержит единичные векторы-строки $(\bar{e}^1)^t, ..., (\bar{e}^r)^t$ пространства $\overset{*}{\mathbb{R}}{}^n$ и нулевые векторы-строки этого пространства в количестве $m-r$ штук. Следовательно, строчный ранг матрицы $\hat{A}$ также равен r. □

12.5. Ранг матрицы и исследование с.л.у. (теорема Кронекера — Капелли). Совпадение трех рангов, введенных в определениях 11.1 и 12.2, позволяет опускать уточняющий эпитет (ступенчатый, столбцовый, строчный) при упоминании ранга и говорить просто о *ранге* матрицы $\operatorname{rank}(A)$. Новая терминология приводит к более лаконичным переформулировкам основных результатов, относящихся к системам линейных уравнений. В частности, простой заменой выражений типа "количество ступенек в ступенчатом виде матрицы" выражением "ранг матрицы" мы можем вывести из теоремы 6.1 следующую знаменитую теорему.

Теорема 12.2 (теорема Кронекера — Капелли). *Рассмотрим систему линейных уравнений (8.5) и ее расширенную матрицу B:*

$$\underset{m\times n}{A}\cdot\underset{n\times 1}{\bar{x}}=\underset{m\times 1}{\bar{b}}\,;\quad \underset{m\times(n+1)}{B}=(\underset{m\times n}{A}\,|\,\underset{m\times 1}{\bar{b}})\,.$$

С.л.у. является совместной тогда и только тогда, когда

$$\operatorname{rank}(A)=\operatorname{rank}(B), \tag{12.9}$$

и является определенной тогда и только тогда, когда оба этих ранга совпадают с количеством неизвестных n.

Доказательство. Внимательно просмотрите формулировку теоремы 6.1. Подробности о структуре решений с.л.у. опустите. □

Замечание 12.2. Практически во всех учебниках алгебры сначала изучается понятие ранга матрицы (для чего предварительно строится теория определителей), а затем оно применяется к системам линейных уравнений. При таком подходе теорема Кронекера — Капелли является содержательным и ключевым результатом.

При нашем подходе с.л.у. изучаются сразу с использованием одного единственного приема — метода Гаусса (и его модификаций). Само понятие ранга вводится на основе уже построенной теории с.л.у. Поэтому "знаменитая теорема" выглядит некоторой тавтологией. Метод Гаусса рекомендует: "Хотите узнать, имеет ли система решения, — решайте!" (Попутно вычислятся все ранги.) Если же есть желание вычислить ранги с самого начала, то (при реализуемом здесь подходе) опять-таки придется приводить матрицы к ступенчатому виду, что сравнимо по трудности с полным анализом системы.

Таким образом, мы оставляем пока за теоремой Кронекера — Капелли лишь "оформительскую" роль. Однако в дальнейшем, после развития теории определителей и изучения четвертого подхода к понятию ранга, эта теорема должна обрести независимый интерес и значение.

Замечание 12.3. Еще одним часто цитируемым результатом является предложение 6.1 — критерий существования нетривиального решения у однородной с.л.у. $A \cdot \bar{x} = \bar{0}$. С помощью понятия ранга матрицы этот критерий также можно сформулировать очень лаконично: однородная с.л.у. имеет нетривиальное решение тогда и только тогда, когда ранг матрицы этой системы меньше количества неизвестных.

§ 13. Алгоритмы построения базисов и вычисления размерностей и рангов

13.1. Два способа задания линейных подпространств в пространстве $\mathbb{R}^n$. Нами уже изучены два типовых приема задания

линейных подпространств в арифметическом линейном пространстве $\mathbb{R}^n$. Оба они связаны с матрицами. И первый из них был рассмотрен в § 3 (см. предложение 3.2). Произвольной $(m \times n)$-матрице A можно сопоставить линейное подпространство решений однородной с.л.у. $A \cdot \bar{x} = \bar{0}$. В § 3 (и далее) мы обозначали это подпространство символом L^0, теперь мы несколько усложним и уточним обозначение: будем писать

$$L^0_A = \{\bar{x} \in \mathbb{R}^n \,:\, A \cdot \bar{x} = \bar{0}\}, \tag{13.1}$$

явно указывая матрицу, по которой строится подпространство решений.

Второй способ мы начали изучать в § 8 (см. предложение 8.1): каждой системе векторов $\mathcal{A}$ в пространстве $\mathbb{R}^n$ (в частности, системе $\mathcal{A} = \mathcal{A}_{\text{стб}}$, составленной из векторов-столбцов матрицы A; см. п. 12.2) сопоставляется линейная оболочка $\langle\mathcal{A}\rangle$, являющаяся линейным подпространством в $\mathbb{R}^m$. Введем обозначение для этого подпространства:

$$R_A = \langle\mathcal{A}_{\text{стб}}\rangle. \tag{13.2}$$

Определение 13.1. Пусть A — матрица размера $m \times n$. Будем называть линейное подпространство $L^0_A \leqslant \mathbb{R}^n$ *нуль-пространством* или *ядром* матрицы A, а линейное подпространство $R_A \leqslant \mathbb{R}^m$ *образом* матрицы A.

Замечание 13.1. Введенные выше понятия ядра и образа по своей природе относятся к теории *линейных отображений*, которую мы будем изучать в § 15. Однако задачи, с ними связанные, можно решать уже сейчас, что мы и собираемся делать. Терминология же введена авансом для краткости изложения. Когда мы доберемся до линейных отображений (операторов), то мы поймем, что такие отображения задаются своими матрицами (относительно некоторых базисов) и что на практике все вычисления проводятся именно с матрицами. В компьютерных математических системах (например, Maple) также имеются команды вычисления ядра (nullspace) и образа (colspace) для матриц.

13.2. Базис и размерность для нуль-пространства матрицы. Пусть линейное подпространство $V \leqslant \mathbb{R}^n$ задано как ядро $(m \times n)$-матрицы A, т. е. $V = L^0_A$. Тогда, решая соответствующую однородную с.л.у., мы сначала приводим матрицу A к виду Жордана — Гаусса A'; в частности, мы удаляем нулевые строки (если они

образовались) и в новой (равносильной исходной) с.л.у. $A' \cdot \bar{x} = \bar{0}$ останется r независимых уравнений, где r — количество ступенек в ступенчатом виде для A, т. е. $r = \mathrm{rank}(A)$.

Так получается другое, "более экономное" представление данного подпространства: $V = L^0_{A'}$.

Продолжая решение, представим это подпространство в виде линейной оболочки базисных частных решений однородной с.л.у. (см. пример 8.3 и предложение 11.3). Это будет уже вторым способом задания подпространства V:

$$V = R_F;\ F = (\bar{f}_{r+1}|...|\,\bar{f}_n), \tag{13.3}$$

причем порождающие векторы (т. е. столбцы фундаментальной матрицы F, или базисные ч.р.о.) образуют базис V. Размерность V, в соответствии с упомянутым предложением 11.3, равна $n - r$, т. е.

$$\dim(L^0_A) = n - \mathrm{rank}(A). \tag{13.4}$$

Пример 13.1. Рассмотрим линейное подпространство V в $\mathbb{R}^n$, заданное однородной с.л.у.

$$x_1 = x_2 = \cdots = x_n.$$

В развернутой записи эта с.л.у. имеет вид

$$\begin{cases} x_1 - x_2 = 0; \\ x_2 - x_3 = 0; \\ \qquad \cdots\cdots \\ x_{n-1} - x_n = 0. \end{cases}$$

Матрица этой системы имеет размеры $(n-1) \times n$ и задается формулой

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

Приведем ее к виду Ж.—Г. Для этого нужно к каждой строке, начиная с предпоследней и продвигаясь вверх, прибавить предыдущую строку. В итоге мы получим матрицу

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

Видим, что последнюю неизвестную x_n можно принять за свободную, остальные неизвестные будут главными (они выразятся через свободную как раз по тем формулам, которые были даны нам изначально). Общее решение системы определится формулой

$$\bar{x} = x_n \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \\ 1 \end{pmatrix}.$$

Здесь мы имеем одно базисное решение

$$\bar{f}_n = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \\ 1 \end{pmatrix}.$$

Следовательно, подпространство решений $V = L_A^0$ является одномерным. Вторым способом оно может быть представлено в виде

$$V = R_F;\ F = (\bar{f}_n),$$

т. е. как линейная оболочка одного вектора (образ одностолбцовой матрицы).

13.3. Базис и размерность для линейной оболочки столбцов матрицы. Рассмотрим теперь линейное подпространство, заданное вторым способом: $V = R_A \leqslant \mathbb{R}^m$, где A — матрица размера $m \times n$. Найдем базис подпространства V из числа порождающих это подпространство векторов (столбцов матрицы A). Для этого приведем матрицу к виду Ж.—Г. A', используя элементарные преобразования типов I — III над строками [см. формулу (5.2)].

В соответствии с предложениями 12.1 и 12.2, при преобразованиях над строками базисные векторы останутся базисными (хотя уже в другом подпространстве!); в матрице A' базисные столбцы будут легко заметны: в качестве базисных можно будет взять ключевые столбцы, являющиеся единичными векторами $\bar{e}_1, \bar{e}_2, ..., \bar{e}_r$, где $r = \operatorname{rank}(A)$.

Базисными векторами в V будут столбцы матрицы A с теми же номерами, что и единичные столбцы в матрице A'. Обозначим B подматрицу размера $m \times r$ в матрице A, составленную из указанных столбцов. Взяв линейную оболочку этих векторов-столбцов, мы получим "более экономное" представление подпространства V (тем же самым, вторым способом), а также значение размерности этого подпространства:

$$V = R_B;\ \dim(V) = \operatorname{rank}(A) = r. \tag{13.5}$$

Замечание 13.2. Номера базисных векторов проясняются уже на этапе ступенчатого вида: если нас интересует лишь базис в подпространстве V, то мы можем остановиться на этом этапе и сформировать матрицу B. Внимание: в эту матрицу следует переписывать (в их исходном виде!) те столбцы из матрицы A , которые в ступенчатом виде "проходят через ключевые элементы" (т. е. являются ключевыми столбцами).

Размерность подпространства V, или, что то же, ранг матрицы A, будет равняться количеству ступенек.

Однако может представлять интерес задача с дополнительным заданием: помимо базиса в линейной оболочке системы векторов (из числа порождающих векторов), определить разложения по найденному базису тех порождающих векторов, которые в этот базис не вошли.

Здесь надо снова применить предложение 12.2, согласно которому линейные соотношения между столбцами матрицы сохраняются при элементарных преобразованиях над строками этой матрицы. Удобно несколько продвинуться в преобразованиях и сгруппировать базисные столбцы в начале матрицы (с применением "меток", аналогичных тем, что мы использовали при решении с.л.у.; см. п. 4.4). В результате мы придем к модифицированному виду Ж.—Г.

В теоретическом рассуждении мы можем считать, что базисными являются первые r столбцов; тогда в матрице A', в левом верхнем углу, будет располагаться единичный блок E_r. (Кстати, будет только легче, если мы выбросим нулевые строки, т. е. применим преобразования типа IV над строками, переходя в пространство меньшей размерности, но с очевидным сохранением линейных соотношений между векторами.)

Для любого столбца $\bar{a}'_j$ матрицы A', не вошедшего в базис, элементы этого столбца служат коэффициентами его разложения по

естественному базису $[\bar{e}_1, ..., \bar{e}_r]$ пространства $\mathbb{R}^r$. Пусть, скажем,

$$\bar{a}'_j = \mu_1 \bar{e}_1 + \cdots + \mu_r \bar{e}_r.$$

Тогда, согласно предложению 12.2, "старый" вектор $\bar{a}_j$ будет разлагаться по "старым" базисным векторам с теми же коэффициентами:

$$\bar{a}_j = \mu_1 \bar{a}_1 + \cdots + \mu_r \bar{a}_r.$$

Пример 13.2. Приведем числовой пример — стандартное упражнение на понятие базиса в линейном подпространстве.

З а д а ч а. Подпространство V пространства $\mathbb{R}^3$ является линейной оболочкой системы векторов $[\bar{a}_1, \bar{a}_2, \bar{a}_3, \bar{a}_4]$, где

$$\bar{a}_1 = \begin{pmatrix} 9 \\ 5 \\ 4 \end{pmatrix};\ \bar{a}_2 = \begin{pmatrix} -1 \\ 1 \\ -2 \end{pmatrix};\ \bar{a}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix};\ \bar{a}_4 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Найти базис подпространства V (из числа порождающих векторов) и вычислить $\dim(V)$. Порождающие векторы, не вошедшие в найденный базис, разложить по этому базису.

Р е ш е н и е. Рассмотрим матрицу, составленную из векторов порождающей системы, и приведем ее к модифицированному виду Жордана — Гаусса. В качестве меток над столбцами проставим их номера в исходной матрице. (В первой главе, в примере 4.1, в качестве меток фигурировали имена неизвестных. Сейчас такие метки были бы "не по задаче".)

$$A = \begin{pmatrix} \boxed{1} & \boxed{2} & \boxed{3} & \boxed{4} \\ 9 & -1 & 1 & 1 \\ 5 & 1 & 1 & 0 \\ 4 & -2 & 0 & 1 \end{pmatrix} \longrightarrow$$

$$\longrightarrow \begin{pmatrix} \boxed{4} & \boxed{3} & \boxed{2} & \boxed{1} \\ 1 & 1 & -1 & 9 \\ 0 & 1 & 1 & 5 \\ 1 & 0 & -2 & 4 \end{pmatrix} \xrightarrow{3^{\text{стр}}+1^{\text{стр}}\cdot(-1)}$$

$$\longrightarrow \begin{pmatrix} \boxed{4} & \boxed{3} & \boxed{2} & \boxed{1} \\ 1 & 1 & -1 & 9 \\ 0 & 1 & 1 & 5 \\ 0 & -1 & -1 & -5 \end{pmatrix} \xrightarrow{3^{\text{стр}}+2^{\text{стр}}}$$

$$\longrightarrow \begin{pmatrix} \boxed{4} & \boxed{3} & \boxed{2} & \boxed{1} \\ 1 & 1 & -1 & 9 \\ 0 & 1 & 1 & 5 \\ 0 & 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} \boxed{4} & \boxed{3} & \boxed{2} & \boxed{1} \\ 1 & 1 & -1 & 9 \\ 0 & 1 & 1 & 5 \end{pmatrix} \longrightarrow ...$$

Достигнут ступенчатый вид, и мы можем (взглянув на метки над столбцами) констатировать, что с.в. $[\bar{a}_4, \bar{a}_3]$ является базисом V и что $\dim(V) = 2$. Сделаем еще один шаг, чтобы достигнуть вида Ж.—Г.:

$$...\xrightarrow{1^{\text{стр}}+2^{\text{стр}}\cdot(-1)} \begin{pmatrix} \boxed{4} & \boxed{3} & \boxed{2} & \boxed{1} \\ 1 & 0 & -2 & 4 \\ 0 & 1 & 1 & 5 \end{pmatrix} = A'.$$

Для матрицы A' базисными столбцами являются $\bar{a}'_4 = \bar{e}_1$ и $\bar{a}'_3 = \bar{e}_2$, где $[\bar{e}_1, \bar{e}_2]$ — естественный базис в $\mathbb{R}^2$. Векторы $\bar{a}'_2$ и $\bar{a}'_1$ разлагаются по этому базису следующим образом:

$$\bar{a}'_2 = (-2) \cdot \bar{e}_1 + 1 \cdot \bar{e}_2 = -2\bar{a}'_4 + \bar{a}'_3;$$

$$\bar{a}'_1 = 4 \cdot \bar{e}_1 + 5 \cdot \bar{e}_2 = 4\bar{a}'_4 + 5\bar{a}'_3.$$

Значит, векторы $\bar{a}_2$ и $\bar{a}_1$ исходной с.в. выражаются по найденному базису $[\bar{a}_4, \bar{a}_3]$ с теми же коэффициентами:

$$\bar{a}_2 = -2\bar{a}_4 + \bar{a}_3;\ \bar{a}'_1 4\bar{a}_4 + 5\bar{a}_3.$$

13.4. Переход от второго способа задания линейных подпространств к первому. В этом пункте мы опишем прием, позволяющий переходить от второго способа задания линейного подпространства V в арифметическом линейном пространстве $\mathbb{R}^n$, т. е. от представления подпространства в виде линейной оболочки некоторой конечной с.в., к первому способу задания, т. е. к представлению V в виде подпространства решений некоторой с.л.у. (см. п. 13.1).

Учтите только, что при одновременном рассмотрении обоих способов (в отношении одного и того же подпространства) в двух описаниях будут фигурировать разные матрицы.

Пусть подпространство $V \leqslant \mathbb{R}^n$ задано вторым способом [см. формулу (13.2)]:

$$V = \langle \mathcal{G} \rangle = R_G, \tag{13.6}$$

где

$$\mathcal{G} = [\bar{g}_1, \bar{g}_2, ..., \bar{g}_k] \tag{13.7}$$

— некоторая с.в. в пространстве $\mathbb{R}^n$, а

$$G = (\bar{g}_1|\ \bar{g}_2|\ ...|\ \bar{g}_k) \tag{13.8}$$

— соответствующая $(n \times k)$-матрица.

Задача состоит в том, чтобы представить это подпространство первым способом [т. е. формулой (13.1)]. Таким образом, подлежит определению $(m \times n)$-матрица A такая, что $V = L_A^0$. Из формул (13.4) и (13.5) следует, что должны выполняться равенства:

$$n - \mathrm{rank}(A) = \dim(V) = \mathrm{rank}(G). \tag{13.9}$$

Поэтому искомая матрица A должна иметь ранг r, равный $n - s$, где $s = \mathrm{rank}(G)$. Следовательно, если мы не хотим, чтобы в матрице A присутствовали "лишние" строки (которые все равно придется выбрасывать при решении с.л.у.), то мы можем считать, что в этой матрице $r = n - s$ линейно независимых строк.

Общим видом для неизвестной вектор-строки матрицы A будет:

$$\bar{\alpha}^t = (\alpha_1\ \alpha_2\ ...\ \alpha_n)\,. \tag{13.10}$$

Далее, каждый вектор $\bar{x} \in V$ должен удовлетворять однородной с.л.у.

$$A \cdot \bar{x} = \bar{0}$$

(с неизвестной пока матрицей A). А поскольку векторы из с.в. $\mathcal{G}$ порождают V, то, для того чтобы это соотношение было справедливым для всех $\bar{x} \in V$, достаточно, чтобы оно выполнялось для всех порождающих векторов:

$$A \cdot \bar{g}_j = \bar{0};\ j = 1, ..., k. \tag{13.11}$$

Вычисления станут только легче, если мы предварительно избавимся от "лишних" порождающих векторов, выбрав (из числа порождающих) какой-либо базис в подпространстве V, как это делалось в предыдущем пункте.

Давайте поэтому считать, что указанное исключение уже произведено и ранг матрицы G совпадает с количеством ее столбцов, т. е. $s = k$.

Для того чтобы матрица A удовлетворяла соотношениям (13.11), необходимо и достаточно (см. определение матричного умножения), чтобы каждая ее строка (13.10) удовлетворяла соотношениям

$$\bar{\alpha}^t \cdot \bar{g}_j = 0;\ j = 1, ..., k. \tag{13.12}$$

Соотношения (13.12), в свою очередь, равносильны матричному равенству

$$\bar{\alpha}^t \cdot G = \bar{0}^t, \tag{13.13}$$

в котором $\bar{\alpha}^t \in \overset{*}{\mathbb{R}}{}^n$ — неизвестная вектор-строка, а $\bar{0}^t \in \overset{*}{\mathbb{R}}{}^r$ — нулевая строка (в ней столько нулей, сколько должно быть строк в неизвестной матрице A).

Равенство (13.13) можно рассматривать как однородную с.л.у., в которой, однако, неизвестные записаны не в столбец, а в строчку. (Здесь это проистекает из существа задачи, поскольку в ней разыскивается не вектор, а, как говорят, *ковектор*, или *линейная форма*; с соответствующей терминологией вы познакомитесь в следующем семестре, при изучении курса линейной алгебры и геометрии.)

Чтобы придать соотношению (13.13) привычную форму, надо его транспонировать [с использованием свойств операции транспонирования (xiv) — (xvii), см. п. 2.3]:

$$G^t \cdot \bar{\alpha} = \bar{0}. \tag{13.14}$$

Решая эту однородную с.л.у. (в пространстве $\mathbb{R}^n$), мы найдем $n - \operatorname{rank}(G^t) = n - s$ базисных ч.р.о., из которых можно составить фундаментальную матрицу F размера $n \times (n-s)$.

Искомая матрица A получится транспонированием F; она будет удовлетворять нужным соотношениям, все ее строки будут линейно независимы, и их количество будет равно $n-s = r$, что и требовалось.

Переход завершен: мы по данной матрице G построили такую матрицу A, что $R_G = L^0_A$.

Пример 13.3. Решим следующую з а д а ч у: линейное подпространство $V \leqslant \mathbb{R}^5$ задано как линейная оболочка системы векторов-столбцов матрицы

$$G = \begin{pmatrix} 1 & 1 & 3 & 0 \\ -1 & 1 & 1 & 2 \\ 1 & 0 & 1 & -1 \\ -1 & 0 & -1 & 1 \\ 1 & 3 & 7 & 2 \end{pmatrix},$$

т. е. $V = R_G$ — образ матрицы G. Задать V как подпространство решений некоторой однородной с.л.у., т. е. представить его в виде $V = L^0_A$ (как ядро некоторой матрицы A).

Р е ш е н и е. Рассмотрим однородную с.л.у. (13.14), матрицей которой служит матрица G^t, транспонированная к данной матрице G, и приведем транспонированную матрицу к модифицированному виду Ж.—Г.:

$$G^t = \begin{pmatrix} \boxed{\alpha_1} & \boxed{\alpha_2} & \boxed{\alpha_3} & \boxed{\alpha_4} & \boxed{\alpha_5} \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 0 & 3 \\ 3 & 1 & 1 & -1 & 7 \\ 0 & 2 & -1 & 1 & 2 \end{pmatrix} \xrightarrow[3^{\text{стр}}+1^{\text{стр}}\cdot(-3)]{2^{\text{стр}}+1^{\text{стр}}\cdot(-1)}$$

$$\longrightarrow \begin{pmatrix} \boxed{\alpha_1} & \boxed{\alpha_2} & \boxed{\alpha_3} & \boxed{\alpha_4} & \boxed{\alpha_5} \\ 1 & -1 & 1 & -1 & 1 \\ 0 & 2 & -1 & 1 & 2 \\ 0 & 4 & -2 & 2 & 4 \\ 0 & 2 & -1 & 1 & 2 \end{pmatrix} \longrightarrow$$

$$\longrightarrow \begin{pmatrix} \boxed{\alpha_1} & \boxed{\alpha_2} & \boxed{\alpha_3} & \boxed{\alpha_4} & \boxed{\alpha_5} \\ 1 & -1 & 1 & -1 & 1 \\ 0 & 2 & -1 & 1 & 2 \end{pmatrix} \longrightarrow ...$$

Видим, что ранг матрицы G^t (или, что то же, ранг матрицы G) равен 2 и что можно было бы, предварительно "обработав" исходную матрицу, оставить только два из четырех порождающих векторов. Но можно начинать вычисления и с "неподготовленной" матрицы, что мы и сделали выше. Продолжим преобразования:

$$... \longrightarrow \begin{pmatrix} \boxed{\alpha_1} & \boxed{\alpha_4} & \boxed{\alpha_2} & \boxed{\alpha_3} & \boxed{\alpha_5} \\ 1 & -1 & -1 & 1 & 1 \\ 0 & 1 & 2 & -1 & 2 \end{pmatrix} \xrightarrow{1^{\text{стр}}+2^{\text{стр}}}$$

$$\longrightarrow \begin{pmatrix} \boxed{\alpha_1} & \boxed{\alpha_4} & \boxed{\alpha_2} & \boxed{\alpha_3} & \boxed{\alpha_5} \\ 1 & 0 & 1 & 0 & 3 \\ 0 & 1 & 2 & -1 & 2 \end{pmatrix}.$$

Модифицированный вид Ж.—Г. достигнут. Выписывая с.л.у., отвечающую преобразованной матрице, и решая эту систему, получим:

$$\bar{\alpha} = \alpha_2 \begin{pmatrix} -1 \\ 1 \\ 0 \\ -2 \\ 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} + \alpha_5 \begin{pmatrix} -3 \\ 0 \\ 0 \\ -2 \\ 1 \end{pmatrix}.$$

Таким образом, фундаментальной матрицей рассматриваемой однородной с.л.у. будет

$$F = \begin{pmatrix} -1 & 0 & -3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Транспонируя эту матрицу, получаем искомую матрицу

$$A = F^t = \begin{pmatrix} -1 & 1 & 0 & -2 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ -3 & 0 & 0 & -2 & 1 \end{pmatrix}.$$

О т в е т. Данное подпространство может быть задано системой линейных уравнений

$$\begin{cases} -x_1 + x_2 - 2x_4 = 0; \\ x_3 + x_4 = 0; \\ -3x_1 - 2x_4 + x_5 = 0. \end{cases}$$

13.5. Вычисление ранга матрицы, зависящей от параметра. О вычислении ранга матрицы мы говорили в предыдущем пункте: остановившись по достижении ступенчатого вида матрицы и подсчитав количество ступенек, мы получим значение ранга матрицы. Задачи с параметром всегда привносят специфические трудности в ход вычислений. В основном они связаны с невозможностью "бесконтрольного" применения элементарных преобразований типов II и III, использующих деление строк на скаляры, представляющие из себя выражения, содержащие параметр. Здесь требуются дополнительные исследования на предмет возможного обращения этих выражений в нуль, в связи с чем возникают так называемые "особые" случаи (см. п. 7.1).

Для анализа всех возможных случаев применим следующую стратегию: пользуясь лишь элементарными преобразованиями типа I и типа II, но без деления прибавляемых строк на выражения, содержащие параметр, приведем данную матрицу $A = A(\lambda)$ к ступенчатому виду. При этом ключевые элементы могут оказаться функциями от параметра λ, которые при некоторых значениях λ могут обращаться в нуль. Отсюда и будут рождаться особые случаи.

Пример 13.4. Дана матрица

$$A = \begin{pmatrix} 1 & \lambda & -1 & \lambda & 1 \\ \lambda & \lambda^2+\lambda-1 & -\lambda+1 & \lambda^2+2 & \lambda+1 \\ 2 & 2\lambda & \lambda^2-\lambda-2 & 2\lambda & 2 \\ 0 & \lambda-1 & 1 & \lambda^2+2\lambda+2 & \lambda+1 \end{pmatrix}.$$

Поставим з а д а ч у: вычислить (при всех значениях параметра λ) ранг матрицы A.

Р е ш е н и е:

$$A \xrightarrow[3^{\text{стр}}+1^{\text{стр}}\cdot(-2)]{2^{\text{стр}}+1^{\text{стр}}\cdot(-\lambda)} \begin{pmatrix} 1 & \lambda & -1 & \lambda & 1 \\ 0 & \lambda-1 & 1 & 2 & 1 \\ 0 & 0 & \lambda^2-\lambda & 0 & 0 \\ 0 & \lambda-1 & 1 & \lambda^2+2\lambda+2 & \lambda+1 \end{pmatrix} \longrightarrow$$

$$\xrightarrow{4^{\text{стр}}+2^{\text{стр}}\cdot(-1)} \begin{pmatrix} 1 & \lambda & -1 & \lambda & 1 \\ 0 & \lambda-1 & 1 & 2 & 1 \\ 0 & 0 & \lambda^2-\lambda & 0 & 0 \\ 0 & 0 & 0 & \lambda^2+2\lambda & \lambda \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & \lambda & -1 & \lambda & 1 \\ 0 & \lambda-1 & 1 & 2 & 1 \\ 0 & 0 & \lambda(\lambda-1) & 0 & 0 \\ 0 & 0 & 0 & \lambda(\lambda+2) & \lambda \end{pmatrix} = A'.$$

Достигнут ступенчатый вид, но три из четырех ступенек занимают многочлены по переменной λ, которые при отдельных значениях λ могут обращаться в нуль.

Видно, что особыми значениями являются $\lambda \in \{0, 1, -2\}$.

В неособом случае ($\lambda \notin \{0, 1, -2\}$) ранг матрицы равен 4.

В первом особом случае ($\lambda = 0$) матрица A' принимает вид:

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

Эта матрица имеет ступенчатый вид с двумя ступеньками.

Во втором особом случае ($\lambda = 1$) получим:

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

После перестановки строк (или после удаления третьей строки) мы придем к ступенчатому виду с тремя ступеньками.

В третьем особом случае ($\lambda = -2$) матрица

$$A' = \begin{pmatrix} 1 & -2 & -1 & -2 & 1 \\ 0 & -3 & 1 & 2 & 1 \\ 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 \end{pmatrix}$$

имеет ступенчатый вид, причем с таким же, как и в неособом случае, количеством ступенек. Производя "слияние" случаев, приводящих к одному и тому же результату, получим следующий

О т в е т: 1) при $\lambda \notin \{0, 1\}$: $\operatorname{rank}(A) = 4$; 2) при $\lambda = 0$: $\operatorname{rank}(A) = 2$; 3) при $\lambda = 1$: $\operatorname{rank}(A) = 3$.

13.6. Решение задач с арифметическими векторами средствами системы Maple. Этот "Клён" настолько буйно разросся, что для большинства решаемых задач системой предусмотрен не один десяток способов решения, а также способов представления исходных данных и результатов.

Здесь мы рассмотрим одну типовую задачу, пример решения которой "вручную" уже приводился в п. 13.3. А именно: будут продемонстрированы возможности системы Maple в задаче об отыскании базиса в линейной оболочке системы векторов в арифметическом линейном пространстве.

Еще раз подчеркнем, что основы синтаксиса Maple вам рекомендуется разобрать самостоятельно. Изучение тонкостей языка — не предмет данного пособия. Автору хотелось бы показать мощь системы, ее интеллект и, может быть, ее отдельные недостатки (чтобы у читателей возникло желание внести свои усовершенствования).

Напомним, что в первой главе приведены протоколы нескольких Maple-сессий. Здесь мы начинаем новую сессию.

После обычных

```
> restart; with ( linalg ) ;
```

введем систему векторов:

```
> v1 := vector ([1, −1, 1, −1, 0]); v2 := vector ([2, −5, −1, −8, −9]);
  v3 := vector ([3, −4, 2, −5, −3]); v4 := vector ([4, 0, 8, 4, 12]);
  v5 := vector ([5, 3, 13, 11, 24]); v6 := vector ([6, 5, 7, 4, 12]);
```

sv := [v1, v2, v3, v4, v5, v6];

$$v1 := [1, -1, 1, -1, 0]$$
$$v2 := [2, -5, -1, -8, -9]$$
$$v3 := [3, -4, 2, -5, -3]$$
$$v4 := [4, 0, 8, 4, 12]$$
$$v5 := [5, 3, 13, 11, 24]$$
$$v6 := [6, 5, 7, 4, 12]$$
$$sv := [v1, v2, v3, v4, v5, v6]$$

Напомним, что Maple воспринимает и сам записывает векторы в строчку, хотя в последующих алгебраических действиях они участвуют как векторы-столбцы. По-видимому, это — не слишком удачный "дизайн", но отменить его можно только "насильно", например оформляя векторы как одностолбцовые матрицы.

Важно также и то, что при определении системы векторов sv мы заключили векторы в квадратные скобки, т. е. по правилам Maple ввели упорядоченную систему ('list'). Предусмотрен вариант оформления с вводом и выводом неупорядоченной системы (множества) векторов; тогда следовало бы заключить векторы, входящие в sv, в фигурные скобки.

Введем теперь команду отыскания базиса в линейной оболочке данной системы векторов.

> bv := basis (sv) ;

$$bv := [v1, v2, v6]$$

Полученный ответ означает, что размерность линейной оболочки данной системы векторов равна 3 и что в качестве базиса в ней может быть выбрана система bv, составленная из указанных трех порождающих векторов.

Составим из этих векторов (произведя конкатенацию) матрицу $Abas$, а затем найдем разложение одного из порождающих векторов $v3$, который не вошел в базис bv, через этот базис (для этого придется решить с.л.у. с матрицей $Abas$ и столбцом правых частей $v3$).

> Abas := concat (v1, v2, v6) ;

$$Abas := \begin{bmatrix} 1 & 2 & 6 \\ -1 & -5 & 5 \\ 1 & -1 & 7 \\ -1 & -8 & 4 \\ 0 & -9 & 12 \end{bmatrix}$$

> linsolve (Abas, v3) ;

$$\left[\frac{7}{3}, \frac{1}{3}, 0\right]$$

Полученный ответ означает, что

$$v3 = \frac{7}{3}v1 + \frac{1}{3}v2.$$

Разберем теперь второй прием решения той же задачи. Сначала запишем все данные векторы в матрицу A, затем (с помощью команды gaussjord) приведем эту матрицу к виду Жордана — Гаусса.

> A := concat (v1, v2, v3, v4, v5, v6) ;

$$A := \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ -1 & -5 & -4 & 0 & 3 & 5 \\ 1 & -1 & 2 & 8 & 13 & 7 \\ -1 & -8 & -5 & 4 & 11 & 4 \\ 0 & -9 & -3 & 12 & 24 & 12 \end{bmatrix}$$

> Ajg := gaussjord (A) ;

$$Ajg := \begin{bmatrix} 1 & 0 & \frac{7}{3} & \frac{20}{3} & \frac{31}{3} & 0 \\ 0 & 1 & \frac{1}{3} & \frac{-4}{3} & \frac{-8}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

В новой матрице Ajg видны ключевые столбцы (первый, второй и шестой) и, таким образом, определены базисные векторы (из числа порождающих) — те же самые $v1$, $v2$ и $v6$. Числа в остальных столбцах являются коэффициентами разложения порождающих векторов,

не вошедших в базис, по этому базису. Например, разложение для $v4$ имеет вид:

$$v4 = \frac{20}{3}v1 - \frac{4}{3}v2.$$

Разберем еще один способ — специальную команду для отыскания базиса в линейной оболочке столбцов матрицы.

> cs := colspace (A) ;

$$cs := \{\left[0, 1, 0, \frac{4}{5}, \frac{9}{10}\right], \left[1, 0, 0, \frac{-7}{5}, \frac{-6}{5}\right], \left[0, 0, 1, \frac{6}{5}, \frac{21}{10}\right]\}.$$

Здесь искомый базис выдан как множество ('set') и, как видите, это уже другой базис, не из числа порождающих векторов.

В заключение добавим, что в задачах с параметром вместо команды gaussjord следует использовать команду ffgausselim, осуществляющую *свободное от дробей гауссово исключение*.

§ 14. Обратимые квадратные матрицы

14.1. Кольцо квадратных матриц заданного размера. В § 2 мы определили алгебраические операции над матрицами, элементы которых принадлежат полю $\mathbb{R}$. Напомним, что как сложение, так и умножение двух матриц осуществимы не всегда: для осуществимости сложения необходимо, чтобы матрицы имели одинаковые размеры, а для того чтобы было возможным умножение, необходимо, чтобы количество столбцов первой матрицы равнялось количеству строк второй.

В этом параграфе мы рассмотрим алгебраические действия над квадратными матрицами фиксированного размера $n \times n$, т. е. алгебраические действия в множестве $\mathrm{Mat}(n, n; \mathbb{R})$, для которого мы введем здесь более короткое обозначение $L(n, \mathbb{R})$. Ясно, что в этом множестве и сложение, и умножение всегда осуществимы. Все законы матричной алгебры (i) — (xvii), доказанные в теореме 2.1, оказываются теперь (без всяких оговорок о размерах матриц) справедливы в множестве $L(n, \mathbb{R})$.

Остановимся сначала на двух операциях: на сложении и умножении матриц. В соответствии с законами (i) — (iv), сложение является коммутативным и ассоциативным, имеет нуль (нулевую матрицу O) и для любой матрицы существует противоположная матрица. Умножение согласовано со сложением двумя дистрибутивными

законами — (ix) и (x), ассоциативно [закон (xii)] и имеет единицу [единичную матрицу E, закон (xiii)]. Умножение (при $n \geqslant 2$) не является коммутативным (см. замечание 2.5).

Множество с заданными на нем двумя алгебраическими операциями — сложением и умножением, которые удовлетворяют всем законам, упомянутым в предыдущем абзаце, называется *кольцом*. С использованием этого термина предыдущий абзац можно выразить так: квадратные матрицы заданного размера образуют кольцо.

Кольцо — это понятие более общее, чем поле. Аксиомы поля $\boxed{1}$ — $\boxed{9}$ (см. п. 2.1) требуют, дополнительно к описанным выше аксиомам кольца, чтобы умножение было коммутативным и чтобы для любого ненулевого элемента существовал обратный элемент.

Хотя кольцо квадратных матриц $L(n, \mathbb{R})$ не является коммутативным, в нем существуют элементы, которые *коммутируют* (т. е. перестановочны при умножении) с любым элементом этого кольца. Такими элементами служат матрицы вида

$$\lambda E = \begin{pmatrix} \lambda & 0 & \ldots & 0 \\ 0 & \lambda & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda \end{pmatrix},$$

где $\lambda \in \mathbb{R}$.

В самом деле,

$$(\lambda E) \cdot A \overset{\text{(xi)}}{=} \lambda(E \cdot A) \overset{\text{(xiii)}}{=\!=} \lambda A \overset{\text{(xiii)}}{=\!=} \lambda(A \cdot E) \overset{\text{(xi)}}{=} A \cdot (\lambda E).$$

Сложнее доказать обратное: если квадратная $(n \times n)$-матрица B коммутирует со всеми $(n \times n)$-матрицами (т. е. удовлетворяет условию $B \cdot A = A \cdot B$ для любой матрицы A), то $B = \lambda E$ для некоторого $\lambda \in \mathbb{R}$. Можете попытаться доказать этот факт в качестве упражнения, воспользовавшись следующим указанием: условие коммутирования обязано выполняться, в частности, для всех матриц E_{ij}, у которых лишь в одной позиции (указанной двумя номерами) стоит единица, а в остальных местах стоят нули.

Рассмотрим теперь применительно к квадратным матрицам фиксированного размера еще одну алгебраическую операцию — умножение матрицы на число (скаляр). Если мы сопоставим скаляру λ матрицу λE, то операция умножения на скаляр может быть сведена к операции умножения на матрицу: $\lambda A = (\lambda E) \cdot A = A \cdot (\lambda E)$.

Можно считать и так: поле $\mathbb{R}$ *вложено* в кольцо квадратных матриц $L(n, \mathbb{R})$ отображением $\lambda \mapsto \lambda E$; образ этого вложения составляют матрицы, пропорциональные единичной (их принято называть *скалярными матрицами*), эти матрицы коммутируют со всеми матрицами.

Замечание 14.1. Мы с вами уже познакомились с двумя важнейшими типами алгебраических систем (т. е. множеств, наделенных алгебраическими операциями): с полями и кольцами. В алгебре изучаются также и более простые объекты, с одной алгебраической операцией (типа сложения или умножения).

Предположим, что эта операция ассоциативна и что для нее существует так называемый *нейтральный элемент* (для сложения — это нуль, для умножения — единица). Предположим, кроме того, что для любого элемента существует противоположный (в случае сложения) или обратный (в случае умножения) элемент. Тогда говорят, что рассматриваемая алгебраическая система является *группой* по сложению или по умножению.

Например, любое кольцо (в частности, любое поле) является группой по сложению, причем эта группа коммутативна. Но кольцо (или поле) по умножению группой не является, т. к. по крайней мере один элемент (нулевой) не имеет обратного.

В случае поля достаточно выбросить один этот элемент, и оставшиеся (ненулевые) элементы будут уже образовывать группу по умножению.

В кольце квадратных матриц, как мы вскоре увидим, существуют ненулевые необратимые элементы.

14.2. Группа обратимых квадратных матриц. Рассмотрим кольцо $L(n, \mathbb{R})$ квадратных $(n \times n)$-матриц.

Определение 14.1. Матрица $A \in L(n, \mathbb{R})$ называется *обратимой*, если существует такая матрица $B \in L(n, \mathbb{R})$, что справедливы равенства

$$A \cdot B = B \cdot A = E. \tag{14.1}$$

Если для матрицы A существует матрица B, удовлетворяющая (14.1), то лишь одна. В самом деле, предположим, что, кроме B, найдется еще матрица B', также удовлетворяющая этому условию (т. е. справедливо $A \cdot B' = B' \cdot A = E$). Тогда

$$B' = B' \cdot E = B' \cdot (A \cdot B) = (B' \cdot A) \cdot B = E \cdot B = B,$$

т. е. матрицы B и B' совпадают. Это наблюдение мотивирует следующее

Определение 14.2. Если существует матрица B, удовлетворяющая условию (14.1), то она называется *обратной матрицей* для матрицы A. Используется обозначение: $B = A^{-1}$.

Множество всех обратимых $(n \times n)$-матриц будем обозначать символом $GL(n, \mathbb{R})$. Отметим следующие свойства обратимых матриц.

Предложение 14.1. *1. Единичная матрица E обратима и является самообратной: $E^{-1} = E$.*

2. Если матрица A обратима, то и обратная к ней матрица обратима, причем $(A^{-1})^{-1} = A$.

3. Если матрицы A и B обратимы, то и произведение $A \cdot B$ есть обратимая матрица, причем

$$(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}. \tag{14.2}$$

Доказательство. Первые два утверждения совершенно очевидны. Третье тоже легко доказать. В самом деле,

$$(A \cdot B) \cdot (B^{-1} \cdot A^{-1}) \overset{\text{(xii)}}{=} (A \cdot (B \cdot B^{-1})) \cdot A^{-1} = (A \cdot E) \cdot A^{-1} = A \cdot A^{-1} = E.$$

Аналогично проверяется второе равенство из условия (14.1). □

Замечание 14.2. Третье утверждение предложения 14.1 коварно: неопытные вычислители, "избалованные" коммутативностью алгебры действительных чисел, забывают о перемене порядка сомножителей-матриц при обращении. Не допускайте таких ошибок.

Замечание 14.3. Если вы не пропустили замечание 14.1 в конце предыдущего пункта, то вы с удовольствием можете сейчас констатировать, что предложение 14.1 равносильно следующему лаконичному утверждению: квадратные матрицы фиксированного размера образуют группу по умножению (у этой группы есть собственное имя: *общая линейная группа*).

А если вы уже приобрели некоторый вкус к абстрактным рассуждениям, то вы легко обобщите определение 14.1 обратимых матриц на случай элементов произвольного кольца и сообразите, что обратимые элементы в любом кольце образуют группу по умножению.

Пример 14.1. Исследуем вопрос о том, когда обратима диагональная квадратная матрица

$$A = \begin{pmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_n \end{pmatrix}.$$

Если все числа α_i $(i = 1, ..., n)$, стоящие на диагонали, отличны от нуля, то такая матрица, очевидно, обратима, причем обратной к ней служит также диагональная матрица с элементами α_i^{-1} $(i = 1, ..., n)$ на диагонали.

Если же в данной матрице A есть хотя бы один нуль среди диагональных элементов, то в ней есть строка, целиком состоящая из нулей. При умножении такой матрицы A на любую $(n \times n)$-матрицу B в произведении $A \cdot B$ также будет нулевая строка, а значит, это произведение не может равняться единичной матрице. Следовательно, матрица A не является обратимой.

Вывод: диагональная матрица обратима тогда и только тогда, когда на диагонали нет нулевых элементов.

14.3. Элементарные матрицы и их связь с элементарными преобразованиями. В этом пункте будут определены и изучены квадратные матрицы специального вида, с помощью умножения на которые можно будет описать рассматривавшиеся ранее (начиная с п. 4.3) элементарные преобразования (типов I — III) над строками и столбцами (прямоугольных) матриц.

Определение 14.3. 1. *Элементарной* $(n \times n)$*-матрицей типа* I называется матрица следующего вида:

$$T_{i,j} = \begin{pmatrix} & & \boxed{i} & & \boxed{j} & & \\ 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \dots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \dots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}, \qquad (14.3)$$

где i, j — два различных номера в пределах от 1 до n, все элементы на главной диагонали равны единице, кроме двух указанных, которые равны нулю, вне главной диагонали имеются только два (указанных) ненулевых элемента.

2. *Элементарной* $(n \times n)$*-матрицей типа* II называется матрица следующего вида:

$$S_{i,j,\lambda} = \begin{pmatrix} & & \boxed{i} & & \boxed{j} & & \\ 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \dots & \lambda & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}, \tag{14.4}$$

где снова i, j — два различных номера в пределах от 1 до n, а λ — произвольный скаляр, он располагается на пересечении i-й строки и j-го столбца, вне главной диагонали нет других ненулевых элементов, а на главной диагонали стоят только единицы.

3. *Элементарной* $(n \times n)$*-матрицей типа* III называется матрица следующего вида:

$$M_{i,\lambda} = \begin{pmatrix} & & \boxed{i} & & \\ 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}, \tag{14.5}$$

где $i \in \{1, ..., n\}$, λ — ненулевой скаляр, все элементы на главной диагонали, кроме указанного, равны единице, вне диагонали все элементы равны нулю.

Замечание 14.4. В формулах (14.3) и (14.4) элементарные матрицы изображены так, как будто $i < j$. Но на самом деле это не обязательно. Точнее: для матриц типа I имеет место равенство

$$T_{i,j} = T_{j,i}$$

при любых значениях $i \neq j$, а для матриц типа II в случае $i > j$ число λ располагается ниже главной диагонали. Можно заметить, что для любых $i \neq j$ справедливо равенство

$$S_{i,j,\lambda} = (S_{j,i,\lambda})^t.$$

Возьмем теперь произвольный вектор $\bar{x} \in \mathbb{R}^n$ и умножим его слева на элементарные матрицы трех описанных выше типов. Всякий, однажды научившийся умножению матриц, уже не потеряет этот навык, и для него будут совершенно очевидными результаты:

$$T_{i,j} \cdot \bar{x} = \begin{pmatrix} & & \boxed{i} & & \boxed{j} & & \\ 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \dots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \dots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \dots \\ x_i \\ \dots \\ x_j \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \dots \\ x_j \\ \dots \\ x_i \\ \dots \\ x_n \end{pmatrix};$$

$$S_{i,j,\lambda} \cdot \bar{x} = \begin{pmatrix} & & \boxed{i} & & \boxed{j} & & \\ 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \dots & \lambda & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \dots \\ x_i \\ \dots \\ x_j \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \dots \\ x_i + \lambda x_j \\ \dots \\ x_j \\ \dots \\ x_n \end{pmatrix};$$

$$M_{i,\lambda} \cdot \bar{x} = \begin{pmatrix} & & \boxed{i} & & \\ 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \dots \\ x_i \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \dots \\ \lambda x_i \\ \dots \\ x_n \end{pmatrix}.$$

Замечаем, что элементарные матрицы не случайно получили свое имя. При умножении на них (слева) произвольного вектора-столбца

происходят следующие преобразования над элементами этого столбца: матрица $T_{i,j}$ переставляет i-ю и j-ю компоненты столбца; матрица $S_{i,j,\lambda}$ оставляет все компоненты столбца неизменными, кроме i-й, к которой прибавляется j-я компонента, домноженная на число λ; матрица $M_{i,\lambda}$ оставляет все компоненты вектора неизменными, кроме i-й, которая умножается на (ненулевое) число λ.

А теперь возьмите произвольную строку $\bar{x}^t \in \overset{*}{\mathbb{R}}{}^n$ и умножьте ее справа на элементарные матрицы. Вы увидите, что при умножении на матрицу типа I снова произойдет перестановка двух компонент (строки), при умножении на матрицу типа II к j-й компоненте строки добавится i-я компонента, домноженная на λ (именно так, в отличие от случая столбцов), умножение на матрицу типа III даст результат, аналогичный результату для столбцов.

Итак, мы приходим к выводу, что элементарные матрицы типов I — III связаны с элементарными преобразованиями соответствующих типов. Точнее, справедливо следующее предложение, относящееся к произвольным (прямоугольным) матрицам.

Предложение 14.2. *Пусть* $A \in \mathrm{Mat}(m, n, \mathbb{R})$.

1. Умножение матрицы A слева на элементарную $(m \times m)$-матрицу $T_{i,j}$ равносильно выполнению над строками A элементарного преобразования типа I: $i^{\text{стр}} \leftrightarrow j^{\text{стр}}$. Умножение A справа на аналогичную матрицу $T_{i,j}$ (но уже размера $n \times n$) равносильно преобразованию $i^{\text{стб}} \leftrightarrow j^{\text{стб}}$ над столбцами A.

2. Умножение матрицы A слева на элементарную $(m \times m)$-матрицу $S_{i,j,\lambda}$ (справа на аналогичную $(n \times n)$-матрицу) равносильно выполнению над строками (над столбцами) A элементарного преобразования типа II: $i^{\text{стр}} + j^{\text{стр}} \cdot \lambda$ (соответственно $j^{\text{стб}} + i^{\text{стб}} \cdot \lambda$).

3. Умножение матрицы A слева на элементарную $(m \times m)$-матрицу $M_{i,\lambda}$ (справа на аналогичную $(n \times n)$-матрицу) равносильно выполнению над строками (над столбцами) A элементарного преобразования типа III: $i^{\text{стр}} \cdot \lambda$ (соответственно $i^{\text{стб}} \cdot \lambda$).

Доказательство немедленно следует из сделанных выше наблюдений и того факта, что при умножении матрицы A слева на какую-либо матрицу B (подходящих размеров) каждый столбец матрицы A по отдельности умножается слева на B (соответственно при умножении A справа на матрицу C, подходящих размеров, каждая строка A по отдельности умножается на C). $\square$

14.4. Обратимость элементарных матриц. Выражение с помощью элементарных матриц приводимости прямоугольной матрицы к скелетному виду

Предложение 14.3. *Все элементарные матрицы (типов I — III) обратимы и матрица, обратная к элементарной, сама является элементарной того же типа, что и исходная:*

$$(T_{i,j})^{-1} = T_{i,j};\ (S_{i,j,\lambda})^{-1} = S_{i,j,-\lambda};\ (M_{i,\lambda})^{-1} = M_{i,\frac{1}{\lambda}}. \qquad (14.6)$$

Доказательство. 1. "Самообратность" матрицы $T_{i,j}$, т. е. равенство $T_{i,j} \cdot T_{i,j} = E$, доказывается следующим образом: умножение матрицы (14.3) слева на себя самое равносильно (в силу предложения 14.2) перестановке в этой матрице i-й и j-й строк, после чего она, очевидно, превратится в единичную матрицу.

2. Второе из равенств (14.6) равносильно двум равенствам:

$$S_{i,j,\lambda} \cdot S_{i,j,-\lambda} = E = S_{i,j,-\lambda} \cdot S_{i,j,\lambda}.$$

Докажем первое из этих равенств. Взгляните на формулу (14.4). Изображенная там матрица умножается справа на такого же типа матрицу, в которой число λ заменено на число $-\lambda$. Согласно предложению 14.2, это равносильно прибавлению к j-му столбцу матрицы (14.4) i-го столбца этой матрицы, домноженного на $-\lambda$. Ясно, что после этого получится матрица E.

Точно так же доказывается второе равенство.

3. Совершенно аналогичное доказательство третьей из формул (14.6) мы оставляем читателям. □

Применим теперь предложения 14.2 и 14.3 к процессу приведения некоторой (прямоугольной) матрицы A к скелетному виду. (Этот процесс был описан при доказательстве теоремы 5.1.)

Предложение 14.4. *Пусть A — произвольная матрица размера $m \times n$, $r = \operatorname{rank}(A)$ — ее ранг. Тогда существуют обратимые квадратные матрицы S (размера $m \times m$) и T (размера $n \times n$), такие, что*

$$A = S \cdot \hat{A} \cdot T, \qquad (14.7)$$

где $\hat{A}$ — скелетная матрица с r единицами в начале главной диагонали.

Доказательство. Согласно теореме 5.1, любая матрица A с помощью элементарных преобразований над строками и над столбцами может быть приведена к скелетному виду. Из первой теоремы о ранге матрицы (см. теорему 12.1) следует, что количество единиц на диагонали в скелетном виде $\hat{A}$ матрицы A совпадает с рангом матрицы A.

В соответствии с предложением 14.2, каждое элементарное преобразование над строками (над столбцами) можно представить как умножение слева (справа) на элементарную матрицу соответствующего типа. Так мы получим формулу

$$P_k \cdot \ldots \cdot P_2 \cdot P_1 \cdot A \cdot Q_1 \cdot Q_2 \cdot \ldots \cdot Q_l = \hat{A}, \tag{14.8}$$

где $P_1, ..., P_k$ — элементарные матрицы размера $m \times m$, а $Q_1, ..., Q_l$ — элементарные матрицы размера $n \times n$.

По предложению 14.3, все эти элементарные матрицы обратимы, а значит, по предложению 14.1, произведения $P = P_k \cdot \ldots \cdot P_2 \cdot P_1$ и $Q = Q_1 \cdot Q_2 \cdot \ldots \cdot Q_l$ также обратимы, и из формулы

$$P \cdot A \cdot Q = \hat{A}, \tag{14.9}$$

домножая ее слева на матрицу $S = P^{-1}$ и справа на матрицу $T = Q^{-1}$, мы получим формулу (14.7). □

14.5. Невырожденные матрицы. Первая теорема об условиях обратимости матрицы. В этом пункте мы выведем необходимое и достаточное условие обратимости квадратной матрицы. Поскольку в заголовке говорится о *первой теореме об условиях обратимости*, то, как вы, наверное, догадываетесь, будет еще и *вторая теорема об условиях обратимости*. Действительно, после изучения теории определителей будет доказано (см. § 28) еще одно условие обратимости. Условие же, которым мы займемся сейчас, будет опираться на понятие ранга матрицы (см. § 12).

Определение 14.4. Квадратная матрица A размера $n \times n$ называется *невырожденной*, если она имеет максимально возможный ранг, т. е. $\operatorname{rank}(A) = n$. В противном случае матрица называется *вырожденной*.

Замечание 14.5. Из определения ранга матрицы и из первой теоремы о ранге (см. теорему 12.1) немедленно следует, что матрица является невырожденной в том и только том случае, если ее столбцы (строки) линейно независимы.

Теорема 14.1 (первая теорема об условиях обратимости матрицы). *Квадратная матрица обратима тогда и только тогда, когда она невырождена.*

Доказательство. 1. Пусть матрица A невырождена, т. е. ее ранг (обозначим его для краткости r) максимален: $r = n$. Это означает, что скелетный вид матрицы A совпадает с единичной матрицей: $\hat{A} = E$. В соответствии с предложением 14.4, данная матрица представляется в виде $A = S \cdot E \cdot T = S \cdot T$, где S и T — обратимые матрицы. Следовательно, матрица A обратима, и в одну сторону утверждение теоремы доказано.

Однако, если вспомнить детали доказательства теоремы 5.1, то можно дополнительно заметить, что в случае квадратной и невырожденной матрицы A для приведения этой матрицы к скелетному виду элементарные преобразования над столбцами вообще не понадобятся. (В самом деле, количество ступенек в ступенчатом виде матрицы A в этом случае совпадает с количеством столбцов, следовательно, ступеньки должны идти подряд и заполнять всю главную диагональ; зоны справа от последней ступеньки (которая обнулялась с помощью элементарных преобразований над столбцами) не будет. К тому же скелетный вид $\hat{A}$ в этом случае совпадает с единичной матрицей E.)

Таким образом, мы получаем $P \cdot A = E$, или с учетом обратимости матрицы P:

$$A = P^{-1} = S = S_1 \cdot S_2 \cdot \ldots \cdot S_k, \tag{14.10}$$

где $S_i = P_i^{-1}$ $(i = 1, ..., k)$.

2. Пусть матрица A обратима и $\operatorname{rank}(A) = r$. По формуле (14.9) с учетом предложения 14.1 скелетный вид этой матрицы $\hat{A} = P \cdot A \cdot Q$ также есть обратимая матрица. Но диагональная $(n \times n)$-матрица $\hat{A}$ может быть обратимой (см. пример 14.1), только если на диагонали отсутствуют нулевые элементы, что возможно лишь в случае $r = n$. А это означает, что данная матрица является невырожденной. □

По ходу доказательства теоремы 14.1 обнаружился следующий (важный и сам по себе) факт:

Предложение 14.5. *Всякая обратимая матрица представляется в виде произведения элементарных матриц.*

Доказательство не требуется: см. формулу (14.10). □

14.6. Алгоритм Жордана — Гаусса вычисления обратной матрицы. Пусть A — квадратная матрица размера $n \times n$. Ниже будет описан алгоритм, выясняющий, является ли данная матрица A обратимой, и, если является, вычисляющий обратную матрицу $B = A^{-1}$.

Описание алгоритма:

— образуем составную матрицу $(A|E)$ размера $n \times 2n$;

— приведем эту матрицу к ступенчатому виду с помощью элементарных преобразований над строками (это равносильно умножению матрицы $(A|E)$ слева на некоторые элементарные матрицы);

— определим число ступенек r, приходящихся на зону матрицы A (т. е. на первые n столбцов); это число будет равняться рангу данной матрицы;

— если $r < n$, то констатируем, что данная матрица не является обратимой и останавливаем работу алгоритма;

— если $r = n$, то констатируем, что матрица A обратима (ступеньки в ступенчатом виде этой матрицы будут располагаться подряд и заполнять всю главную диагональ);

— продолжим преобразования и доведем ступенчатый вид до вида Ж.—Г.; в данном случае левая половина полученной матрицы будет совпадать с единичной матрицей E;

— считываем из преобразованной составной матрицы ее правую половину B, образовавшуюся на месте единичной матрицы E;

— записываем ответ: $A^{-1} = B$.

Схематически работу алгоритма можно представить следующим образом:

$$(A|E) \xrightarrow[\text{типов } I - III \text{ над строками}]{\text{Элементарные преобразования}} (E|A^{-1}).$$

Обоснование алгоритма: вывод об обратимости (необратимости) исходной матрицы обосновывается теоремой 14.1; в случае невырожденности для матрицы A можно найти такие элементарные матрицы $P_1, P_2, ..., P_k$, что для их произведения $P = P_k \cdot \ldots \cdot P_2 \cdot P_1$ справедливо равенство $A = P^{-1}$ и, следовательно, равенство $P = A^{-1}$; значит, при умножении на P слева составной матрицы $(A|E)$ мы получим:

$$P \cdot (A|\,E) = (P \cdot A \,|\, P \cdot E) = (E|\,P) = (E|A^{-1}).$$

Пример 14.2. Решим следующую з а д а ч у: выяснить, при каких значениях параметра λ матрица

$$A = \begin{pmatrix} 1 & 1 & -1 \\ 0 & -1 & -1 \\ \lambda & 0 & 1 \end{pmatrix}$$

является обратимой. Найти (при этих значениях λ) обратную матрицу A^{-1}.

Р е ш е н и е. Выпишем составную матрицу $(A|E)$ и приведем ее к ступенчатому виду (с помощью элементарных преобразований над строками; подчеркнем, что преобразования над столбцами в этом алгоритме не допустимы).

$$(A\,|\,E) = \left(\begin{array}{ccc|ccc} 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 1 & 0 \\ \lambda & 0 & 1 & 0 & 0 & 1 \end{array}\right) \longrightarrow$$

$$\xrightarrow{3^{\text{стр}}+1^{\text{стр}}\cdot(-\lambda)} \left(\begin{array}{ccc|ccc} \boxed{1} & 1 & -1 & 1 & 0 & 0 \\ 0 & \boxed{-1} & -1 & 0 & 1 & 0 \\ 0 & -\lambda & \lambda+1 & -\lambda & 0 & 1 \end{array}\right) \longrightarrow$$

$$\xrightarrow{3^{\text{стр}}+2^{\text{стр}}\cdot(-\lambda)} \left(\begin{array}{ccc|ccc} \boxed{1} & 1 & -1 & 1 & 0 & 0 \\ 0 & \boxed{-1} & -1 & 0 & 1 & 0 \\ 0 & 0 & 2\lambda+1 & -\lambda & -\lambda & 1 \end{array}\right) \longrightarrow \ldots$$

Замечаем, что уже обозначились две ступеньки и что в особом случае $\boxed{\lambda = -1/2}$ третья ступенька придется на четвертый столбец, а в зоне матрицы A будут только две ступеньки и поэтому $\operatorname{rank}(A)$ будет равен 2, и следовательно, матрица A будет в этом случае необратимой.

В неособом случае $\boxed{\lambda \neq -1/2}$ три ступеньки будут приходиться на зону матрицы A, эта матрица будет невырожденной и, значит, обратимой.

Продолжим преобразования до достижения вида Ж.—Г.:

$$\ldots \xrightarrow[2^{\text{стр}}\cdot(-1)]{1^{\text{стр}}+2^{\text{стр}}} \left(\begin{array}{ccc|ccc} \boxed{1} & 0 & -2 & 1 & 1 & 0 \\ 0 & \boxed{1} & 1 & 0 & -1 & 0 \\ 0 & 0 & \boxed{2\lambda+1} & -\lambda & -\lambda & 1 \end{array}\right) \longrightarrow$$

$$\xrightarrow{3^{\text{стр}}\cdot\frac{1}{2\lambda+1}} \left(\begin{array}{ccc|ccc} \boxed{1} & 0 & -2 & 1 & 1 & 0 \\ 0 & \boxed{1} & 1 & 0 & -1 & 0 \\ 0 & 0 & \boxed{1} & -\frac{\lambda}{2\lambda+1} & -\frac{\lambda}{2\lambda+1} & \frac{1}{2\lambda+1} \end{array}\right) \longrightarrow$$

$$\xrightarrow[2^{\text{стр}}+3^{\text{стр}}\cdot(-1)]{1^{\text{стр}}+3^{\text{стр}}\cdot 2} \left(\begin{array}{ccc|ccc} \boxed{1} & 0 & 0 & \frac{1}{2\lambda+1} & \frac{1}{2\lambda+1} & \frac{2}{2\lambda+1} \\ 0 & \boxed{1} & 0 & \frac{\lambda}{2\lambda+1} & -\frac{\lambda+1}{2\lambda+1} & -\frac{1}{2\lambda+1} \\ 0 & 0 & \boxed{1} & -\frac{\lambda}{2\lambda+1} & -\frac{\lambda}{2\lambda+1} & \frac{1}{2\lambda+1} \end{array}\right).$$

Преобразования завершены, из зоны матрицы E можно считатывать сформировавшуюся там матрицу A^{-1}.

О т в е т. Матрица A обратима тогда и только тогда, когда параметр $\lambda \neq -1/2$. В этом случае

$$A^{-1} = \frac{1}{2\lambda+1} \begin{pmatrix} 1 & 1 & 2 \\ \lambda & -\lambda-1 & -1 \\ -\lambda & -\lambda & 1 \end{pmatrix}.$$

Замечание 14.6. Maple легко справляется с задачей вычисления обратной матрицы: можно использовать команды inverse(A) или evalm(A ^ (-1)). В случае матриц, зависящих от параметра, рекомендуется команда ffgausselim(concat(A,E)).

Еще одна небольшая Maple-сессия:

> restart; with (linalg) :

Сначала вычислим матрицу, обратную к числовой матрице:

> M := matrix ([[2, 0, 0, 1],[3, 1, −1, 0],[−1, −1, 0, 0],[0, 0, −2, 2]]);

$$M := \begin{bmatrix} 2 & 0 & 0 & 1 \\ 3 & 1 & -1 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & -2 & 2 \end{bmatrix}$$

> evalm (M ^ (−1)) ;

$$\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{8} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{5}{4} & \frac{1}{8} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{4} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{4} \end{bmatrix}$$

Теперь введем (исследованную выше вручную) матрицу с параметром:

> A := matrix ([[1, 1, −1], [0, −1, −1], [lambda, 0, 1]]);

$$A := \begin{bmatrix} 1 & 1 & -1 \\ 0 & -1 & -1 \\ \lambda & 0 & 1 \end{bmatrix}$$

Введем единичную матрицу:

> E := evalm (array (1..3, 1..3, identity)) ;

$$E := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Приведем к ступенчатому виду (без дробей) конкатенацию $(A|E)$:

> G := ffgausselim (concat (A, E)) ;

$$G := \begin{bmatrix} 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1-2\lambda & \lambda & \lambda & -1 \end{bmatrix}$$

Замечаем, что при $\lambda = -1/2$ матрица A необратима. В случае, когда $\lambda \neq -1/2$, можем продолжить преобразования (символ % в следующей команде обозначает результат предыдущего действия):

> gaussjord (%) ;

$$\begin{bmatrix} 1 & 0 & 0 & \frac{1}{1+2\lambda} & \frac{1}{1+2\lambda} & \frac{2}{1+2\lambda} \\ 0 & 1 & 0 & \frac{\lambda}{1+2\lambda} & -\frac{1+\lambda}{1+2\lambda} & -\frac{1}{1+2\lambda} \\ 0 & 0 & 1 & -\frac{\lambda}{1+2\lambda} & -\frac{\lambda}{1+2\lambda} & \frac{1}{1+2\lambda} \end{bmatrix}$$

Из правой зоны полученной матрицы считываем ответ. Естественно, он совпал с полученным вручную.

§ 15. Линейные операторы в арифметических линейных пространствах

15.1. Линейные отображения арифметических линейных пространств и алгебраические действия над ними. На интуитивном уровне представление о свойстве линейности для отображений было дано уже в самом начале настоящего пособия, в замечании 1.1. Сейчас мы дадим строгое определение понятия линейного отображения из одного арифметического линейного пространства в другое.

Определение 15.1. Отображение $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ называется *линейным отображением* (или *линейным оператором*), если оно согласовано с алгебраическими действиями сложения векторов и умножения вектора на скаляр, т. е. если выполнены следующие два условия:

— согласованность со сложением:

$$\varphi(\bar{a} + \bar{b}) = \varphi(\bar{a}) + \varphi(\bar{b}); \tag{15.1}$$

— согласованность с умножением на скаляр (или *правило вынесения скаляра из-под знака линейного оператора*):

$$\varphi(\lambda\bar{a}) = \lambda\varphi(\bar{a}), \tag{15.2}$$

для любых векторов $\bar{a}, \bar{b} \in \mathbb{R}^n$ и любого скаляра $\lambda \in \mathbb{R}$.

Множество всех линейных операторов из пространства $\mathbb{R}^n$ в пространство $\mathbb{R}^m$ обозначается $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

Замечание 15.1. В некоторых учебниках по алгебре термин "линейное отображение" считается более общим, а термин "линейный оператор" закрепляется за линейными отображениями из одного линейного пространства в само это линейное пространство. Такое словоупотребление расходится с практикой других математических дисциплин (например, функционального анализа), и мы здесь от него отказываемся, считая в дальнейшем упомянутые термины синонимами.

Остановимся еще на одном варианте терминологии, характерном для так называемой *общей алгебры* (науки, изучающей множества с заданными на них алгебраическими действиями: группы, кольца, поля, линейные пространства и т. п.).

Отображение какого-либо алгебраического объекта в однотипный объект называется *гомоморфизмом*, если оно согласовано со всеми алгебраическими действиями, характерными для данного типа объектов. Так, гомоморфизмы линейных пространств – это не что иное, как линейные отображения. Бывают еще гомоморфизмы групп, колец и т.д. На этом языке гомоморфизмы алгебраического объекта в себя именуются *эндоморфизмами*.

Из языка общей алгебры нам вскоре понадобится еще один термин. Гомоморфизм (каких-либо алгебраических объектов) называется *изоморфизмом*, если он является взаимно однозначным отображением (биекцией). Легко доказать, что в этом случае отображение, обратное к данному гомоморфизму, также будет являться гомоморфизмом (а значит, и изоморфизмом; получается пара взаимно обратных изоморфизмов). (Для некоторых частных типов алгебраических объектов, например при доказательстве теоремы 15.1, это будет проделано ниже.)

В частности, изоморфизм линейных пространств — это взаимно однозначное линейное отображение одного пространства на другое. В п. 15.6 мы значительно подробнее остановимся на линейных изоморфизмах; докажем что между различными арифметическими линейными пространствами не может быть изоморфизма.

С точки зрения общей алгебры изоморфные объекты неразличимы (их возможные различия лежат вне сферы, изучаемой этой наукой).

Слово "изоморфизм" является, наверное, самым важным словом во всей математике. Вам еще предстоит изучить много так называемых *теорем об изоморфизме* и познакомиться (в связи с этим понятием) с афоризмами типа: "Математика — это искусство называть разные вещи одинаковыми именами".

Замечание 15.2. Отметим некоторые непосредственные следствия из определения 12.1.

Во-первых, линейное отображение переводит нулевой вектор снова в нулевой: $\varphi(\bar{0}) = \bar{0}$. (Это следует из возможности вынести скалярный нуль из-под знака φ; однако можно рассуждать и иначе: равенство $\bar{0} + \bar{0} = \bar{0}$ влечет равенство $\varphi(\bar{0}) + \varphi(\bar{0}) = \varphi(\bar{0})$, добавляя к обеим частям которого вектор, противоположный $\varphi(\bar{0})$, мы получаем требуемое.)

Во-вторых, комбинируя свойства (15.1) и (15.2), легко доказать,

что линейное отображение "сохраняет" линейные комбинации, т. е.

$$\varphi\left(\sum_{i=1}^{k}\lambda_i\bar{a}_i\right)=\sum_{i=1}^{k}\lambda_i\varphi(\bar{a}_i) \tag{15.3}$$

для любых векторов $\bar{a}_i\in\mathbb{R}^n$ и любых скаляров $\lambda_i\in\mathbb{R}$ $(i=1,...,k)$.

Переходим к описанию трех алгебраических действий над линейными операторами. Так же как и алгебраические действия над прямоугольными матрицами, они бывают осуществимы не всегда. Сложить можно будет только такие два линейных оператора φ и ψ, которые действуют из одного и того же арифметического линейного пространства $\mathbb{R}^n$ в одно и то же пространство $\mathbb{R}^m$. Точное определение:

Определение 15.2. Суммой двух линейных операторов $\varphi,\psi\in\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$ называется *поточечная сумма* отображений φ и ψ :

$$\varphi+\psi:\mathbb{R}^n\to\mathbb{R}^m;\ (\varphi+\psi)(\bar{x})=\varphi(\bar{x})+\psi(\bar{x});\ \bar{x}\in\mathbb{R}^n. \tag{15.4}$$

Легко доказать, что сумма $\varphi+\psi$ двух линейных операторов сама является линейным оператором, принадлежащим $\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$. В самом деле,

$$\begin{aligned}&(\varphi+\psi)(\bar{x}+\bar{y})=\varphi(\bar{x}+\bar{y})+\psi(\bar{x}+\bar{y})=\\ =&(\varphi(\bar{x})+\varphi(\bar{y}))+(\psi(\bar{x})+\psi(\bar{y}))=\\ =&(\varphi(\bar{x})+\psi(\bar{x}))+(\varphi(\bar{y})+\psi(\bar{y}))=\\ =&(\varphi+\psi)(\bar{x})+(\varphi+\psi)(\bar{y}),\end{aligned}$$

и свойство (15.1) доказано.

Аналогично доказывается второе свойство, входящее в определение линейности отображения [свойство (15.2)].

Умножение линейного оператора на скаляр также определяется поточечно:

Определение 15.3. Произведением линейного отображения $\varphi\in\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$ на скаляр $\lambda\in\mathbb{R}$ называется отображение

$$\lambda\varphi:\mathbb{R}^n\to\mathbb{R}^m;\ (\lambda\varphi)(\bar{x})=\lambda\cdot\varphi(\bar{x});\ \bar{x}\in\mathbb{R}^n. \tag{15.5}$$

Столь же просто, как и выше, для отображения (15.4) проверяются свойства (15.1) и (15.2) и, таким образом, доказывается, что $\lambda\varphi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

Роль умножения для линейных операторов играет композиция последовательно действующих отображений. Если $\varphi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ и $\psi \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^p)$ — два последовательно действующих линейных оператора в арифметических линейных пространствах, то их композиция

$$\psi \circ \varphi : \mathbb{R}^n \to \mathbb{R}^p;\ (\psi \circ \varphi)(\bar{x}) = \psi(\varphi(\bar{x}));\ \bar{x} \in \mathbb{R}^n \tag{15.6}$$

также является линейным оператором, принадлежащим $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$.

Для разнообразия проверим свойство (15.2) [оставляя проверку свойства (15.1) в качестве упражнения]:

$$(\psi \circ \varphi)(\lambda\bar{x}) = \psi(\varphi(\lambda\bar{x})) = \psi(\lambda \cdot \varphi(\bar{x})) = \lambda \cdot \psi(\varphi(\bar{x})) = \lambda \cdot (\psi \circ \varphi)(\bar{x}).$$

Замечание 15.3 (для служебного пользования). С обозначением композиции отображений в алгебре царит разнобой. Имеются две непримиримые партии. "Левые" записывают композицию двух отображений множеств

$$\varphi : X \to Y;\ \psi : Y \to Z$$

следующим образом:

$$\psi \circ \varphi : X \to Z;\ (\psi \circ \varphi)(x) = \psi(\varphi(x));\ x \in X.$$

"Правые" придерживаются противоположных принципов. Во-первых, они знак отображения помещают справа от элемента, на который действует отображение, во-вторых, по возможности, стараются избегать скобок и, в-третьих, используют противоположный порядок записи сомножителей (отображений):

$$\varphi \circ \psi : X \to Z;\ x(\varphi \circ \psi) = (x\varphi)\psi;\ x \in X.$$

По большому счету, конечно, правосторонняя и левосторонняя записи совершенно равноправны. "Правые" могут привести примеры из других разделов математики, где успешно используется бесскобочная правосторонняя запись и будут, как обычно, правы.

Но если говорить о *линейной алгебре*, то, по-видимому, нет другой области математики, в которой правосторонняя система обозначений была бы менее удобна. Уже в следующем пункте при изучении матричного задания линейных операторов мы сможем убедиться

в том, что если векторы записывать в столбик (а их надо записывать именно так; см. замечание 8.6), то (при общепринятом правиле умножения матриц) матрицу, умножаемую на вектор, надо располагать слева от вектора, а следовательно, и знак линейного оператора естественно ставить слева, и значит, для композиции естественно использовать левостороннюю запись.

Забегая вперед, можно привести и такой аргумент: векторы-строки понадобятся нам для представления линейных форм (ковекторов), мы будем умножать их на матрицы справа (что будет диктоваться самой природой "двойственной" задачи).

Несколько слов об обозначениях в других разделах математики. В учебниках по анализу и геометрии левосторонняя запись практически общепринята. (Алгебраисты здесь существенно выбиваются из общей традиции.)

При переходе к алгоритмическим языкам и к математическим компьютерным системам мы также будем, как правило, иметь дело с левосторонней записью функций и операторов, а отказ от скобок будет чаще всего невозможен (скобок, скорее, добавится).

Пример 15.1. Простейшими примерами линейных операторов являются *нулевые операторы*

$$o : \mathbb{R}^n \to \mathbb{R}^m;\ o(\bar{x}) = \bar{0};\ \bar{x} \in \mathbb{R}^n \tag{15.7}$$

и *тождественные операторы*

$$\varepsilon : \mathbb{R}^n \to \mathbb{R}^n;\ \varepsilon(\bar{x}) = \bar{x};\ \bar{x} \in \mathbb{R}^n. \tag{15.8}$$

Свойства (15.1) и (15.2) для нулевых операторов тривиальны, а для тождественных — тавтологичны.

Пример 15.2. Рассмотрим теперь произвольную $(m \times n)$-матрицу A и определим следующее отображение:

$$\varphi = \varphi_A : \mathbb{R}^n \to \mathbb{R}^m;\ \varphi(\bar{x}) = A \cdot \bar{x};\ \bar{x} \in \mathbb{R}^n. \tag{15.9}$$

(Просто φ — это краткое обозначение; обозначение φ_A с явным указанием матрицы, по которой строится отображение, будет использоваться в случае необходимости, например, если придется рассматривать сразу несколько матриц и соответствующих отображений.)

Линейность отображения (15.9) немедленно следует из законов матричной алгебры, и она фактически уже установлена по ходу доказательства предложения 3.1.

Заметьте, в частности, что нулевой матрице отвечает нулевой оператор, а единичной матрице — тождественный оператор.

15.2. Матрица линейного оператора (относительно естественных базисов в арифметических линейных пространствах). В последнем примере мы по произвольной прямоугольной матрице строили линейный оператор. Теперь мы проделаем путь в противоположном направлении: начнем с произвольного линейного оператора в арифметических линейных пространствах и сопоставим ему матрицу.

Рассмотрим линейный оператор $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ и естественные базисы $\mathcal{E}_n$ и $\mathcal{E}_m$ в пространствах $\mathbb{R}^n$ и $\mathbb{R}^m$ (если забыли этот термин — вернитесь к примеру 10.2).

Всякий вектор $\bar{x} \in \mathbb{R}^n$ можно разложить по базису $\mathcal{E}_n$:

$$\bar{x} = \sum_{j=1}^{n} x_j \bar{e}_j.$$

Пользуясь тем, что линейное отображение φ сохраняет линейные комбинации [см. формулу (15.3)], мы можем расписать значение этого отображения на векторе $\bar{x}$ следующим образом:

$$\varphi(\bar{x}) = \varphi\left(\sum_{j=1}^{n} x_j \bar{e}_j\right) = \sum_{j=1}^{n} x_j \varphi(\bar{e}_j).$$

Вводя обозначение

$$\bar{a}_j = \varphi(\bar{e}_j);\ j = 1, ..., n$$

для векторов пространства $\mathbb{R}^m$, являющихся образами единичных векторов пространства $\mathbb{R}^n$, мы получим формулу

$$\varphi(\bar{x}) = \sum_{j=1}^{n} x_j \bar{a}_j. \tag{15.10}$$

Составим из векторов $\bar{a}_j$ матрицу размера $m \times n$:

$$A = A_\varphi = (\bar{a}_1 | \, \bar{a}_2 \, | \, ... \, | \, \bar{a}_n). \tag{15.11}$$

(Просто A — это краткое обозначение; в более подробном обозначении A_φ явно указывается оператор, по которому построена матрица; это обозначение будет применяться в случае необходимости, например, если мы будем иметь дело с несколькими операторами и соответствующими матрицами.)

Определение 15.4. Матрица (15.11) называется *матрицей линейного отображения* φ относительно естественных базисов в пространствах $\mathbb{R}^n$ и $\mathbb{R}^m$.

С помощью понятия матрицы линейного оператора действие этого оператора на вектор может быть записано формулой

$$\varphi(\bar{x}) = A \cdot \bar{x};\ \bar{x} \in \mathbb{R}^n, \tag{15.12}$$

следующей из (15.10), с учетом сделанного в начале п. 8.6 замечания о векторной форме записи матричного произведения $A \cdot \bar{x}$.

Пример 15.3. Приведем важнейший (геометрический) пример линейного оператора. Рассмотрим двумерное арифметическое линейное пространство (плоскость) $\mathbb{R}^2$ и отображение $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$, состоящее в том, что всякий вектор $\bar{x} \in \mathbb{R}^2$ поворачивается (вокруг начала координат) на один и тот же угол α. Это отображение в самом деле является линейным, поскольку при повороте параллелограммы переходят в параллелограммы, и следовательно, отображение φ согласовано со сложением векторов; еще очевиднее согласованность φ с умножением на скаляр.

Рассмотрим естественный базис $[\bar{e}_1, \bar{e}_2]$ и "новые", повернутые векторы $\bar{a}_1 = \varphi(\bar{e}_1)$ и $\bar{a}_2 = \varphi(\bar{e}_2)$. Легко доказать (и вы это в других обозначениях уже делали в курсе аналитической геометрии), что

$$\begin{array}{rcrcrcl} \bar{a}_1 & = & \cos\alpha & \bar{e}_1 & + & \sin\alpha & \bar{e}_2; \\ \bar{a}_2 & = & -\sin\alpha & \bar{e}_1 & + & \cos\alpha & \bar{e}_2. \end{array}$$

Следовательно, матрица (15.11) имеет в данном случае хорошо знакомый вам вид:

$$A = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}.$$

15.3. Теорема об изоморфизме для алгебраической системы линейных операторов в арифметических линейных пространствах и алгебраической системы прямоугольных матриц. Пришел черед первой серьезной теоремы, устанавливающей изоморфизм между двумя алгебраическими объектами (или, как еще говорят, алгебраическими системами). Перечитайте замечание 15.1, именно там было описано общее понятие изоморфизма алгебраических объектов (систем), здесь оно будет конкретизировано.

Одной из алгебраических систем будет множество всех прямоугольных матриц с алгебраическими действиями сложения, умножения на скаляр и умножения, описанными в п. 1.2.

Второй алгебраической системой будет описанное в п. 15.1 множество всевозможных линейных операторов в арифметических линейных пространствах, наделенное алгебраическими действиями сложения, умножения на скаляр и композиции.

Сопоставляя каждому линейному оператору $\varphi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ соответствующую ему матрицу $A_\varphi \in \mathrm{Mat}(m, n; \mathbb{R})$, мы получаем отображение

$$\mu : \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \to \mathrm{Mat}(m, n; \mathbb{R});\ \mu(\varphi) = A_\varphi. \tag{15.13}$$

[Вообще-то, формула (15.13) задает не одно отображение, а семейство отображений: при различных значениях размерностей m и n получаются различные отображения. В таких случаях можно вводить "различающие индексы" и писать $\mu_{m,n}$, но сейчас нам этого делать не хочется. Громоздкие обозначения не способствуют пониманию.]

А теперь вернемся к примеру 15.2. В этом примере мы с помощью формулы (15.9) по матрице строили линейный оператор. Так возникает отображение (встречное по отношению к μ)

$$\nu : \mathrm{Mat}(m, n; \mathbb{R}) \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m);\ \nu(A) = \varphi_A. \tag{15.14}$$

Убедимся, что отображения

$$\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \underset{\nu}{\overset{\mu}{\rightleftarrows}} \mathrm{Mat}(m, n; \mathbb{R})$$

не просто действуют навстречу друг другу, но являются *взаимно обратными отображениями*, т. е. справедливы два соотношения

$$\text{а)}\ \nu \circ \mu = \varepsilon;\ \text{б)}\ \mu \circ \nu = \varepsilon, \tag{15.15}$$

где одним и тем же символом ε обозначены тождественные отображения в двух различных множествах (в первом из соотношений в правой части фигурирует тождественное отображение множества линейных операторов, действующих из $\mathbb{R}^n$ в $\mathbb{R}^m$, а во втором соотношении — тождественное отображение множества $(m \times n)$-матриц).

В самом деле, возьмем произвольный линейный оператор $\varphi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Отображение μ сопоставляет ему матрицу A, заданную формулой (15.11), а соотношение (15.12) выражает исходный оператор φ через эту матрицу, причем по тому же принципу, что и в формуле (15.9), задающей отображение (15.14). Таким образом, для любого линейного оператора φ получается $\nu(\mu(\varphi)) = \varphi$, т. е. соотношение (15.15а) доказано.

Обратно, возьмем произвольную $(m \times n)$-матрицу A. Отображение ν сопоставляет ей линейный оператор φ, заданный формулой (15.9). Выясним, какая матрица будет соответствовать этому оператору при отображении μ. Для этого надо [см. формулу (15.11)] применить оператор φ к единичным векторам $\bar{e}_j$ $(j = 1, ..., n)$, т. е. вычислить произведения $A \cdot \bar{e}_j$. Но указанное произведение в точности равно вектору $\bar{a}_j$ (j-му столбцу матрицы A). Следовательно, мы вернулись к исходной матрице и тем самым доказали равенство $\mu(\nu(A)) = A$ для любой матрицы A. Соотношение (15.15б) также доказано.

Другими словами, между множеством всех линейных операторов в арифметических линейных пространствах и множеством всех прямоугольных матриц установлено взаимно однозначное соответствие. В следующей теореме будет доказана согласованность этого соответствия с алгебраическими действиями над операторами и над матрицами.

Теорема 15.1. *Взаимно однозначное соответствие, заданное отображениями (15.13) и (15.14), является изоморфизмом алгебраических систем, а именно: эти отображения согласованы*

1) со сложением, т. е. справедливы соотношения

$$\text{а) } \mu(\varphi + \psi) = \mu(\varphi) + \mu(\psi); \ \text{б) } \nu(A + B) = \nu(A) + \nu(B) \qquad (15.16)$$

для любых линейных операторов $\varphi, \psi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ и любых матриц $A, B \in \mathrm{Mat}(m, n; \mathbb{R})$;

2) с умножением на скаляр, т. е. справедливы соотношения

$$\text{а) } \mu(\lambda\varphi) = \lambda\mu(\varphi); \ \text{б) } \nu(\lambda A) = \lambda\nu(A) \qquad (15.17)$$

для любого линейного оператора φ, любой матрицы A и любого скаляра $\lambda \in \mathbb{R}$;

3) с действиями композиции (для линейных операторов) и умножения (для матриц), т. е. справедливы соотношения

$$\text{а) } \mu(\psi \circ \varphi) = \mu(\psi) \cdot \mu(\varphi);\ \text{б) } \nu(B \cdot A) = \nu(B) \circ \nu(A) \qquad (15.18)$$

для любых линейных операторов $\varphi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, $\psi \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^p)$ и любых матриц $A \in \mathrm{Mat}(m, n; \mathbb{R})$, $B \in \mathrm{Mat}(p, m; \mathbb{R})$.

Доказательство. Требуется доказать три пары соотношений.

Первые соотношения в каждой из этих пар выражают согласованность (с одним из трех алгебраических действий) для отображения μ; вторые соотношения выражают аналогичное свойство для обратного отображения ν.

Объясним сначала, что достаточно доказать свойство согласованности лишь для одного из двух взаимно обратных отображений. Пусть, например, установлено второе из соотношений (15.18б). Тогда первое соотношение доказывается следующим образом:

$$\mu(\psi \circ \varphi) \overset{(15.15\text{а})}{=\!=} \mu(\,\nu(\mu(\psi)) \circ \nu(\mu(\varphi))\,) \overset{(15.18\text{б})}{=\!=}$$
$$= \mu(\nu(\,\mu(\psi) \cdot \mu(\varphi)\,)) = (\mu \circ \nu)\,(\mu(\psi) \cdot \mu(\varphi)) \overset{(15.15\text{б})}{=\!=} \mu(\psi) \cdot \mu(\varphi).$$

[Разумеется, можно было бы начинать с соотношения (15.18а) и выводить соотношение (15.18б). Точно такое же рассуждение годится и для двух других пар соотношений. И вообще, как уже отмечалось в замечании 15.1, это — общий факт общей алгебры: если одно из двух взаимно обратных отображений согласовано с алгебраическими действиями, то и другое отображение с ними согласовано. Отсюда возникает некоторая "свобода": можно доказывать то из двух соотношений, доказательство которого проще.]

Приступаем к доказательству вторых соотношений в каждой из пар (т. е. доказываем согласованность с алгебраическими действиями для отображения ν).

Доказательство (15.16б). Пусть $A, B \in \mathrm{Mat}(m, n; \mathbb{R})$ — две матрицы одного и того же размера; им [по формуле (15.9)] соответствуют два линейных оператора, действующих из $\mathbb{R}^n$ в $\mathbb{R}^m$ и выражаемых формулами $\varphi(\bar{x}) = A \cdot \bar{x}$; $\psi(\bar{x}) = B \cdot \bar{x}$ ($\bar{x} \in \mathbb{R}^n$). Сумма этих операторов будет выражаться формулой

$$(\varphi + \psi)(\bar{x}) \overset{(15.4)}{=\!=} \varphi(\bar{x}) + \psi(\bar{x}) = A \cdot \bar{x} + B \cdot \bar{x} \overset{(\text{x})}{=} (A + B) \cdot \bar{x},$$

т. е. будет соответствовать матрице $A + B$, что и утверждается в доказываемой формуле.

Доказательство (15.17б). Пусть матрице A соответствует линейный оператор φ. Надо доказать, что матрице λA (где λ — скаляр) будет соответствовать оператор $\lambda\varphi$. Это следует из цепочки равенств

$$(\lambda\varphi)(\bar{x}) \overset{(15.5)}{=\!=} \lambda \cdot (\varphi(\bar{x})) = \lambda \cdot (A \cdot \bar{x}) \overset{\text{(xi)}}{=} (\lambda A) \cdot \bar{x}.$$

Доказательство (15.18б). Пусть матрице A размера $m \times n$ соответствует линейный оператор $\varphi : \mathbb{R}^n \to \mathbb{R}^m$; $\varphi(\bar{x}) = A \cdot \bar{x}$ $(\bar{x} \in \mathbb{R}^n)$, а матрице B размера $p \times m$ соответствует линейный оператор $\psi : \mathbb{R}^m \to \mathbb{R}^p$; $\psi(\bar{y}) = B \cdot \bar{y}$ $(\bar{y} \in \mathbb{R}^m)$. Надо доказать, что матрице $B \cdot A$ будет соответствовать оператор $\psi \circ \varphi : \mathbb{R}^n \to \mathbb{R}^p$. Это следует из цепочки равенств

$$(\psi \circ \varphi)(\bar{x}) = \psi(\varphi(\bar{x})) = \psi(A \cdot \bar{x}) = B \cdot (A \cdot \bar{x}) \overset{\text{(xii)}}{=} (B \cdot A) \cdot \bar{x}.$$

Теорема доказана. □

Замечание 15.4. Доказанная выше теорема позволяет (см. замечание 15.1) "не различать" изоморфные алгебраические системы линейных операторов и прямоугольных матриц. Все, что доказано для одной из них, будет справедливо и для другой. Может показаться, что в такой ситуации стоит ограничиться изучением лишь одной системы (например, только алгебраической системы матриц, как более простой для понимания).

Однако этот подход оказывается неплодотворным по многим причинам. Первая причина утилитарна: часто доказательства значительно проще проводятся "на языке линейных операторов", чем "на языке матриц" (см. следующий пункт).

Вторая причина более существенна: само понятие линейного оператора значительно глубже элементарного "бухгалтерского" объекта — матрицы. Это понятие будет развиваться далее и переноситься на так называемые *абстрактные* линейные пространства (в том числе и *бесконечномерные*). Именно линейные операторы являются основным предметом изучения как в линейной алгебре, так и в линейном функциональном анализе.

А матрицу можно рассматривать как "портрет" линейного оператора. Этот портрет получается с помощью вспомогательных "приборов". В случае арифметических линейных пространств такими

"приборами" могут служить естественные базисы. Но могут использоваться и другие "приборы" (базисы). И, что очень важно, для того же самого оператора в других базисах может получиться лучшая "фотография", по которой мы сможем непосредственно увидеть интересующие нас свойства оператора.

15.4. Законы для алгебраических действий над линейными операторами. В первой главе, в п. 2.3, были доказаны законы (i) — (xiii) для алгебраических действий над матрицами. Выше, в п. 15.3, было доказано, что существует изоморфизм между алгебраической системой прямоугольных матриц и алгебраической системой линейных операторов в арифметических линейных пространствах, при котором сложению матриц соответствует сложение линейных операторов, умножению матрицы на скаляр — умножение оператора на скаляр, умножению матриц — композиция операторов.

Всякому истинному утверждению, относящемуся к элементам алгебраической системы (и результатам алгебраических действий над ними), соответствует истинное утверждение, относящееся к элементам и результатам алгебраических действий в изоморфной алгебраической системе. Поэтому "автоматически" оказываются справедливыми 13 законов, аналогичных законам (i) — (xiii), для алгебраических действий над линейными операторами в арифметических линейных пространствах.

Например, первому дистрибутивному закону (ix) будет соответствовать следующее утверждение.

Пусть $\alpha, \beta \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ — два линейных оператора, действующих из $\mathbb{R}^n$ в $\mathbb{R}^m$, $\gamma \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^n)$ — еще один линейный оператор, действующий из $\mathbb{R}^p$ в $\mathbb{R}^n$. Тогда имеет место равенство

$$(\alpha + \beta) \circ \gamma = \alpha \circ \gamma + \beta \circ \gamma.$$

Очень полезным для "вживания" в проблематику общей алгебры было бы следующее упражнение: выписать подробно формулировки всех 13 законов для алгебры линейных операторов.

Замечание 15.5. В силу теоремы 15.1, законы для алгебраических действий над линейными операторами не требуют доказательства.

Однако, если вы попробуете все-таки провести для этих законов *независимые* доказательства (не опирающиеся на связь операторов с матрицами), то вы скоро убедитесь, что новые доказательства будут значительно проще тех, которые проводились (для матриц) в п. 2.3.

Скажем, ассоциативный закон для композиции линейных операторов вообще не требует отдельного доказательства, поскольку он справедлив для любых отображений любых множеств. Об этом вам говорили в подготовительном курсе "Введение в математику". Здесь мы только "освежим память" небольшой выкладкой.

Рассмотрим цепочку из трех последовательно действующих (произвольных) отображений (произвольных) множеств

$$X \xrightarrow{\alpha} Y \xrightarrow{\beta} Z \xrightarrow{\gamma} U.$$

Докажем ассоциативный закон

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

Как левая, так и правая части этого равенства являются отображениями из множества X в множество U. Равенство отображений понимается как их равенство на каждом элементе $x \in X$.

Вычислим на элементе x значение левой части:

$$(\,\gamma \circ (\beta \circ \alpha)\,)\,(x) = \gamma(\,(\beta \circ \alpha)(x)\,) = \gamma(\beta(\alpha(x))).$$

Вычислим теперь на элементе x значение правой части:

$$(\,(\gamma \circ \beta) \circ \alpha)\,)\,(x) = (\gamma \circ \beta)\,(\alpha(x)) = \gamma(\beta(\alpha(x))).$$

Получились одинаковые результаты. Значит, равенство отображений доказано.

Ранее вам было рекомендовано упражнение на правильное формулирование законов (для случая алгебры линейных операторов). Теперь вам предлагается дать независимые доказательства для всех сформулированных законов. Особое внимание обращайте на те этапы доказательств, где используется линейность отображений. [В вышепроведенном рассуждении для закона (xii) линейность совсем не использовалась. Не будет она использоваться и при доказательстве закона (xiii), выражающего свойства тождественных отображений ε. В других случаях (в каких?) без предположения линейности отображений не обойтись.]

И наконец, проясним возможно возникшее у внимательных читателей недоумение: а нельзя ли было с самого начала вести изложение на языке линейных операторов, а потом, при решении задач, перейти к вычислениям с матрицами?

Можно было бы, конечно. Но кто бы это понял? Без предварительного знакомства с системами линейных уравнений, со структурой их общих решений абстрактные определения и теоремы были бы, на наш взгляд, трудно воспринимаемы.

А вот "при втором заходе" (во втором семестре, в курсе линейной алгебры и геометрии) такой вариант изложения будет уже вполне уместен. В этом курсе многие из изученных понятий будут обобщены, причем некоторые рассуждения придется повторить в новой ситуации и на новом уровне понимания.

15.5. Линейные операторы в одном пространстве. Обратимые линейные операторы и обратимые матрицы. В этом пункте мы рассмотрим множество $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ всех линейных операторов, действующих из арифметического линейного пространства $\mathbb{R}^n$ в само это пространство (т. е., как иначе говорят, множество *эндоморфизмов* пространства $\mathbb{R}^n$; см. замечание 15.1).

Введем для этого множества более короткое обозначение $\mathcal{L}(\mathbb{R}^n)$.

Для любых двух линейных операторов $\varphi, \psi \in \mathcal{L}(\mathbb{R}^n)$ определены их сумма $\varphi + \psi$ и композиция $\varphi \circ \psi$ (играющая роль произведения), причем, как показано в предыдущем пункте, выполнены все законы, позволяющие назвать множество $\mathcal{L}(\mathbb{R}^n)$ кольцом (см. п. 14.1).

При отображении μ, описанном в п. 15.3, каждому линейному оператору из $\mathbb{R}^n$ в $\mathbb{R}^n$ будет соответствовать квадратная матрица размера $n \times n$. В силу теоремы 15.1, это отображение будет изоморфизмом кольца $\mathcal{L}(\mathbb{R}^n)$ на кольцо квадратных матриц

$$L(n, \mathbb{R}) = \mathrm{Mat}(n, n; \mathbb{R}).$$

Описанное выше кольцо линейных операторов, так же как и кольцо квадратных матриц, является (при $n \geqslant 2$) некоммутативным. Но в нем, как и в кольце квадратных матриц, существуют элементы, коммутирующие со всеми элементами. Такими элементами будут так называемые *скалярные операторы* вида $\lambda\varepsilon$ ($\lambda \in \mathbb{R}$), соответствующие скалярным матрицам λE.

Подобно тому как в определении 14.1 в кольце $L(n, \mathbb{R})$ всех квадратных $(n \times n)$-матриц выделялись обратимые матрицы, образующие (см. замечание 14.3) группу $GL(n, \mathbb{R})$ относительно умножения, так и в изоморфном кольце $\mathcal{L}(\mathbb{R}^n)$ линейных операторов выделяются *обратимые линейные операторы*.

Определение 15.5. Линейный оператор $\varphi \in \mathcal{L}(\mathbb{R}^n)$ называется *обратимым*, если существует оператор $\psi \in \mathcal{L}(\mathbb{R}^n)$ такой, что

$$\psi \circ \varphi = \varphi \circ \psi = \varepsilon. \tag{15.19}$$

Замечание 15.6. Так же как и в случае с матрицами, если оператор ψ, удовлетворяющий (15.19), существует, то — только один. Он называется оператором, обратным к φ, и обозначается $\psi = \varphi^{-1}$.

Заметим, что обратимость φ как элемента кольца равносильна обратимости отображения $\varphi : \mathbb{R}^n \to \mathbb{R}^n$.

В самом деле, соотношения (15.19) влекут, очевидно, биективность φ (и ψ); с другой стороны, если отображение φ биективно, то оно и обратное отображение ψ связаны соотношением (15.19), причем отображение ψ тоже будет линейным (соответствующее рассуждение — в более общей ситуации — проведено в начале доказательства теоремы 15.1).

Введем обозначение $\mathcal{GL}(\mathbb{R}^n)$ для множества обратимых линейных операторов, действующих в пространстве $\mathbb{R}^n$. В силу теоремы 15.1, линейный оператор обратим тогда и только тогда, когда обратима соответствующая ему квадратная матрица, множество обратимых операторов $\mathcal{GL}(\mathbb{R}^n)$ является группой относительно композиции, причем эта группа изоморфна группе обратимых матриц $GL(n, \mathbb{R})$. Имеют место свойства обратимых операторов, аналогичные свойствам обратимых матриц, перечисленным в предложении 14.1. Например, аналогом формулы (14.2) будет следующая формула:

$$(\varphi \circ \psi)^{-1} = \psi^{-1} \circ \varphi^{-1}. \tag{15.20}$$

15.6. Линейные эпиморфизмы и линейные мономорфизмы. Равносильность мономорфности и эпиморфности для эндоморфизмов. Данный пункт можно при первом чтении пропустить без ущерба для понимания "основного корпуса" материала. Здесь автор обращается к читателям, обнаружившим в себе некоторую склонность к абстрактным вопросам общей алгебры.

Для начала напомним из курса введения в математику понятия инъективного отображения и сюръективного отображения.

Рассмотрим некоторое отображение

$$\varphi : X \to Y$$

множества X в множество Y; введем обозначение

$$Y' = \varphi(X) = \{\varphi(x) : x \in X\}$$

для *образа* отображения φ.

Отображение φ называется *сюръективным* (или, как еще говорят, "отображением *на*" множество Y), если $Y' = Y$.

Для сюръективного отображения φ можно построить *обратное справа* отображение

$$\chi : Y \to X,$$

такое, что

$$\varphi \circ \chi = \varepsilon_Y.$$

В самом деле, для любого элемента $y \in Y$ найдется элемент $x \in X$ такой, что $\varphi(x) = y$. Выбирая для каждого y один (любой) из таких элементов x, мы получим требуемое отображение χ.

В другую сторону, из существования отображения χ, обратного справа для отображения φ, следует сюръективность отображения φ. Это вытекает из равенства $\varphi(\chi(y)) = y$, справедливого для любого $y \in Y$: в элемент y под действием отображения φ переходит элемент $\chi(y) \in X$.

Таким образом, сюръективность отображения равносильна его обратимости справа.

Отображение φ называется *инъективным* (или, как еще говорят, *взаимно однозначным* отображением множества X *в* множество Y), если под действием отображения φ различные элементы множества X переходят в различные элементы множества Y. Этот факт можно также выразить импликацией:

$$(\varphi(u) = \varphi(v)) \Rightarrow (u = v)$$

для любых $u, v \in X$.

Если отображение φ инъективно, то существует такое отображение

$$\psi' : Y' \to X;\ Y' = \varphi(X),$$

что $\psi' \circ \varphi = \varepsilon_X$; достаточно для любого $y' \in Y'$ положить $\psi'(y') = x$, где x — такой (однозначно определенный) элемент множества X, который переходит в y' под действием φ. Отображение ψ' допускает *продолжение* с (возможно, более узкого) подмножества $Y' \subseteq Y$ на все множество Y, т. е. существует отображение

$$\psi : Y \to X,$$

совпадающее с отображением ψ' на Y'. Для этого достаточно любой элемент $y \in Y \setminus Y'$ отправить в произвольный элемент множества X (можно, например, для любого $y \in Y \setminus Y'$ положить $\psi(y) = x_0$, где x_0 — фиксированный элемент X).

Так построенное отображение ψ будет *левым обратным* для данного отображения φ, т. е. будет выполняться равенство

$$\psi \circ \varphi = \varepsilon_X.$$

В другую сторону, если отображение φ обратимо слева, то оно инъективно. В самом деле, применяя к обеим частям условия $\varphi(u) = \varphi(v)$ отображение ψ, обратное слева для отображения φ, мы получим $u = v$.

Таким образом, инъективность отображения равносильна его обратимости слева.

Отображение φ называется *биективным* (или взаимно однозначным отображением множества X *на* множество Y), если оно сюръективно и инъективно.

Отображение φ является биективным тогда и только тогда, когда оно обратимо как справа, так и слева, причем из описанного выше построения отображений χ и ψ (*односторонних обратных* для φ) ясно, что в данном случае эти два отображения совпадают между собой: $\chi = \psi$. Можно утверждать, что имеется "настоящее" (*двустороннее*) обратное отображение ψ, удовлетворяющее двум условиям:

$$\varphi \circ \psi = \varepsilon_Y;\ \psi \circ \varphi = \varepsilon_X.$$

Таким образом, биективность отображения равносильна его обратимости.

Рассмотрим теперь отображение $\varphi : X \to X$ множества X в себя. Это отображение будет биективным тогда и только тогда, когда для него существует обратное отображение $\psi : X \to X$, которое должно удовлетворять условиям

$$\varphi \circ \psi = \psi \circ \varphi = \varepsilon_X.$$

Легко убедиться на примерах, что из односторонней обратимости, вообще говоря, не следует настоящая (двусторонняя) обратимость.

[Рассмотрите два отображения φ и ψ множества $\mathbb{N}$ натуральных чисел в себя, заданные формулами $\varphi(n) = n+1$; $n \in \mathbb{N}$ и $\psi(n) = n-1$

(для $n \geqslant 2$); $\psi(1) = 1$. Отображение φ инъективно, но не сюръективно; отображение ψ сюръективно, но не инъективно, причем ψ служит левым обратным для φ, а φ служит правым обратным для ψ.]

Замечание 15.7. В следующем параграфе мы сосредоточимся на изучении отображений конечных множеств. В этой ситуации односторонняя обратимость отображения конечного множества X в себя уже влечет настоящую обратимость (биективность) этого отображения.

А пока мы перейдем из области теории множеств в область алгебры: вместо произвольных множеств и произвольных отображений будем рассматривать арифметические линейные пространства и линейные отображения. (Точнее выражаясь, мы возвращаемся в область *линейной* алгебры. Можно было бы рассматривать и другие алгебраические объекты, например кольца, и гомоморфизмы этих объектов; см. замечание 15.1.)

В общей алгебре применяется особая терминология: линейные отображения, как уже говорилось в указанном выше замечании, именуются линейными *гомоморфизмами*; сюръективные и инъективные гомоморфизмы получают специальные названия.

Определение 15.6. Сюръективные гомоморфизмы называются *эпиморфизмами*; инъективные гомоморфизмы называются *мономорфизмами*.

Определение 15.7. Гомоморфизм, для которого существует обратное отображение, также являющееся гомоморфизмом, называется *изоморфизмом*.

Замечание 15.8. Это понятие уже встречалось нам — см. замечание 15.1 — причем даже в более общей ситуации. Далее будет доказано, что если гомоморфизм является одновременно эпиморфизмом и мономорфизмом, т. е. если он является биективным и, следовательно, обратимым отображением, то обратное к нему отображение также является гомоморфизмом. (См. ниже вывод 3, а также сделанное ранее замечание 15.6, на которое этот вывод будет опираться.)

Таким образом, фактически окажется, что понятие изоморфизма будет совпадать с понятием биективного гомоморфизма.

Согласно теореме 15.1, каждый линейный гомоморфизм

$$\varphi : \mathbb{R}^n \to \mathbb{R}^m$$

однозначно определяется своей матрицей A (размера $m \times n$), т. е. может быть выражен формулой

$$\varphi(\bar{x}) = A \cdot \bar{x};\ \bar{x} \in \mathbb{R}^n.$$

Проанализируем вопрос о том, когда гомоморфизм φ является эпиморфизмом (мономорфизмом, изоморфизмом).

Прежде всего опишем образ $\varphi(\mathbb{R}^n)$ данного гомоморфизма (и введем для него обозначение R_φ). Этот образ состоит из всевозможных векторов вида $\bar{y} = \varphi(\bar{x})$, где $\bar{x}$ пробегает все пространство $\mathbb{R}^n$. Но всякий вектор $\bar{x} \in \mathbb{R}^n$ может быть разложен по естественному базису $\mathcal{E}_n = \{\bar{e}_1, \bar{e}_2, ..., \bar{e}_n\}$:

$$\bar{x} = \sum_{j=1}^{n} x_j \bar{e}_j.$$

Следовательно, в силу линейности φ, всякий вектор из R_φ будет иметь вид

$$\bar{y} = \sum_{j=1}^{n} x_j \bar{a}_j,$$

где $\bar{a}_j = \varphi(\bar{e}_j)$ есть не что иное, как векторы-столбцы матрицы A. Таким образом, вектор $\bar{y}$ должен быть линейной комбинацией столбцов матрицы A или, другими словами, должен принадлежать линейной оболочке

$$R_A = \langle \bar{a}_1, \bar{a}_2, ..., \bar{a}_n \rangle$$

векторов-столбцов матрицы A. (Линейная оболочка R_A уже рассматривалась в п. 13.1, она является некоторым линейным подпространством в пространстве $\mathbb{R}^m$. Подпространство R_A называлось образом матрицы A.) Выше нами установлено включение $R_\varphi \subseteq R_A$.

Для доказательства обратного включения предположим, что вектор $\bar{y}$ принадлежит подпространству R_A. Этот вектор представляется в виде линейной комбинации векторов $\bar{a}_j$ $(j = 1, ..., n)$. Взяв линейную комбинацию векторов $\bar{e}_j$ (с теми же скалярными коэффициентами), мы построим некоторый вектор $\bar{x} \in \mathbb{R}^n$, переходящий в $\bar{y}$ под действием гомоморфизма φ. Значит, $\bar{y} \in R_\varphi$.

Теперь полностью доказано, что образ R_φ линейного гомоморфизма φ совпадает с образом R_A матрицы A, соответствующей этому гомоморфизму (т. е. с линейной оболочкой векторов-столбцов матрицы A).

Для эпиморфности φ необходимо и достаточно, чтобы для любого вектора $\bar{b} \in \mathbb{R}^m$ существовал такой вектор $\bar{x} \in \mathbb{R}^n$, что $\varphi(\bar{x}) = \bar{b}$, т. е. матрица A, соответствующая φ, должна удовлетворять следующему условию:

(epi) с.л.у. $A \cdot \bar{x} = \bar{b}$ имеет хотя бы одно решение (т. е. совместна) при любой правой части $\bar{b} \in \mathbb{R}^m$.

С другой стороны, эпиморфность равносильна тому, что образ R_φ совпадает со всем пространством $\mathbb{R}^m$, или, в терминах матрицы A : $R_A = \mathbb{R}^m$. Вспоминая свойства размерности (см. предложение 11.2), мы можем заключить, что это утверждение равносильно равенству $\dim(R_A) = m$, или (см. п. 12.2) равенству $\operatorname{rank}(A) = m$.

В ы в о д 1: линейный гомоморфизм является эпиморфизмом тогда и только тогда, когда ранг соответствующей матрицы равен числу строк этой матрицы.

В частности, при $n < m$ не существует линейных эпиморфизмов из $\mathbb{R}^n$ на $\mathbb{R}^m$.

Теперь перейдем на "язык односторонних обратных". Если существует линейный гомоморфизм $\chi : \mathbb{R}^m \to \mathbb{R}^n$, обратный справа к гомоморфизму φ, т. е. удовлетворяющий условию $\varphi \circ \chi = \varepsilon_m$, где ε_m — тождественный гомоморфизм пространства $\mathbb{R}^m$, то данный гомоморфизм φ будет сюръективным отображением (это усматривается точно так же, как и в случае произвольных множеств и их отображений). Следовательно, φ будет эпиморфизмом.

А вот обратное утверждение менее тривиально. Дело в том, что, хотя сюръективность данного гомоморфизма влечет (как и в случае множеств) существование обратного справа *отображения*, теперь нам требуется обратное справа *линейное* отображение. Существование такого отображения мы докажем позже, во втором семестре, а пока ограничимся следующим дополнением к выводу 1.

В ы в о д 1 а: если для линейного гомоморфизма существует обратный справа линейный гомоморфизм, то данный гомоморфизм является эпиморфизмом.

Обратимся к исследованию условий мономорфности линейного гомоморфизма.

Гомоморфизм φ является мономорфизмом, если под действием φ различные векторы пространства $\mathbb{R}^n$ переходят в различные векторы пространства $\mathbb{R}^m$. В частности, в нулевой вектор $\bar{0} \in \mathbb{R}^m$ должен переходить *только* нулевой вектор $\bar{0} \in \mathbb{R}^n$. Последнее условие является не только необходимым, но и достаточным для мономорфно-

сти φ. В самом деле, если в нулевой вектор переходит лишь нулевой вектор, то из равенства $\varphi(\bar{u}) = \varphi(\bar{v})$ $(\bar{u}, \bar{v} \in \mathbb{R}^n)$, в силу линейности φ, следует:

$$\varphi(\bar{u} - \bar{v}) = \varphi(\bar{u}) - \varphi(\bar{v}) = \bar{0},$$

т. е. вектор $\bar{u} - \bar{v}$ отображается в нулевой вектор и, следовательно, сам равняется нулю; значит, $\bar{u} = \bar{v}$.

В терминах матрицы A полученное выше условие мономорфности выглядит следующим образом:

(mono) однородная с.л.у. $A \cdot \bar{x} = \bar{0}$ имеет только тривиальное (нулевое) решение.

Введем понятие *ядра* (или *нуль-пространства*) для линейного гомоморфизма .

Определение 15.8. Ядром линейного гомоморфизма φ называется множество

$$L^0_\varphi = \{\, \bar{x} \in \mathbb{R}^n \,:\, \varphi(\bar{x}) = \bar{0} \,\}.$$

Очевидно, это множество совпадает с ядром L^0_A матрицы A (определенным в п. 13.1), которое, как известно, является некоторым линейным подпространством в пространстве $\mathbb{R}^n$.

Приведенное выше условие (mono) сводится к тому, что подпространство L_A является нулевым: $L^0_A = O$. Это, в свою очередь, равносильно (см. предложение 12.2) равенству нулю размерности $\dim(L^0_A) = 0$, или, в силу предложения 11.3, равенству $\mathrm{rank}(A) = n$.

В ы в о д 2: линейный гомоморфизм является мономорфизмом тогда и только тогда, когда ранг соответствующей матрицы равен числу столбцов этой матрицы.

В частности, при $n > m$ не существует линейных мономорфизмов из $\mathbb{R}^n$ в $\mathbb{R}^m$.

Снова обсудим версию полученного вывода, выраженную с помощью односторонних обратных гомоморфизмов.

Если для гомоморфизма φ существует обратный слева гомоморфизм $\psi : \mathbb{R}^m \to \mathbb{R}^n$; $\psi \circ \varphi = \varepsilon_n$, то, рассуждая так же, как в случае произвольных множеств и их отображений, мы заключим, что φ является инъективным отображением и, следовательно, мономорфизмом. И опять для доказательства обратного утверждения потребуется более "тонкое" исследование, которое мы пока отложим. А сейчас приведем следующее дополнение к выводу 2.

В ы в о д 2 а: если для линейного гомоморфизма существует обратный слева линейный гомоморфизм, то данный гомоморфизм является мономорфизмом.

Переходим к исследованию вопроса о том, когда линейный гомоморфизм является изоморфизмом.

Из сделанных выше выводов 1 и 2 сразу следует, что при $m \neq n$ не существует изоморфизма между пространствами $\mathbb{R}^n$ и $\mathbb{R}^m$. Поэтому рассмотрим случай $m = n$, т. е. будем изучать линейные *эндоморфизмы* $\varphi : \mathbb{R}^n \to \mathbb{R}^n$.

Соответствующая такому эндоморфизму матрица A будет квадратной матрицей (размера $n \times n$).

Из тех же выводов 1 и 2 теперь вытекает, что если эндоморфизм φ является эпиморфизмом, то ранг его матрицы A равняется числу строк, а значит, и числу столбцов, откуда следует, что φ является мономорфизмом. В частности, φ оказывается биективным отображением. Поэтому существует обратное отображение ψ, которое, в силу замечания 15.6, также будет линейным отображением, т. е. эндоморфизмом. Тем самым доказано, что φ является обратимым эндоморфизмом, т. е. изоморфизмом.

Совершенно аналогично доказывается, что мономорфность φ влечет его изоморфность.

Кроме того, вспомним, что условие $\operatorname{rank}(A) = n$ есть не что иное, как условие невырожденности матрицы A (см. определение 14.4).

Таким образом, справедливо следующее утверждение.

В ы в о д 3: для линейного эндоморфизма свойства мономорфности, эпиморфности и изоморфности равносильны между собой и равносильны тому, что соответствующая эндоморфизму квадратная матрица невырождена.

Отчасти последний вывод для нас не нов, поскольку, в силу теоремы 14.1, обратимость квадратной матрицы равносильна обратимости (и значит — биективности) соответствующего линейного оператора.

Добавления 1а и 2а к первым двум выводам позволяют вывести следующее добавление к выводу 3.

В ы в о д 3 а: для линейного эндоморфизма равносильны обратимость справа, обратимость слева и двусторонняя обратимость.

В самом деле, если эндоморфизм φ обратим, скажем, справа, то, в силу вывода 1а, он является эпиморфизмом и, в силу вывода 3, изоморфизмом, а значит, он является обратимым (имеет двусторонний

обратный). Аналогично обратимость слева также влечет двустороннюю обратимость.

Замечание 15.9. Из последнего вывода и из теоремы 15.1 вытекает утверждение, аналогичное этому выводу, но относящееся к матрицам: если для квадратной матрицы A существует обратная справа (слева) матрица B, т. е. такая, что $A \cdot B = E$ (соответственно $B \cdot A = E$), то матрица B является двусторонней обратной к матрице A.

Позже (см. замечание 28.1) это утверждение о матрицах получит другое обоснование, использующее теорию определителей.

Соберем теперь наши выводы 1 — 3 и дополнения к ним в одну теорему.

Теорема 15.2. *Рассмотрим гомоморфизм $\varphi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ и соответствующую матрицу $A \in \mathrm{Mat}(m, n; \mathbb{R})$.*

1. Гомоморфизм φ является эпиморфизмом тогда и только тогда, когда $n \geqslant m$ и $\mathrm{rank}(A) = m$. Гомоморфизм, обратимый справа, является эпиморфизмом.

2. Гомоморфизм φ является мономорфизмом тогда и только тогда, когда $n \leqslant m$ и $\mathrm{rank}(A) = n$. Гомоморфизм, обратимый слева, является мономорфизмом.

3. Гомоморфизм φ является изоморфизмом тогда и только тогда, когда $n = m$ и $\mathrm{rank}(A) = n$, т. е. тогда и только тогда, когда он является эндоморфизмом и его матрица является невырожденной.

Для эндоморфизмов условия эпиморфности, мономорфности и изоморфности, а также условия обратимости справа, обратимости слева и двусторонней обратимости — равносильны.

Доказательство проведено выше, перед формулировкой. □

Глава 3

ТЕОРИЯ ПЕРЕСТАНОВОК

§ 16. Перестановки и алгебраические действия над ними

16.1. Биекции конечного множества. Сейчас нам снова предстоит заняться вопросами, относящимися к теории множеств и *общей алгебре*.

Рассмотрим конечное множество X, состоящее из n элементов. Поскольку для математика природа этих элементов не важна (он же не ботаник!), то занумеруем элементы множества X натуральными числами и будем считать, что множество X состоит из этих номеров:

$$X = \{1, 2, \ldots, n\}. \tag{16.1}$$

Определение 16.1. *Перестановкой степени* n будем называть биективное отображение (биекцию) n-элементного множества X на себя:

$$\varphi : X \longrightarrow X;\ i \mapsto \varphi(i);\ i = 1, 2, ..., n.$$

Множество всех перестановок степени n будем обозначать символом S_n и для произвольной перестановки $\varphi \in S_n$ будем применять *двустрочную запись*:

$$\varphi = \begin{pmatrix} 1 & 2 & \ldots & n \\ \varphi(1) & \varphi(2) & \ldots & \varphi(n) \end{pmatrix}, \tag{16.2}$$

указывая во второй строке, куда переходит каждый элемент множества X (представленного в первой строке).

Замечание 16.1. Подчеркнем, что перестановки суть отображения и равенство двух перестановок $\varphi, \psi \in S_n$ есть равенство отображений:

$$[\,\varphi = \psi\,] \iff (\forall i = 1, ..., n)[\varphi(i) = \psi(i)].$$

Символы в первой строке записи (16.2) разрешается как угодно переставлять, меняя при этом (соответствующим образом) порядок элементов во второй строке. Скажем, элемент

$$\varphi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 4 & 5 & 1 \end{pmatrix} \in S_5$$

можно записать в виде

$$\varphi = \begin{pmatrix} 5 & 1 & 3 & 4 & 2 \\ 1 & 3 & 4 & 5 & 2 \end{pmatrix}.$$

Замечание 16.2. Широко распространен другой термин — "подстановки" (вместо "перестановки"), но в учебнике [2] убедительно показана бóльшая адекватность термина "перестановки". При чтении других книг по алгебре (например, [3]) следует иметь в виду указанное расхождение в терминологии.

Замечание 16.3. Для отображения $\varphi : X \to X$ конечного множества X в себя справедливы эквивалентности

$$(\varphi \text{ — инъективно}) \iff (\varphi \text{ — сюръективно}) \iff (\varphi \text{ — биективно}).$$

Попробуйте сами доказать эти утверждения, предварительно заметив, что количество элементов в образе $\varphi(X)$ не превышает количества элементов X. (Можно обратиться также к замечанию 15.7.)

Подсчитаем теперь, сколько всего существует перестановок степени n.

Во второй строке записи (16.2) все элементы попарно различны; они такие же, как в первой строке, но расположены в произвольном порядке.

Имеется n способов выбора элемента $\varphi(1)$. Если этот элемент уже выбран, то для выбора $\varphi(2)$ остается $n-1$ способов. Каждый из способов выбора первого элемента может сочетаться с любым способом выбора второго. Поэтому два элемента $\varphi(1)$, $\varphi(2)$ могут быть выбраны $n \cdot (n-1)$ способами. Каждый из этих способов может сочетаться с любым из способов выбора $\varphi(3)$ и т. д. Если элементы $\varphi(1)$, $\varphi(2)$, $\ldots$, $\varphi(n-1)$ уже выбраны, то элемент $\varphi(n)$ определяется однозначно (один способ выбора). Итоговая формула для определения количества перестановок степени n:

$$n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1 = n!$$

Замечание 16.4. Мощность множества X (см. п. 11.1) обычно обозначается $|X|$. Это очень неудобно. "Палочки" сильно перегружены; они обозначают как модуль числа и длину вектора, так и определитель квадратной матрицы (о котором речь пойдет в следующей главе). Но ничего лучшего пока никто не предложил. Придется и нам смириться с записью

$$|S_n| = n!$$

Пример 16.1. Приведем списки всех перестановок ctпени $n \leqslant 4$. Обратите внимание на *лексикографический* (словарный) способ их упорядочения.

$$S_1 = \left\{ \varepsilon = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\};$$

$$S_2 = \left\{ \varepsilon = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}; \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \right\};$$

$$S_3 = \left\{ \varepsilon = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}; \right.$$
$$\left. \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \right\};$$

$$S_4 = \left\{ \varepsilon = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix}; \right.$$
$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 2 & 3 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix};$$
$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix};$$
$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix};$$
$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix};$$
$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix};$$
$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix};$$
$$\left. \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 3 & 1 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix}; \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} \right\}.$$

16.2. Умножение перестановок. Рассмотрим множество S_n перестановок степени n. Действия над перестановками — это действия над отображениями (в данном случае — биективными отображениями множества (16.1) на себя).

Умножение перестановок — это не что иное, как их композиция: по двум перестановкам $\varphi, \psi \in S_n$ определяется их произведение $\varphi\psi = \varphi \circ \psi$. (Ради краткости записи кружок $\circ$, знак композиции, обычно опускается.) Ясно, что композиция двух биекций сама является биекцией, т. е. $\varphi\psi \in S_n$.

Пример 16.2. Покажем один из возможных способов "оформления" умножения перестановок. При минимальном навыке это оформление делается в уме.

$$\varphi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}; \quad \psi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}$$

$$\begin{array}{cccccc} & 1 & 2 & 3 & 4 & \qquad & & 1 & 2 & 3 & 4 \\ \varphi & \downarrow & \downarrow & \downarrow & \downarrow & & \psi & \downarrow & \downarrow & \downarrow & \downarrow \\ & 4 & 1 & 3 & 2 & & & 3 & 4 & 1 & 2 \\ \psi & \downarrow & \downarrow & \downarrow & \downarrow & & \varphi & \downarrow & \downarrow & \downarrow & \downarrow \\ & 2 & 3 & 1 & 4 & & & 3 & 2 & 4 & 1 \end{array}$$

$$\psi\varphi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}; \quad \varphi\psi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix}$$

Надеемся, вы обратили внимание на тот факт, что умножение (композиция) перестановок не удовлетворяет коммутативному закону: перестановки-сомножители, вообще говоря, нельзя переставлять.

Замечание 16.5. Как и в общем случае действий над отображениями (см. замечание 15.3), непримиримые партии "правых" и "левых" используют противоположный порядок действия сомножителей в произведении перестановок. Мы, следуя [2], придерживаемся *левосторонней* записи: в произведении $\varphi\psi$ сначала действует ψ, а потом φ. В курсе [3] используется правосторонняя запись.

16.3. Тождественная (единичная) перестановка. Символом ε будет обозначаться (фактически это уже делалось в примере 16.1) *тождественная* (*единичная*) перестановка

$$\varepsilon = \begin{pmatrix} 1 & 2 & \cdots & n \\ 1 & 2 & \cdots & n \end{pmatrix} \in S_n. \tag{16.3}$$

При умножении ε "ведет себя как единица":

$$\varphi\varepsilon = \varepsilon\varphi = \varphi \tag{16.4}$$

для любой перестановки $\varphi \in S_n$.

16.4. Обратная перестановка. Всякая перестановка $\varphi \in S_n$ имеет *обратную*

$$\psi = \varphi^{-1} \in S_n,$$

являющуюся биекцией, обратной к биекции φ (см. п. 15.6).

Если перестановка φ задана двустрочной записью (16.2), то перестановку φ^{-1} можно задать двустрочной записью

$$\varphi^{-1} = \begin{pmatrix} \varphi(1) & \varphi(2) & \dots & \varphi(n) \\ 1 & 2 & \dots & n \end{pmatrix}, \tag{16.5}$$

где элементы в первой строке можно *упорядочить* (соответствующим образом переставляя элементы второй строки).

Перестановки φ и $\psi = \varphi^{-1}$ связаны соотношениями

$$\varphi\psi = \psi\varphi = \varepsilon, \tag{16.6}$$

фактически уже встречавшимися нам (в более общей ситуации) в п. 15.6.

Пример 16.3. Вычисление обратной перестновки:

$$\varphi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 1 & 5 & 6 & 2 \end{pmatrix} \in S_6;$$

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

16.5. Группа перестановок. Нам знакомо уже понятие *группы* (см. замечание 14.1). Рассматривались также конкретные примеры групп: группа обратимых матриц заданного размера, группа обратимых линейных операторов, действующих в заданном пространстве.

Умножение (композиция) перестановок удовлетворяет ассоциативному закону; в множестве S_n существует нейтральный элемент ε и для любого элемента существует обратный элемент. Поэтому можно сделать следующий

В ы в о д: Множество S_n перестановок фиксированной степени n является группой относительно умножения.

16.6. Область действия перестановки. Независимые перестановки

Определение 16.2. *Областью действия* перестановки $\varphi \in S_n$ называется множество номеров $i \in X$, "реально перемещаемых" перестановкой φ. Обозначение:

$$D_\varphi = \{\, i \in X \,:\, \varphi(i) \neq i \,\}.$$

Заметим, что у тождественной перестановки область действия оказывается пустой: $D_\varepsilon = \varnothing$.

Дополнение $X \setminus D_\varphi$ множества D_φ состоит из номеров, неподвижных под действием φ. Множество D_φ *инвариантно* относительно φ, т. е. $i \in D_\varphi$ влечет $\varphi(i) \in D_\varphi$.

В самом деле, если бы $\varphi(i) \notin D_\varphi$, то элемент $\varphi(i)$ был бы неподвижен под действием φ, т. е. выполнялось бы равенство $\varphi(\varphi(i)) = \varphi(i)$, которое, в силу биективности φ, влекло бы $\varphi(i) = i$, что противоречит предположению $i \in D_\varphi$.

Определение 16.3. Перестановки $\varphi, \psi \in S_n$ называются *независимыми*, если

$$D_\varphi \cap D_\psi = \varnothing.$$

Пример 16.4. Рассмотрим две перестановки:

$$\varphi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 5 & 3 & 4 & 1 & 6 \end{pmatrix};\ \psi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 6 & 4 & 5 & 3 \end{pmatrix}.$$

Их области действия

$$D_\varphi = \{1, 2, 5\};\ D_\psi = \{3, 6\}$$

не пересекаются; значит, φ и ψ независимы.

16.7. Коммутирующие перестановки. В примере 16.2 мы убедились в том, что умножение перестановок не удовлетворяет коммутативному закону, т. е., вообще говоря, $\varphi\psi \neq \psi\varphi$. Однако для некоторых $\varphi, \psi \in S_n$ равенство $\varphi\psi = \psi\varphi$ может выполняться. В связи с этим дается

Определение 16.4. Перестановки $\varphi, \psi \in S_n$ называются *коммутирующими*, если $\varphi\psi = \psi\varphi$.

Очевидно, единичная перестановка коммутирует со всеми перестановками.

Выведем далее достаточное условие коммутирования двух перестановок.

Предложение 16.1. *Две независимые перестановки коммутируют.*

Доказательство. Если $\varphi, \psi \in S_n$ независимы, то множество $X = \{1, 2, ..., n\}$ представляется в виде объединения трех попарно не пересекающихся подмножеств:

$$X = D_\varphi \cup D_\psi \cup C,$$

причем φ реально переставляет лишь номера из D_φ и тождественна на $D_\psi \cup C$, а ψ переставляет элементы D_ψ и тождественна на $D_\varphi \cup C$.

Докажем, что $\varphi\psi = \psi\varphi$, т. е. для любого $i \in X$

$$\varphi(\psi(i)) = \psi(\varphi(i)). \tag{16.7}$$

Если $i \in D_\varphi$, то $\varphi(i) \in D_\varphi$ и $\psi(\varphi(i)) = \varphi(i)$, а с другой стороны, $\psi(i) = i$, и следовательно, $\varphi(\psi(i)) = \varphi(i)$. Таким образом, равенство (16.7) выполняется.

Если $i \in D_\psi$, то рассуждение совершенно аналогично.

И уж совсем тривиальным образом равенство (16.7) выполняется в случае $i \in C$: на таких i тождественны обе перестановки. □

16.8. Отображения левого (правого) сдвига и обращения на группе перестановок. Рассмотрим группу перестановок S_n и следующие три типа отображений группы S_n в себя.

1. *Левые сдвиги.* Зафиксируем перестановку $\alpha \in S_n$ и рассмотрим отображение

$$l_\alpha : S_n \longrightarrow S_n;\ l_\alpha(\varphi) = \alpha\varphi;\ \varphi \in S_n, \tag{16.8}$$

состоящее в левом умножении всех перестановок на фиксированную перестановку α.

2. *Правые сдвиги* определяются аналогично:

$$r_\alpha : S_n \longrightarrow S_n;\ r_\alpha(\varphi) = \varphi\alpha;\ \varphi \in S_n. \tag{16.9}$$

3. *Отображение обращения* задается формулой

$$\nu : S_n \longrightarrow S_n;\ \nu(\varphi) = \varphi^{-1};\ \varphi \in S_n, \tag{16.10}$$

т. е. сопоставляет каждой перестановке степени n обратную к ней перестановку.

Предложение 16.2. *Левые (правые) сдвиги и отображение обращения являются биекциями.*

Доказательство. Как известно (см. п. 15.6), биективность отображения равносильна его обратимости. Для исследуемых отображений l_α, r_α и ν легко указать обратные к ним отображения:

$$(l_\alpha)^{-1} = l_{\alpha^{-1}};\ (r_\alpha)^{-1} = r_{\alpha^{-1}};\ \nu^{-1} = \nu. \tag{16.11}$$

Скажем, первая из формул (16.11) немедленно следует из ассоциативного закона для умножения перестановок и из свойств обратной перестановки: если перестановку φ сначала умножить слева на α, а потом результат умножить слева на α^{-1}, то мы вернемся к исходной перестановке φ и т. д. "Самообратность" отображения ν равносильна формуле $(\varphi^{-1})^{-1} = \varphi$. $\square$

Замечание 16.6. Доказанное выше предложение понадобится нам в п. 20.3 и в следующей главе, при изучении теории определителей.

Замечание 16.7. Разумеется, утверждение, аналогичное предложению 16.2, остается справедливым для любой группы по умножению.

§ 17. Циклические перестановки. Разложение перестановки в произведение независимых циклов

17.1. Циклы и частичные циклы

Определение 17.1. Перестановка $\varphi \in S_n$, имеющая в двустрочной записи вид

$$\varphi = \begin{pmatrix} 1 & 2 & 3 & \dots & n-1 & n \\ 2 & 3 & 4 & \dots & n & 1 \end{pmatrix}, \tag{17.1}$$

называется *циклической* (или *циклом* длины n) и обозначается

$$\varphi = (1, 2, 3, \dots, n-1, n). \tag{17.2}$$

Если $\{i_1, i_2, ..., i_s\}$ — какое-либо s-элементное ($2 \leqslant s \leqslant n$) подмножество множества (16.1) (номера попарно различны, но не обязательно идут по порядку), то символом $(i_1, i_2, ..., i_s)$ обозначается перестановка

$$(i_1, i_2, ..., i_s) = \left(\begin{array}{ccccc|cccc} i_1 & i_2 & \dots & i_{s-1} & i_s & j_1 & j_2 & \dots & j_{n-s} \\ i_2 & i_3 & \dots & i_s & i_1 & j_1 & j_2 & \dots & j_{n-s} \end{array}\right), \tag{17.3}$$

где "в первой зоне" показаны "реально переставляемые" по циклу номера (s штук), а во "второй зоне" — "неподвижные" номера (оставшиеся $n - s$ штук). Перестановка (17.3) называется *(частичным) циклом* длины s; перестановку (17.2) можно назвать *полным циклом* длины n.

Замечание 17.1. Обозначения (17.2) и (17.3) называются *однострочной записью* для циклов.

С этими обозначениями связан ряд условностей.

Например, цикл $(5, 1, 4) \in S_6$ является перестановкой вида

$$(5,1,4) = \begin{pmatrix} \boxed{1} & 2 & 3 & \boxed{4} & \boxed{5} & 6 \\ \boxed{4} & 2 & 3 & \boxed{5} & \boxed{1} & 6 \end{pmatrix},$$

где квадратиками выделены реально переставляемые номера. Цикл $(5, 1, 4) \in S_8$ выглядит аналогично, но с добавлением еще двух неподвижных номеров (7 и 8).

Таким образом, однострочная запись цикла определяет его однозначно, если задана *степень* перестановки (не меньшая наибольшего из участвующих в цикле номеров).

Другая особенность: записи $(5, 1, 4)$, $(1, 4, 5)$ и $(4, 5, 1)$ определяют один и тот же цикл.

В общем виде правило выражается так: цикл (17.3) можно начинать с любого i_k $(k = 1, ..., s)$, т. е.

$$(i_1, i_2, \ldots, i_s) = (i_k, i_{k+1}, \ldots, i_s, i_1, i_2, \ldots, i_{k-1}).$$

Определение 17.2. Цикл длины 2 называется *транспозицией.*

Общий вид транспозиции:

$$\tau = (i_1, i_2), \tag{17.4}$$

где $i_1 \neq i_2$, причем, в силу замечания 17.1,

$$\tau = (i_1, i_2) = (i_2, i_1).$$

Замечание 17.2. Как определить цикл длины 1, который следовало бы обозначить (i_1)?

Это, очевидно, — тождественная перестановка ε (ибо она должна оставлять неподвижными как номер i_1, так и все остальные номера). Такие циклы мы будем называть *тривиальными* и будем выбрасывать их (без ущерба для истинности формул). Обратим, однако, внимание на то, что область действия цикла (i_1) пуста, а отнюдь не совпадает с одноэлементным множеством $\{i_1\}$.

17.2. Теорема о разложении перестановки на независимые циклы. Докажем теперь, что всякую перестановку можно разложить в произведение независимых циклов.

Теорема 17.1. *Для любой перестановки $\varphi \in S_n$ существует представление в виде*

$$\varphi = \sigma_1 \sigma_2 \dots \sigma_m, \tag{17.5}$$

где перестановки σ_j $(j = 1, ..., m)$ являются попарно независимыми (частичными) циклами.

Представление (17.5) *определено однозначно с точностью до порядка сомножителей.*

Доказательство. Возьмем любой номер $i_1 \in X$ (на практике естественно брать $i_1 = 1$). Под действием φ этот номер перейдет в элемент $\varphi(i_1) \in X$. Если $\varphi(i_1) = i_1$, то первым циклом, входящим в разложение (17.5), будет тривиальный цикл

$$\sigma_1 = (i_1) = \varepsilon.$$

Если же $\varphi(i_1) = i_2 \neq i_1$, то начинается нетривиальный цикл

$$\sigma_1 = (i_1, i_2, \dots, i_{s_1}) \tag{17.6}$$

некоторой длины $s_1 \leqslant n$.

Этот цикл образуется следующим образом. Рассмотрим бесконечную последовательность элементов

$$i_1 \overset{\varphi}{\mapsto} i_2 \overset{\varphi}{\mapsto} i_3 \overset{\varphi}{\mapsto} \dots \overset{\varphi}{\mapsto} i_{k-1} \overset{\varphi}{\mapsto} i_k \overset{\varphi}{\mapsto} i_{k+1} \overset{\varphi}{\mapsto} \dots \tag{17.7}$$

Так как множество X конечно, то в последовательности (17.7) неизбежны повторения. Рассмотрим первое из них. Пусть, скажем,

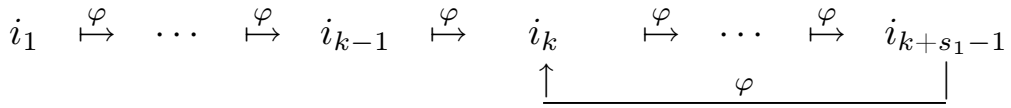
$$i_{k+s_1} = i_k; \;\; k, s_1 \geqslant 1; \tag{17.8}$$

а все элементы

$$i_1, i_2, \dots, i_{k-1}, i_k, i_{k+1}, \dots, i_{k+s_1-1}$$

попарно различны. (Другими словами, на шаге с номером $k + s_1$ обнаруживается первое повторение, т. е. совпадение с одним из предыдущих элементов, номер которого равен k.)

Так возникает цикл (может быть, "с хвостиком")



Легко, однако, убедиться в том, что на самом деле "хвостика" не будет, т. е. обязательно будет $k = 1$. Действительно, в противном случае ($k > 1$) в элемент i_k под действием φ будут переходить два (различных по предположению) элемента i_{k-1} и i_{k+s_1-1}, что противоречит биективности φ.

Итак, "зацикливание" происходит начиная с элемента i_1:

$$i_{1+s_1} = i_1;$$

далее образы $i_k = \varphi^k(i_1)$ повторяются с периодом s_1:

$$i_{k+s_1} = i_k$$

(для любого натурального k); возникает цикл (17.6) с областью действия

$$D_1 = \{i_1, i_2, \ldots, i_{s_1}\}.$$

Эта область действия инвариантна относительно φ, и на ней φ действует так же, как и циклическая перестановка σ_1.

Важно заметить, что если бы мы начали цикл с какого-то другого номера i'_1, но по ходу построения цикла в него попал бы элемент i_1, то цикл, начинающийся с i'_1, и цикл, начинающийся с i_1, были бы одинаковыми (это вытекает из замечания 17.1).

Иначе говоря, каждый элемент множества X содержится в одном и только одном (частичном) цикле относительно перестановки φ.

Образуем множество G_1 для учета просмотренных элементов. Если первый цикл σ_1 не тривиален, то положим $G_1 = D_1$, если же $\sigma_1 = (i_1)$, то возьмем $G_1 = \{i_1\}$, хотя область действия D_1 в этом случае пуста. (Используется такой термин: множество G_1 называется *орбитой* элемента i_1 под действием перестановки φ.)

Если длина первого цикла $s_1 = n$, то разложение (17.5) получено (и содержит всего один множитель). Если же $s_1 < n$, то выберем любой номер $j_1 \notin G_1$. (На практике, т. е. при *алгоритмической* организации процесса, можно в качестве j_1 выбирать наименьший из номеров, не вошедших в G_1.)

Ни один из элементов, не вошедших в G_1, не может перейти ни в какой из элементов, входящих в G_1, т. к. это противоречило бы биективности φ. Поэтому с j_1 начнется новый цикл

$$\sigma_2 = (j_1, j_2, \ldots, j_{s_2})$$

некоторой длины s_2, независимый с циклом σ_1.

Если $s_1 + s_2 < n$, то процесс продолжается аналогично, вплоть до нахождения последнего цикла

$$\sigma_m = (k_1, k_2, \ldots, k_{s_m})$$

некоторой длины s_m.

Все найденные циклы по построению попарно независимы. Объединение орбит G_j должно совпадать с множеством всех номеров:

$$G_1 \cup G_2 \cup \cdots \cup G_m = X;$$

т. е. сумма длин σ_j должна равняться степени перестановки (но это — с учетом тривиальных циклов!):

$$s_1 + s_2 + \cdots + s_m = n.$$

Каждая из орбит G_j $(j = 1, ..., m)$ инвариантна относительно φ, причем действие φ на элементах G_j совпадает с действием цикла σ_j. Далее, цикл σ_j на множестве G_i $(i \neq j)$ действует тождественно; попарная независимость циклов σ_j влечет их попарное коммутирование (см. предложение 16.1).

Докажем теперь справедливость разложения (17.5). Возьмем любой элемент (номер) $p \in X$. Перестановка φ [левая часть (17.5)] принимает на этом элементе значение $\varphi(p)$. Вычислим значение на p правой части. Элемент p принадлежит одной и только одной из орбит. Пусть, скажем, $p \in G_j$. Циклы σ_i (в силу их коммутирования) можно применять в любом порядке. Ни один из них, кроме σ_j, не сдвинет элемент p с места; и только под действием σ_j этот элемент перейдет в

$$\sigma_j(p) = \varphi(p),$$

т. е. значение правой части формулы (17.5) на произвольном элементе $p \in X$ будет таким же, как и значение левой части. Равенство перестановок (17.5) доказано.

Остается обосновать единственность полученного разложения (с точностью до порядка сомножителей).

Пусть есть еще одно разложение такого же типа, что и (17.5), и пусть p — произвольный элемент множества X.

Тогда p принадлежит области действия одного и только одного из циклов первого разложения и аналогично — одного из циклов второго разложения. Как выше объяснялось, отсюда будет следовать совпадение указанных циклов.

В качестве логического упражнения доведите до конца рассуждение, приводящее к выводу о совпадении двух разложений. (Указание: сначала объясните, почему два разложения обязательно имеют одинаковое количество сомножителей.) После этого можно будет считать, что доказательство теоремы завершено. □

Замечание 17.3. По ходу доказательства уже отмечалось, что среди сомножителей, входящих в разложение (17.5), могут встречаться циклы длины 1, равные ε. Их можно исключить и считать, что в (17.5) присутствуют лишь нетривиальные циклы $\sigma_1, \sigma_2, \ldots, \sigma_l$ (но тогда сумма их длин $s_1 + s_2 + \ldots + s_l$ не обязательно будет равняться n). Единственным исключением при таком подходе будет перестановка $\varphi = \varepsilon$; в ее разложении участвуют лишь тривиальные циклы:

$$\varepsilon = (1)(2)\ldots(n).$$

Пример 17.1. Разложим заданную перестановку на нетривиальные независимые циклы:

$$\varphi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 3 & 7 & 9 & 4 & 10 & 1 & 2 & 5 & 6 & 8 \end{pmatrix} =$$
$$= (1,3,9,6)\,(2,7)\,(4)\,(5,10,8) =$$
$$= (1,3,9,6)\,(2,7)\,(5,10,8).$$

Пример 17.2. Если данные перестановки разложены на циклы (даже не обязательно независимые), то их можно перемножить, "следя за судьбой" каждого из номеров и продвигаясь справа налево.

Пусть даны две перестановки $\varphi, \psi \in S_9$:

$$\varphi = (2,7,3,8,4)\,(1,5,2);$$

$$\psi = (8,6,1,2)\,(1,2).$$

В силу ассоциативности умножения, произведения можно записывать без скобок, указывающих на порядок действий:

$$\varphi\psi = (2,7,3,8,4)\,(1,5,2)\,(8,6,1,2)\,(1,2).$$

Найдем двустрочную запись для $\varphi\psi$.

Пусть $i = 1$. Применим самую правую перестановку — транспозицию $(1,2)$, она отобразит 1 в 2. Применим второй справа цикл $(8,6,1,2)$, он отобразит 2 в 8. Подействуем третьим справа циклом $(1,5,2)$, в нем 8 отсутствует, значит, этот цикл оставляет 8 на месте. И наконец, четвертый цикл $(2,7,3,8,4)$ отобразит 8 в 4. В итоге 1 отобразится в 4. Переходим к $i = 2$ и т. д.

Дозаполните двустрочную запись

$$\varphi\psi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & ? & ? & ? & ? & ? & ? & ? & ? \end{pmatrix}.$$

17.3. Декремент перестановки. Для всякой перестановки может быть определена особая числовая характеристика, связанная с разложением на независимые циклы.

Определение 17.3. *Декрементом* перестановки $\varphi \in S_n$ называется (целое неотрицательное) число

$$\delta(\varphi) = n - m, \tag{17.10}$$

где m — количество циклов в разложении (17.5).

Замечание 17.4. В формуле (17.10) учитываются все циклы, в том числе и тривиальные. В частности, для тождественной перестановки $\varphi = \varepsilon$ (и только для нее) имеет место равенство $\delta(\varphi) = 0$.

Замечание 17.5. Если среди циклов в разложении (17.5) имеется l нетривиальных:

$$\varphi = \sigma_1\sigma_2\ldots\sigma_l, \tag{17.11}$$

а остальные $m - l$ циклов — тривиальные, то степень перестановки n складывается из суммы длин нетривиальных циклов и количества тривиальных циклов:

$$n = \sum_{j=1}^{l} s_j + m - l,$$

и следовательно,

$$\delta(\varphi) = \sum_{j=1}^{l} s_j - l, \tag{17.12}$$

т. е. декремент перестановки φ равен разности между суммой длин нетривиальных циклов в разложении φ и количеством этих циклов.

§ 18. Степени перестановки. Порядок перестановки

18.1. Целые степени перестановки и их свойства. Пусть $\varphi \in S_n$. Возведение перестановки φ в натуральную степень определяется по индукции:

$$\varphi^1 = \varphi;\ \varphi^{s+1} = \varphi\varphi^s;\ s = 1, 2, ... \tag{18.1}$$

Отдельно определяется нулевая степень:

$$\varphi^0 = \varepsilon. \tag{18.2}$$

Отрицательные степени вводятся с помощью обратной перестановки:

$$\varphi^{-s} = (\varphi^{-1})^s;\ s = 1, 2, ... \tag{18.3}$$

Целые степени обладают двумя привычными свойствами, знакомыми по школьному курсу математики. (Но, разумеется, там они относились не к перестановкам, а к положительным действительным числам и строго не доказывались.) Вот эти свойства:

$$\varphi^s\varphi^t = \varphi^{s+t}; \tag{18.4}$$

$$(\varphi^s)^t = \varphi^{st}, \tag{18.5}$$

где φ — произвольная перестановка; s, t — любые целые числа.

Строгое доказательство формул (18.4) и (18.5) может быть получено, исходя из *ассоциативного* закона для умножения перестановок с помощью *метода математической индукции*. Если вы "дозрели" до осознания необходимости таких доказательств, то загляните в учебник [2, гл. 4, § 1, 2] (там эта формула доказывается в более общей ситуации — для элементов произвольных групп).

Замечание 18.1. Из формулы (18.4) вытекает, что степени одной и той же перестановки коммутируют:

$$\varphi^s \varphi^t = \varphi^t \varphi^s.$$

Есть еще одно привычное свойство степеней:

$$(\varphi\psi)^s = \varphi^s \psi^s;\ s \in \mathbb{Z}. \tag{18.6}$$

Но это свойство включает в себя *коммутативный* закон для умножения (точнее, оно равносильно коммутированию перестановок φ и ψ). В самом деле, при $s = 2$ формула (18.6) принимает вид

$$(\varphi\psi)^2 = \varphi^2 \psi^2.$$

Или (с учетом ассоциативности, без скобок):

$$\varphi\psi\varphi\psi = \varphi\varphi\psi\psi,$$

что (умножением слева на φ^{-1} и справа на ψ^{-1}) сводится к равносильному равенству

$$\varphi\psi = \psi\varphi. \tag{18.7}$$

Как мы помним из п. 16.7, перестановки коммутируют не всегда. Поэтому формула (18.6) выполняется (для всех $s \in \mathbb{Z}$) также не всегда, а в том и только том случае, когда выполняется (18.7).

Пример 18.1. Чтобы возвести перестановку $\varphi \in S_n$, заданную двустрочной записью, в целую положительную степень s, совсем не обязательно выписывать s сомножителей, равных φ. Достаточно по данной записи "проследить судьбу" каждого из номеров $i \in \{1, 2, ..., n\}$ за s шагов.

Пусть

$$\varphi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 2 & 4 & 6 & 3 & 1 \end{pmatrix} \in S_6$$

и требуется вычислить φ^7.

Возьмем $i = 1$; за 7 шагов получим:

$$1 \overset{\varphi}{\mapsto} 5 \overset{\varphi}{\mapsto} 3 \overset{\varphi}{\mapsto} 4 \overset{\varphi}{\mapsto} 6 \overset{\varphi}{\mapsto} 1 \overset{\varphi}{\mapsto} 5 \overset{\varphi}{\mapsto} 3,$$

таким образом, $\varphi^7(1) = 3$. Для неподвижного номера $i = 2$ сразу видим: $\varphi^7(2) = 2$. Заполните до конца двустрочную запись:

$$\varphi^7 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & ? & ? & ? & ? \end{pmatrix}.$$

При возведении в отрицательную степень надо начинать с номера i во *второй* строке и идти "против стрелок". Доведите до конца вычисления:

$$\varphi^{-7} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ ? & 2 & 1 & ? & ? & ? \end{pmatrix}.$$

18.2. Порядок перестановки. Поскольку множество S_n всех перестановок степени n содержит лишь конечное число (а именно $n!$) элементов, то среди степеней $\{\varphi^k\}_{k\in\mathbb{Z}}$ заданной перестановки φ обязательно будут одинаковые. Следовательно, найдутся $k, l \in \mathbb{Z}$ (пусть, скажем, $k < l$), такие, что $\varphi^k = \varphi^l$.

Домножим последнее равенство на φ^{-k} (не важно, с какой стороны; см. замечание 18.1). С учетом свойства (18.4) получим:

$$\varphi^{l-k} = \varepsilon.$$

Таким образом, найдется *положительная* степень перестановки φ, равная тождественной перестановке ε. Можно выбрать наименьшую из таких степеней.

Определение 18.1. Наименьшая из целых положительных степеней s, таких, что $\varphi^s = \varepsilon$, называется *порядком* перестановки $\varphi \in S_n$ и обозначается

$$o(\varphi) = \min\{\, s \in \mathbb{N} \,:\, \varphi^s = \varepsilon \,\}. \tag{18.8}$$

Замечание 18.2. Обозначение $o(\varphi)$ происходит от слова "order". Используется еще обозначение $|\varphi|$ (связанное с аналогичным обозначением для мощности множества; см. замечание 16.4).

Замечание 18.3. Совершенно очевидно следующее свойство порядка:

$$[\, o(\varphi) = 1 \,] \iff [\, \varphi = \varepsilon \,]. \tag{18.9}$$

Замечание 18.4. Перескажем определение 18.1 иначе: равенство $o(\varphi) = m$, где $m \in \mathbb{N}$, равносильно тому, что, во-первых, $\varphi^m = \varepsilon$, а во-вторых, из равенства $\varphi^s = \varepsilon$ для некоторого $s \in \mathbb{N}$ следует $m \leqslant s$.

Ниже, в предложении 18.1, последнее утверждение будет еще уточнено.

Предложение 18.1. *Пусть $\varphi \in S_n$. Тот факт, что*

$$o(\varphi) = m, \tag{18.10}$$

равносилен следующему утверждению: для целого числа s равенство

$$\varphi^s = \varepsilon \tag{18.11}$$

справедливо тогда и только тогда, когда $m|s$.

Доказательство. 1. Пусть перестановка φ имеет порядок m, т. е. выполняется (18.10).

Целое число s поделим с остатком на m (см. во введении сводку информации из теории чисел), т. е. найдем такие $q, r \in \mathbb{Z}$, что

$$s = q \cdot m + r;\ 0 \leqslant r < m. \tag{18.12}$$

Теперь равенство (18.11) можно записать в виде

$$\varphi^{m \cdot q + r} = \varepsilon,$$

что с учетом свойств (18.4) и (18.5) будет равносильно равенству

$$(\varphi^m)^q \varphi^r = \varepsilon.$$

В силу (18.10), имеем $\varphi^m = \varepsilon$, и значит,

$$\varphi^r = \varepsilon. \tag{18.13}$$

Последнее равенство не может выполняться при $r > 0$, т. к. это противоречило бы определению порядка.

Значит, равенство (18.13) [и вместе с ним равенство (18.11)] равносильно обращению остатка в нуль ($r = 0$) или, другими словами, факту делимости нацело ($s = q \cdot m$, т. е. $m|s$).

2. Пусть справедливо утверждение

$$[\ \varphi^s = \varepsilon\ ] \Longleftrightarrow [\ m|s\ ].$$

Докажем, что m является порядком φ.

Поскольку $m|m$ истинно, то $\varphi^m = \varepsilon$. Далее, $\varphi^s = \varepsilon$ равносильно $m|s$, что для натуральных s влечет $m \leqslant s$, т. е. m является минимальным из натуральных показателей, дающих единичную степень для φ. Значит, справедливо (18.10). $\square$

Предложение 18.2. *1. Значения степеней перестановки повторяются с периодом, равным порядку m этой перестановки:*

$$[\,\varphi^s = \varphi^t\,] \iff [\,s \equiv t\,(\bmod\, m\,)\,]. \tag{18.14}$$

2. Перестановки

$$\varphi^0 = \varepsilon,\ \varphi,\ \varphi^2, \ldots, \varphi^{m-1}$$

попарно различны.

Доказательство. 1. Равенство $\varphi^s = \varphi^t$ домножением на φ^{-s} сводится к равносильному равенству $\varphi^{t-s} = \varepsilon$, которое, в силу предложения 18.1, равносильно делимости $m|t-s$, или, что то же самое (см. сводку информации из теории чисел в предисловии), сравнимости $s \equiv t\,(\bmod\, m\,)$.

2. Если $s, t \in \{1, 2, ..., m-1\}$ и, например, $s < t$, то натуральное число $t - s$ не может делиться на m, и следовательно, $\varphi^s \neq \varphi^t$. □

18.3. Степени и порядок для циклической перестановки. Пусть

$$\sigma = (i_1, i_2, \ldots, i_k) \tag{18.15}$$

— цикл длины k $(2 \leqslant k \leqslant n)$. Напомним, что он оставляет неподвижными все номера, не входящие в цикл, а всякий номер, входящий в цикл (кроме последнего), переводит в соседний справа номер (последний номер переходит в первый). [Говорят, что перестановка (18.15) осуществляет "циклический сдвиг" на одну позицию вправо.]

Ясно, что перестановка σ^s $(s \geqslant 2)$ будет осуществлять циклический сдвиг элементов множества $\{i_1, i_2, ..., i_k\}$ на s единиц вправо. (Номера, не входящие в это множество, будут оставаться на месте.)

При $s = k$ (и не раньше!) все номера окажутся неподвижными, т. е. получится $\sigma^k = \varepsilon$ и $\sigma^s \neq \varepsilon$ при $1 \leqslant s \leqslant k-1$. Следовательно,

$$o(\sigma) = k. \tag{18.16}$$

Тем самым доказано

Предложение 18.3. *Порядок цикла равен его длине.* □

Пример 18.2. При возведении цикла в степень он может распадаться в произведение циклов меньшей длины:

$$\sigma = (2,8,6,4,1,5) \in S_8;$$

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 8 & 3 & 1 & 2 & 4 & 7 & 6 \end{pmatrix};$$

$$\sigma^2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 6 & 3 & 5 & 8 & 1 & 7 & 4 \end{pmatrix} = (1,2,6)(4,5,8);$$

$$\sigma^{-2} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 6 & 1 & 3 & 8 & 4 & 2 & 7 & 5 \end{pmatrix} = (1,6,2)(4,8,5).$$

18.4. Вычисление порядка перестановки с помощью ее разложения на независимые циклы. Всякую перестановку $\varphi \in S_n$ можно (согласно теореме 17.1) разложить в произведение независимых циклов (17.5). Это разложение позволяет легко вычислить порядок $o(\varphi)$.

Предложение 18.4. *Порядок перестановки $\varphi \in S_n$, заданной своим разложением на независимые циклы* (17.5)*, равен наименьшему общему кратному длин этих циклов:*

$$o(\varphi) = \text{НОК}(s_1, s_2, \ldots, s_m), \tag{18.17}$$

где s_j — длина цикла σ_j $(j = 1, ..., m)$.

Доказательство. Прежде всего вы должны четко представлять себе, что такое наименьшее общее кратное (НОК) двух и нескольких натуральных чисел. Если это не совсем так, то немедленно прочитайте параграф, посвященный арифметике целых чисел в книге [2] (или [1]). Напомним также, что в предисловии имеется краткая сводка соответствующего материала.

Пусть

$$q = \text{НОК}(s_1, s_2, \ldots, s_m), \tag{18.18}$$

т. е., во-первых, q делится на каждое из чисел s_j $(j = 1, ..., m)$, другими словами, является общим кратным этих чисел, и во-вторых, q делит любое общее кратное чисел s_j.

Пусть G_j $(j = 1, ..., m)$ — орбиты для перестановки φ, т. е. множества номеров, входящих в независимые циклы σ_j. (Для нетривиальных циклов эти множества совпадают с их областями действия.)

Множества G_j попарно не пересекаются и в объединении составляют все множество $X = \{1, 2, ..., n\}$.

При возведении перестановки в произвольную целую степень s можно воспользоваться коммутированием σ_j $(j = 1, ..., m)$ и свойством (18.6):

$$\varphi^s = \sigma_1^s \sigma_2^s \dots \sigma_m^s. \tag{18.19}$$

Ясно, что возведение перестановок в степень не расширяет их области действия, и следовательно, независимые перестановки останутся таковыми и после возведения в степень. Таким образом, перестановки σ_j^s будут независимыми (хотя уже не обязательно будут циклическими).

Столь же очевидно, что произведение нескольких независимых перестановок может быть тождественной перестановкой в том и только том случае, когда все сомножители этого произведения будут тождественными. Поэтому

$$[\,\varphi^s = \varepsilon\,] \iff [\,(\forall j = 1, ..., m)(\sigma_j^s = \varepsilon)\,]. \tag{18.20}$$

В силу предложений 18.3 и 18.1, равенство $\sigma_j^s = \varepsilon$ равносильно делимости $s_j | s$. Таким образом,

$$[\,\varphi^s = \varepsilon\,] \iff [\,(\forall j = 1, ..., m)(s_j | s)\,]. \tag{18.21}$$

Правая часть эквивалентности (18.21) равносильна утверждению, что s является общим кратным чисел s_j $(j = 1, ..., m)$, а это, в силу определения НОК, равносильно тому, что s делится на q.

Итак,

$$[\,\varphi^s = \varepsilon\,] \iff [\,q | s\,]. \tag{18.22}$$

Согласно предложению 18.1, утверждение (18.22) равносильно тому, что $o(\varphi) = q$. □

Пример 18.3. Вернемся к примеру 17.1. Данная там перестановка $\varphi \in S_{10}$ разложена в произведение независимых циклов (один из которых — тривиальный). Значит,

$$o(\varphi) = \text{НОК}(4, 2, 1, 3) = 12.$$

Разумеется, единички (длины тривиальных циклов) под знаком НОК можно не писать.

Пример 18.4. Возведем перестановку из предыдущего примера в степень $s = -1001$.

Сначала поделим с остатком число $s = -1001$ на порядок $m = 12$:

$$-1001 = 12 \cdot (-84) + 7.$$

В силу предложения 18.2,

$$\varphi^{-1001} = \varphi^7.$$

Перестановка φ представлена как произведение трех нетривиальных циклов:

$$\varphi = (1, 3, 9, 6)\,(2, 7)\,(5, 10, 8).$$

Эти циклы попарно независимы и, следовательно, попарно коммутируют. В силу замечания 18.1,

$$\varphi^7 = \sigma_1^7 \sigma_2^7 \sigma_3^7.$$

Порядки циклических перестановок равны их длинам (по предложению 18.4):

$$o(\sigma_1) = 4;\ o(\sigma_2) = 2;\ o(\sigma_3) = 3.$$

Поделим с остатком показатель степени 7 на эти порядки:

$$7 = 1 \cdot 4 + 3;\ \ 7 = 3 \cdot 2 + 1;\ \ 7 = 2 \cdot 3 + 1.$$

Снова воспользуемся "периодичностью степеней":

$$\sigma_1^7 = \sigma_1^3 = (1, 3, 9, 6)^3 = (1, 6, 9, 3);$$

$$\sigma_2^7 = \sigma_2^1 = (2, 7);$$

$$\sigma_3^7 = \sigma_3^1 = (5, 10, 8).$$

Окончательно получим:

$$\varphi^{-1001} = \varphi^7 = (1, 6, 9, 3)\,(2, 7)\,(5, 10, 8) =$$

$$= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 6 & 7 & 1 & 4 & 10 & 9 & 2 & 5 & 3 & 8 \end{pmatrix}.$$

§ 19. Разложение перестановки в произведение транспозиций

19.1. Разложение циклической перестановки в произведение транспозиций. Напомним (см. определение 17.2), что транспозиция — это частичный цикл вида

$$\tau = (i_1, i_2), \tag{19.1}$$

где i_1, i_2 — различные номера из множества $X = \{1, 2, ..., n\}$.

Как уже отмечалось, запись (i_2, i_1) задает ту же самую транспозицию (19.1). Кроме того, очевидна *самообратность* транспозиций

$$\tau^{-1} = \tau, \tag{19.2}$$

которая может быть выражена и иначе — с помощью понятия порядка перестановки:

$$o(\tau) = 2. \tag{19.3}$$

Рассмотрим теперь произвольный частичный цикл

$$\sigma = (i_1, i_2, i_3, \ldots, i_{s-1}, i_s) \tag{19.4}$$

длины $s \geqslant 2$.

Предложение 19.1. *Цикл (19.4) разлагается в произведение* $s - 1$ *транспозиций:*

$$(i_1, i_2, i_3, \ldots, i_{s-1}, i_s) = (i_1, i_s)\,(i_1, i_{s-1}) \ldots (i_1, i_3)\,(i_1, i_2). \tag{19.5}$$

Доказательство. Не забыли ли вы, что произведение перестановок мы "читаем" справа налево [т. е. сначала применяется самая правая из транспозиций-сомножителей в правой части (19.5)]?

Равенство (19.5) подлежит проверке на произвольном элементе $j \in X$.

Если этот номер $j \notin D_\sigma = \{i_1, i_2, ..., i_s\}$, то все участвующие в (19.5) перестановки оставляют j на месте, и значит, равенство справедливо.

Если $j = i_1$, то левая часть переводит j в i_2. В правой части транспозиция (i_1, i_2) тоже переводит j в i_2, а все последующие транспозиции не содержат i_2 и, значит, оставляют i_2 на месте. Таким образом, значение правой части равно i_2.

Если $j = i_2$, то левая часть дает i_3. В правой части "срабатывают" две транспозиции: первая справа переводит i_2 в i_1, а соседняя переводит i_1 в i_3. В последующих транспозициях i_3 отсутствует и значение "стабилизируется".

Номер $j = i_3$ левая часть переводит в i_4. В правой части этот номер сначала [при действии (i_1, i_2)] остается на месте, затем [под действием (i_1, i_3)] переходит в номер i_1, который [под действием (i_1, i_4)] переходит в i_4, после чего наступает стабилизация, и т. д.

Последний номер $j = i_s$ левая часть переводит в i_1, а в правой все транспозиции "бездействуют", кроме самой последней [справа] (i_1, i_s), которая переводит $j = i_s$ в i_1. □

19.2. Разложение произвольной перестановки в произведение транспозиций. Согласно теореме 17.1, произвольную перестановку $\varphi \in S_n$ можно разложить в произведение независимых циклов.

Если $\varphi \neq \varepsilon$, то в разложении можно оставить только нетривиальные циклы [см. формулу (17.1)].

Каждый из циклов σ_j $(j = 1, .., l)$ длины s_j разложим (в соответствии с предложением 19.1) в произведение $s_j - 1$ транспозиций.

Тогда данная перестановка разложится в произведение

$$\sum_{j=1}^{l}(s_j - 1) = \sum_{j=1}^{l} s_j - l$$

транспозиций. Используя понятие декремента перестановки (см. определение 17.3 и замечание 17.5), можем сформулировать

Предложение 19.2. *Всякую перестановку $\varphi \in S_n$ можно разложить в произведение $\delta(\varphi)$ транспозиций, где $\delta(\varphi)$ — декремент перестановки φ.*

Доказательство см. выше. □

Замечание 19.1. В отличие от разложения на независимые циклы, которое определено однозначно (с точностью до порядка сомножителей), разложение перестановки на транспозиции определено отнюдь не однозначно.

Скажем, цикл $\sigma = (1, 2, 3)$ можно [см. (19.5)] представить в виде

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

а можно в виде

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

или в виде

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

(при этом используется тот факт, что квадрат транспозиции равен тождественной перестановке ε), или в виде

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

(почему?) и т. д.

Ниже будет доказано, что хотя разложение на транспозиции и даже количество сомножителей в нем определены не однозначно, *четность количества сомножителей* определена однозначно (инвариантно).

В качестве информации сообщим (без доказательства; подробнее см. в задачнике [4, гл. 1, § 3]): в разложении перестановки на транспозиции количество сомножителей не может быть меньше, чем декремент этой перестановки.

И еще: позволим себе (условно) считать, что тождественная перестановка разлагается в произведение 0 транспозиций (таков ее декремент!), хотя, кроме этого "разложения", имеется и множество других: $\varepsilon = (1,2)(1,3)(1,3)(1,2)$ и т. п.

Пример 19.1. Разложим перестановку φ из примера 17.1 на транспозиции (в указанном примере она уже была разложена на независимые циклы):

$$\begin{aligned}\varphi &= (1,3,9,6)\,(2,7)\,(5,10,8) = \\ &= (1,6)\,(1,9)\,(1,3)\,(2,7)\,(5,8)\,(5,10).\end{aligned}$$

19.3. Коммутационные соотношения для транспозиций. В дальнейшем нам понадобятся простые формулы, связанные с перемножением транспозиций.

Предложение 19.3. *1. Для транспозиции (i,j) справедливо*

$$(i,j)^2 = \varepsilon. \tag{19.6}$$

2. Пусть i, j, k — три попарно различных номера. Тогда справедливы формулы:

$$(i,k)\,(i,j) = (j,k)\,(i,k); \tag{19.7}$$

$$(j,k)\,(i,j) = (i,j)\,(i,k). \tag{19.8}$$

3. Пусть i, j, k, l — четыре попарно различных номера. Тогда

$$(k,l)\,(i,j) = (i,j)\,(k,l). \tag{19.9}$$

Доказательство. Первая формула уже не раз встречалась [см. (19.3)].

Второе соотношение получается с помощью разложения цикла на транспозиции (предложение 19.1; см. также замечание 17.1):

$$\begin{aligned}(i,k)\,(i,j) = (i,j,k) = (j,k,i) = \\ =(k,i,j) = (k,j)\,(k,i) = (j,k)\,(i,k).\end{aligned}$$

Третье соотношение получается из второго заменой j на k и обратно.

Четвертое соотношение является частным случаем предложения 16.1: две независимые транспозиции коммутируют. □

19.4. Теорема о разложениях перестановки на транспозиции. В этом пункте мы докажем объявленную в замечании 19.1 инвариантность четности количества транспозиций в разложении перестановки.

Предварительно мы установим вспомогательное предложение о разложениях на транспозиции тождественной перестановки. Обратите внимание на его доказательство. На наш взгляд, оно совершенно замечательно, в нем — особый алгебраический шарм.

Предложение 19.4. *Тождественную перестановку нельзя представить в виде произведения нечетного числа транспозиций.*

Доказательство. Рассмотрим тождественную перестановку $\varepsilon \in S_n$ $(n \geqslant 2)$. Предположим, что ε *можно* представить в виде произведения нечетного числа транспозиций. Выберем число m как *наименьшее* из таких нечетных чисел s, что существует разложение ε в произведение s транспозиций.

Заметим, что $m \neq 1$. В самом деле, равенство $\varepsilon = \tau_1$, где τ_1 — транспозиция, ложно, поскольку правая часть перемещает (переставляет) два номера и, следовательно, не тождественна.

Значит, $m \geqslant 3$, причем существует некоторое разложение

$$\varepsilon = \tau_m \tau_{m-1} \dots \tau_3 \tau_2 \tau_1, \tag{19.10}$$

содержащее m транспозиций-сомножителей, и не существует никакого разложения, содержащего $m-2$ транспозиции.

Возьмем действующую первой транспозицию

$$\tau_1 = (i, j) = (j, i)$$

и пометим один из номеров в ней, скажем $\boxed{j}$.

После этого обратимся к соседней транспозиции

$$\tau_2 = (k, l) = (l, k)$$

и рассмотрим четыре возможных случая соотношения областей действия транспозиций τ_2 и τ_1.

1. Области действия совпадают: $\{k, l\} = \{i, j\}$. Тогда сами транспозиции тоже совпадают:

$$\tau_2 = \tau_1 = (i, j),$$

и, в силу (19.6), в разложении (19.10) произойдет "сокращение":

$$\tau_2 \tau_1 = \tau_1^2 = \varepsilon,$$

после чего мы придем к "более короткому" разложению

$$\varepsilon = \tau_m \tau_{m-1} \dots \tau_3,$$

содержащему $m-2$ транспозиции, а это противоречит предположению.

2. Области действия пересекаются по элементу i:

$$\{k, l\} \cap \{i, j\} = \{i\}.$$

В этом случае можно считать, что $l = i$ и номера i, j, k попарно различны. Применим коммутационное соотношение (19.7):

$$\tau_2 \tau_1 = (i, k)\,(i, \boxed{j}\,) = (\boxed{j}\,, k)\,(i, k) = \tau_2' \tau_1'.$$

Снова получилось произведение двух транспозиций, но *помеченный номер* $\boxed{j}$ переместился во вторую справа транспозицию τ_2', а новая первая транспозиция τ_1' номер $\boxed{j}$ не содержит.

3. Аналогичное явление происходит и в случае

$$\{k,l\} \cap \{i,j\} = \{\boxed{j}\}.$$

Можем считать $l = j$; с помощью (19.8) получим:

$$\tau_2\tau_1 = (\boxed{j}, k)\,(i, \boxed{j}) = (i, \boxed{j})\,(i,k) = \tau_2'\tau_1'.$$

Снова $\boxed{j}$ переместился в τ_2' и отсутствует в τ_1'.

4. В случае

$$\{k,l\} \cap \{i,j\} = \varnothing$$

еще проще:

$$\tau_2\tau_1 = \tau_1\tau_2 = (i, \boxed{j})\,(k,l)$$

с тем же эффектом.

Итак, на первом шаге либо происходит сокращение и получается противоречие, либо отмеченный номер перемещается с первой позиции на вторую, правее которой он отсутствует.

На втором шаге (если он понадобится) мы рассматриваем "новых соседей" (τ_3 и τ_2') и проводим аналогичный анализ; и т. д.

Ясно, что в этом процессе

— либо на некотором шаге произойдет сокращение и получится противоречие;

— либо помеченный номер $\boxed{j}$ "благополучно доберется" до последней (крайней слева) позиции и будет отсутствовать правее этой позиции.

В последнем случае получим разложение

$$\varepsilon = \tilde{\tau}_m\tilde{\tau}_{m-1}\dots\tilde{\tau}_3\tilde{\tau}_2\tilde{\tau}_1, \tag{19.11}$$

где $\tilde{\tau}_m = (h, \boxed{j})$ [а остальные транспозиции не содержат $\boxed{j}$], что также приводит к противоречию, ибо левая часть (19.11) оставляет все номера на месте, а правая переводит $\boxed{j}$ в $h \neq \boxed{j}$ (поскольку транспозиции $\tilde{\tau}_{m-1}, \dots, \tilde{\tau}_2, \tilde{\tau}_1$ не содержат и, следовательно, не перемещают $\boxed{j}$).

Итог: во всех случаях получается противоречие. Значит, сделанное предположение не верно и предложение доказано. □

Переходим к формулировке основного результата настоящего параграфа.

Теорема 19.1. *Если перестановка $\varphi \in S_n$ представляется двумя способами в виде произведения транспозиций:*

$$\varphi = \tau_p \tau_{p-1} \dots \tau_2 \tau_1; \tag{19.12}$$

$$\varphi = \tau'_q \tau'_{q-1} \dots \tau'_2 \tau'_1, \tag{19.13}$$

то числа p и q (количества сомножителей) имеют одинаковую четность.

Доказательство. Из условия следует равенство

$$\tau_p \tau_{p-1} \dots \tau_2 \tau_1 = \tau'_q \tau'_{q-1} \dots \tau'_2 \tau'_1. \tag{19.14}$$

Умножим (19.14) справа на произведение $\tau_1 \tau_2 \dots \tau_{p-1} \tau_p$ и учтем самообратность транспозиций

$$\tau_1^2 = \tau_2^2 = \dots = \tau_p^2 = \varepsilon.$$

Получим

$$\varepsilon = \tau'_q \tau'_{q-1} \dots \tau'_2 \tau'_1 \tau_1 \tau_2 \dots \tau_{p-1} \tau_p, \tag{19.15}$$

что является разложением тождественной перестановки в произведение $p + q$ транспозиций.

В силу предложения 19.4, число $p + q$ должно быть *четным*, а значит, числа p и q должны быть либо оба четными, либо оба нечетными.

Теорема доказана. □

§ 20. Знак и четность перестановки

20.1. Знак перестановки. Пусть дана перестановка $\varphi \in S_n$. Разложим ее (каким-либо образом) в произведение транспозиций (19.12).

Определение 20.1. Знаком перестановки φ называется число

$$\operatorname{sgn}(\varphi) = (-1)^p, \tag{20.1}$$

где p — количество транспозиций в разложении (19.12).

Замечание 20.1. Мы помним о неоднозначности разложения на транспозиции. Но помним также, что если найдется другое разложение (19.13), содержащее q сомножителей, то число q имеет такую же четность, что и p, и следовательно, $(-1)^p = (-1)^q$.

Этим доказана *корректность* определения 20.1.

Замечание 20.2. Отметим простейшие следствия из данного выше определения:

1) для транспозиции τ знак

$$\operatorname{sgn}(\tau) = -1; \tag{20.2}$$

2) для тождественной перестановки

$$\operatorname{sgn}(\varepsilon) = 1. \tag{20.3}$$

Замечание 20.3. Из предложения 19.2 вытекает, что

$$\operatorname{sgn}(\varphi) = (-1)^{\delta(\varphi)}, \tag{20.4}$$

где $\delta(\varphi)$ — декремент перестановки φ.

Замечание 20.4. Предостережем вас от неграмотного словоупотребления типа "знак перестановки $+$".

Знак — это число (1 или -1).

Пример 20.1. Вычислим знак перестановки φ из примера 19.1 (в том примере перестановка была разложена в произведение 6 транспозиций; см. также пример 19.2, где констатировалось, что декремент этой перестановки $\delta(\varphi) = 6$). По определению 20.1,

$$\operatorname{sgn}(\varphi) = (-1)^{\delta(\varphi)} = (-1)^6 = 1.$$

20.2. Свойства знака. В п. 16.5 мы установили, что множество S_n является *группой по умножению*. Заметим теперь, что множество (состоящее из двух целых чисел)

$$\mathbb{Z}^* = \{\, 1\,, -1\,\} \tag{20.5}$$

также является группой по умножению.

Мы уже знакомы и с понятием гомоморфизма одного алгебраического объекта в другой, однотипный алгебраический объект (см.

замечание 15.1). Сейчас нашими объектами являются группы. Отображение

$$f : G_1 \longrightarrow G_2$$

называется *гомоморфизмом групп*, если оно согласовано с операциями (умножения) в этих группах, т. е. если для любых элементов $g, h \in G_1$ справедливо равенство

$$f(g \cdot h) = f(g) \cdot f(h). \tag{20.6}$$

Из условия (20.6) легко выводятся следующие свойства гомоморфизмов:

$$f(1) = 1; \tag{20.7}$$

$$f(g^{-1}) = (f(g))^{-1}. \tag{20.8}$$

Заметьте, что в формуле (20.6) один и тот же символ " $\cdot$ " обозначает, вообще говоря, разные умножения в разных группах, а в формуле (20.7) один и тот же символ "1" обозначает разные единицы. (Скажем, в группе S_n : умножение — это композиция, единица — это тождественная перестановка, а в группе $\mathbb{Z}^*$: и умножение, и единица — обычные, числовые.)

Предложение 20.1. *Отображение вычисления знака*

$$\operatorname{sgn} : S_n \longrightarrow \mathbb{Z}^*; \varphi \mapsto \operatorname{sgn}(\varphi); \varphi \in S_n \tag{20.9}$$

является гомоморфизмом групп, т. е. для любых перестановок φ, ψ справедливо равенство

$$\operatorname{sgn}(\varphi\psi) = \operatorname{sgn}(\varphi)\operatorname{sgn}(\psi). \tag{20.10}$$

Доказательство. Обе данные перестановки можно разложить на транспозиции. Пусть в разложении для φ имеется p транспозиций, а в разложении для ψ — q транспозиций.

Для произведения $\varphi\psi$ получится разложение, содержащее $p + q$ транспозиций. Поэтому

$$\operatorname{sgn}(\varphi\psi) = (-1)^{p+q} = (-1)^p(-1)^q = \operatorname{sgn}(\varphi)\operatorname{sgn}(\psi).$$

Предложение доказано. □

Замечание 20.5. Будучи гомоморфизмом, отображение sgn обладает еще свойствами $\operatorname{sgn}(\varepsilon) = 1$ (что уже отмечалось) и

$$\operatorname{sgn}(\varphi^{-1}) = (\operatorname{sgn}(\varphi))^{-1} \tag{20.11}$$

для любой перестановки $\varphi \in S_n$.

Но поскольку элементы группы $\mathbb{Z}^*$ самообратны, то последняя формула приобретает вид

$$\operatorname{sgn}(\varphi^{-1}) = \operatorname{sgn}(\varphi). \tag{20.12}$$

20.3. Четность перестановки. Подгруппа четных перестановок

Определение 20.2. Перестановка $\varphi \in S_n$ называется *четной*, если $\operatorname{sgn}(\varphi) = 1$, и *нечетной*, если $\operatorname{sgn}(\varphi) = -1$.

Множество четных перестановок обозначается

$$A_n = \{\, \varphi \in S_n \,:\, \operatorname{sgn}(\varphi) = 1 \,\}. \tag{20.13}$$

Сделаем небольшое отвлечение о подобъектах в алгебраических объектах. Подмножество в каком-либо алгебраическом объекте называется *подобъектом*, если оно *устойчиво* относительно всех алгебраических операций, характерных для рассматриваемого типа объектов. (В таком случае это подмножество само становится объектом того же типа.)

Например, мы уже активно работали с линейными подпространствами в линейных пространствах (см. п. 3.2). Теперь мы познакомимся с понятием подгруппы.

Подгруппой является такое подмножество H в группе G, которое

1) вместе с любыми двумя элементами $h_1, h_2 \in H$ содержит их произведение $h_1 \cdot h_2 \in H$;

2) содержит единичный элемент $1 \in H$;

3) вместе с каждым элементом $h \in H$ содержит обратный элемент $h^{-1} \in H$.

Применяется обозначение $H \leqslant G$ (такое же, как и для линейных подпространств, но здесь оно читается соответственно: *H является подгруппой в группе G*).

Предложение 20.2. *Множество четных перестановок A_n является подгруппой в группе всех перестановок S_n, т. е.*

$$A_n \leqslant S_n. \tag{20.14}$$

Доказательство. 1. Если $\varphi, \psi \in A_n$, т. е. $\operatorname{sgn}(\varphi) = \operatorname{sgn}(\psi) = 1$, то, в силу предложения 20.1, $\operatorname{sgn}(\varphi\psi) = \operatorname{sgn}(\varphi)\operatorname{sgn}(\psi) = 1$, и следовательно, $\varphi\psi \in A_n$.

2. Тождественная перестановка $\varepsilon \in A_n$, т. к. $\operatorname{sgn}(\varepsilon) = 1$.

3. Из свойства (20.12) следует, что $\varphi \in A_n$ влечет $\varphi^{-1} \in A_n$. □

Замечание 20.6. Подмножество всех нечетных перестановок $B_n = S_n \setminus A_n$ подгруппой в S_n не является. (Почему?)

Подсчитаем теперь количество четных перестановок степени n. Для этого рассмотрим биекцию

$$l_\tau : S_n \longrightarrow S_n;\ l_\tau(\varphi) = \tau\varphi;\ \varphi \in S_n \tag{20.15}$$

левого сдвига на произвольную (фиксированную) транспозицию $\tau \in S_n$ (см. п. 16.8).

Из свойства (20.10) с учетом (20.2) имеем:

$$\operatorname{sgn}(\tau\varphi) = -\operatorname{sgn}(\varphi). \tag{20.16}$$

Следовательно, биекция l_τ переводит четные перестановки в нечетные и наоборот. Стало быть, количество четных перестановок равно количеству нечетных:

$$|A_n| = |B_n| = \frac{|S_n|}{2} = \frac{n!}{2}. \tag{20.17}$$

Пример 20.2. Вернемся к примеру 16.1, в котором были перечислены все перестановки в группах S_2, S_3 и S_4.

Соблюдая принятый в списках примера 16.1 порядок перечисления перестановок, мы каждую из этих перестановок разложим на (нетривиальные) независимые циклы, а затем — на транспозиции, после чего определим ее **порядок** и знак.

(А вы обратите внимание на омонимию: выше слово "порядок" употреблено в двух различных математических смыслах.)

S_2 :
$\varepsilon;\ \mathbf{1};\ 1;$
$(1,2);\ \mathbf{2};\ -1.$

S_3 :
$\varepsilon;\ \mathbf{1};\ 1;$
$(2,3);\ \mathbf{2};\ -1;$
$(1,2);\ \mathbf{2};\ -1;$
$(1,2,3) = (1,3)\,(1,2);\ \mathbf{3};\ 1;$
$(1,3,2) = (1,2)\,(1,3);\ \mathbf{3};\ 1;$
$(1,3);\ \mathbf{2};\ -1.$

S_4 :
$\varepsilon;\ \mathbf{1};\ 1;$
$(3,4);\ \mathbf{2};\ -1;$
$(2,3);\ \mathbf{2};\ -1;$
$(2,3,4) = (2,4)\,(2,3);\ \mathbf{3};\ 1;$
$(2,4,3) = (2,3)\,(2,4);\ \mathbf{3};\ 1;$
$(2,4);\ \mathbf{2};\ -1;$
$(1,2);\ \mathbf{2};\ -1;$
$(1,2)\,(3,4);\ \mathbf{2};\ 1;$
$(1,2,3) = (1,3)\,(1,2);\ \mathbf{3};\ 1;$
$(1,2,3,4) = (1,4)\,(1,3)\,(1,2);\ \mathbf{4};\ -1;$
$(1,2,4,3) = (1,3)\,(1,4)\,(1,2);\ \mathbf{4};\ -1;$
$(1,2,4) = (1,4)\,(1,2);\ \mathbf{3};\ 1;$
$(1,3,2) = (1,2)\,(1,3);\ \mathbf{3};\ 1;$
$(1,3,4,2) = (1,2)\,(1,4)\,(1,3);\ \mathbf{4};\ -1;$
$(1,3);\ \mathbf{2};\ -1;$
$(1,3,4) = (1,4)\,(1,3);\ \mathbf{3};\ 1;$
$(1,3)\,(2,4);\ \mathbf{2};\ 1;$
$(1,3,2,4) = (1,4)\,(1,2)\,(1,3);\ \mathbf{4};\ -1;$
$(1,4,3,2) = (1,2)\,(1,3)\,(1,4);\ \mathbf{4};\ -1;$
$(1,4,2) = (1,2)\,(1,4);\ \mathbf{3};\ 1;$
$(1,4,3) = (1,3)\,(1,4);\ \mathbf{3};\ 1;$
$(1,4);\ \mathbf{2};\ -1;$
$(1,4,2,3) = (1,3)\,(1,2)\,(1,4);\ \mathbf{4};\ -1;$
$(1,4)\,(2,3);\ \mathbf{2};\ 1.$

§ 21. Число инверсий в перестановке. Второй способ определения знака перестановки

21.1. Инверсии в перестановке. Рассмотрим перестановку [см. (16.2)]

$$\varphi = \begin{pmatrix} 1 & 2 & \dots & n \\ \varphi(1) & \varphi(2) & \dots & \varphi(n) \end{pmatrix} \tag{21.1}$$

и всевозможные пары номеров (i, j), такие, что $1 \leqslant i < j \leqslant n$. Сколько существует таких пар? Если вы немного знакомы с комбинаторикой, то легко ответите: их

$$C_n^2 = \frac{n(n-1)}{2} \tag{21.2}$$

штук. Формула (21.2) выражает *число сочетаний из n элементов по* 2". Но если вы таких слов пока не знаете, то можно объяснить по-другому: последовательность i, j двух различных номеров можно выбрать $n(n-1)$ способами; лишь одна из двух последовательностей i, j и j, i является *возрастающей* (либо $i < j$, либо $j < i$); число возрастающих последовательностей дается формулой (21.2).

Определение 21.1. Будем говорить, что перестановка φ *осуществляет инверсию* (или: имеет инверсию; или: допускает *беспорядок*; и т. п.) на возрастающей последовательности i, j $(i < j)$, если $\varphi(i) > \varphi(j)$.

Количество инверсий в перестановке φ будем обозначать $\iota(\varphi)$ (ι — греческая буква "йота").

Замечание 21.1. Наш подсчет в начале данного пункта показывает, что число инверсий $\iota(\varphi)$ заключено в пределах от 0 до C_n^2, причем $\iota(\varphi) = 0$ для перестановки $\varphi = \varepsilon$ (и только для нее), а наибольшее возможное число инверсий $\iota(\varphi) = C_n^2$ имеет перестановка

$$\varphi = \begin{pmatrix} 1 & 2 & \dots & n-1 & n \\ n & n-1 & \dots & 2 & 1 \end{pmatrix}. \tag{21.3}$$

Можно сказать, что $\varphi = \varepsilon : X \to X$ является единственной *строго возрастающей функцией*, а перестановка (21.3) — единственной *строго убывающей*.

Пример 21.1. Возьмем снова перестановку φ из серии примеров 17.1, 17.2, 19.1, 20.1:

$$\varphi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 3 & 7 & 9 & 4 & 10 & 1 & 2 & 5 & 6 & 8 \end{pmatrix}.$$

Инверсии (беспорядки) наблюдаются по нижней строке в двустрочной записи.

Рассмотрим номер 1 во второй строке; слева от него расположено 5 номеров; все они, естественно, больше 1.

Переходим к номеру 2; слева от него расположено 5 номеров, больших 2.

Слева от 3 вообще ничего нет.

Слева от 4 имеем 2 номера, больших 4.

И так далее, до номера 9. (Номер 10 можно не рассматривать, т. к. больших, чем он, просто нет.)

Набегает следующая сумма:

$$\iota(\varphi) = 5 + 5 + 0 + 2 + 3 + 3 + 0 + 1 = 16.$$

21.2. Второй способ определения знака перестановки

Предложение 21.1. *Знак перестановки $\varphi \in S_n$ может быть подсчитан по формуле*

$$\operatorname{sgn}(\varphi) = (-1)^{\iota(\varphi)}, \tag{21.4}$$

где $\iota(\varphi)$ — количество инверсий в перестановке φ.

Доказательство. В соответствии с определением 20.1, знак φ может быть вычислен по формуле

$$\operatorname{sgn}(\varphi) = (-1)^p,$$

где p — количество сомножителей в каком-либо (все равно каком!) разложении перестановки φ на транспозиции.

Формула (21.4) будет доказана, если мы сможем подобрать такое разложение φ на транспозиции, в котором число сомножителей p будет равно числу инверсий $\iota(\varphi)$.

Если $\varphi = \varepsilon$, то инверсий нет и можно считать (см. замечание 19.1), что в разложении на транспозиции 0 сомножителей.

Если $\varphi \neq \varepsilon$, то инверсии есть; пусть их количество $\iota(\varphi) = k$, где $k \geqslant 1$.

Среди инверсий обязательно найдется такая, которая реализуется на *соседних* номерах $i, i+1$ $(1 \leqslant i \leqslant n-1)$. (В противном случае перестановка φ была бы строго возрастающей функцией и, следовательно, равнялась бы ε; см. замечание 21.1.)

Выберем какую-либо из таких инверсий, т. е. пару соседних номеров $i, i+1$, для которой $\varphi(i) > \varphi(i+1)$.

Рассмотрим перестановку

$$\varphi_1 = \varphi\,(i, i+1). \tag{21.5}$$

На номера j, отличные от i и $i+1$, перестановка φ_1 действует так же, как и φ : $\varphi_1(j) = \varphi(j)$. На номерах $i, i+1$ получим:

$$\varphi_1(i) = \varphi(i+1);\ \ \varphi_1(i+1) = \varphi(i),$$

и следовательно, $\varphi_1(i) < \varphi_1(i+1)$, т. е. в перестановке φ_1 на этих соседних номерах инверсии не будет.

Другие же инверсии сохранятся. (Почему?)

Стало быть, перестановка φ_1 будет иметь на одну инверсию меньше, чем φ.

Если φ_1 еще не равна ε, то и в ней найдется "инверсия соседей" и процесс можно будет продолжить.

(Чем он закончится? Завершите рассуждение.) □

Пример 21.2. Знак перестановки из примера 21.1:

$$\operatorname{sgn}(\varphi) = (-1)^{\iota(\varphi)} = (-1)^{16} = 1.$$

Замечание 21.2 (для служебного пользования). В подавляющем числе учебников по алгебре (например, в [3]) второй способ определения знака излагается как основной. Мы предпочли последовать методике курса [2] по следующим причинам.

1. Нам представляется очень красивым, поучительным (и непривычным для школярского мышления) доказательство предложения 19.4.

2. При избранном нами подходе обоснование главного свойства знака (см. предложение 20.1) совершенно тривиально, в то время как при втором способе именно этот факт вызывает наибольшие (хотя, конечно, преодолимые) трудности.

Кстати, есть еще и третий способ, совершенно заформализованный и вряд ли приемлемый для изложения первокурсникам (см., например, [7]).

Знак перестановки $\varphi \in S_n$ можно определить формулой

$$\operatorname{sgn}(\varphi) = \prod_{\substack{\{i,j\}\subseteq X \\ i\neq j}} \operatorname{sign}\left(\frac{j-i}{\varphi(j)-\varphi(i)}\right), \tag{21.6}$$

где произведение берется по всем двухэлементным подмножествам $\{i,j\}$ множества $X = \{1,2,...,n\}$ (это произведение содержит C_n^2 сомножителей) и где аббревиатура sign обозначает известную из математического анализа функцию *знак действительного числа* x:

$$\operatorname{sign}(x) = \begin{cases} -1, & \text{если x<0;} \\ 0, & \text{если x=0;} \\ 1, & \text{если x>0.} \end{cases}$$

(Иная аббревиатура sgn для знака перестановки специально выбрана слегка отличной от более привычного sign.)

Поясним, что в (21.6) произведение берется по *неупорядоченным наборам* $\{i,j\}$, т. е. считается, что $\{i,j\} = \{j,i\}$. Однако, поскольку знак дроби не изменится при замене i на j и наоборот, это определение является корректным.

Для "любознательных читателей" будет полезным упражнением доказательство того, что третий способ определения знака перестановки эквивалентен двум предыдущим.

А есть еще и четвертый способ, близкий по идее к третьему:

$$\operatorname{sgn}(\varphi) = \prod_{1\leqslant i<j\leqslant n} \frac{j-i}{\varphi(j)-\varphi(i)}. \tag{21.7}$$

В формуле (21.7) произведение берется по *возрастающим* последовательностям i,j; символ sign становится излишним. (Почему?)

§ 22. Вычисления с перестановками в системе Maple

22.1. Форматы задания перестановок. В математической системе Maple вычисления с перестановками можно осуществлять в рамках пакета group, который требуется подгрузить командой

```
> with ( group ) ;
```

Данный пакет предназначен для решения некоторых вычислительных задач теории *конечных групп*; перестановки в этой теории играют ключевую роль.

Двустрочную запись перестановки (16.5) Maple представляет следующим образом: верхнюю строку он, естественно, не выводит ("держит в уме"); выводит только нижнюю (как *список* особого рода, именуемый 'permlist'; от permutation = перестановка).

Например, перестановку

$$\phi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 3 & 7 & 9 & 4 & 10 & 1 & 2 & 5 & 6 & 8 \end{pmatrix}$$

можно ввести командой

```
> phi := [ 3, 7, 9, 4, 10, 1, 2, 5, 6, 8 ] ;
```

Maple среагирует:

$$\phi := [3, 7, 9, 4, 10, 1, 2, 5, 6, 8];$$

Тождественная перестановка (степени 10) вводится командой

```
> epsilon := [ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ] ;
```

$$\varepsilon := [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]$$

Вторым (основным!) форматом представления перестановок является их разложение на независимые (непересекающиеся = disjoint) циклы (cycles). Обозначение формата: 'disjcyc'.

Например, дав команду

```
> psi := convert ( phi, 'disjcyc' ) ;
```

мы получим *список списков* ('listlist'):

$$\psi := [[1, 3, 9, 6], [2, 7], [5, 10, 8]]$$

(Команда convert играет весьма важную роль в системе Maple. Средства этой системы строго ориентированы на определенные *типы* данных. Некоторые из типов допускают *конвертацию* из одного типа в другой и обратно.)

Обратный переход от типа 'disjcyc' к типу 'permlist' возможен, если уточнить *степень* перестановки (по разложению на циклы она не видна). Команда

```
> convert ( psi, 'permlist', 10 ) ;
```

должна вернуть нас к исходной перестановке:

$$[3, 7, 9, 4, 10, 1, 2, 5, 6, 8]$$

Еще один пример:

```
> convert ( epsilon, 'disjcyc' ); convert ( [ ], 'permlist' , 10 );
```

$$[\,]$$

$$[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]$$

Видим, что тождественная перестановка в формате 'disjcyc' представляется *пустым списком.*

22.2. Умножение перестановок в пакете group. К сожалению, общая несогласованность в обозначениях композиции (см. замечание 15.3) коснулась и системы Maple.

Порядок перемножения перестановок в Maple противоположен порядку, принятому в настоящем пособии. Это не слишком страшно. Если вы хотите перемножить перестановки ϕ и ψ (т. е. вычислить $\mu = \phi\psi$ в нашем "левом" смысле), то вам придется набрать:

```
> mu := mulperms ( psi, phi ) ;
```

(Таким образом, первой указывается та перестановка, которая на самом деле действует первой.)

Важно другое: обе данные перестановки должны быть — предварительно! — переведены в формат 'disjcyc'.

Ниже мы изначально задаем сомножители в этом формате.

```
> alpha := [ [ 4, 5, 8 ], [ 10, 3 ], [ 2, 6, 12 ] ] ;
  beta := [ [ 1, 11 ], [ 3, 8, 9, 2, 5 ] ] ;
```

$$\alpha := [[4, 5, 8], [10, 3], [2, 6, 12]]$$

$$\beta := [[1, 11], [3, 8, 9, 2, 5]]$$

> mu := mulperms (alpha, beta) ;
  nu := mulperms (beta, alpha) ;

$$\mu := [[1, 11], [2, 6, 12, 5, 9], [3, 10, 8, 4]]$$

$$\nu := [[1, 11], [2, 8, 9, 6, 12], [3, 4, 5, 10]]$$

Ответы $\mu = \beta\alpha$ и $\nu = \alpha\beta$ также возвращаются в формате 'disjcyc'. Проверим свойство тождественной перестановки:

> mulperms ([ ], alpha) ;

$$[[2, 6, 12], [3, 10], [4, 5, 8]]$$

Мы ожидали, что в ответе получится перестановка α. Именно ее мы и получили, только 1) циклы в списке идут в другом порядке и 2) внутри одного из циклов номера идут в ином порядке. (Это вполне согласуется с тем, что мы уже знаем: 1) разложение перестановки на независимые циклы определено однозначно — лишь с точностью до порядка сомножителей; 2) запись цикла может начинаться с любого входящего в этот цикл номера.)

22.3. Обращение перестановок в пакете group. Перед вычислением обратной перестановки данная также должна быть представлена в "рабочем" формате. Обращение осуществляется командой

> lambda := invperm (mu) ;

$$\lambda := [[1, 11], [2, 9, 5, 12, 6], [3, 4, 8, 10]]$$

Проверим результат умножением:

> mulperms (lambda, mu) ;

$$[\,]$$

22.4. Вычисление порядка перестановки. Пакет group позволяет вычислять порядок перестановки в формате 'disjcyc', если дополнительно указана ее степень:

```
> m := grouporder ( permgroup ( 10, { psi } ) ;
```

$$m := 12$$

(Синтаксис последней команды несколько усложнен. Дело в том, что она и предназначена для решения более сложной задачи: с ее помощью можно найти порядок подгруппы в группе перестановок указанной степени, порожденной несколькими перечисляемыми в фигурных скобках перестановками.)

22.5. Вычисление знака перестановки. Знак (четность = parity) перестановки вычисляется командой:

```
> parity ( psi ) ;
```

1

Глава 4

ТЕОРИЯ ОПРЕДЕЛИТЕЛЕЙ

§ 23. Определение определителя квадратной матрицы. Определитель треугольной матрицы. Определитель транспонированной матрицы

23.1. Определение определителя. Пусть A — квадратная матрица размера $n \times n$ с элементами из поля $\mathbb{R}$, т. е. $A \in L(n, \mathbb{R})$ (см. обозначение в п. 14.1).

Рассмотрим какую-либо перестановку

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & n-1 & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n-1) & \sigma(n) \end{pmatrix} \in S_n. \tag{23.1}$$

Составим с помощью перестановки (23.1) следующее произведение элементов матрицы A:

$$\operatorname{sgn}(\sigma) \cdot a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdot \ldots \cdot a_{(n-1)\sigma(n-1)} \cdot a_{n\sigma(n)}. \tag{23.2}$$

В произведении (23.2) элементы матрицы берутся по одному из каждой строки матрицы и по одному из каждого столбца, причем из строки с номером i выбирается элемент, расположенный в столбце с номером $\sigma(i)$ $(i = 1, ..., n)$. Произведение снабжается знаком, равным знаку перестановки σ.

Определение 23.1. Выражение (23.2) называется *членом определителя* матрицы A, отвечающим перестановке (23.1).

Определителем матрицы A называется сумма всех таких членов, отвечающих всевозможным перестановкам (23.1).

Прокомментируем данное выше определение, а затем перескажем его более полно и строго.

Пусть, например, матрица A имеет порядок 5:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix}.$$

Рассмотрим — также для примера — перестановку

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 2 & 5 & 1 \end{pmatrix}.$$

Знак этой перестановки (см. § 20) $\operatorname{sgn}(\sigma) = -1$.

Пометим в матрице A элементы, отвечающие перестановке σ:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \boxed{a_{14}} & a_{15} \\ a_{21} & a_{22} & \boxed{a_{23}} & a_{24} & a_{25} \\ a_{31} & \boxed{a_{32}} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & \boxed{a_{45}} \\ \boxed{a_{51}} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix};$$

и образуем соответствующий член определителя:

$$(-1) \cdot \boxed{a_{14}} \cdot \boxed{a_{23}} \cdot \boxed{a_{32}} \cdot \boxed{a_{45}} \cdot \boxed{a_{51}}.$$

Всего определитель будет содержать $5! = 120$ членов такого вида. В частности, тождественной перестановке $\varepsilon \in S_5$ будет соответствовать произведение

$$a_{11} \cdot a_{22} \cdot a_{33} \cdot a_{44} \cdot a_{55}.$$

Определение 23.1а. *Определителем* (*детерминантом*) матрицы A называется действительное число, обозначаемое $\det(A)$ или $|A|$, равное сумме $n!$ слагаемых, каждое из которых

1) отвечает некоторой перестановке σ [вида (23.1)];

2) является произведением [вида (23.2)], составленным из n элементов матрицы A.

Элементы произведения (23.2) берутся по одному из каждой строки матрицы A и по одному из каждого ее столбца [в строке с номером i выбирается элемент из столбца с номером $\sigma(i)$].

Произведение (23.2) снабжается знаком, равным знаку $\operatorname{sgn}(\sigma)$ перестановки (23.1).

Формула для определителя:

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdot \ldots \cdot a_{n\sigma(n)}. \tag{23.3}$$

Замечание 23.1. Происхождение такого громоздкого определения будет прокомментировано позже.

Несколько слов о терминологии. Словосочетание "определение определителя" может показаться досадной тавтологией. Но это не совсем так, поскольку данная "тавтология" является всего лишь следствием не очень удачной русификации.

На самом же деле "определение" = definition, а "определитель" = determinant. И "заимствованный" термин *детерминант* был бы, возможно, лучшим вариантом, нежели привычный термин-перевод. Однако приходится считаться с устоявшимися традициями.

Уточним далее, что если говорится об элементах (строках или столбцах) определителя, то имеются в виду элементы (строки или столбцы) соответствующей квадратной матрицы.

Стоит упомянуть также о том, что с первых занятий по курсу аналитической геометрии студенты-первокурсники знакомятся с определителями *второго и третьего порядков*, т. е. с определителями для матриц размеров 2×2 и 3×3. (О матрицах при этом речь не идет. Определители выступают лишь в роли удобных сокращений для громоздких выражений.)

Данное выше определение значительно глубже и сложнее утилитарных обозначений из элементарной аналитической геометрии. Поэтому мы предпочли отложить (фактически на полсеместра) знакомство с алгебраической сущностью определителей.

До сих пор мы обходились при решении систем линейных уравнений и в других задачах методом Гаусса. (И он прекрасно справлялся!) Но теперь пришло время разобрать другой подход.

Замечание 23.2. Формула (23.3) определяет отображение

$$\det : L(n, \mathbb{R}) \longrightarrow \mathbb{R};\ A \mapsto \det(A);\ A \in L(n, \mathbb{R}). \tag{23.4}$$

Свойства этого отображения будут изучаться в следующих параграфах. Но некоторые из них совершенно очевидны уже сейчас.

Например:

$$\det(\underset{n \times n}{O}) = 0; \tag{23.5}$$

$$\det(\underset{n\times n}{E}) = 1. \tag{23.6}$$

Свойство (23.5) пояснений не требует, а свойство (23.6) — требует: надо понять, что для $A = E$ в сумме (23.3) все члены обращаются в 0, кроме одного, соответствующего $\sigma = \varepsilon$ члена: $a_{11}\cdot a_{22}\cdot\ldots\cdot a_{nn} = 1$.

Подобная ситуация будет иметь место и в более общем случае, рассматриваемом в п. 23.3.

23.2. Определители малых порядков. Рассмотрим определители первого, второго и третьего порядков (хорошо уже вам знакомые). Как они получаются исходя из общего определения 23.1?

Обратимся еще раз к примерам 16.2 и 20.2 со списками всех перестановок малой степени.

Квадратная матрица размера 1×1 содержит всего один элемент:

$$A = (\, a_{11} \,).$$

И группа перестановок S_1 содержит всего один элемент — тождественную перестановку. Сумма (23.3) сводится к единственному слагаемому, совпадающему с единственным элементом матрицы:

$$\det(A) = a_{11}.$$

В этом тривиальном случае одно из двух обозначений определителя (использующее "палочки") малопригодно (ввиду неизбежной путаницы со знаком модуля).

Возьмем теперь квадратную матрицу размера 2×2:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Группа S_2 содержит два элемента; сумма (23.3) включает два слагаемых:

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Рассмотрим (3×3)-матрицу

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Группа S_3 содержит 6 элементов (3 четные перестановки и 3 нечетные); столько же слагаемых включает сумма (23.3); половина из них снабжены знаком плюс, остальные — знаком минус:

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \\ = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + \\ + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{21}.$$

Остается дать совет, которому вряд ли кто последует: выписать полное разложение (все 24 члена) для определителя четвертого порядка (см. перечисление перестановок степени 4 в примере 16.2).

Впрочем, можно поручить эту задачу Maple.

23.3. Определитель треугольной матрицы

Определение 23.2. Квадратная матрица называется *верхней* (*нижней*) *треугольной*, если все ее элементы, расположенные ниже (выше) главной диагонали, равны нулю.

Общий вид верхней треугольной матрицы:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1(n-1)} & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2(n-1)} & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3(n-1)} & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & a_{(n-1)(n-1)} & a_{(n-1)n} \\ 0 & 0 & 0 & \cdots & 0 & a_{nn} \end{pmatrix}.$$

Диагональная матрица

$$A = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

относится как к верхним, так и к нижним треугольным. Частным случаем диагональных являются *скалярные* матрицы:

$$A = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda \end{pmatrix} = \lambda E.$$

Предложение 23.1. *Определитель верхней (нижней) треугольной, в частности диагональной, матрицы равен произведению диагональных элементов. Определитель скалярной матрицы (порядка n)*

$$\det(\lambda E) = \lambda^n. \tag{23.7}$$

Доказательство. Пусть, для определенности, рассматривается верхняя треугольная матрица. (Доказательство для случая нижних треугольных матриц отличается лишь "направлением движения": "сверху вниз" вместо "снизу вверх".)

Каждый член определителя отвечает некоторой перестановке σ степени n и является произведением вида (23.2).

Для того чтобы это произведение могло быть отличным от нуля, необходимо, чтобы из последней строки выбирался последний элемент, т. е. чтобы $\sigma(n) = n$. Если это так, то последний столбец уже "занят", и снова, для того чтобы произведение могло быть ненулевым, необходимо, чтобы из предпоследней строки выбирался предпоследний (диагональный) элемент, т. е. должно выполняться $\sigma(n-1) = n-1$; и т. д. Из второй строки будет выбираться элемент a_{22}, из первой — элемент a_{11}.

Получается, что ненулевым может быть лишь член определителя, отвечающий тождественной перестановке $\sigma = \varepsilon$. Окончательная формула (справедливая для обоих типов треугольных и, в частности, для диагональных матриц):

$$\det(A) = a_{11} \cdot a_{22} \cdot \ldots \cdot a_{(n-1)(n-1)} \cdot a_{nn}. \tag{23.8}$$

Формула (23.7) вытекает из (23.8). □

23.4. Определитель транспонированной матрицы. Одним из основных свойств определителя является то, что он не меняется при транспонировании матрицы.

Предложение 23.2. *Для любой квадратной матрицы A (размера $n \times n$)*

$$\det(A^t) = \det(A). \tag{23.9}$$

Доказательство. Пусть $B = A^t$. По определению транспонированной матрицы [см. (2.1)], $b_{ij} = a_{ji}$ $(i, j = 1, ..., n)$. По определению определителя,

$$\begin{aligned}\det(B) &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \dots b_{n\sigma(n)} = \\ &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1} a_{\sigma(2)\,2} \dots a_{\sigma(n)\,n} = \\ &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma^{-1}) a_{1\sigma^{-1}(1)} a_{2\sigma^{-1}(2)} \dots a_{n\sigma^{-1}(n)} = \\ &= \sum_{\lambda\in S_n} \operatorname{sgn}(\lambda) a_{1\lambda(1)} a_{2\lambda(2)} \dots a_{n\lambda(n)} = \det(A).\end{aligned}$$

Выше, в цепочке равенств, на предпоследнем шаге мы заметили, что, в силу свойства знака (20.12), общий член второй суммы является членом определителя матрицы A, отвечающим перестановке

$$\begin{aligned}\sigma^{-1} &= \begin{pmatrix} \sigma(1) & \sigma(2) & \cdots & \sigma(n) \\ 1 & 2 & \cdots & n \end{pmatrix} = \\ &= \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma^{-1}(1) & \sigma^{-1}(2) & \cdots & \sigma^{-1}(n) \end{pmatrix}.\end{aligned}$$

(После того как мы расположили по порядку номера в верхней строке двустрочной записи перестановки σ^{-1}, нам пришлось переставить скалярные сомножители в произведении, что законно в силу коммутативного закона для умножения скаляров.)

На последнем шаге мы произвели *замену переменной суммирования*: $\lambda = \sigma^{-1}$, что также законно в силу того, что отображение обращения является биекцией группы S_n на себя (см. предложение 16.2). □

Замечание 23.3. Доказанное выше предложение можно понимать как утверждение *равноправия строк и столбцов* в определителе: все, что можно будет сказать о столбцах определителя, будет справедливо и в отношении его строк. В самом деле, при транспонировании строки и столбцы меняются ролями, а определитель не изменяется.

Мы неоднократно будем пользоваться этим обстоятельством. В частности, немедленно выпишем вторую формулу для вычисления определителей, равносильную формуле (23.3):

$$\det(A) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \cdot a_{\sigma(1)\,1} \cdot a_{\sigma(2)\,2} \cdot \ldots \cdot a_{\sigma(n)\,n}. \tag{23.3t}$$

В отличие от формулы (23.3), в формуле (23.3t) сомножители упорядочены так, что подряд идут номера столбцов, а номера строк задаются перестановкой σ.

§ 24. Определитель квадратной матрицы как полилинейная и антисимметрическая функция ее столбцов (строк)

24.1. Функции от векторов-столбцов (векторов-строк) квадратной матрицы. Скалярная функция от n-мерного вектора есть отображение $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, обладающее свойствами

$$1)\ f(\bar{a}+\bar{b}) = f(\bar{a}) + f(\bar{b});\quad 2)\ f(\lambda\bar{a}) = \lambda f(\bar{a})$$

для любых векторов $\bar{a}, \bar{b} \in \mathbb{R}^n$ и любого скаляра $\lambda \in \mathbb{R}$. (Это — не новое определение, а частный случай общего понятия линейного отображения; см. определение 15.1.)

Рассмотрим теперь функцию от n штук n-мерных векторов, т. е. отображение

$$f : \underbrace{\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n \text{ раз}} \longrightarrow \mathbb{R};$$

$$(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n) \mapsto f(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n), \quad (24.1)$$

сопоставляющее каждому набору из n векторов $(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n)$ определенный скаляр $f(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n)$.

Определение 24.1. Функция (24.1) называется *полилинейной*, если она линейна по каждому из своих аргументов (при фиксированных остальных), т. е. если для любого номера $j = 1, ..., n$ (и для любых значений участвующих в формулах векторов и скаляров) справедливы свойства:

$$f(\bar{a}_1, \ldots, \bar{a}'_j + \bar{a}''_j, \ldots, \bar{a}_n) = \\ = f(\bar{a}_1, \ldots, \bar{a}'_j, \ldots, \bar{a}_n) + f(\bar{a}_1, \ldots, \bar{a}''_j, \ldots, \bar{a}_n); \quad (24.2)$$

$$f(\bar{a}_1, \ldots, \lambda\bar{a}_j, \ldots, \bar{a}_n) = \lambda f(\bar{a}_1, \ldots, \bar{a}_j, \ldots, \bar{a}_n). \quad (24.3)$$

Замечание 24.1. Из полилинейности функции [точнее, из (24.3)] следует, что если какой-либо из аргументов функции принимает нулевое значение (равен нулевому вектору), то и значение функции равно нулю.

Определение 24.2. Функция (24.1) называется *антисимметрической*, если она меняет знак при перестановке любых двух своих аргументов, т. е. если для любых двух различных номеров $j, k = 1, ..., n$ $(j \neq k)$ и для любых значений участвующих в формуле векторов справедливо свойство

$$f(\bar{a}_1, \ldots, \underset{(j)}{\bar{a}_k}, \ldots, \underset{(k)}{\bar{a}_j}, \ldots, \bar{a}_n) = -f(\bar{a}_1, \ldots, \underset{(j)}{\bar{a}_j}, \ldots, \underset{(k)}{\bar{a}_k}, \ldots, \bar{a}_n). \quad (24.4)$$

Замечание 24.2. Из антисимметричности функции следует, что если какие-либо два из аргументов функции принимают одинаковые значения, то значение функции равно нулю.

В самом деле, пусть $\bar{a}_j = \bar{a}_k$ $(j \neq k)$. Тогда, с одной стороны, при перестановке этих векторов значение функции не изменится, а с другой стороны, оно должно изменить знак, что возможно лишь в случае, когда значение равно нулю.

Замечание 24.3. Обратим внимание на то, что в определении антисимметрической функции фактически фигурируют *транспозиции* аргументов (или их номеров). Каждая такая транспозиция меняет знак значения функции на противоположный.

Можно рассмотреть произвольную перестановку $\sigma \in S_n$ и значение функции $f(\bar{a}_{\sigma(1)}, \bar{a}_{\sigma(2)}, \ldots, \bar{a}_{\sigma(n)})$ на наборе векторов, полученном из исходного набора с помощью перестановки σ.

Каждую перестановку можно разложить в произведение транспозиций (см. предложение 19.2); пусть, скажем, имеется разложение σ в произведение p транспозиций. Действие каждой из них на аргументы фунции изменяет знак значения функции. В итоге набежит множитель $(-1)^p$, равный (см. определение 20.1) знаку $\operatorname{sgn}(\sigma)$.

Тем самым справедливо следующее свойство антисимметрических функций:

$$f(\bar{a}_{\sigma(1)}, \bar{a}_{\sigma(2)}, \ldots, \bar{a}_{\sigma(n)}) = \operatorname{sgn}(\sigma) f(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n). \quad (24.5)$$

Выше рассматривались векторы-столбцы из пространства $\mathbb{R}^n$, их наборы и функции от наборов из n векторов-столбцов. Разумеется, совершенно аналогичным образом можно рассматривать векторы-строки из пространства $\overset{*}{\mathbb{R}}{}^n$ (см. замечание 2.2) и функции

$$f : \underbrace{\overset{*}{\mathbb{R}}{}^n \times \overset{*}{\mathbb{R}}{}^n \times \cdots \times \overset{*}{\mathbb{R}}{}^n}_{n \text{ раз}} \longrightarrow \mathbb{R};$$

$$((\bar{a}^1)^t, (\bar{a}^2)^t, \ldots, (\bar{a}^n)^t) \mapsto f((\bar{a}^1)^t, (\bar{a}^2)^t, \ldots, (\bar{a}^n)^t) \quad (24.1\text{t})$$

от наборов из n таких векторов. (Заметьте, что номера векторов-строк мы размещаем как *верхние индексы*.) Все данные выше определения (и следствия из них) переносятся на этот случай без каких-либо изменений.

Оба типа векторов и функций возникают, если рассматривать множество $L(n,\mathbb{R})$ квадратных матриц порядка n и определенные на этом множестве функции

$$f : L(n,\mathbb{R}) \longrightarrow \mathbb{R};\ A \mapsto f(A);\ A \in L(n,\mathbb{R}). \tag{24.6}$$

Всякую функцию (24.6) можно представить как функцию (24.1), если матрицу A представить разбитой на векторы-столбцы:

$$A = (\bar{a}_1\ |\bar{a}_2\ |...\ |\bar{a}_n)\,, \tag{24.7}$$

а можно — как функцию (24.1t), если разбить матрицу A на векторы-строки:

$$A = \left(\begin{array}{c} (\bar{a}^1)^t \\ \hline (\bar{a}^2)^t \\ \hline \cdots \\ \hline (\bar{a}^n)^t \end{array}\right). \tag{24.7t}$$

В частности, определенную в предыдущем параграфе функцию (23.4), сопоставляющую квадратной матрице ее определитель, можно рассматривать еще двумя способами: как функцию от набора столбцов и как функцию от набора строк этой матрицы.

24.2. Полилинейность и антисимметричность функции $\mathbf{A \longmapsto \det(A)}$

Теорема 24.1. *Функция (23.4) является полилинейной и антисимметрической функцией столбцов (строк) матрицы.*

Доказательство. Строки и столбцы определителя равноправны (см. замечание 23.3). Поэтому теорему достаточно доказать, скажем, в отношении столбцов, и тогда можно будет утверждать, что она справедлива и в отношении строк.

Доказательство для столбцов состоит в проверке свойств (24.2) — (24.4), которые для функции $f = \det$ приобретают вид

$$\det(\bar{a}_1|...|\bar{a}'_j + \bar{a}''_j|...|\bar{a}_n) = \\ = \det(\bar{a}_1|...|\bar{a}'_j|...|\bar{a}_n) + \det(\bar{a}_1|...|\bar{a}''_j|...|\bar{a}_n); \tag{24.8}$$

$$\det(\bar{a}_1|...|\lambda\bar{a}_j|...|\bar{a}_n) = \lambda\det(\bar{a}_1|...|\bar{a}_j|...|\bar{a}_n); \tag{24.9}$$

$$\det(\bar{a}_1|...|\underset{(j)}{\bar{a}_k}|...|\underset{(k)}{\bar{a}_j}|...|\bar{a}_n) = -\det(\bar{a}_1|...|\underset{(j)}{\bar{a}_j}|...|\underset{(k)}{\bar{a}_k}|...|\bar{a}_n). \tag{24.10}$$

Поскольку мы работаем со столбцами, то нам удобнее будет применять вторую формулу для определителя [т. е. формулу (23.3t)].

Пусть матрица A задана формулой (24.7), т. е. разбита на столбцы, и один из этих столбцов представлен в виде суммы двух столбцов:

$$\bar{a}_j = \bar{a}'_j + \bar{a}''_j \tag{24.11}$$

— или в координатах:

$$a_{ij} = a'_{ij} + a''_{ij} \tag{24.11a}$$

для любого $i = 1, ..., n$.

Введем матрицы:

$$A' = (\bar{a}_1\ |...\ \left|\bar{a}'_j\right.\ |...\ |\bar{a}_n)\,;\ A'' = (\bar{a}_1\ |...\ \left|\bar{a}''_j\right.\ |...\ |\bar{a}_n)\,.$$

Равенство (24.8) в этих обозначениях принимает вид

$$\det(A) = \det(A') + \det(A''). \tag{24.8a}$$

Докажем его:

$$\begin{aligned}
\det(A) &\overset{(23.3t)}{=} \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1}...a_{\sigma(j)\,j}...a_{\sigma(n)\,n} = \\
&\overset{(24.11a)}{=} \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1}...(a'_{\sigma(j)\,j} + a''_{\sigma(j)\,j})...a_{\sigma(n)\,n} = \\
&= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1}...a'_{\sigma(j)\,j}...a_{\sigma(n)\,n} + \\
&+ \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1}...a''_{\sigma(j)\,j}...a_{\sigma(n)\,n} = \\
&\overset{(23.3t)}{=} \det(A') + \det(A'').
\end{aligned}$$

Аналогично, но несколько проще доказывается формула (24.9). [Проведите вычисления сами. Можно порекомендовать свести (24.9) к формуле $\det(A') = \lambda\det(A)$, где матрица A' отличается от A одним столбцом: $\bar{a}'_j = \lambda\bar{a}_j$.]

Приступаем к доказательству формулы (24.10), которой можно придать вид

$$\det(A') = -\det(A), \tag{24.10a}$$

где матрица A' получается из матрицы A перестановкой j-го и k-го столбцов. Этот факт можно выразить равенствами

$$\bar{a}'_j = \bar{a}_k;\ \bar{a}'_k = \bar{a}_j;\ \bar{a}'_s = \bar{a}_s\ (s \neq j, k) \tag{24.12}$$

или в координатах (для любого $i = 1, ..., n$):

$$a'_{ij} = a_{ik};\ a'_{ik} = a_{ij};\ a'_{is} = a_{is}\ (s \neq j, k). \tag{24.12a}$$

Снова используем формулу (23.3t) для определителя:

$$\det(A') = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a'_{\sigma(1)\,1} ... a'_{\sigma(j)\,j} ... a'_{\sigma(k)\,k} ... a'_{\sigma(n)\,n} =$$

$$\overset{(24.12a)}{=} \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1} ... a_{\sigma(j)\,k} ... a_{\sigma(k)\,j} ... a_{\sigma(n)\,n} = ...$$

Во второй сумме номера столбцов идут не по порядку, что легко исправить, т. к. скалярные множители можно переставлять. Но и после этого останется "нарушенным" соответствие между номерами столбцов и строк, установленное перестановкой σ.

Это мы исправим введением *новой переменной перестановки* $\rho = \sigma\tau$, где τ — фиксированная транспозиция: $\tau = (j, k)$. Перестановка ρ действует следующим образом:

$$\rho(j) = \sigma(k);\ \rho(k) = \sigma(j);\ \rho(s) = \sigma(s)\ (s \neq j, k). \tag{24.13}$$

По свойствам знака, $\operatorname{sgn}(\sigma) = -\operatorname{sgn}(\rho)$. По предложению 16.2, если переменная перестановка σ пробегает всю группу S_n, то и новая переменная $\rho = r_\tau(\sigma)$ пробегает S_n. Можем продолжить вычисления:

$$... = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1} ... a_{\sigma(k)\,j} ... a_{\sigma(j)\,k} ... a_{\sigma(n)\,n} =$$

$$\overset{(24.13)}{=} \sum_{\rho \in S_n} (-\operatorname{sgn}(\rho)) a_{\rho(1)\,1} ... a_{\rho(j)\,j} ... a_{\rho(k)\,k} ... a_{\sigma(n)\,n} = -\det(A).$$

Что и требовалось доказать. □

Замечание 24.4. Термин *полилинейность* в рассмотренной выше ситуации может быть уточнен (конкретизирован): речь идет о n-линейных функциях. (Может быть, вы помните, что при изучении матричной алгебры — см. п. 2.3 — уже использовался термин *билинейность*.)

Предостережем от возможных иллюзий: полилинейность отнюдь не влечет *линейность*. (Скажем, определитель суммы двух матриц, за исключением тривиальных случаев, не равен сумме определителей.)

Свойство определителя (24.8) (и аналогичное утверждение для строк) иногда читают так: *общий скалярный множитель из столбца (строки) определителя можно выносить за знак определителя.*

Если общий множитель λ имеется у всех элементов определителя (порядка n), то за знак определителя выносится λ^n.

Замечание 24.5. Наряду с антисимметрическими (они еще называются *кососимметрическими*) функциями важную роль играют в алгебре *симметрические* функции. Имеется симметрический аналог определителя, называемый *перманентом.*

24.3. Следствия из свойств полилинейности и антисимметричности определителя. Из двух основных свойств определителя, доказанных в теореме 24.1, и из предложения 23.2 вытекает целый ряд других свойств.

Предложение 24.1. *Функция* $\det(A)$ *обладает следующими свойствами:*

(1) Если в матрице A имеется нулевой столбец (нулевая строка), то $\det(A) = 0$.

(2) Если в матрице A имеются два одинаковых (пропорциональных) столбца [две одинаковых (пропорциональных) строки], то $\det(A) = 0$.

(3) Определитель матрицы не изменится, если матрицу подвергнуть элементарным преобразованиям второго типа над столбцами (над строками).

(4) Для любой квадратной матрицы A и любого скаляра λ справедлива формула

$$\det(\lambda A) = \lambda^n \det(A);\ A \in L(n, \mathbb{R}). \tag{24.14}$$

Доказательство. Свойство (1) справедливо для любых полилинейных функций (см. замечание 24.1).

Свойство (2) справедливо для любых полилинейных антисимметрических функций, но рассуждение мы проведем для функции det, применительно к столбцам.

Пусть в матрице A один столбец пропорционален другому. Пользуясь полилинейностью, вынесем коэффициент пропорциональности за знак определителя. В оставшемся определителе окажутся два одинаковых столбца. В силу антисимметричности (см. замечание 24.2) этот определитель будет равен нулю.

Аналогичным образом дело обстоит и со свойством (3). На этот раз мы оформим рассуждение подробнее.

Элементарное преобразование типа II над столбцами:

$$A = (\bar{a}_1|...|\bar{a}_j|...|\bar{a}_k|...|\bar{a}_n) \xrightarrow{j^{\text{стб}}+k^{\text{стб}}\cdot\lambda} (\bar{a}_1|...|\bar{a}_j + \lambda\bar{a}_k|...|\bar{a}_k|...|\bar{a}_n) = A'.$$

Значение определителя:

$$\det(A') \underset{(24.9)}{\overset{(24.8)}{=\!=}} \det(A) + \lambda \det(\bar{a}_1|...|\bar{a}_k|...|\bar{a}_k|...|\bar{a}_n) =$$

$$\overset{(24.10)}{=\!=} \det(A) + \lambda \cdot 0 = \det(A).$$

Свойство (5) объяснено в замечании 24.4. □

24.4. Метод Гаусса вычисления определителей. При практическом вычислении определителей *формулы полного разложения* (23.3) и (23.3t) применяются очень редко ввиду их чрезвычайно высокой вычислительной трудоемкости.

(В каждой из этих формул $n!$ слагаемых, в каждом из которых n сомножителей; функция $n!$ с увеличением n растет быстрее показательной. Точный смысл последних терминов станет вам известен значительно позже, однако интуитивно вы должны чувствовать вычислительную неприемлемость формул, определяющих определитель.)

Впрочем, из изученных выше свойств определителей ясно, что при их вычислении можно применять хорошо знакомый нам метод Гаусса, хотя в данной ситуации применение этого метода имеет ряд очень существенных особенностей.

Основу метода Гаусса составляют элементарные преобразования типов I — III над строками и столбцами матрицы.

Можно ли применять преобразования типа I (перестановку строк или столбцов) при вычислении определителей? Да, но с непременным учетом антисимметричности: при перестановке строк (столбцов) надо менять знак определителя на противоположный.

Элементарные преобразования типа II (как над строками, так и над столбцами) применимы безоговорочно [см. выше свойство (3) в предложении 24.1].

Преобразования типа III выступают теперь в форме вынесения за знак определителя общего множителя из строки (столбца) (см. замечание 24.4; лишний раз подчеркнем, что общий множитель можно вынести, но не выбросить).

С помощью описанных выше версий элементарных преобразований квадратную матрицу следует привести к треугольному виду, а затем вычислить последний определитель как произведение диагональных элементов (см. предложение 24.1), не забыв при этом про возможно набежавший перед знаком определителя скалярный множитель.

Пример 24.1.

$$\begin{vmatrix} 5 & 6 & 13 \\ 3 & 4 & 5 \\ 1 & 2 & 3 \end{vmatrix} \xrightarrow{1^{\text{стр}} \leftrightarrow 3^{\text{стр}}} - \begin{vmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 5 & 6 & 13 \end{vmatrix} \xrightarrow[3^{\text{стр}}+1^{\text{стр}}\cdot(-5)]{2^{\text{стр}}+1^{\text{стр}}\cdot(-3)}$$

$$= - \begin{vmatrix} 1 & 2 & 3 \\ 0 & -2 & -4 \\ 0 & -4 & -2 \end{vmatrix} = -(-2)\cdot(-2)\cdot \begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{vmatrix} \xrightarrow{3^{\text{стр}}+2^{\text{стр}}\cdot(-2)}$$

$$= (-4)\cdot \begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & -3 \end{vmatrix} = 12.$$

§ 25. Вычисление определителя с помощью разложения по столбцу (строке)

25.1. Алгебраические дополнения к элементам квадратной матрицы. Рассмотрим квадратную матрицу A размера $n \times n$. Выберем какой-либо элемент a_{ij} из этой матрицы.

Определение 25.1. *Алгебраическим дополнением* к элементу a_{ij} матрицы A называется скаляр A_{ij}, равный взятому со знаком

$(-1)^{i+j}$ определителю матрицы размера $(n-1)\times(n-1)$, полученной из матрицы A вычеркиванием i-й строки и j-го столбца.

Замечание 25.1. Данное выше определение можно выразить формулой:

$$A_{ij} = (-1)^{i+j} M_{ij};$$

$$M_{ij} = \begin{vmatrix} a_{11} & \dots & a_{1(j-1)} & a_{1(j+1)} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{(i-1)1} & \dots & a_{(i-1)(j-1)} & a_{(i-1)(j+1)} & \dots & a_{(i-1)n} \\ a_{(i+1)1} & \dots & a_{(i+1)(j-1)} & a_{(i+1)(j+1)} & \dots & a_{(i+1)n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{n(j-1)} & a_{n(j+1)} & \dots & a_{nn} \end{vmatrix}. \quad (25.1)$$

Заметьте, что A_{ij} никак не зависит ни от элемента a_{ij}, ни от других элементов, расположенных в той же строке или в том же столбце, что a_{ij}. Можно сказать иначе: алгебраическое дополнение к элементу зависит не от самого элемента, а от его позиции в матрице.

В формуле (25.1) фигурирует также определитель M_{ij} (без знакового множителя). У него есть свое название — *минор* $(n-1)$-го порядка для (квадратной) матрицы A. Позже (см. § 30) мы познакомимся с минорами различных порядков (для произвольных прямоугольных матриц).

Сейчас мы сформулируем и докажем вспомогательное утверждение об определителях специального вида. Рассмотрим $(n\times n)$-матрицу A, имеющую в первом столбце все нулевые элементы, начиная со второго:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & a_{n2} & \dots & a_{nn} \end{pmatrix}. \quad (25.2)$$

Рассмотрим также блок размера $(n-1)\times(n-1)$

$$A' = \begin{pmatrix} a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots \\ a_{n2} & \dots & a_{nn} \end{pmatrix}. \quad (25.3)$$

Если нумеровать строки и столбцы матрицы A' как обычно, начиная с 1, то мы получим:

$$a'_{ij} = a_{(i+1)(j+1)};\ i,j = 1,...,n-1. \quad (25.4)$$

Определитель блока (25.3) есть не что иное, как алгебраическое дополнение к элементу a_{11} в определителе матрицы A:

$$\det(A') = A_{11}.$$

Предложение 25.1. *Определитель матрицы (25.2) равен произведению элемента a_{11} на его алгебраическое дополнение:*

$$\det(A) = a_{11}A_{11}. \tag{25.5}$$

Доказательство. По общей формуле (23.3t) определитель матрицы (25.2) равен сумме $n!$ членов вида

$$\operatorname{sgn}(\sigma)a_{\sigma(1)\,1}a_{\sigma(2)\,2}\dots a_{\sigma(n)\,n}, \tag{25.6}$$

каждый из которых соответствует некоторой перестановке $\sigma \in S_n$. Но член (25.6) будет нулевым, если $\sigma(1) \neq 1$. Поэтому реальный вклад в сумму могут внести лишь $(n-1)!$ перестановок вида

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ 1 & \sigma(2) & \dots & \sigma(n) \end{pmatrix}. \tag{25.7}$$

Перестановка (25.7) (степени n) однозначно определяется следующей перестановкой (степени $n-1$):

$$\sigma' = \begin{pmatrix} 1 & \dots & n-1 \\ \sigma'(1) & \dots & \sigma'(n-1) \end{pmatrix}, \tag{25.7a}$$

где

$$\sigma'(j) = \sigma(j+1) - 1;\ j = 1, \dots, n-1. \tag{25.8}$$

При этом σ' пробегает всю группу S_{n-1} и $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma')$ (хотя бы потому, что все инверсии перестановки σ сосредоточены в зоне номеров от 2 до n).

С учетом (25.4) и (25.8) получим:

$$\begin{aligned} \det(A) &= \sum_{\substack{\sigma\in S_n \\ \sigma(1)=1}} \operatorname{sgn}(\sigma)a_{11}a_{\sigma(2)\,2}\dots a_{\sigma(n)\,n} = \\ &= a_{11} \sum_{\sigma'\in S_{n-1}} \operatorname{sgn}(\sigma')a'_{\sigma'(1)\,1}\dots a'_{\sigma'(n-1)\,(n-1)} = \\ &= a_{11}\det(A') = a_{11}A_{11}. \end{aligned}$$

Формула (25.5) доказана. □

25.2. Теорема Лапласа о вычислении определителя разложением по строке (столбцу). Сейчас мы опишем (в форме теоремы) давно знакомую вам индуктивную процедуру представления определителя n-го порядка в виде линейной комбинации n определителей $(n-1)$-го порядка.

Теорема 25.1 (теорема Лапласа). *Определитель квадратной матрицы $A \in L(n, \mathbb{R})$ равен сумме произведений элементов какого-либо столбца (какой-либо строки) определителя на соответствующие алгебраические дополнения, т. е.*

$$\det(A) = \sum_{i=1}^{n} a_{ij} A_{ij} \tag{25.9}$$

для любого столбца с номером $j = 1, ..., n$, и

$$\det(A) = \sum_{j=1}^{n} a_{ij} A_{ij} \tag{25.9t}$$

для любой строки с номером $i = 1, ..., n$.

Доказательство. Докажем утверждение для разложения по столбцу. (Случай разложения по строке рассматривается совершенно аналогично.)

Столбец $\bar{a}_j \in \mathbb{R}^n$, входящий в матрицу A, разложим по естественному базису в $\mathbb{R}^n$:

$$\bar{a}_j = \sum_{i=1}^{n} a_{ij} \bar{e}_i; \;\; i = 1, ..., n \tag{25.10}$$

— и воспользуемся полилинейностью определителя [см. формулы (24.8) и (24.9)]:

$$\det(A) = \det(\bar{a}_1 | ... | \bar{a}_j | ... | \bar{a}_n) =$$

$$= \det(\bar{a}_1 | ... | \sum_{i=1}^{n} a_{ij} \bar{e}_i | ... | \bar{a}_n) = \sum_{i=1}^{n} a_{ij} \det(\bar{a}_1 | ... | \bar{e}_i | ... | \bar{a}_n) =$$

$$= \sum_{i=1}^{n} a_{ij} \det \begin{pmatrix} a_{11} & ... & 0 & ... & a_{1n} \\ ... & ... & ... & ... & ... \\ a_{i1} & ... & 1 & ... & a_{in} \\ ... & ... & ... & ... & ... \\ a_{n1} & ... & 0 & ... & a_{nn} \end{pmatrix} = ...$$

В последнем определителе переставим i-ю строку с каждой из предыдущих, пока она не займет первую позицию; затем аналогичным образом переставим j-й столбец, пока он не станет первым. По свойству антисимметричности определителя [см. формулу (24.10)] при этом "набежит" знаковый множитель

$$(-1)^{i-1}(-1)^{j-1} = (-1)^{i+j-2} = (-1)^{i+j};$$

в левом верхнем углу определителя окажется единица, алгебраическое дополнение к которой (в новом определителе) будет совпадать с минором M_{ij} [для старого определителя; см. формулу (25.1)].

На последнем шаге останется применить предложение 25.1:

$$\ldots = \sum_{i=1}^{n} a_{ij} \cdot (-1)^{i+j} \det \left(\begin{array}{ccc|c|cc} 1 & a_{i1} & \ldots & & \ldots & a_{in} \\ 0 & a_{11} & \ldots & & \ldots & a_{1n} \\ \ldots & \ldots & \ldots & & \ldots & \ldots \\ \hline\hline \ldots & \ldots & \ldots & & \ldots & \ldots \\ 0 & a_{n1} & \ldots & & \ldots & a_{nn} \end{array} \right) =$$

$$= \sum_{i=1}^{n} a_{ij} \cdot (-1)^{i+j} M_{ij} = \sum_{i=1}^{n} a_{ij} A_{ij}.$$

Формула (25.10) доказана. □

Замечание 25.2. Доказанная выше теорема является простейшим частным случаем принадлежащей Лапласу теоремы о разложении определителя *по нескольким строкам* (см., например, [9, гл. 4, § 5]).

25.3. Еще одно свойство алгебраических дополнений. Если взять сумму произведений элементов какого-либо столбца (какой-либо строки) определителя на соответствующие алгебраические дополнения, то, как утверждает теорема Лапласа, получится значение этого определителя. Оказывается важным выяснить, что получится, если взять сумму произведений элементов некоторого столбца (некоторой строки) определителя на алгебраические дополнения к соответствующим элементам другого столбца (другой строки) определителя.

Точный результат дает следующее

Предложение 25.2. *Сумма произведений элементов какого-либо столбца определителя на алгебраические дополнения к соответствующим элементам другого столбца равна нулю. Аналогичное утверждение справедливо для строк определителя.*

Доказательство. Проведем рассуждение для столбцов; рассуждение для строк — совершенно аналогично.

Рассмотрим сумму элементов j-го столбца на ("чужие") алгебраические дополнения к соответствующим элементам k-го столбца ($k \neq j$).

Эту сумму можно рассматривать как разложение по k-му столбцу определителя (с двумя одинаковыми столбцами $\bar{a}_k = \bar{a}_j$):

$$a_{1j}A_{1k} + a_{2j}A_{2k} + ... + a_{nj}A_{nk} = \det(\bar{a}_1|...|\underset{(j)}{\bar{a}_j}|...|\underset{(k)}{\bar{a}_j}|...|\bar{a}_n).$$

(Здесь мы пользуемся тем фактом, что алгебраическое дополнение к элементу определителя зависит не от этого элемента, а от его позиции в матрице; см. замечание 25.1.)

В силу предложения 24.1, последний определитель равен нулю. □

Замечание 25.3. Можно выписать формулы, охватывающие как случай, когда элементы столбца (строки) умножаются на "свои" алгебраические дополнения, так и случай, когда они умножаются на "чужие" алгебраические дополнения:

$$\sum_{i=1}^{n} a_{ij}A_{ik} = \delta_{jk}\det(A); \tag{25.11}$$

$$\sum_{j=1}^{n} a_{ij}A_{kj} = \delta_{ik}\det(A). \tag{25.11t}$$

25.4. Индуктивный алгоритм вычисления определителя. Предположим, требуется вычислить некоторый определитель n-го порядка. С помощью теоремы Лапласа разложим его по какой-либо строке (или по какому-либо столбцу), используя формулу (25.9) или (25.9t). (На практике, при ручном счете, приходится выбирать строку или столбец "попроще", в которых "побольше нулей".) В результате получится линейная комбинация не более чем n определителей $(n-1)$-го порядка. С каждым из них поступаем аналогичным образом, пока не дойдем до определителей второго порядка (а если угодно — то и до первого).

Вот и весь "алгоритм". Беда в том, что его вычислительная сложность — того же уровня, что и у формул полного разложения (23.3) и (23.3t). Таким способом можно подсчитать только определители невысокого порядка. В этом плане разобранный в предыдущем параграфе алгоритм Гаусса — вне конкуренции.

Тем не менее, ввиду теоретической важности индуктивного алгоритма, пересчитаем с его помощью пример 24.1.

Пример 25.1.

$$\begin{vmatrix} 5 & 6 & 13 \\ 3 & 4 & 5 \\ 1 & 2 & 3 \end{vmatrix} \overset{\text{по } 1^{\text{стр}}}{=\!=\!=} 5 \cdot \begin{vmatrix} 4 & 5 \\ 2 & 3 \end{vmatrix} - 6 \cdot \begin{vmatrix} 3 & 5 \\ 1 & 3 \end{vmatrix} + 13 \cdot \begin{vmatrix} 3 & 4 \\ 1 & 2 \end{vmatrix} =$$
$$= 5 \cdot (4 \cdot 3 - 5 \cdot 2) - 6 \cdot (3 \cdot 3 - 5 \cdot 1) + 13 \cdot (3 \cdot 2 - 4 \cdot 1) = 12.$$

Замечание 25.4. Maple легко справляется с задачей вычисления числовых определителей; достаточно дать простую команду:

> det (A);

Гораздо важнее то, что можно вычислять "символьные" определители.

Введем, например, матрицу

> A := matrix ([[1, 1, 1], [alpha, beta, gamma],
[alpha ^ 2, beta ^ 2, gamma ^ 2]]);

$$A := \begin{bmatrix} 1 & 1 & 1 \\ \alpha & \beta & \gamma \\ \alpha^2 & \beta^2 & \gamma^2 \end{bmatrix}$$

(Если у вас смутные воспоминания: "где-то эта матрица уже попадалась", то загляните в первую главу и перечитайте замечание 7.3.)

> det (A);

$$\beta\gamma^2 - \gamma\beta^2 - \alpha\beta^2 + \alpha^2\gamma - \alpha^2\beta$$

> factor(%);

$$-(-\gamma + \beta)(\alpha - \gamma)(\alpha - \beta)$$

Maple вычислил определитель Вандермонда третьего порядка.

После команды "разложить на множители ответ" стала ясна роль этого замечательного определителя: он обращается в нуль тогда и только тогда, когда в списке чисел α, β, γ имеются одинаковые.

Этот результат остается в силе и в случае определителя Вандермонда n-го порядка. Попробуйте сами записать его, а затем — вычислить (в общем виде, с помощью метода Гаусса).

Если не получится, то загляните в дополнительный § 30а.

§ 26. Описание всех полилинейных и антисимметрических функций от столбцов (строк) квадратной матрицы

26.1. Теорема о полилинейных и антисимметрических функциях от столбцов (строк) квадратной матрицы. В теореме 24.1 мы доказали, что функция $\det(\underset{n\times n}{A})$ является полилинейной и антисимметрической, если ее рассматривать как функцию от n векторов-столбцов (векторов-строк) матрицы A.

Оказывается, однако, что свойства полилинейности и антисимметричности "почти однозначно" (а именно однозначно с точностью до скалярного множителя) определяют функцию от столбцов квадратной матрицы. (Далее мы будем говорить только о столбцах, но вы должны понимать, что все сказанное может быть отнесено и к строкам матрицы.)

Теорема 26.1. *Пусть $f : L(n,\mathbb{R}) \longrightarrow \mathbb{R}$ полилинейная и антисимметрическая функция от столбцов квадратной матрицы. Тогда для любой матрицы $A \in L(n,\mathbb{R})$ справедлива формула*

$$f(A) = \det(A) \cdot f(E), \tag{26.1}$$

т. е. функции f и det *отличаются лишь постоянным скалярным множителем, равным значению функции f на единичной матрице $E \in L(n,\mathbb{R})$.*

Доказательство. Разобьем данную матрицу $A \in L(n,\mathbb{R})$ на столбцы $\bar{a}_k$ $(k = 1, ..., n)$. Функция f линейна по каждому из $\bar{a}_k$.

Представим первый столбец его разложением по естественному базису $\mathcal{E}$ в пространстве $\mathbb{R}^n$, т. е. запишем его в виде линейной комбинации:

$$\bar{a}_1 = \sum_{i_1=1}^{n} a_{i_1 1}\bar{e}_{i_1}.$$

В силу полилинейности f, получим:

$$f(A) = f(\sum_{i_1=1}^{n} a_{i_1 1}\bar{e}_{i_1}, \bar{a}_2, ..., \bar{a}_n) = \sum_{i_1=1}^{n} a_{i_1 1} f(\bar{e}_{i_1}, \bar{a}_2, ..., \bar{a}_n).$$

Аналогичным образом разложим второй столбец:

$$\bar{a}_2 = \sum_{i_2=1}^{n} a_{i_2 2}\bar{e}_{i_2}.$$

После подстановки полученного разложения в предыдущее выражение для $f(A)$ и внесения множителя $a_{i_1 1}$ под знак второй суммы получим *двойную сумму* (содержащую уже n^2 слагаемых):

$$f(A) = \sum_{i_1=1}^{n} a_{i_1 1} \sum_{i_2=1}^{n} a_{i_2 2} f(\bar{e}_{i_1}, \bar{e}_{i_2}, ..., \bar{a}_n) =$$
$$= \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} a_{i_1 1} a_{i_2 2} f(\bar{e}_{i_1}, \bar{e}_{i_2}, ..., \bar{a}_n).$$

И так далее. (Обратите внимание на то, что для корректной записи двойной суммы индексы суммирования в двух суммах должны быть разными: при разложении первого столбца мы использовали индекс i_1, а при разложении второго — индекс i_2. Так же мы будем поступать и далее.)

После n-кратного применения описанной процедуры мы получим *n-кратную* сумму (содержащую n^n слагаемых):

$$f(A) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} a_{i_1 1} a_{i_2 2} ... a_{i_n n} f(\bar{e}_{i_1}, \bar{e}_{i_2}, ..., \bar{e}_{i_n}) =$$
$$= \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} a_{i_1 1} a_{i_2 2} ... a_{i_n n} f(E'), \qquad (26.2)$$

где

$$E' = (\bar{e}_{i_1} | \bar{e}_{i_2} | ... | \bar{e}_{i_n}) \qquad (26.3)$$

— матрица, составленная из векторов базиса $\mathcal{E}$ (не обязательно идущих по порядку и не обязательно различных).

Из формулы (26.2) следует, что значение функции f на произвольной $(n \times n)$-матрице A будет однозначно определено, если будут определены значения f на матрицах вида (26.3).

Если в матрице (26.3) имеются одинаковые столбцы, то, в силу антисимметричности функции f (см. замечание 24.2), значение $f(E') = 0$.

В противном случае, т. е. если номера $i_1, i_2, ..., i_n$ попарно различны и, следовательно, задают некоторую перестановку

$$\sigma = \begin{pmatrix} 1 & 2 & ... & n \\ i_1 & i_2 & ... & i_n \end{pmatrix};\ \sigma(k) = i_k;\ k = 1, ..., n, \qquad (26.4)$$

свойство (24.6) кососимметрических функций позволит нам заключить, что

$$f(E') = \operatorname{sgn}(\sigma) f(E). \tag{26.5}$$

Выбросив в n-кратной сумме (26.2) нулевые слагаемые, отвечающие матрицам E' с повторяющимися столбцами, мы получим сумму, содержащую $n!$ слагаемых, каждое из которых соответствует некоторой перестановке (26.4):

$$f(A) = \sum_{\sigma \in S_n} a_{i_1 1} a_{i_2 2} \ldots a_{i_n n} \operatorname{sgn}(\sigma) f(E) =$$

$$= \left(\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1} a_{\sigma(2)\,2} \ldots a_{\sigma(n)\,n} \right) \cdot f(E) \overset{(23.3t)}{=} \det(A) \cdot f(E).$$

Формула (26.1) доказана. □

26.2. Аксиоматический подход к определению определителя. Итак, произвольная полилинейная и антисимметрическая функция $f(A)$ от столбцов квадратной матрицы A отличается от определителя $\det(A)$ лишь постоянным (не зависящим от A) множителем, равным значению $f(E)$.

В частности, если выполнено *условие нормировки*

$$f(E) = 1, \tag{26.6}$$

то функция f однозначно определена — она тождественно совпадает с функцией $\det$.

Это служит основанием для *аксиоматического* подхода к построению теории определителей.

Можно определить определитель как такую функцию от столбцов квадратной матрицы, которая

1) полилинейна,

2) антисимметрична,

3) удовлетворяет условию нормировки.

Далее, рассуждением, практически идентичным доказательству теоремы 26.1, можно установить, что такая функция существует и однозначно определена.

Формулы полного разложения (23.3) и (23.3t), фигурировавшие ранее как определения, при этом подходе получались бы как следствия.

Замечание 26.1. Корни теории определителей во многом уходят в плодородную почву геометрии. Определители n-го порядка служат для вычисления n-мерных *ориентированных объемов*. Точнее, $\det(\bar{a}_1|\bar{a}_2|...|\bar{a}_n)$ равняется ориентированному объему n-мерного параллелепипеда, построенного на векторах $\bar{a}_1, \bar{a}_2, ..., \bar{a}_n$ в n-мерном евклидовом пространстве.

(Одномерный параллелепипед — это отрезок на прямой, его *одномерный ориентированный объем* — это длина, взятая со знаком плюс или минус в зависимости от ориентации отрезка; двумерный параллелепипед — это параллелограмм на плоскости, его *двумерный ориентированный объем* — это площадь, тоже со знаком, зависящим от ориентации пары векторов, задающих параллелограмм; и т. д.)

Очень хорошо о применении определителей для вычисления объемов рассказано в учебниках [1, 5]. Прочитайте!

§ 27. Определитель блочно-треугольной матрицы. Определитель произведения матриц

27.1. Определитель блочно-треугольной матрицы. С блочными матрицами мы уже встречались в п. 5.1 [см. формулу (5.5)].

Теперь мы рассмотрим квадратную матрицу (размера $n \times n$, где $n = m + l$)

$$\underset{n\times n}{A} = \left(\begin{array}{c|c} \underset{m\times m}{B} & \underset{m\times l}{D} \\ \hline \underset{l\times m}{O} & \underset{l\times l}{C} \end{array}\right), \tag{27.1}$$

разбитую на четыре блока, два из которых (диагональные) являются квадратными. Нижний внедиагональный блок является нулевым.

Определение 27.1. Матрица вида (27.1) называется *верхней блочно-треугольной*.

Нижняя блочно-треугольная матрица имеет аналогичный вид:

$$\underset{n\times n}{A} = \left(\begin{array}{c|c} \underset{m\times m}{B} & \underset{m\times l}{O} \\ \hline \underset{l\times m}{D} & \underset{l\times l}{C} \end{array}\right), \tag{27.1t}$$

В частном случае $D = O$ матрица (27.1) [или (27.1t)] называется *блочно-диагональной*.

Теорема 27.1. *Определитель верхней (нижней) блочно-треугольной матрицы равен произведению определителей диагональных блоков:*

$$\det(A) = \det(B) \cdot \det(C). \tag{27.2}$$

Доказательство. При транспонировании верхняя блочно-треугольная матрица превращается в нижнюю блочно-треугольную и наоборот. А определитель при транспонировании не меняется. Поэтому достаточно доказать теорему лишь для одного из типов блочно-треугольных матриц.

Рассмотрим матрицы типа (27.1t). Зафиксируем блоки B и D и будем представлять $\det(A)$ как функцию квадратной $(l \times l)$-матрицы C:

$$f(C) = \det \left(\begin{array}{c|c} \underset{m\times m}{B} & \underset{m\times l}{O} \\ \hline \underset{l\times m}{D} & \underset{l\times l}{C} \end{array} \right). \tag{27.3}$$

Докажем, что функция (27.3) является полилинейной и антисимметрической.

Пусть какой-либо столбец $\bar{c}$ в матрице C представлен как сумма двух столбцов: $\bar{c} = \bar{c}' + \bar{c}''$. Рассмотрим соответствующие матрицы C' и C'', введя их по тому же принципу, что и при доказательстве теоремы 24.1.

Требуется доказать равенство

$$f(C) = f(C') + f(C''). \tag{27.4}$$

Поскольку столбцы матрицы C в блочной матрице A продолжаются (вверх) нулями, то можно считать, что и в матрице A соответствующий столбец $\bar{a}$ представлен как сумма столбцов: $\bar{a} = \bar{a}' + \bar{a}''$ (верхние части столбцов $\bar{a}'$ и $\bar{a}''$ — нулевые). Сформируем матрицы A' и A'' (напомним, что они отличаются от A лишь по одному столбцу).

В силу теоремы 24.1, определитель является полилинейной функцией. Поэтому $\det(A) = \det(A') + \det(A'')$, что, по определению функции f, равносильно (27.4).

Пусть теперь в каком-либо столбце матрицы C имеется общий скалярный множитель λ. Тогда можно считать, что этот же множитель является общим для соответствующего столбца в A. В силу полилинейности определителя, λ можно вынести из-под знака $\det(A)$ и тем самым из-под знака $f(C)$.

Полилинейность $f(C)$ доказана. Докажем антисимметричность.

Если в матрице C переставить два столбца, то и соответствующие столбцы в матрице A переставятся. В силу антисимметричности определителя, $\det(A)$ поменяет знак, а значит, поменяет знак и $f(C)$.

Антисимметричность функции $f(C)$ тоже доказана.

Вычислим значение $f(\underset{l\times l}{E})$, т. е. вычислим определитель

$$f(E) = \det \left(\begin{array}{c|c} \underset{m\times m}{B} & \underset{m\times l}{O} \\ \hline \underset{l\times m}{D} & \underset{l\times l}{E} \end{array} \right). \tag{27.5}$$

Определитель (27.4) является суммой $n!$ слагаемых ($n = m + l$), каждое из которых отвечает некоторой перестановке степени n

$$\sigma = \left(\begin{array}{ccc|ccc} 1 & ... & m & m+1 & ... & n \\ \sigma(1) & ... & \sigma(m) & \sigma(m+1) & ... & \sigma(n) \end{array} \right). \tag{27.6}$$

Однако если хотя бы для одного из номеров $j = m + 1, ..., n$ будет $\sigma(j) \neq j$, то соответствующий член определителя обратится в нуль. Поэтому в сумме можно оставить лишь $m!$ членов, отвечающих перестановкам, тождественным в указанной зоне номеров, т. е. имеющим вид:

$$\sigma = \left(\begin{array}{ccc|ccc} 1 & ... & m & m+1 & ... & n \\ \sigma(1) & ... & \sigma(m) & m+1 & ... & n \end{array} \right). \tag{27.7}$$

Перестановки такого вида однозначно определяются соответствующими перестановками (степени m):

$$\sigma' = \left(\begin{array}{ccc} 1 & ... & m \\ \sigma'(1) & ... & \sigma'(m) \end{array} \right);\ \sigma'(j) = \sigma(j);\ j = 1, ..., m. \tag{27.8}$$

Поскольку все инверсии перестановки (27.7) сосредоточены в зоне номеров $j = 1, ..., m$, то знак этой перестановки совпадает со знаком перестановки (27.8).

Таким образом,

$$f(E) \overset{(23.3t)}{=\!=} \sum_{\sigma' \in S_m} a_{\sigma'(1)\,1} ... a_{\sigma'(m)\,m} a_{(m+1)(m+1)} ... a_{nn}.$$

Но для номеров $i, j \in \{1, ..., m\}$ имеем $a_{ij} = b_{ij}$. Кроме того, все диагональные элементы a_{ii} при $i = m+1, ..., n$ равны единице. С учетом этого

$$f(E) = \sum_{\sigma' \in S_m} b_{\sigma'(1)\,1} ... b_{\sigma'(m)\,m} \overset{(23.3t)}{=\!=} \det(B).$$

С помощью теоремы 26.1 окончательно получается:

$$f(C) = \det(C) \cdot f(E) = \det(C) \cdot \det(B).$$

Формула (27.2) доказана. □

Замечание 27.1. Стандартным индуктивным рассуждением формула (27.2) обобщается на случай матриц A более сложного блочного строения, разбитых, скажем, на s^2 блоков, из которых s (диагональных) являются квадратными. В предположении, что все блоки, расположенные ниже (выше) главной диагонали, являются нулевыми, справедливо утверждение: определитель матрицы A равен произведению определителей диагональных блоков.

27.2. Мультипликативное свойство определителя. Рассмотрим две квадратные матрицы $A, B \in L(n, \mathbb{R})$ и их произведение

$$C = A \cdot B. \tag{27.9}$$

Теорема 27.2. *Определитель произведения матриц равен произведению определителей матриц-сомножителей:*

$$\det(A \cdot B) = \det(A) \cdot \det(B). \tag{27.10}$$

Доказательство. По правилу умножения матриц (см. п. 1.2), при умножении матрицы B слева на матрицу A каждый столбец матрицы B умножается на A:

$$C = A \cdot (\bar{b}_1 | \bar{b}_2 | ... | \bar{b}_n) = (A\bar{b}_1 | A\bar{b}_2 | ... | A\bar{b}_n). \tag{27.11}$$

Зафиксируем матрицу A и рассмотрим функцию

$$f(B) = \det(A \cdot B) = \det(A\bar{b}_1 | A\bar{b}_2 | ... | A\bar{b}_n) \tag{27.12}$$

как функцию от столбцов матрицы B. Докажем, что эта функция полилинейна и антисимметрична.

Пусть $\bar{b}_j$ (один из столбцов матрицы B) представлен в виде суммы $\bar{b}_j = \bar{b}'_j + \bar{b}''_j$. Рассмотрим, как обычно, матрицы B' и B'', отличающиеся от B по j-му столбцу. (В первой из них, B', вместо $\bar{b}_j$ фигурирует $\bar{b}'_j$; и т. д.)

В матрице $A \cdot B$ (в силу дистрибутивности матричного умножения; см. теорему 2.1) j-й столбец тоже разобьется на два слагаемых:

$$A \cdot \bar{b}_j = A \cdot \bar{b}'_j + A \cdot \bar{b}''_j;$$

и так же, как и выше, можно будет рассмотреть две матрицы (отличающиеся от $A \cdot B$ по j-му столбцу), причем они будут, очевидно, равняться $A \cdot B'$ и $A \cdot B''$.

В силу полилинейности определителя, будет иметь место равенство

$$\det(A \cdot B) = \det(A \cdot B') + \det(A \cdot B''),$$

или для функции (27.12):

$$f(B) = f(B') + f(B''). \tag{27.13}$$

Если в столбце $\bar{b}_j$ выделен общий скалярный множитель λ, то его с помощью матричного закона (xi) (см. теорему 2.1) можно будет вынести и поставить перед матрицей A:

$$A \cdot (\lambda \cdot \bar{b}_j) = \lambda \cdot (A \cdot \bar{b}_j),$$

а затем вынести и из-под знака определителя.

В итоге мы получим вынесение за знак функции $f(B)$ общего скалярного множителя из j-го столбца.

Полилинейность функции $f(B)$ доказана.

Ее антисимметричность легко следует из антисимметричности определителя и того (очевидного) факта, что перестановка столбцов в матрице B влечет перестановку соответствующих столбцов в $A \cdot B$.

Вычислим значение функции (27.12) на единичной матрице:

$$f(E) = \det(A \cdot E) = \det(A). \tag{27.14}$$

Применим теорему 26.1:

$$f(B) = \det(B) \cdot f(E) = \det(B) \cdot \det(A).$$

Формула (27.10) доказана. □

Замечание 27.2. Свойство (27.10) можно интерпретировать как *согласованность* отображения

$$\det : L(n, \mathbb{R}) \longrightarrow \mathbb{R}$$

с операциями умножения в кольце квадратных матриц $L(n, \mathbb{R})$ и в поле $\mathbb{R}$.

§ 28. Присоединенная матрица. Выражение обратной матрицы через присоединенную

28.1. Определение и основное свойство присоединенной матрицы. Рассмотрим квадратную матрицу $\underset{n\times n}{A}$. Из алгебраических дополнений A_{ij} к элементам a_{ij} матрицы A можно составить новую матрицу:

$$\underset{n\times n}{\tilde{A}} = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ A_{n1} & A_{An2} & \ldots & A_{nn} \end{pmatrix}. \tag{28.1}$$

Определение 28.1. Матрица, транспонированная к матрице алгебраических дополнений (28.1), называется *присоединенной* матрицей для матрицы A. Обозначение:

$$A^{\vee} = (\tilde{A})^t = \begin{pmatrix} A_{11} & A_{21} & \ldots & A_{n1} \\ A_{12} & A_{22} & \ldots & A_{n2} \\ \ldots & \ldots & \ldots & \ldots \\ A_{1n} & A_{2n} & \ldots & A_{nn} \end{pmatrix}. \tag{28.2}$$

Предложение 28.1. *Для любой квадратной матрицы A справедливы равенства*

$$A \cdot A^{\vee} = A^{\vee} \cdot A = \det(A) \cdot E. \tag{28.3}$$

Доказательство. Докажем одно из равенств (28.3), второе доказывается абсолютно аналогично.

Надо перемножить матрицы A и $A^\vee$. Произведение i-й строки матрицы A на k-й столбец матрицы $A^\vee$ есть не что иное, как сумма произведений элементов i-й строки матрицы A на алгебраические дополнения к элементам ее k-й строки. В силу формулы (25.11t), эта сумма равна $\det(A)\delta_{ik}$.

Теперь все доказано: результат перемножения оказывается скалярной матрицей, равной произведению единичной матрицы (с элементами δ_{ik}) на скаляр $\det(A)$. □

28.2. Неособые матрицы. Вторая теорема об условиях обратимости матрицы. Вторую теорему об условиях обратимости квадратной матрицы мы сформулируем так, чтобы она включала в себя первую (см. теорему 14.1). Сначала дадим следующее

Определение 28.2. Квадратная матрица называется *неособой*, если ее определитель отличен от нуля, и *особой* — в противном случае (если определитель равен нулю).

Теорема 28.1 (вторая теорема об условиях обратимости матрицы). *Пусть A — квадратная матрица. Следующие три утверждения равносильны:*

(1) матрица A обратима;

(2) матрица A невырождена;

(3) матрица A неособа.

Доказательство. Равносильность утверждений (1) и (2) составляла содержание первой теоремы об условиях обратимости. Докажем равносильность (1) и (3).

Пусть матрица A обратима, т. е. существует такая матрица B, что $A \cdot B = B \cdot A = E$. Применяя мультипликативное свойство определителя (см. теорему 27.2), получим:

$$\det(A) \cdot \det(B) = 1. \tag{28.4}$$

Из (28.4) следует, что $\det(A) \neq 0$, т. е. матрица A является неособой.

Обратно, пусть A — неособа. Рассмотрим присоединенную матрицу $A^\vee$ и поделим ее на скаляр $\det(A) \neq 0$, т. е. образуем матрицу

$$B = \frac{1}{\det(A)} A^\vee. \tag{28.5}$$

С использованием закона (xi) матричной алгебры и предложения 28.1 получим:

$$A \cdot B = A \cdot (\frac{1}{\det(A)} \cdot A^{\vee}) = \frac{1}{\det(A)}(A \cdot A^{\vee}) = \frac{1}{\det(A)}(\det(A)E) = E.$$

Совершенно аналогично получается $B \cdot A = E$. Значит, матрица A обратима (и $B = A^{-1}$). $\square$

Замечание 28.1. В замечании 15.9 указывалось, что существование односторонней (правой или левой) обратной матрицы B для квадратной матрицы A влечет тот факт, что B оказывается настоящей (двусторонней) обратной матрицей. Другими словами, из равенства $A \cdot B = E$ вытекает равенство $B \cdot A = E$ и наоборот.

Тот же самый вывод легко следует из теоеремы 28.1. [В самом деле, для получения формулы (28.4) достаточно одного из указанных выше равенств. Второе будет выполняться автоматически, в силу теоремы.]

В качестве еще одного следствия из доказанной выше теоремы приведем *критерий обращения в нуль определителя* квадратной матрицы.

Предложение 28.2. *Определитель квадратной матрицы обращается в нуль тогда и только тогда, когда столбцы (строки) этой матрицы линейно зависимы.*

Доказательство. Согласно теореме 28.1, обращение в нуль определителя квадратной $(n \times n)$-матрицы A равносильно *вырожденности* матрицы A, т. е. тому факту, что ранг этой матрицы (скажем — столбцовый; см. теорему 12.1) меньше n, что, в свою очередь, равносильно линейной зависимости столбцов A. (Со строками — все точно так же, поскольку строчный и столбцовый ранги совпадают.) $\square$

Еще одним (очевидным) следствием теорем 28.1 и 27.1 является

Предложение 28.3. *Если матрица A обратима, то*

$$\det(A^{-1}) = \frac{1}{\det(A)}. \tag{28.6}$$

Доказательство вытекает из соотношения $A \cdot A^{-1} = E$. $\square$

Замечание 28.2. Из замечания 27.2 следует, что отображение вычисления определителя согласовано с операциями умножения матриц и скаляров. В частности, можно заключить, что определитель оратимой матрицы является обратимым (или, что равносильно, ненулевым) скаляром. Так возникает *гомоморфизм* (определение гомоморфизма см. в п. 20.2)

$$\det : GL(n, \mathbb{R}) \longrightarrow \mathbb{R}^*$$

группы обратимых $(n \times n)$-матриц (см. замечание 14.3) в группу ненулевых действительных чисел.

28.3. Алгоритм вычисления обратной матрицы с помощью присоединенной. В п. 14.6 был изложен алгоритм Ж.—Г. исследования квадратной матрицы на обратимость и отыскания обратной матрицы (в случае обратимости данной). Сейчас мы изложим и покажем на примере второй алгоритм. В вычислительном отношении он отнюдь не лучше первого, однако обладает (теоретическим) преимуществом — для обратной матрицы выписывается *явная* формула (тогда как в первом алгоритме реализуется другой подход: хочешь вычислить — вычисляй; что получится — то и ответ).

Пусть дана квадратная матрица A размера $n \times n$.

1. Для каждого элемента a_{ij} вычислим алгебраическое дополнение A_{ij} (оно является определителем $(n-1)$-го порядка). Составим из алгебраических дополнений матрицу $\tilde{A}$.

2. Вычислим определитель $\det(A)$, используя ранее найденные алгебраические дополнения и разлагая определитель по какой-либо строке (какому-либо столбцу). Если этот определитель равен нулю, то данная матрица необратима.

3. Если $\det(A) \neq 0$, то матрица A обратима. Вычислим присоединенную матрицу $A^\vee$, транспонировав матрицу $\tilde{A}$.

4. Вычислим обратную матрицу, поделив присоединенную матрицу на $\det(A)$:

$$A^{-1} = \frac{1}{\det(A)} A^\vee. \tag{28.7}$$

Пример 28.1. Перерешаем с помощью второго алгоритма пример 14.2.

$$A = \begin{pmatrix} 1 & 1 & -1 \\ 0 & -1 & -1 \\ \lambda & 0 & 1 \end{pmatrix};$$

$$A_{11}=\begin{vmatrix}-1 & -1\\ 0 & 1\end{vmatrix}=-1;\ A_{12}=-\begin{vmatrix}0 & -1\\ \lambda & 1\end{vmatrix}=-\lambda;\ A_{13}=\begin{vmatrix}0 & -1\\ \lambda & 0\end{vmatrix}=\lambda;$$

$$A_{21}=-\begin{vmatrix}1 & -1\\ 0 & 1\end{vmatrix}=-1;\ A_{22}=\begin{vmatrix}1 & -1\\ \lambda & 1\end{vmatrix}=\lambda+1;\ A_{23}=-\begin{vmatrix}1 & 1\\ \lambda & 0\end{vmatrix}=\lambda;$$

$$A_{31}=\begin{vmatrix}1 & -1\\ -1 & -1\end{vmatrix}=-2;\ A_{32}=-\begin{vmatrix}1 & -1\\ 0 & -1\end{vmatrix}=1;\ A_{33}=\begin{vmatrix}1 & 1\\ 0 & -1\end{vmatrix}=-1;$$

$$\tilde{A}=\begin{pmatrix}-1 & -\lambda & \lambda\\ -1 & \lambda+1 & \lambda\\ -2 & 1 & -1\end{pmatrix};$$

$$\det(A)\overset{\text{по }2^{\text{стр}}}{=\!=\!=}0\cdot(-\lambda)+(-1)\cdot(\lambda+1)+(-1)\cdot\lambda=-(2\lambda+1);$$

$$(A\text{ обратима})\Longleftrightarrow(\lambda\neq-\frac{1}{2});\ A^{\vee}=\begin{pmatrix}-1 & -1 & -2\\ -\lambda & \lambda+1 & 1\\ \lambda & \lambda & -1\end{pmatrix};$$

$$A^{-1}=-\frac{1}{2\lambda+1}\begin{pmatrix}-1 & -1 & -2\\ -\lambda & \lambda+1 & 1\\ \lambda & \lambda & -1\end{pmatrix}=\frac{1}{2\lambda+1}\begin{pmatrix}1 & 1 & 2\\ \lambda & -\lambda-1 & -1\\ -\lambda & -\lambda & 1\end{pmatrix}.$$

Замечание 28.3. В системе Maple имеется команда, вычисляющая присоединенную матрицу:

> adj (A) ;

Можно попробовать сравнить результаты, получаемые двумя способами:

A := matrix ([[1, 1, −1], [0, −1, −1], [lambda, 0, 1]]);

$$A:=\begin{bmatrix}1 & 1 & -1\\ 0 & -1 & -1\\ \lambda & 0 & 1\end{bmatrix}$$

> evalm (A ^ (−1)) = evalm (adj(A) / det(A)) ;

(Обратите внимание на "равенство без двоеточия". Это — не присваивание. Равенство будет просто записано и должно оказаться верным.)

$$\begin{bmatrix}\frac{1}{1+2\lambda} & \frac{1}{1+2\lambda} & \frac{2}{1+2\lambda}\\ \frac{\lambda}{1+2\lambda} & -\frac{1+\lambda}{1+2\lambda} & -\frac{1}{1+2\lambda}\\ -\frac{\lambda}{1+2\lambda} & -\frac{\lambda}{1+2\lambda} & \frac{1}{1+2\lambda}\end{bmatrix}=\begin{bmatrix}-\frac{1}{-1-2\lambda} & -\frac{1}{-1-2\lambda} & -\frac{2}{-1-2\lambda}\\ -\frac{\lambda}{-1-2\lambda} & \frac{1+\lambda}{-1-2\lambda} & \frac{1}{-1-2\lambda}\\ \frac{\lambda}{-1-2\lambda} & \frac{\lambda}{-1-2\lambda} & -\frac{1}{-1-2\lambda}\end{bmatrix}$$

Равенство, очевидно, верное, но правая часть выдана неупрощенной. Предвидя это, можно было сразу набрать правую часть с командой на упрощение:

> evalm(A ^ (−1)) = map(simplify, evalm(adj(A) / det(A)));

Тогда правая часть окажется идентична левой.

§ 29. Решение квадратных систем линейных уравнений с помощью обратной матрицы и по формулам Крамера

29.1. Решение квадратных с.л.у. с помощью обратной матрицы. Рассмотрим квадратную с.л.у.

$$\underset{n\times n}{A}\cdot\underset{n\times 1}{\bar{x}}=\underset{n\times 1}{\bar{b}}. \tag{29.1}$$

Предложение 29.1. *Если матрица A обратима (или, что равносильно, неособа), то с.л.у. (29.1) является совместной и определенной и ее единственное решение может быть найдено по формуле*

$$\bar{x}=A^{-1}\cdot\bar{b}. \tag{29.2}$$

Доказательство. Матричное равенство (29.1) заменится на равносильное, если его домножить слева на (существующую по предположению) обратную матрицу A^{-1}:

$$A^{-1}\cdot(A\cdot\bar{x})=A^{-1}\cdot\bar{b}.$$

В силу ассоциативности матричного умножения и определения обратной матрицы, последнее равенство равносильно равенству (29.2), дающему единственное решение системы (29.1). □

Замечание 29.1. В отличие от метода Гаусса, который применим к любым с.л.у., *метод обратной матрицы* применим лишь к квадратным системам с обратимыми матрицами. У него есть единственное (теоретическое) преимущество: явное выражение для ответа.

Но и это преимущество может оказаться мнимым, если для вычисления обратной матрицы используется не дающий явного ответа

алгоритм Ж.—Г. (Если вычислять по Гауссу A^{-1}, то разумнее сразу решать по Гауссу систему.)

Поэтому для нахождения A^{-1} в данном случае естественно применять второй, явный алгоритм, использующий присоединенную матрицу. (См. соответствующий комментарий в п. 28.3.)

Такой подход позволит нам (в следующем пункте) сделать "еще более явной" формулу (29.2).

29.2. Решение квадратных с.л.у. по формулам Крамера. Обозначим

$$\Delta = \det(A) \tag{29.3}$$

так называемый *главный определитель* квадратной системы (29.1). Предположим, что он отличен от нуля. Тогда существует обратная матрица A^{-1} и ее элементы могут быть определены (через алгебраические дополнения к элементам A) по формулам

$$[A^{-1}]_{ij} = \frac{1}{\Delta}[A^{\vee}]_{ij} = \frac{1}{\Delta}A_{ji};\ i,j = 1,...,n. \tag{29.4}$$

Координаты вектора (29.2), являющегося решением с.л.у., могут быть выражены формулами

$$x_i = \sum_{j=1}^{n}[A^{-1}]_{ij}b_j = \sum_{j=1}^{n}\frac{1}{\Delta}A_{ji}b_j = \frac{1}{\Delta}\sum_{j=1}^{n}A_{ji}b_j. \tag{29.5}$$

Рассмотрим результат, полученный в (29.5), с точки зрения теоремы Лапласа 25.1. Фигурирующая в правой части сумма является не чем иным, как разложением по i-му столбцу следующего определителя:

$$\Delta_i = \det(\bar{a}_1|...|\bar{a}_{i-1}|\,\bar{b}\,|\bar{a}_{i+1}|...|\bar{a}_n), \tag{29.6}$$

где $i = 1,...,n$. Этот определитель получается из главного определителя Δ заменой его i-го столбца $\bar{a}_i$ на столбец правых частей $\bar{b}$.

[Напомним (см. замечание 25.1), что алгебраическое дополнение к элементу определителя зависит лишь от позиции этого элемента. Так, алгебраическим дополнением к элементу b_j определителя (29.6) служит алгебраическое дополнение к элементу a_{ji}, стоявшему на этом месте в исходной матрице A.]

Таким образом, решение с.л.у. (29.1) (в случае $\Delta \neq 0$) может быть найдено по следующим формулам (называемым *формулами Крамера*):

$$x_i = \frac{\Delta_i}{\Delta};\ i = 1,...,n, \tag{29.7}$$

или в векторном виде:

$$\bar{x} = \frac{1}{\Delta} \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \dots \\ \Delta_n \end{pmatrix}. \tag{29.8}$$

Подводим итог: доказано следующее

Предложение 29.2. *Рассмотрим квадратную с.л.у. вида (29.1). Пусть Δ — ее главный определитель (29.6), а Δ_i — определители, получаемые из определителя Δ заменой его i-го столбца на столбец правых частей $\bar{b}$.*

Если $\Delta \neq 0$, то система (29.1) является совместной и определенной и ее единственное решение может быть найдено по формуле (29.8). □

Замечание 29.2. Продолжая "идеологию" замечания 29.1, констатируем малую пригодность формул Крамера для серьезных практических вычислений. Однако эти формулы дают явный ответ, чем объясняется их теоретическое значение.

Пример 29.1. Вернемся к примеру 7.1 и перерешаем приведенную там квадратную с.л.у. (содержащую параметр) с использованием формул Крамера.

Вычислим главный определитель системы:

$$\Delta = \begin{vmatrix} \lambda & 1 & 1 \\ 1 & \lambda & 1 \\ 1 & 1 & \lambda \end{vmatrix} = \lambda^3 - 3\lambda + 2.$$

Корни полученного кубического многочлена легко найти подбором или группировкой:

$$\lambda^3 - 3\lambda + 2 = (\lambda^3 - 2\lambda^2 + \lambda) + (2\lambda^2 - 4\lambda + 2) = \\ = \lambda(\lambda-1)^2 + 2(\lambda-1)^2 = (\lambda-1)^2(\lambda+2).$$

Особые случаи $\boxed{\lambda = 1}$ и $\boxed{\lambda = -2}$ анализируются по Гауссу. (Крамер здесь — "не помощник".)

При рассмотрении неособого случая $\boxed{\lambda \notin \{1, -2\}}$ понадобятся три определителя

$$\Delta_1 = \begin{vmatrix} 1 & 1 & 1 \\ 1 & \lambda & 1 \\ 1 & 1 & \lambda \end{vmatrix}; \ \Delta_2 = \begin{vmatrix} \lambda & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & \lambda \end{vmatrix}; \ \Delta_3 = \begin{vmatrix} \lambda & 1 & 1 \\ 1 & \lambda & 1 \\ 1 & 1 & 1 \end{vmatrix}.$$

Легко видеть, что эти определители одинаковы и равны $(\lambda-1)^2$. По формулам Крамера получаем в этом случае ответ:

$$x_i = \frac{\Delta_i}{\Delta} = \frac{1}{\lambda+2};\ i = 1,\ 2,\ 3.$$

§ 30. Минорный ранг матрицы. Вторая теорема о ранге матрицы

30.1. Миноры матрицы. Пусть A — прямоугольная матрица размера $m \times n$, k — натуральное число, не превышающее $\min(m, n)$.

Определение 30.1. Рассмотрим какие-либо k строк матрицы A с номерами $i_1 < i_2 < \ldots < i_k$ и какие-либо k столбцов с номерами $j_1 < j_2 < \ldots < j_k$, а также $(k \times k)$-матрицу, получаемую на пересечении указанных строк и столбцов. Определитель этой матрицы называется минором порядка k для матрицы A и обозначается

$$\overset{k}{M}{}^{i_1,i_2,\ldots,i_k}_{j_1,j_2,\ldots,j_k} = \det \begin{pmatrix} a_{i_1j_1} & a_{i_1j_2} & \ldots & a_{i_1j_k} \\ a_{i_2j_1} & a_{i_2j_2} & \ldots & a_{i_2j_k} \\ \ldots & \ldots & \ldots & \ldots \\ a_{i_kj_1} & a_{i_kj_2} & \ldots & a_{i_kj_k} \end{pmatrix}. \qquad (30.1)$$

Замечание 30.1. Сложное обозначение (30.1), содержащее нижние индексы (номера столбцов), верхние индексы (номера строк), а также *указатель порядка минора*, располагающийся над буквой M, мы будем часто сокращать до более короткого обозначения $\overset{k}{M}$ (говоря, например, о произвольном миноре указанного порядка или о существовании некоторого минора такого порядка с предписанным свойством).

Ниже нам придется также говорить о *включении* одного минора в другой, более высокого порядка. Мы будем понимать это так: к последовательностям строк и столбцов, задающим исходный минор, добавляется еще несколько новых строк и столько же новых столбцов (последовательности остаются упорядоченными по возрастанию).

Замечание 30.2. Подсчитаем количество миноров порядка k. Снова нам понадобится комбинаторная формула для числа сочетаний

(из n элементов по k). (В замечании 21.1 мы эту формулу уже использовали в частном случае сочетаний из n элементов по 2.)

Из n номеров столбцов можно $n(n-1)...(n-k+1)$ способами выбрать k-элементный упорядоченный набор $j_1, j_2, ..., j_k$. Существует $k!$ различных способов упорядочения набора мощности k. Только один из этих способов дает набор, упорядоченный по возрастанию.

Поэтому количество упорядоченных по возрастанию наборов из k столбцов будет равно:

$$C_n^k = \frac{n(n-1)...(n-k+1)}{k!}. \tag{30.2}$$

Аналогично количество упорядоченных по возрастанию наборов $i_1, i_2, ..., i_k$ из номеров строк будет равно C_m^k.

При образовании минора k-го порядка любой из подсчитанных выше наборов строк может сочетаться с любым из наборов столбцов. Поэтому общее количество миноров порядка k будет равно произведению $C_m^k \cdot C_n^k$.

Миноры первого порядка суть элементы данной матрицы. Их количество равно mn.

Замечание 30.3. Термин "миноры" уже "звучал" ранее (см. замечание 25.1) в специальном частном случае, когда данная матрица A была квадратной и мы, вычеркивая из нее i-ю строку и j-й столбец и беря определитель оставшейся квадратной матрицы, получали миноры $(n-1)$-го порядка M_{ij}.

Заметьте, что система обозначений § 25 отличается от принятой в настоящем параграфе. Так, описанный выше минор порядка $n-1$ по принципу, сформулированному в определении 30.1, получил бы обозначение

$$M_{ij} = {}^{n-1}M_{1,...,i-1,i+1,...,n}^{1,...,j-1,j+1,...,n}.$$

30.2. Минорный ранг матрицы. Ранее были введены (см. определения 11.2 и 12.2) три типа ранга: ступенчатый, столбцовый и строчный. Ниже дается определение еще одного, четвертого типа.

Определение 30.2. *Минорным рангом* матрицы $\underset{m\times n}{A}$ называется наивысший порядок миноров этой матрицы, отличных от нуля. (Если таковых вообще нет, то минорный ранг считается равным нулю.) Обозначение: $\mathrm{rank}_{\text{мин}}(A)$.

Замечание 30.4. Всегда находятся зубрилки, тараторящие последнее определение, "слегка" путая окончания: "отличный" вместо "отличных". Если вы — не из их числа, то (в качестве упражнения) подумайте: что определяет искаженное определение?

Замечание 30.5. Используя краткое обозначение для миноров матрицы A (см. замечание 30.1), данное выше определение 30.2 можно пересказать следующим образом.

Равенство

$$\operatorname{rank}_{\text{мин}}(A) = r, \tag{30.3}$$

равносильно тому, что

1) в матрице A существует минор $\overset{r}{M}$, отличный от нуля;

2) все миноры $\overset{k}{M}$ порядка $k > r$ обращаются в нуль (или их вообще нет).

Если вы уже свыклись с условностями математической логики, то должны понимать, что во втором утверждении уточнение в скобках можно без ущерба для смысла высказывания исключить. (Дело в том, что об элементах пустого множества, например об отсутствующих минорах, можно сказать что угодно.)

Если придерживаться описанной выше "софистики", то и случай нулевого ранга (имеющий место тогда и только тогда, когда $A = O$) будет включаться в сформулированное выше правило. (Действительно, при $r = 0$ первое утверждение правила будет истинным, поскольку миноров нулевого порядка просто нет, а второе утверждение будет означать прежде всего, что все миноры порядка 1, т. е. все элементы матрицы, равны нулю.)

Теперь мы еще раз вернемся ко второму утверждению обсуждаемого правила и сформулируем вспомогательное

Предложение 30.1. *Если в матрице A все миноры некоторого порядка k обращаются в нуль, то и все миноры более высокого порядка $l > k$ также обращаются в нуль.*

Доказательство практически очевидно в силу теоремы Лапласа: всякий минор $\overset{k+1}{M}$, как определитель, можно разложить по какой-либо строке, причем алгебраические дополнения к элементам строки будут (с точностью до знака) совпадать с минорами k-го порядка (в исходной матрице), а они по предположению все равны нулю; получается, что и все миноры порядка $k + 1$ также равны нулю; и т. д. вверх по порядку миноров. $\square$

С учетом предложения 30.1 правило из замечания 30.5 допускает более лаконичную формулировку.

Предложение 30.2. *Минорный ранг матрицы A равен (неотрицательному целому) числу r тогда и только тогда, когда выполняются условия: 1)* $(\exists \overset{r}{M})\,[\,\overset{r}{M} \neq 0\,]$ *и 2)* $(\forall \overset{r+1}{M})\,[\,\overset{r+1}{M} = 0\,]$. $\square$

Определение 30.3. Если $\operatorname{rank}_{\text{мин}}(A) = r$, то всякий ненулевой минор порядка r (т. е. $\overset{r}{M} \neq 0$) называется *ранговым минором* для матрицы A.

30.3. Вторая теорема о ранге матрицы. Согласно теореме 12.1 (первой теореме о ранге матрицы), ступенчатый, столбцовый и строчный ранги совпадают. Сейчас мы присоединим к этой "компании" минорный ранг.

Теорема 30.1 (вторая теорема о ранге матрицы). *Все четыре ранга (ступенчатый, столбцовый, строчный и минорный) совпадают.*

Доказательство. Пусть $\operatorname{rank}(A) = r$ (здесь имеется в виду любой из трех первых рангов, про которые уже известно, что они совпадают). Докажем, что минорный ранг матрицы A также равен r.

Для случая $r = 0$ утверждение очевидно: для каждого из четырех рангов его обращение в нуль равносильно обращению в нуль самой матрицы. Поэтому далее мы считаем, что $r > 0$. [Кроме того, мы помним, что $r \leqslant \min(m, n)$.]

Нам надо доказать равенство (30.3). Согласно предложению 30.2, оно равносильно конъюнкции двух утверждений.

1. Докажем существование минора порядка r, отличного от нуля. Поскольку $\operatorname{rank}_{\text{стб}}(A) = r$, в матрице A найдется r линейно независимых столбцов. Соберем их в подматрицу B размера $m \times r$. Столбцовый ранг подматрицы B, очевидно, тоже равен r (все столбцы B линейно независимы). В силу теоремы 12.1, $\operatorname{rank}_{\text{стр}}(B) = r$, и следовательно, в матрице B найдутся r линейно независимых строк. Соберем эти строки в подматрицу C размера $r \times r$. Определитель квадратной подматрицы C будет минором порядка r для исходной матрицы A. В силу предложения 28.2, этот минор будет отличен от нуля.

2. Докажем, что всякий минор порядка $r+1$ равен нулю. Предположим противное, т. е. предположим, что найдется некоторый $\overset{r+1}{M} \neq 0$. Согласно предложению 28.2, столбцы этого минора являются линейно независимыми векторами. Рассмотрим столбцы исходной матрицы A, являющиеся продолжениями столбцов минора $\overset{r+1}{M}$. Эти (более длинные) столбцы также линейно независимы (в силу замечания 9.3 об "укорачивании" системы векторов). Последний факт свидетельствует о том, что (столбцовый) ранг матрицы A не может быть ниже $r+1$, что противоречит предположению. □

Замечание 30.6. После второй теоремы о ранге мы имеем право всюду в дальнейшем употреблять обозначение $\operatorname{rank}(A)$, не уточняя, о каком типе ранга идет речь. (В нужный момент будут использоваться различные характеризации ранга.)

В качестве следствия из теоремы 30.1 докажем

Предложение 30.3. *Всякий ненулевой минор $\overset{k}{M}$ матрицы A можно включить в некоторый ранговый минор.*

Доказательство. Пусть $\operatorname{rank}(A) = r$.

Столбцы данного минора линейно независимы (в пространстве $\mathbb{R}^k$). Являющиеся их продолжениями столбцы данной матрицы будут линейно независимыми векторами в пространстве $\mathbb{R}^m$ (рассуждаем так же, как и во втором пункте доказательства теоремы 30.1).

Составим из продолженных столбцов подматрицу G; ее столбцовый ранг будет равен k.

В соответствии со свойством продолжения базисов (см. предложение 10.2) система векторов-столбцов матрицы G может быть включена в некоторый базис в линейной оболочке столбцов A (причем добавочные $r-k$ векторов могут быть выбраны из числа столбцов матрицы A).

Базисные столбцы будут образовывать (более широкую, чем G) подматрицу B размера $m \times r$ (и ранга r).

Рассмотрим теперь те строки матрицы B, которые "пересекают" исходный минор. Они линейно независимы и могут быть дополнены до базиса в линейной оболочке строк матрицы B.

Рассмотрим далее подматрицу C размера $r \times r$ в матрице B, образованную базисными строками. Ее определитель будет ранговым минором и будет включать исходный минор. □

30.4. Метод окаймляющих миноров для вычисления ранга матрицы. Как должно быть понятно из п. 30.1, количество миноров заданного порядка может быть очень большим и их непосредственный перебор может оказаться затруднительным. В этом пункте мы опишем алгоритм вычисления минорного ранга матрицы (с отысканием рангового минора), использующий некоторое сокращение перебора миноров.

Определение 30.4. *Окаймляющим* минором для некоторого минора порядка k называется всякий минор порядка $k+1$, включающий (см. замечание 30.1) данный.

Замечание 30.7. Легко подсчитать, что количество миноров, окаймляющих заданный минор порядка k, равно $(m-k)(n-k)$.

Предложение 30.4. *Если все миноры $(k+1)$-го порядка, окаймляющие заданный минор $\overset{k}{M} \neq 0$, равны нулю, то вообще все миноры $(k+1)$-го порядка равны нулю.*

Доказательство. Докажем, что $\operatorname{rank}(A) = k$; тогда утверждение предложения получится автоматически.

Предположим противное, т. е. предположим, что $\operatorname{rank}(A) = r > k$.

Дальше рассуждение аналогично проведенному при доказательстве предложения 30.3, но мы добавим к столбцам, пересекающим минор, не $r-k$ столбцов (вплоть до достижения базиса), а лишь один из них. Таким же образом мы поступим при добавлении строк.

В итоге мы получим матрицу размера $(k+1)\times(k+1)$, определитель которой будет ненулевым минором, окаймляющим данный минор, что противоречит условию. □

Переходим к описанию алгоритма отыскания ранга матрицы методом окаймляющих миноров.

Д а н о: матрица A размера $m \times n$.

Т р е б у е т с я: вычислить $\operatorname{rank}(A)$ и указать какой-либо ранговый минор.

О п и с а н и е р а б о т ы а л г о р и т м а:

1. Если $A = O$, то $\operatorname{rank}(A) = 0$, рангового минора не существует.

2. Если $A \neq O$, то выбираем какой-либо минор первого порядка $\overset{1}{M}$ (т. е. элемент матрицы A), отличный от нуля.

3. Перебираем миноры второго порядка, окаймляющие выбранный минор первого порядка, до тех пор пока не переберем все (и они все окажутся нулевыми) или пока не обнаружим ненулевой минор $\overset{2}{M}$. В первом случае $\operatorname{rank}(A) = 1$ и в качестве рангового можно предъявить выбранный на предыдущем шаге ненулевой минор $\overset{1}{M}$. Во втором случае перебираем миноры третьего порядка, окаймляющие выбранный на предыдущем шаге ненулевой минор $\overset{2}{M}$; и т. д.

4. Останов происходит, либо когда найденный на шаге с номером r ненулевой минор $\overset{r}{M}$ не имеет окаймляющих [это бывает, если $r = \min(m, n)$], либо когда все миноры $\overset{r+1}{M}$, окаймляющие минор $\overset{r}{M}$, равны нулю. В обоих случаях выводим

О т в е т: $\operatorname{rank}(A) = r$; $\overset{r}{M}$ — ранговый минор.

Замечание 30.8. Здесь мы не уточняем способ выбора первого ненулевого минора и — на каждом шаге — способ перебора окаймляющих миноров, но если программисту потребуется довести метод, изложенный в виде общей идеи, до уровня работающей программы, то все необходимые уточнения ему придется сделать.

Впрочем, данный конкретный метод имеет скорее теоретическое (и познавательное) значение. При ручном счете (для небольших матриц) его можно применять, но чуть более громоздкая задача вызовет уже большие затруднения.

Ранг матрицы легко вычисляется методом Гаусса (который можно дополнить таким образом, чтобы выдавался не только ранг, но и ранговый минор).

Пример 30.1. Вычислим с помощью метода окаймляющих миноров ранг матрицы

$$A = \begin{pmatrix} -6 & 4 & 8 & -1 & 6 \\ -5 & 2 & 4 & 1 & 3 \\ 7 & 2 & 4 & 1 & 3 \\ 2 & 4 & 8 & -7 & 6 \\ 3 & 2 & 4 & -5 & 3 \end{pmatrix}.$$

Ненулевой минор первого порядка: $\overset{1}{M}{}^1_1 = -6$.

Окаймляющие миноры будем перебирать по такой системе: добавим к списку строк первую из в него не вошедших и аналогично со списком столбцов.

Минор

$$\overset{2}{M}{}^{1,2}_{1,2} = \begin{vmatrix} -6 & 4 \\ -5 & 2 \end{vmatrix} = 8$$

оказался ненулевым.

Снова начинаем перебор по той же системе. Добавляем третью строку и третий столбец и вычисляем минор

$$\overset{3}{M}{}^{1,2,3}_{1,2,3} = \begin{vmatrix} -6 & 4 & 8 \\ -5 & 2 & 4 \\ 7 & 2 & 4 \end{vmatrix} = 0.$$

Он получился нулевым. Оставим пока в числе добавленных третью строку, а третий столбец заменим на (новый добавленный) четвертый столбец. Вычислим минор

$$\overset{3}{M}{}^{1,2,3}_{1,2,4} = \begin{vmatrix} -6 & 4 & -1 \\ -5 & 2 & 1 \\ 7 & 2 & 1 \end{vmatrix} = 72.$$

Он оказался ненулевым. Снова ищем окаймление. Первая не вошедшая в предыдущий список строка — четвертая, а первый не вошедший столбец — третий, т. е. нам надо начать перебор с минора $\overset{4}{M}{}^{1,2,3,4}_{1,2,3,4}$.

Он окажется равным нулю, как и следующие в переборе: $\overset{4}{M}{}^{1,2,3,4}_{1,2,4,5}$; $\overset{4}{M}{}^{1,2,4,5}_{1,2,3,4}$; $\overset{4}{M}{}^{1,2,4,5}_{1,2,3,5}$. (Проверьте!)

О т в е т: $\operatorname{rank}(A) = 3$; $\overset{3}{M}{}^{1,2,3}_{1,2,4}$ — ранговый минор.

§ 30a. Рекуррентности и определители

30a.1. Понятие о рекуррентностях. Данный параграф является вставкой; он не относится напрямую к изучаемой теме, однако представляется автору абсолютно необходимым для развития математической культуры будущих компьютерщиков.

Сначала мы познакомимся с понятием *рекуррентности* (*рекуррентной формулы*), а затем применим методы *решения рекуррентностей* к вычислению некоторых важных определителей.

Мы не можем здесь стремиться к максимальной общности и строгости. Материал будет излагаться описательно, и в основном мы будем интересоваться конкретными приложениями.

Определение 30а.1. Под *рекуррентностью порядка* d мы будем понимать числовую последовательность $\{u_k\}_{k=1}^{\infty}$, каждый член которой, начиная с u_{d+1}, определяется d предыдущими членами с помощью *рекуррентной формулы*

$$u_k = F(u_{k-1}, u_{k-2}, ..., u_{k-d});\ k \geqslant d+1, \tag{30а.1}$$

где F — заданная функция от d переменных.

Любая последовательность, удовлетворяющая всем соотношениям (30а.1), называется *решением* рекуррентности.

Замечание 30а.1. Ясно, что все члены последовательности $\{u_k\}$ будут однозначно определены, если произвольным образом задать первые d ее членов. Обычно это делается в форме так называемых *начальных условий*:

$$u_1 = u_1^0;\ u_2 = u_2^0;\ ...;\ u_d = u_d^0. \tag{30а.2}$$

Решить рекуррентность (30а.1) — это значить найти все ее решения. Формула, содержащая все *частные решения* рекуррентности [т. е. решения всевозможных начальных задач (30а.1)—(30а.2)] называется ее *общим решением*. Такая формула, как правило, содержит d произвольных констант $C_1, C_2, ..., C_d$, которые могут быть определены, если заданы конкретные начальные значения $u_i^0, ..., u_d^0$.

Рекуррентности очень часто встречаются как в математике, так и в компьютерных науках. Далеко не всегда удается решить их *явно*, т. е. получить ответ в виде формулы, определяющей u_k как функцию от номера k. Чаще всего приходится удовлетвориться *процедурой* (программой), позволяющей, в принципе, вычислить любой член последовательности с наперед заданным номером. (Разумеется, все такие "принципиальные возможности" на практике ограничены быстродействием и объемом памяти компьютера.)

Еще одна необходимая оговорка: нумерация членов последовательности совсем не обязательно начинается с единицы; часто начальным номером служит 0, но допустимо и любое другое целое число.

Пример 30а.1. Рекуррентная формула (порядка $d = 1$)

$$u_k = ku_{k-1};\ k \geqslant 1;\ u_0 = 1 \tag{30а.3}$$

задает не что иное, как факториал: $u_k = k!$

[Не подумайте только, что рекуррентность (30а.3) здесь "решена"! Мы лишь пересказали определение факториала. А рекуррентность задает процедуру для его вычисления.]

Пример 30а.2. Решим теперь следующую *линейную однородную* рекуррентность *первого порядка*

$$u_k = au_{k-1};\ k \geqslant 1 \tag{30а.4}$$

с начальным условием

$$u_0 = \alpha, \tag{30а.5}$$

где a, α — действительные числа; $a \neq 0$.

Рассматриваемая рекуррентность названа линейной и однородной потому, что члены искомой последовательности входят в уравнение исключительно в первой степени. Легко понять, что сумма двух решений (30а.4), а также произведение решения на число снова являются решениями (30а.4). В частности, всегда существует (*тривиальное*) решение $u_k = 0$. Попытаемся найти *нетривиальное* решение.

Рассуждение начнем с весьма популярного в математике оборота: "Будем искать решение задачи в виде (...)". Эта фраза означает, что наши славные предшественники путем экспериментов и/или рассуждений пришли к выводу, что некоторая функция от k, содержащая неопределенный параметр, удовлетворяет исследуемым соотношениям, но не при любых значениях параметра, а лишь при некоторых, подлежащих определению.

Итак, будем искать решение рекуррентности в виде

$$u_k = \lambda^k, \tag{30а.6}$$

где λ — подлежащий определению (ненулевой) параметр. Подставляя (30а.6) в (30а.4) и сокращая на λ^{k-1}, получим $\lambda = a$. Только при таком значении параметра последовательность (30а.6) подходит в рекуррентную формулу.

Таким образом, найдено одно нетривиальное решение рекуррентности:

$$u_k^{(1)} = a^k. \tag{30а.7}$$

При любом значении константы C последовательность

$$u_k = Cu_k^{(1)} = Ca^k \tag{30а.8}$$

также будет решением (30а.4).

Докажем, что формула (30а.8) представляет из себя общее решение рекуррентности, т. е. что эта формула содержит в себе все решения всевозможных начальных задач вида (30а.5).

В самом деле, полагая в (30а.8) $k = 0$ и приравнивая результат числу α, мы получаем:

$$C = \alpha.$$

Найденное значение константы подставим в (30а.8). Формула

$$u_k^* = \alpha a^k \tag{30а.9}$$

определит частное решение рекуррентности (30а.4), удовлетворяющее начальному условию (30а.5). Значит, в (30а.8) содержатся все решения всевозможных начальных задач для (30а.4), что и требовалось.

Итог: (30а.8) есть общее решение рекуррентности (30а.4), а (30а.9) есть ее частное решение, удовлетворяющее (30а.5).

30а.2. Линейные однородные рекуррентности второго порядка. В этом пункте мы исследуем рекуррентность второго порядка, являющуюся обобщением рекуррентности из примера 30а.2. А именно: рассмотрим линейную однородную рекуррентность

$$u_k = au_{k-1} + bu_{k-2};\ k \geqslant 2, \tag{30а.10}$$

где a, b — действительные числа, $b \neq 0$. [При $b = 0$ получилась бы рекуррентность первого порядка (30а.4).]

Для рекуррентностей второго порядка начальная задача должна содержать два условия:

$$\begin{cases} u_0 = \alpha; \\ u_1 = \beta, \end{cases} \tag{30а.11}$$

где $\alpha, \beta \in \mathbb{R}$.

Будем искать решение (30а.10) в том же "показательном" виде (30а.6). Подстановка этого выражения в рекуррентность, сокращение на λ^{k-2} и перенос всех оставшихся членов в левую часть приводят к следующему квадратному уравнению для параметра λ:

$$\lambda^2 - a\lambda - b = 0. \tag{30а.12}$$

Уравнение (30а.12) принято называть *характеристическим уравнением* для рекуррентности (30а.10).

Мы условились рассматривать задачу над полем $\mathbb{R}$, поэтому при решении характеристического уравнения придется исследовать три случая, в зависимости от дискриминанта

$$D = a^2 + 4b. \tag{30а.13}$$

В случае $D > 0$ характеристическое уравнение имеет два различных действительных корня:

$$\lambda_1 = \frac{1}{2}(a - \sqrt{D});\ \lambda_2 = \frac{1}{2}(a + \sqrt{D}). \tag{30a.14}$$

Оба корня отличны от нуля, т. к., по теореме Виета, их произведение равняется $-b$.

Следовательно, мы получаем два различных нетривиальных решения для рекуррентности (30a.10):

$$u_k^{(1)} = \lambda_1^k;\ u_k^{(2)} = \lambda_2^k. \tag{30a.15}$$

В силу того что рекуррентность является линейной и однородной, линейная комбинация

$$u_k = C_1 u_k^{(1)} + C_2 u_k^{(2)} = C_1 \lambda_1^k + C_2 \lambda_2^k \tag{30a.16}$$

при любых значениях констант C_1 и C_2 также будет ее решением. Надо доказать, что (30a.16) есть общее решение этой рекуррентности.

Подберем значения C_1 и C_2 так, чтобы (30a.16) удовлетворяло начальным условиям (30a.11). Для этого надо решить (относительно C_1 и C_2) систему линейных уравнений

$$\begin{cases} C_1 + C_2 = \alpha; \\ \lambda_1 C_1 + \lambda_2 C_2 = \beta. \end{cases} \tag{30a.17}$$

Определитель этой (квадратной) с.л.у. равен $\lambda_2 - \lambda_1$ и, следовательно, отличен от нуля. Решение существует и единственно при любых правых частях; его можно найти по формулам Крамера.

Этим уже доказано, что (30a.17) есть общее решение рекуррентности.

Переходим ко второму случаю, $D = 0$. Здесь характеристическое уравнение имеет единственный корень

$$\lambda_0 = \frac{a}{2}, \tag{30a.18}$$

который отличен от нуля (в противном случае коэффициент a равнялся бы нулю, и с ним вместе, в силу условия $D = a^2 + 4b = 0$, обращался бы в нуль коэффициент b, что противоречит предположению).

Корню λ_0 соответствует одно нетривиальное решение рекуррентности

$$u_k^{(1)} = \lambda_0^k. \tag{30a.19}$$

Имеется, однако, возможность построить еще одно нетривиальное решение, не пропорциональное первому. Рассмотрим последовательность

$$u_k^{(2)} = k\lambda_0^k \tag{30a.20}$$

и докажем, что она является решением рекуррентности (30a.10).

В самом деле,

$$\begin{aligned} &u_k^{(2)} - au_{k-1}^{(2)} - bu_{k-2}^{(2)} = \\ &= k\lambda_0^k - a(k-1)\lambda_0^{k-1} - b(k-2)\lambda_0^{k-2} = \lambda_0^{k-2}[k\lambda_0^2 - a(k-1)\lambda_0 - b(k-2)] = \\ &\qquad = \lambda_0^{k-2}[k(\lambda_0^2 - a\lambda_0 - b) + (a\lambda_0 + 2b)] = 0, \end{aligned}$$

поскольку, во-первых, $\lambda_0^2 - a\lambda_0 - b = 0$, т. к. λ_0 есть корень характеристического уравнения и, во-вторых, $a\lambda_0 + 2b = \frac{1}{2}(a^2 + 4b) = 0$.

Образуем теперь линейную комбинацию двух найденных нетривиальных решений

$$u_k = C_1 u_k^{(1)} + C_2 u_k^{(2)} = C_1\lambda_0^k + C_2 k\lambda_0^k = \lambda_0^k(C_1 + C_2 k) \tag{30a.21}$$

и докажем, что эта комбинация является общим решением рекуррентности.

Начальные условия приводят к с.л.у.

$$\begin{cases} C_1 & = & \alpha; \\ \lambda_0(C_1 + C_2) & = & \beta, \end{cases} \tag{30a.22}$$

которая, в силу того что $\lambda_0 \neq 0$, имеет единственное решение (при любых правых частях).

Значит, и в этом случае получено общее решение рекуррентности (32a.10).

И наконец, случай $D < 0$. Хотя он и не является менее интересным, но нам пока не по зубам, поскольку мы только в начале следующей главы собираемся приступать к изучению *комплексных чисел*, а действительных корней в этом случае нет.

Пример 30a.3. Решим рекуррентность

$$u_k = 3u_{k-1} - 2u_{k-2};\ k \geqslant 4;\ u_2 = 7;\ u_3 = 15.$$

Характеристическое уравнение: $\lambda^2 - 3\lambda + 2 = 0$.
Корни: $\lambda_1 = 1,\ \lambda_2 = 2$.
Общее решение рекуррентности: $u_k = C_1 + C_2 2^k$.
С.л.у. для отыскания неопределенных констант:

$$\begin{cases} C_1 + 4C_2 = 7; \\ C_1 + 8C_2 = 15. \end{cases}$$

Ее решение: $C_1 = -1;\ C_2 = 2$.
Частное решение рекуррентности, удовлетворяющее начальным условиям: $u_k^* = -1 + 2^{k+1}$.

Пример 30a.4 (без комментариев).

$$u_k = 2u_{k-1} - u_{k-2};\ u_2 = 3;\ u_3 = 4;$$
$$\lambda^2 - 2\lambda + 1 = 0;\ D = 0;\ \lambda_0 = 1;$$
$$u_k = C_1 + C_2 k;\ C_1 = C_2 = 1;\ u_k^* = 1 + k.$$

Замечание 27a.2. Maple умеет решать рекуррентности. В специальной команде rsolve надо (в фигурных скобках) набрать рекуррентное уравнение и начальные условия (аргумент последовательности заключается в круглые скобки); кроме того, следует явно указать, относительно какой переменной решается рекуррентность.

```
> rsolve(  u(k) = 3*u(k-1) - 2*u(k-2), u(2) = 7, u(3) = 15 , u );
```

$$2\ 2^k - 1$$

```
> rsolve(  u(k) = 2*u(k-1) - u(k-2), u(2) = 3, u(3) = 4 , u(k) );
```

$$k + 1$$

30a.3. Последовательность Фибоначчи. Один из примеров, посвященный так называемым *числам Фибоначчи*, в силу его исключительной важности, автору показалось уместным выделить в особый пункт.

Пример 30а.5. Последовательность чисел Фибоначчи определяется рекуррентностью:

$$f_k = f_{k-1} + f_{k-2};\ f_0 = 1;\ f_1 = 1. \tag{30а.23}$$

Несколько ее первых членов таковы:

$$1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ ...$$

Характеристическое уравнение и его корни:

$$\lambda^2 - \lambda - 1 = 0;\ \lambda_{1,2} = \frac{1 \pm \sqrt{5}}{2}.$$

Явная формула для чисел Фибоначчи:

$$f_k = \frac{1}{\sqrt{5}} \left(- \left(\frac{1-\sqrt{5}}{2} \right)^{k+1} + \left(\frac{1+\sqrt{5}}{2} \right)^{k+1} \right). \tag{30а.24}$$

Впечатляет тот факт, что в явной формуле для членов заведомо целочисленной последовательности фигурирует иррациональное число $\sqrt{5}$.

Замечание 30а.3. Формулу для чисел Фибоначчи Maple представляет следующим образом:

```
> rsolve(  f(k) = f(k−1) + f(k−2), f(0) = 1, f(1) = 1, f(k) );
```

$$\left(\frac{\sqrt{5}}{10} + \frac{1}{2} \right) \left(\frac{1}{2} + \frac{\sqrt{5}}{2} \right)^k + \left(-\frac{\sqrt{5}}{10} + \frac{1}{2} \right) \left(\frac{1}{2} - \frac{\sqrt{5}}{2} \right)^k$$

30а.4. Определители некоторых трехдиагональных матриц. Рассмотрим определитель k-го порядка следующего специального вида:

$$d_k = \begin{vmatrix} a & c & 0 & 0 & ... & 0 & 0 & 0 \\ b & a & c & 0 & ... & 0 & 0 & 0 \\ 0 & b & a & c & ... & 0 & 0 & 0 \\ 0 & 0 & b & a & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & ... & a & c & 0 \\ 0 & 0 & 0 & 0 & ... & b & a & c \\ 0 & 0 & 0 & 0 & ... & 0 & b & a \end{vmatrix}, \tag{30а.25}$$

где $a, b, c \in \mathbb{R};\ k \geqslant 2$.

При $k = 2$ имеем: $d_2 = a^2 - bc$, а при $k = 3$ получим: $d_3 = a^3 - 2abc$.

Чтобы вычислить этот определитель при $k \geqslant 4$, разложим его по первой строке.

Алгебраическим дополнением к элементу $a_{11} = a$ будет, очевидно, $A_{11} = d_{k-1}$, а алгебраическим дополнением к $a_{12} = c$ будет

$$A_{12} = - \begin{vmatrix} b & c & 0 & \dots & 0 & 0 & 0 \\ 0 & a & c & \dots & 0 & 0 & 0 \\ 0 & b & a & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a & c & 0 \\ 0 & 0 & 0 & \dots & b & a & c \\ 0 & 0 & 0 & \dots & 0 & b & a \end{vmatrix} = -b\, d_{k-2},$$

где на последнем шаге последний определитель, $(k-1)$-го порядка, разложен по первому столбцу. При этом алгебраическим дополнением к (единственному в столбце ненулевому) угловому элементу оказывается определитель $(k-2)$-го порядка, такого же вида, что и исходный.

Получаем рекуррентную формулу

$$d_k = a d_{k-1} - bc\, d_{k-2}. \tag{30a.26}$$

Характеристическим уравнением будет

$$\lambda^2 - a\lambda + bc = 0.$$

Если дискриминант $D = a^2 - 4bc \geqslant 0$, то эту рекуррентность можно решить описанным в п. 30a.3 способом, и тем самым получить выражение для определителя d_k при любом $k \geqslant 2$.

Пример 30a.6. 1. При $a = 3$, $b = 1$, $c = 2$ получается рекуррентность из примера 30a.3 и определитель находится по формуле $d_k = 2^{k+1} - 1$.

2. При $a = 2$, $b = c = 1$ получается рекуррентность из примера 30a.4 и определитель находится по формуле $d_k = k + 1$.

3. При $a = c = 1$, $b = -1$ получаем представление в виде определителя чисел Фибоначчи:

$$f_k = \begin{vmatrix} 1 & 1 & 0 & 0 & ... & 0 & 0 & 0 \\ -1 & 1 & 1 & 0 & ... & 0 & 0 & 0 \\ 0 & -1 & 1 & 1 & ... & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & ... & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & ... & -1 & 1 & 1 \\ 0 & 0 & 0 & 0 & ... & 0 & -1 & 1 \end{vmatrix}. \qquad (30a.27)$$

30a.5. Определитель Вандермонда. В этом пункте мы рассмотрим еще один "знаменитый" определитель, который носит имя, указанное в заголовке пункта, и определяется формулой

$$\Delta_n = \Delta_n(x_1, x_2, ..., x_n) = \\ = \begin{vmatrix} 1 & 1 & 1 & ... & 1 & 1 \\ x_1 & x_2 & x_3 & ... & x_{n-1} & x_n \\ x_1^2 & x_2^2 & x_3^2 & ... & x_{n-1}^2 & x_n^2 \\ ... & ... & ... & ... & ... & \\ x_1^{n-2} & x_2^{n-2} & x_3^{n-2} & ... & x_{n-1}^{n-2} & x_n^{n-2} \\ x_1^{n-1} & x_2^{n-1} & x_3^{n-1} & ... & x_{n-1}^{n-1} & x_n^{n-1} \end{vmatrix}, \qquad (30a.28)$$

где $n \geqslant 2$.

Как видите, этот определитель зависит от n параметров (переменных) $x_1, ..., x_n$. Поскольку при вычислении определителя используются лишь алгебраические действия сложения и умножения, то можно заранее утверждать, что $\Delta_n(x_1, ..., x_n)$ является многочленом от n своих переменных. (Теории многочленов от нескольких переменных будут посвящены § 48, 49. Там нам предстоит еще одна встреча с определителем Вандермонда.)

Начальным значением для рекуррентности (первого порядка), которую мы далее выведем, будет

$$\Delta_2(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ x_1 & x_2 \end{vmatrix} = x_2 - x_1. \qquad (30a.29)$$

Чтобы вывести рекуррентную формулу, воспользуемся тем фактом, что значение определителя не меняется при элементарных преобразованиях второго типа над строками.

В определителе (30a.28) прибавим к каждой строке, начиная с последней, предыдущую строку, домноженную на $(-x_1)$; затем полученный определитель разложим по первому столбцу (он содержит

всего один элемент), а в получившемся определителе порядка $n-1$ из каждого столбца вынесем образовавшийся в этом столбце общий множитель. Получим:

$$\Delta_n(x_1, x_2, ..., x_n) =$$

$$= \begin{vmatrix} 1 & 1 & 1 & ... & 1 \\ 0 & x_2 - x_1 & x_3 - x_1 & ... & x_n - x_1 \\ 0 & x_2^2 - x_2 x_1 & x_3^2 - x_3 x_1 & ... & x_n^2 - x_n x_1 \\ ... & ... & ... & ... & ... \\ 0 & x_2^{n-1} - x_2^{n-2} x_1 & x_3^{n-1} - x_3^{n-2} x_1 & ... & x_n^{n-1} - x_n^{n-2} x_1 \end{vmatrix} =$$

$$= \begin{vmatrix} x_2 - x_1 & x_3 - x_1 & ... & x_n - x_1 \\ x_2^2 - x_2 x_1 & x_3^2 - x_3 x_1 & ... & x_n^2 - x_n x_1 \\ ... & ... & ... & ... \\ x_2^{n-1} - x_2^{n-2} x_1 & x_3^{n-1} - x_3^{n-2} x_1 & ... & x_n^{n-1} - x_n^{n-2} x_1 \end{vmatrix} =$$

$$= \begin{vmatrix} x_2 - x_1 & x_3 - x_1 & ... & x_n - x_1 \\ x_2(x_2 - x_1) & x_3(x_3 - x_1) & ... & x_n(x_n - x_1) \\ ... & ... & ... & ... \\ x_2^{n-2}(x_2 - x_1) & x_3^{n-2}(x_3 - x_1) & ... & x_n^{n-2}(x_n - x_1) \end{vmatrix} =$$

$$= (x_2 - x_1)(x_3 - x_1)...(x_n - x_1) \begin{vmatrix} 1 & 1 & ... & 1 \\ x_2 & x_3 & ... & x_n \\ ... & ... & ... & ... \\ x_2^{n-2} & x_3^{n-2} & ... & x_n^{n-2} \end{vmatrix} =$$

$$= (x_2 - x_1)(x_3 - x_1)...(x_n - x_1)\, \Delta_{n-1}(x_2, x_3, ..., x_n).$$

Объединяя начало и конец цепочки вычислений, получаем:

$$\Delta_n(x_1, x_2, ..., x_n) = \\ = (x_2 - x_1)(x_3 - x_1)...(x_n - x_1)\, \Delta_{n-1}(x_2, x_3, ..., x_n). \quad (30a.30)$$

Теперь можно сформулировать окончательный результат.

Предложение 30а.1. *Определитель Вандермонда (30а.28) может быть вычислен по формуле*

$$\Delta_n(x_1, x_2, ..., x_n) = \prod_{1 \leqslant i < j \leqslant n} (x_j - x_i). \qquad (30а.31)$$

Этот определитель отличен от нуля тогда и только тогда, когда все значения $x_1, x_2, ..., x_n$ попарно различны.

Доказательство. При $n = 2$ формула (30а.31) верна, поскольку сводится к (30а.29). Индуктивный шаг обеспечивает рекуррентная формула (30а.30).

Второе утверждение предложения очевидным образом следует из формулы (30а.31). □

Замечание 30а.4. Именно тем фактом, что определитель Вандермонда служит "индикатором" наличия (отсутствия) совпадений в списке чисел $x_1, x_2, ..., x_n$, объясняется его важная роль во многих разделах математики.

Например, в § 49 с помощью этого определителя будет введено понятие *дискриминанта* для многочлена степени n, который отвечает за наличие (отсутствие) у многочлена кратных корней.

Глава 5

ПОЛЕ КОМПЛЕКСНЫХ ЧИСЕЛ

§ 31. Векторная модель поля комплексных чисел

31.1. Интуитивное представление о комплексных числах. Квадратное уравнение с действительными коэффициентами (случай отрицательного дискриминанта). "Алгебраические недостатки" поля действительных чисел $\mathbb{R}$ начинают давать о себе знать при изучении алгебраической операции извлечения квадратного корня.

Квадратным корнем из действительного числа a называется такое действительное число x_0, что $x_0^2 = a$. Иначе говоря, квадратный корень из a — это решение уравнения

$$x^2 = a. \tag{31.1}$$

Квадратный корень существует лишь для $a \geqslant 0$ и принимает (при $a > 0$) два различных значения:

$$x_{1,2} = \pm\sqrt{a}; \tag{31.2}$$

положительное значение квадратного корня $\sqrt{a}$ называется *арифметическим квадратным корнем* из a.

Квадратный корень из 0 принимает одно значение $x_0 = 0$; можно, однако, условиться, что в этом случае имеются "два одинаковых значения" корня (два одинаковых решения $x_1 = x_2 = 0$ для уравнения $x^2 = 0$).

При $a < 0$ уравнение (31.1) действительных решений не имеет.

В этом случае мы представим число a в виде произведения

$$a = (-a) \cdot (-1),$$

где $-a > 0$, и заметим: для того чтобы научиться извлекать квадратные корни из отрицательных чисел, достаточно определить, чему равен корень квадратный из -1.

Введем в рассмотрение новый *символ*

$$i = \sqrt{-1}, \tag{31.3}$$

который будем называть *мнимой единицей*. По построению получим:

$$i^2 = -1. \tag{31.4}$$

Таким образом, "новое число" i будет удовлетворять уравнению $x^2 = -1$. Число $-i$ также будет удовлетворять этому уравнению, и поскольку соблюдение обычных алгебраических законов (в новой, более широкой числовой области) потребует, чтобы квадратное уравнение имело не более двух решений, то формула $x_{1,2} = \pm i$ будет исчерпывать множество решений уравнения $x^2 = -1$.

Теперь мы сумеем найти множество решений уравнения (31.1) с отрицательной правой частью:

$$x_{1,2} = \pm\sqrt{-a} \cdot i, \tag{31.5}$$

где под $\sqrt{-a}$ понимается арифметический квадратный корень.

Выражения вида $y \cdot i$, где $y \in \mathbb{R}$, $y \neq 0$, называются *(чисто) мнимыми числами*. В этих терминах можно заявить, что уравнение $x^2 = a$ с отрицательной правой частью имеет два (различных) чисто мнимых решения или, другими словами, существует два чисто мнимых значения для квадратного корня из отрицательного действительного числа.

Комплексным числом называется выражение вида

$$z = x + yi, \tag{31.6}$$

где x и y — действительные числа, а i — мнимая единица.

При $y = 0$ комплексное число (31.6) является действительным, а при $x = 0$ и $y \neq 0$ — чисто мнимым. Действительное число x называется *действительной частью* комплексного числа z; вводится обозначение

$$\mathrm{Re}(z) = x. \tag{31.7}$$

Мнимой частью комплексного числа z называется (также действительное!) число y; обозначение:

$$\mathrm{Im}(z) = y. \tag{31.8}$$

Два комплексных числа $z = x + yi$ и $z' = x' + y'i$ считаются равными, если равны их действительные и мнимые части: $x = x'$ и $y = y'$.

Множество всех комплексных чисел определяется формулой

$$\mathbb{C} = \{\, z = x + yi \,:\, x, y \in \mathbb{R};\, i^2 = -1 \,\}. \tag{31.9}$$

Рассмотрим квадратное уравнение с действительными коэффициентами ($a, b, c \in \mathbb{R};\ a \neq 0$):

$$ax^2 + bx + c = 0. \tag{31.10}$$

В случае положительности дискриминанта

$$D = b^2 - 4ac \tag{31.11}$$

это уравнение имеет два различных действительных решения:

$$x_{1,2} = \frac{-b \pm \sqrt{D}}{2a}. \tag{31.12}$$

В случае $D = 0$ уравнение (31.10) имеет одно действительное решение

$$x_0 = -\frac{b}{2a}. \tag{31.13}$$

Допустимо также говорить о двух одинаковых решениях:

$$x_1 = x_2 (= x_0).$$

В случае $D < 0$ формуле (31.12) можно придать смысл, если извлечь корень квадратный из дискриминанта с использованием мнимых чисел [см. формулу (31.5)]:

$$x_{1,2} = \frac{-b \pm i\sqrt{-D}}{2a}. \tag{31.14}$$

Если мы убедимся предварительно, что над комплексными числами можно производить все те же алгебраические действия, что и над действительными, с соблюдением всех девяти аксиом поля (см. п. 2.1), то [поскольку при выводе формулы (31.12) ничего, кроме

этих аксиом, не используется] мы сможем заключить, что эта формула останется справедливой и в случае отрицательного дискриминанта. В этом случае формула (31.12) примет вид (31.14) и будет определять два *комплексных решения* для уравнения (31.10):

$$x_1 = \alpha + \beta i;\ x_2 = \alpha - \beta i, \tag{31.15}$$

где действительные числа α и β определяются формулами

$$\alpha = -\frac{b}{2a};\ \beta = \frac{\sqrt{-D}}{2a}. \tag{31.16}$$

Сопряженным (или *комплексно сопряженным*) к комплексному числу (31.6) называется комплексное число

$$\bar{z} = x - yi. \tag{31.17}$$

Действительные числа (и только они) являются "самосопряженными".

В случае $D < 0$ комплексные решения уравнения (31.10) (с действительными коэффициентами) являются сопряженными друг другу:

$$x_2 = \overline{x_1}. \tag{31.18}$$

Замечание 31.1. Приведенные выше "наивные" (наводящие) соображения по поводу введения комплексных чисел являются нестрогими и требуют уточнения и детализации. В особенности нас должна беспокоить *возможность* введения такого "символа", как мнимая единица. (Откуда берется эта "мифическая сущность"?) Единственным строгим способом обосновать введение числа i и комплексных чисел является построение модели для множества $\mathbb{C}$ в рамках уже развитой теории "действительных объектов" (например, векторов или матриц с элементами из поля $\mathbb{R}$).

Этому будут посвящены пп. 31.3 и 31.4, а пока мы займемся определением алгебраических действий над комплексными числами, оставаясь в рамках наивного представления о них. Руководящим принципом будет представление о комплексных числах как о выражениях (двучленах) вида (31.6), которые можно складывать (вычитать, перемножать и т. д.) по законам элементарной алгебры (используя правило раскрытия скобок, приведение подобных членов и т. п.) с единственным дополнительным предположением (31.4).

31.2. Алгебраические действия над комплексными числами. *Сумма* двух комплексных чисел

$$z = x + yi;\ w = u + vi \tag{31.19}$$

определяется естественной формулой

$$z + w = (x + yi) + (u + vi) = (x + u) + (y + v)i, \tag{31.20}$$

т. е. при сложении комплексных чисел по отдельности складываются их действительные и мнимые части. Роль нуля в множестве $\mathbb{C}$ будет играть (действительное) число $0 = 0 + 0 \cdot i$. Противоположным к числу z будет число $-z = (-x) + (-y)i$. Вычитание $z - w$ сводится к сложению $z + (-w)$.

Произведение чисел (31.19) определяется с помощью следующей выкладки [использующей (31.4)]:

$$z \cdot w = (x + yi) \cdot (u + vi) = x \cdot u + x \cdot (vi) + (yi) \cdot u + (yi) \cdot (vi) = \\ = xu + (yv)i^2 + (xv + yu)i = (xu - yv) + (xv + yu)i,$$

или в окончательном виде:

$$z \cdot w = (xu - yv) + (xv + yu)i. \tag{31.21}$$

Роль единичного элемента в множестве комплексных чисел играет обычная (действительная) единица. В частном случае, когда один из сомножителей в формуле (31.21) является действительным числом, умножение приобретает "покомпонентный" характер: действительная и мнимая части комплексного числа по отдельности умножаются на действительное число:

$$x \cdot w = (xu) + (xv)i;\ x \in \mathbb{R}. \tag{31.22}$$

Важным особым случаем является произведение комплексного числа на сопряженное число:

$$z \cdot \bar{z} = (x + yi) \cdot (x - yi) = x^2 + y^2. \tag{31.23}$$

Результат является действительным числом, причем неотрицательным. Более того, можно заметить, что произведение $z\bar{z}$ положительно (соответственно равно 0) тогда и только тогда, когда $z \neq 0$ (соответственно $z = 0$).

Для любого ненулевого комплексного числа z существует *обратное* число: оно получается, если сопряженное число $\bar{z}$ домножить на (положительное) действительное число $(z\bar{z})^{-1}$:

$$z^{-1} = \frac{1}{z\bar{z}} \cdot \bar{z} = \frac{1}{x^2+y^2} \cdot (x - yi) = \frac{x}{x^2+y^2} - i\frac{y}{x^2+y^2}. \qquad (31.24)$$

[Убедитесь в том, что произведение данного числа z на последнее выражение в цепочке равенств (31.24) в самом деле дает единицу.]

Деление комплексного числа w на комплексное число $z \neq 0$ осуществляется как умножение $w \cdot z^{-1}$. На практике деление оформляют следующим образом:

$$\frac{w}{z} = \frac{u+vi}{x+yi} = \frac{(u+vi)(x-yi)}{(x+yi)(x-yi)} = \frac{(ux+vy)+(vx-uy)i}{x^2+y^2}.$$

Это вычисление часто сопровождают словами: *домножим числитель и знаменатель дроби на число, сопряженное к знаменателю.* Окончательный результат выглядит следующим образом:

$$\frac{w}{z} = \frac{ux+vy}{x^2+y^2} + \frac{vx-uy}{x^2+y^2}i. \qquad (31.25)$$

Для алгебраических действий в $\mathbb{C}$ сохраняются все "полевые законы" (см. п. 2.1) и их следствия (такие, например, как "самые школьные" *формулы сокращенного умножения*; кстати, раньше эти формулы школьники помнили лучше, поскольку их писали на ученических портфелях).

Остаются справедливыми и более сложные формулы, скажем *биномом Ньютона*:

$$(z+w)^n = \sum_{k=0}^{n} C_n^k z^{n-k} w^k, \qquad (31.26)$$

где $z, w \in \mathbb{C}$; $n \in \mathbb{N}$; $C_n^k = \frac{n!}{k!(n-k)!}$ — биномиальный коэффициент (число сочетаний из n элементов по k).

По формуле (31.26) удобно производить возведение комплексного числа в натуральную степень:

$$(x+yi)^n = \sum_{k=0}^{n} C_n^k x^{n-k} y^k i^k.$$

Следует только учесть периодическое, с периодом 4, повторение степеней мнимой единицы:

$$i^0 = 1;\ i^1 = i;\ i^2 = -1;\ i^3 = -i;\ i^4 = 1;\ i^5 = i; \dots \qquad (31.27)$$

В окончательном ответе можно перегруппировать слагаемые, отделив действительную часть от мнимой; получатся две суммы (в каждой из которых чередуются знаки):

$$(x+yi)^n = \\ = \sum_{\substack{l\geqslant 0\\ 2l\leqslant n}} (-1)^l C_n^{2l} x^{n-2l} y^{2l} + i \sum_{\substack{l\geqslant 0\\ 2l+1\leqslant n}} (-1)^l C_n^{2l+1} x^{n-2l-1} y^{2l+1}. \qquad (31.28)$$

[Не пугайтесь формулы (31.28) — ее не следует запоминать: в конкретных примерах будут конкретные значения n и слагаемые "сами собой" перегруппируются.]

Пример 31.1. Рассмотрим комплексные числа $z = 2 + 3i$ и $w = -5 + 7i$. Выполним над ними четыре арифметических действия:

$$z + w = -3 + 10i;\ z - w = 7 - 4i;$$

$$z \cdot w = (2+3i)(-5+7i) = -10 + 21i^2 + 14i - 15i = -31 - i;$$

$$\frac{w}{z} = \frac{-5+7i}{2+3i} = \frac{(-5+7i)(2-3i)}{(2+3i)(2-3i)} = \frac{11+29i}{13} = \frac{11}{13} + \frac{29}{13}i.$$

Возведем число z в степень $n = 5$:

$$z^5 = (2+3i)^5 = \\ = C_5^0 2^5 3^0 i^0 + C_5^1 2^4 3^1 i^1 + C_5^2 2^3 3^2 i^2 + C_5^3 2^2 3^3 i^3 + C_5^4 2^1 3^4 i^4 + C_5^5 2^0 3^5 i^5 = \\ = 1\cdot 32\cdot 1\cdot 1 + 5\cdot 16\cdot 3\cdot i + 10\cdot 8\cdot 9\cdot(-1) + 10\cdot 4\cdot 27\cdot(-i) + 5\cdot 2\cdot 81\cdot 1 + 1\cdot 1\cdot 243\cdot i = \\ = (32 - 720 + 810) + (240 - 1080 + 243)i = 122 - 597i.$$

Пример 31.2. Система Maple работает с комплексными числами столь же легко, как и с действительными. Для мнимой единицы в Maple зарезервировано обозначение I.

$>$ z := $2 + 3 * \mathrm{I}$; w := $-5 + 7 * \mathrm{I}$;

$$z := 2 + 3I$$

$$w := -5 + 7I$$

> [z + w, z − w, z ∗ w, w/z, z^5];

$$[-3 + 10I,\ 7 - 4I,\ -31 - I,\ \frac{11}{13} + \frac{29}{13}I,\ 122 - 597I]$$

> (1 + I)^103;

$$2251799813685248 - 2251799813685248I$$

Maple знает команды вычисления действительной и мнимой части, сопряженного числа:

> [Re(z), Im(w), conjugate(w), w ∗ conjugate(w)];

$$[2,\ 7,\ -5 - 7I,\ 74]$$

При вычислениях с символьными выражениями понадобится команда evalc (вычислить в поле $\mathbb{C}$):

> evalc((a + b ∗ I)^5);

$$a^5 - 10a^3b^2 + 5ab^4 + (5a^4b - 10a^2b^3 + b^5)I$$

31.3. Комплексные числа как двумерные действительные векторы. Формулы (31.20) и (31.22) наводят на мысль, что комплексные числа складываются и умножаются на действительные числа так же, как двумерные действительные арифметические векторы. Две компоненты такого вектора суть действительная и мнимая части комплексного числа.

Рассмотрим в связи с этим арифметическое линейное пространство $\mathbb{R}^2$, составленное из векторов-строк вида

$$z = (x,\ y);\ x, y \in \mathbb{R}. \tag{31.29}$$

Замечание 31.2. В данном конкретном случае для обозначения таких "маленьких" векторов-строк мы не будем использовать черточки (и знаки транспонирования). Делается это с целью упрощения символики и в связи с тем, что черточки нам понадобятся для обозначения комплексно сопряженных чисел. (Тем самым мы временно, в пределах данного пункта, отступаем от наших общих договоренностей; см. п. 1.3 и замечание 2.2.)

Над векторами рассматриваемого вида, (31.29) и $w = (u, v)$, выполнимы обычные *линейные* алгебраические действия: сложение и умножение на (действительный) скаляр. Эти действия производятся покомпонентно:

$$z + w = (x + u, y + v); \; xw = (xu, xv). \tag{31.30}$$

Для сложения векторов выполняются четыре аксиомы, идентичные аксиомам поля $\boxed{1}$ — $\boxed{4}$ (см. п. 2.1).

В пространстве $\mathbb{R}^2$ имеется естественный базис. Для элементов этого базиса мы также выберем "специфические" обозначения, отличные от принятых в предыдущих главах, $\bar{e}_1^t, \bar{e}_2^t$. А именно: положим

$$1 = (1, 0); \; i = (0, 1). \tag{31.31}$$

Эти векторы будут далее играть роль "обычной" и мнимой единиц. Теперь всякий вектор вида (31.29) может быть представлен своим разложением по базису:

$$z = (x, y) = x(1, 0) + y(0, 1) = x \cdot 1 + y \cdot i,$$

т. е. в форме, идентичной записи (31.6) комплексного числа.

Условимся отождествлять в записи действительные числа $x \in \mathbb{R}$ и их векторные представления $(x, 0) \in \mathbb{R}^2$, т. е. будем считать, что $x = (x, 0)$.

Теперь нам предстоит сделать самый главный шаг в построении модели — определить алгебраическое действие умножения векторов $z, w \in \mathbb{R}^2$. Сделаем мы это "по подсказке" полученной выше с помощью интуитивных соображений формулы (31.21):

$$z \cdot w = (xu - yv, xv + yu). \tag{31.32}$$

Проверим аксиомы поля $\boxed{5}$ — $\boxed{9}$.

Коммутативный закон для умножения $\boxed{7}$ практически очевиден (ниже, в цепочке вычислений, используются законы для сложения и умножения в поле $\mathbb{R}$):

$$z \cdot w = (x, y) \cdot (u, v) = (xu - yv, xv + yu) = \\ = (ux - vy, uy + vx) = (u, v) \cdot (x, y) = w \cdot z.$$

Докажем аксиому $\boxed{5}$. (В силу коммутативности умножения, все равно в какой форме, левой или правой, доказывать дистрибутивный закон.)

$$\begin{aligned} z_1 \cdot (z_2 + z_3) = \\ = (x_1, y_1) \cdot ((x_2, y_2) + (x_3, y_3)) = (x_1, y_1) \cdot (x_2 + x_3, y_2 + y_3) = \\ = (\, x_1(x_2 + x_3) - y_1(y_2 + y_3),\, x_1(y_2 + y_3) + y_1(x_2 + x_3)\,) = \\ = (\, (x_1x_2 - y_1y_2) + (x_1x_3 - y_1y_3),\, (x_1y_2 + y_1x_2) + (x_1y_3 + y_1x_3)\,) = \\ = (x_1x_2 - y_1y_2,\, x_1y_2 + y_1x_2) + (x_1x_3 - y_1y_3,\, x_1y_3 + y_1x_3) = \\ = z_1 \cdot z_2 \,+\, z_1 \cdot z_3. \end{aligned}$$

Попробуйте сами доказать ассоциативный закон (аксиому $\boxed{6}$)

$$(z_1 \cdot z_2) \cdot z_3 = z_1 \cdot (z_2 \cdot z_3).$$

Справедливость аксиомы существования единицы $\boxed{8}$ очевидна: эту роль играет действительная единица $1 = (1,\, 0)$.

Если мы умножим мнимую единицу $i = (0, 1)$ на себя [по правилу (31.32)], то, как легко видеть, получим:

$$i^2 = (0, 1) \cdot (0, 1) = (-1, 0) = -1.$$

Таким образом, "таинственное" тождество (31.4) оказалось следствием определения умножения в $\mathbb{R}^2$, а мнимая единица обрела вполне реалистическое представление.

Перед тем как доказывать последнюю из аксиом поля $\boxed{9}$, определим (в векторной модели) операцию сопряжения комплексных чисел:

$$z = (x,\, y) \;\mapsto\; (x,\, -y) = \bar{z}. \tag{31.33}$$

Формула $z\bar{z} = x^2 + y^2$ [ср. с (31.23)] немедленно получается из (31.32).

Существование обратного элемента для ненулевого вектора

$$z = (x, y) \in \mathbb{R}^2 \;\; (x^2 + y^2 > 0)$$

доказывается теперь непосредственно умножением этого вектора на вектор $\frac{1}{x^2+y^2}(x, -y)$. Получается единица. Значит, для вектора z существует обратный вектор:

$$z^{-1} = \left(\frac{x}{x^2 + y^2},\, \frac{-y}{x^2 + y^2} \right). \tag{31.34}$$

Подведем итоги. Справедлива следующая

Теорема 31.1. *Арифметическое линейное пространство $\mathbb{R}^2$ с заданной на нем операцией умножения (31.32) является полем. Это поле*

$$\mathbb{C} = \{(x,\, y) : x, y \in \mathbb{R}\} \tag{31.35}$$

содержит в качестве подполя поле действительных чисел

$$\mathbb{R} = \{(x,\, 0) : x \in \mathbb{R}\}$$

и содержит элемент $i = (0, 1)$, удовлетворяющий условию $i^2 = -1$. Всякий элемент $z \in \mathbb{C}$ представляется в виде $z = x + yi$, где $x, y \in \mathbb{R}$.

Доказательство см. выше. □

Замечание 31.3. Мы построили новое поле $\mathbb{C}$, являющееся *расширением* поля $\mathbb{R}$ и содержащее мнимую единицу. Нетрудно понять (и затем строго доказать) тот факт, что построенное поле $\mathbb{C}$ обладает *свойством минимальности*: если P — произвольное поле, содержащее подполе, *изоморфное* $\mathbb{R}$, а также — мнимую единицу [т. е. элемент i, удовлетворяющий (31.4)], то поле P содержит подполе, изоморфное $\mathbb{C}$.

После этого доказывается, что *минимальное расширение*, содержащее $\mathbb{R}$ и мнимую единицу, определено однозначно (с точностью до изоморфизма) или, другими словами, оно изоморфно построенному выше полю $\mathbb{C}$.

Для дальнейшего нам понадобятся свойства операции комплексного сопряжения, которые мы здесь оформим в виде следующего утверждения.

Предложение 31.1. *Отображение*

$$^{-} : \mathbb{C} \longrightarrow \mathbb{C};\ z \mapsto \bar{z};\ z \in \mathbb{C}, \tag{31.36}$$

сопоставляющее комплексному числу сопряженное число, обладает следующими свойствами:

$$\overline{z \pm w} = \overline{z} \pm \overline{w};\ \overline{z \cdot w} = \overline{z} \cdot \overline{w};\ \overline{\left(\frac{z}{w}\right)} = \frac{\overline{z}}{\overline{w}};\ \overline{z^n} = \overline{z}^{\,n} \tag{31.37}$$

для любых $z, w \in \mathbb{C}$ (в случае деления $w \neq 0$) и любого $n \in \mathbb{N}$.

Кроме того, 1) равенство $\bar{z} = z$ имеет место тогда и только тогда, когда число z является действительным; 2) произведение $z\bar{z}$ является неотрицательным (при $z \neq 0$ — положительным) действительным числом; 3) для любого $z \in \mathbb{C}$ справедливо: $\overline{\overline{z}} = z$.

Доказательство частично проведено в предыдущем пункте. Остальное рассматривайте как простое упражнение. (Причем вычисления можете вести либо на векторном языке, записывая комплексные числа в форме $z = (x, y)$, либо пользуясь записью $z = x + yi$. Мы уже доказали равносильность этих двух форм представления комплексных чисел.) $\square$

Замечание 31.4. Свойства (31.37) обычно выражают следующей лаконичной формулировкой: отображение сопряжения (31.36) является изоморфизмом поля $\mathbb{C}$ на себя (или *автоморфизмом* поля $\mathbb{C}$).

31.4. Матричная модель для поля комплексных чисел. Изложенная в предыдущем пункте векторная модель $\mathbb{C} = \mathbb{R}^2$ является строгой, но, может быть, оставляет некоторую неудовлетворенность ввиду недостаточной мотивированности определения умножения (31.32).

Сейчас мы кратко изложим лучшую модель — матричную.

Рассмотрим кольцо $K = L(2, \mathbb{R})$ квадратных матриц порядка 2 с элементами из поля $\mathbb{R}$ (см. п. 14.1). Одновременно K является четырехмерным линейным пространством над $\mathbb{R}$. Относительно умножения кольцо K некоммутативно.

Рассмотрим в K подмножество P, составленное из матриц вида

$$Z = \begin{pmatrix} x & y \\ -y & x \end{pmatrix}, \tag{31.38}$$

где $x, y \in \mathbb{R}$. Каждая из таких матриц Z задается двумя действительными числами. Множество P является двумерным линейным подпространством в линейном пространстве K. Оно изоморфно арифметическому линейному пространству $\mathbb{C} = \mathbb{R}^2$, рассматривавшемуся в предыдущем пункте (матрице Z соответствует вектор $z = (x, y)$; при сложении матриц складываются и соответствующие векторы). В подпространстве P можно выбрать базис, составленный из матриц

$$E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix};\ J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix};$$

он будет соответствовать естественному базису в $\mathbb{C}$ (единичная матрица E будет соответствовать единице 1, а матрица J — мнимой единице i).

Перемножим теперь две матрицы из P: матрицу (31.38) и матрицу

$$W = \begin{pmatrix} u & v \\ -v & u \end{pmatrix}.$$

Получим:

$$Z \cdot W = \begin{pmatrix} x & y \\ -y & x \end{pmatrix} \cdot \begin{pmatrix} u & v \\ -v & u \end{pmatrix} = \begin{pmatrix} xu - yv & xv + yu \\ -(xv + yu) & xu - yv \end{pmatrix}.$$

Замечаем, что, во-первых, произведение матриц $Z \cdot W$ снова принадлежит P (т. е. подпространство P является подкольцом в K) а во-вторых, этому произведению отвечает произведение $z \cdot w$ векторов из $\mathbb{C}$. Ясно также, что подкольцо P является коммутативным.

Кроме того, для матрицы вида (31.38) определитель

$$\det(Z) = x^2 + y^2,$$

и следовательно, при $Z \neq O$ матрица Z обратима, а значит, подкольцо P является на самом деле *подполем* в кольце K; обратной матрице Z^{-1} соответствует обратное число z^{-1}.

Приходим к выводу, что соответствие $Z \mapsto z$ задает изоморфизм полей $P \cong \mathbb{C}$. При этом проясняется происхождение формулы, задающей умножение комплексных чисел в векторной модели: умножение (31.32) отвечает обычному умножению матриц в P.

Убедитесь в том, что $J^2 = -E$, что соответствует тождеству $i^2 = -1$. Заметьте также, что сопряженное комплексное число $\bar{z}$ связано с транспонированной матрицей Z^t, действительным числам отвечают скалярные матрицы, а чисто мнимым — антисимметрические матрицы.

31.5. Вычисления с комплексными числами в алгебраической форме. Под *алгебраической формой* записи комплексного числа понимается его обычная запись в виде $z = x + yi$ [см. формулу (31.6)]. Ниже мы изучим еще две формы записи комплексных чисел, а пока займемся очередной серией типовых задач над полем $\mathbb{C}$.

З а д а ч а 1. Решить с.л.у. [вида (1.10)]

$$\underset{m\times n}{A} \cdot \underset{n\times 1}{\bar{z}} = \underset{m\times 1}{\bar{b}},$$

где заданы матрица $A \in \mathrm{M}at(m, n; \mathbb{C})$ и вектор $\bar{b} \in \mathbb{C}^m$ и неизвестен вектор $\bar{z} \in \mathbb{C}^n$. (Ничего нельзя поделать с "перегруженностью" черточки: здесь черта над буквой говорит о том, что рассматривается арифметический вектор, а не комплексно сопряженное число. Когда эти понятия встретятся рядом, придется "выкручиваться", например заменять черту, обозначающую сопряжение, волной.)

Р е ш е н и е типовой задачи сразу продемонстрируем на конкретном примере.

Пример 31.3. Пусть

$$A = \begin{pmatrix} 1+i & 1 & 2-i & 1 \\ -1+i & i & i & -1-i \\ 2i & 1+i & 2 & -i \end{pmatrix};\ \bar{b} = \begin{pmatrix} 1 \\ i \\ 1+i \end{pmatrix}.$$

Составим расширенную матрицу для с.л.у.:

$$B = \left(\begin{array}{cccc|c} 1+i & 1 & 2-i & 1 & 1 \\ -1+i & i & i & -1-i & i \\ 2i & 1+i & 2 & -i & 1+i \end{array}\right).$$

Над полем комплексных чисел пропорциональные строки "невооруженным глазом" не видны. Поэтому начать преобразования полезно с деления каждой из строк на ее первый элемент (если он отличен от нуля). После этого будем "делать нули", переходить к следующему столбцу и т. д. (освежите ваш первый практический навык решения с.л.у.; см. пример 4.1).

$$B \xrightarrow[2^{\text{стр}}\cdot\frac{1}{-1+i};\ 3^{\text{стр}}\cdot\frac{1}{2i}]{1^{\text{стр}}\cdot\frac{1}{1+i}} \left(\begin{array}{cccc|c} 1 & \frac{1-i}{2} & \frac{1-3i}{2} & \frac{1-i}{2} & \frac{1-i}{2} \\ 1 & \frac{1-i}{2} & \frac{1-i}{2} & i & \frac{1-i}{2} \\ 1 & \frac{1-i}{2} & -\frac{1}{2} & \frac{1-i}{2} & \frac{1-i}{2} \end{array}\right) \longrightarrow$$

$$\xrightarrow[3^{\text{стр}}+1^{\text{стр}}\cdot(-1)]{2^{\text{стр}}+1^{\text{стр}}\cdot(-1)} \left(\begin{array}{cccc|c} 1 & \frac{1-i}{2} & \frac{1-3i}{2} & \frac{1-i}{2} & \frac{1-i}{2} \\ 0 & 0 & i & \frac{-1+3i}{2} & 0 \\ 0 & 0 & \frac{-1+i}{2} & \frac{-2+i}{2} & 0 \end{array}\right) \longrightarrow$$

$$\xrightarrow[3^{\text{стр}}+2^{\text{стр}}\cdot\frac{1-i}{2}]{2^{\text{стр}}\cdot\frac{1}{i};\ 1^{\text{стр}}+2^{\text{стр}}\cdot\frac{-1+3i}{2}} \left(\begin{array}{cccc|c} 1 & \frac{1-i}{2} & 0 & \frac{-2+3i}{2} & \frac{1-i}{2} \\ 0 & 0 & 1 & \frac{3+i}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right);$$

$$\bar{z} = z_2 \begin{pmatrix} \frac{-1+i}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix} + z_4 \begin{pmatrix} \frac{2-3i}{2} \\ 0 \\ -\frac{3+i}{2} \\ 1 \end{pmatrix} + \begin{pmatrix} \frac{-1+i}{2} \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

"За кадром" остались достаточно кропотливые вычисления (пересчет элементов в ходе элементарных преобразований). Вообще, линейная алгебра над $\mathbb{C}$ значительно сложнее (для "ручных" вычислений), нежели линейная алгебра над $\mathbb{R}$, хотя теоретически ничем от нее не отличается (см. замечание 2.1).

З а д а ч а 2. Исследовать на обратимость квадратную матрицу $A \in L(n, \mathbb{C})$; вычислить A^{-1} (если она существует).

Р е ш е н и е продемонстрируем на примере.

Пример 31.4. Пусть

$$A = \begin{pmatrix} 1+i & 2+i & -2 \\ 2 & i & 1-i \\ -1-i & -i & -1 \end{pmatrix}.$$

Вычислим обратную матрицу с помощью присоединенной (см. п. 28.3). Для этого сначала подсчитаем алгебраические дополнения ко всем элементам A и составим из них (не забудьте про транспонирование!) матрицу

$$A^{\vee} = \begin{pmatrix} 1 & 2+3i & 3+i \\ 0 & -3-3i & -6 \\ -1-i & -2-2i & -5-i \end{pmatrix}.$$

Используя (уже готовые) алгебраические дополнения, вычислим определитель $\det(A) = 3+3i$. Определитель отличен от нуля, поэтому матрица A обратима; выполняя деление матрицы A на число $3(1+i)$, получим:

$$A^{-1} = \frac{1}{6}\begin{pmatrix} 1-i & 5+i & 4-2i \\ 0 & -6 & -6+6i \\ -2 & -4 & -6+4i \end{pmatrix}.$$

Замечание 31.5. Для системы Maple комплексная линейная алгебра не представляет каких-либо затруднений. Работают все команды, уже встречавшиеся нам в предыдущих главах. (Иногда, правда, требуется дополнительное упрощение результатов с помощью команды evalc.)

З а д а ч а 3. Извлечь квадратный корень из комплексного числа $w = a+bi$ (т. е. найти в поле $\mathbb{C}$ решение $z = x+yi$ уравнения $z^2 = w$).

Р е ш е н и е. Уравнение представим в виде $(x+yi)^2 = a+bi$ или, после возведения в квадрат: $x^2 - y^2 + 2xyi = a+bi$. Комплексные числа равны тогда и только тогда, когда равны их действительные и мнимые части. Поэтому последнее уравнение (в поле $\mathbb{C}$) сводится к следующей системе уравнений (в поле $\mathbb{R}$):

$$\begin{cases} x^2 - y^2 = a; \\ 2xy = b. \end{cases}$$

Решение таких систем есть типичная школьная задача. Можно выписать решение в общем виде, но это будет довольно громоздко ввиду наличия различных случаев. Мы рекомендуем решать ее не по готовым формулам (даже если вы их найдете в справочниках); в каждом конкретном (числовом) примере будет лишь один случай (см. ниже образец).

Пример 31.5. Извлечь квадратный корень из комплексного числа $w = -32 - 126i$.

Ищем число $z = x + yi$ такое, что $(x + yi)^2 = -32 - 126i$, или, что равносильно, решаем систему

$$\begin{cases} x^2 - y^2 = -32; \\ 2xy = -126. \end{cases}$$

Из второго уравнения следует, что неизвестные $x, y \in \mathbb{R}$ отличны от нуля и имеют противоположные знаки. Можно выразить y через x:

$$y = -\frac{63}{x}.$$

Подставив это выражение в первое уравнение системы, получим для отыскания x уравнение

$$x^2 - \frac{3969}{x^2} = -32,$$

которое сводится к биквадратному:

$$x^4 + 32x^2 - 3969 = 0.$$

Введем новую неизвестную $t = x^2$ $(t \geqslant 0)$. Для t получится уравнение

$$t^2 + 32t - 3969 = 0;$$

оно имеет дискриминант 1690 и корни $t_1 = 49$ и $t_2 = -81$. Второй корень не подходит, поскольку является отрицательным. Остается $t = 49$, или $x^2 = 49$, откуда: $x_1 = 7$; $x_2 = -7$.

(Здесь важно не "переусердствовать"! Помните, что x, y и t являются действительными; решать уравнение $x^2 = -81$ не надо: в $\mathbb{R}$ оно решений не имеет.)

Находим соответствующие значения y: $y_1 = -9$; $y_2 = 9$. И наконец, выписываем

О т в е т. Корень квадратный принимает два значения:

$$z_{1,2} = \pm(7 - 9i).$$

Замечание 31.6. Принято одно из значений корня квадратного из числа $w = a + bi$ рассматривать в качестве *главного* (и обозначать символом $\sqrt{w}$), а именно: если мнимая часть $b \neq 0$, выбирается то из значений корня $z = x + yi$, у которого мнимая часть y имеет такой же знак, что и b; если же $b = 0$ (т. е. $w = a$ является действительным числом), то при $a \geqslant 0$ считается, что $\sqrt{w} = \sqrt{a}$ есть арифметический корень из a, а при $a < 0$ полагаем, что $\sqrt{w} = i\sqrt{b}$ (главное значение корня имеет положительную мнимую часть).

Именно это главное значение квадратного корня возвращается Maple-функцией sqrt. Например:

> sqrt(−32 − 126 ∗ I), sqrt(−32 + 126 ∗ I), sqrt(I), sqrt(−1);

$$7 - 9I,\ 7 + 9I,\ \frac{\sqrt{2}}{2} + \frac{1}{2}I\sqrt{2},\ I$$

Оба значения корня можно получить с помощью команды solve:

> solve(z^2 = −32 − 126I);

$$7 - 9I,\ -7 + 9I$$

З а д а ч а 4. Решить в поле $\mathbb{C}$ квадратное уравнение (31.10) с комплексными коэффициентами $a, b, c \in \mathbb{C}$ $(a \neq 0)$.

Р е ш е н и е дается обычной ("школьной") формулой (31.12) с той лишь разницей, что вычисление квадратного корня $\sqrt{D}$ осуществляется по алгоритму задачи 3.

Пример 31.6. Решить уравнение

$$(2 + i)z^2 - (7 + i)z + 15 + 10i = 0.$$

Вычислим дискриминант:

$$D = (-(7 + i))^2 - 4 \cdot (2 + i) \cdot (15 + 10i) = -32 - 126i.$$

Квадратный корень из этого числа мы извлекали в предыдущем примере: $\sqrt{D} = 7 - 9i$. По формуле (31.12) получим:

$$z_{1,2} = \frac{7 + i \pm (7 - 9i)}{2(2 + i)},$$

или

$$z_1 = \frac{7+i+7-9i}{2(2+i)} = \frac{7-4i}{2+i} = \frac{(7-4i)(2-i)}{(2+i)(2-i)} = \frac{10-15i}{5} = 2-3i,$$

и аналогично: $z_2 = 1+2i$.

О т в е т: $2-3i;\ 1+2i$.

Замечание 31.7. Решим ту же задачу средствами Maple. Для этого сначала зададим уравнение (присвоим переменной *eq* значение, "равное равенству"; затем применим универсальную команду solve):

> eq := (2 + I) * z^2 - (7 + I) * z + 15 + 10 * I = 0;

$$eq := (2+I)z^2 - (7+I)z + 15 + 10I = 0$$

> solve(eq);

$$2-3I,\ 1+2I$$

Можно разложить левую часть (lhs) уравнения *eq* на множители:

> factor(lhs(eq));

$$(2+I)(z-2+3I)(z-1-2I)$$

§ 32. Комплексные числа в тригонометрической форме

32.1. Геометрическое представление комплексного числа. Модуль, аргумент и тригонометрическая форма комплексного числа. Векторная модель (п. 31.3) с неизбежностью привносит в алгебру комплексных чисел геометрию: длины, углы.

Плоскость $\mathbb{C} = \mathbb{R}^2$ мы будем отождествлять с евклидовой плоскостью, которую вы изучали в школьном курсе математики и изучаете в курсе аналитической геометрии.

Двумерный вектор $z = (x,\, y) = x + yi$ изображается как радиус-вектор с началом в начале координат (см. рис. 32.1 в прил. 1).

Это и есть "геометрическое представление" комплексного числа.

Две координатные оси плоскости $\mathbb{C}$ получают специальные названия: ось, на которой располагаются все действительные числа [векторы $x = (x, 0)$], называется *действительной осью*, а ось, на которой

располагаются чисто мнимые числа [векторы $yi = (0, y)$], — *мнимой осью*.

Мы уже отмечали (см. п. 31.3), что сложение комплексных чисел производится по тем же законам, что и сложение двумерных векторов (т. е. по правилу параллелограмма; см. рис. 32.2), а умножение комплексного числа на действительное число производится так же, как умножение вектора на скаляр. Но в геометрии рассматриваются и другие понятия (которые, собственно, и выделяют геометрию из прочих математических дисциплин). Это прежде всего понятие *длины вектора*. В алгебре комплексных чисел ему соответствует понятие модуля комплексного числа.

Определение 32.1. *Модулем* комплексного числа $z \in \mathbb{C}$ называется неотрицательное действительное число

$$|z| = \sqrt{z \cdot \bar{z}} = \sqrt{x^2 + y^2}, \tag{32.1}$$

равное арифметическому квадратному корню из неотрицательного действительного числа $z \cdot \bar{z} = x^2 + y^2$.

Очевидно, модуль комплексного числа z равен длине r вектора z.

Расстояние между комплексными числами (векторами) $z, w \in \mathbb{C}$ (может быть, вам удобнее будет говорить о расстоянии между *концами* этих векторов) полагается равным длине разности этих векторов, т. е. модулю разности $|w - z|$ данных комплексных чисел.

Пользуясь школьной геометрией (или, если угодно, векторной алгеброй из курса аналитической геометрии; см. рис. 32.2), легко убедиться в справедливости *неравенств треугольника*:

$$||z| - |w|| \leqslant |z + w| \leqslant |z| + |w|. \tag{32.2}$$

(Для случая действительных чисел вы проходили эти неравенства в курсе математического анализа. Напомним, что неравенств треугольника три, поскольку длина *каждой* из сторон треугольника не превышает суммы длин двух других. Напомним также, что треугольник может "вырождаться" в случае коллинеарности векторов z и w.)

Есть еще четыре свойства модуля (первые три из которых являются "привычными", поскольку встречались в курсе математического анализа для действительных чисел):

$$|z \cdot w| = |z| \cdot |w|;\ \left|\frac{w}{z}\right| = \frac{|w|}{|z|};\ |z^n| = |z|^n;\ |\overline{z}| = |z|, \tag{32.3}$$

где $z, w \in \mathbb{C}$, $n \in \mathbb{N}$ и во втором равенстве $z \neq 0$.

Доказательство четвертого свойства очевидно. Проверим первую из формул (32.3); вторая и третья будут из нее следовать.

Используя свойства операции сопряжения (см. предложение 31.1), а также "полевые законы" (в поле $\mathbb{C}$) и свойства корня квадратного (в поле $\mathbb{R}$), получим:

$$|z \cdot w| = \sqrt{(zw) \cdot \overline{(zw)}} = \sqrt{z \cdot w \cdot \overline{z} \cdot \overline{w}} = \\ = \sqrt{(z \cdot \overline{z}) \cdot (w \cdot \overline{w})} = \sqrt{z \cdot \overline{z}} \cdot \sqrt{w \cdot \overline{w}} = |z| \cdot |w|.$$

Подведем итог.

Предложение 32.1. *Отображение*

$$|\,.\,| : \mathbb{C} \longrightarrow \mathbb{R};\ z \mapsto |z|;\ z \in \mathbb{C}, \tag{32.4}$$

сопоставляющее комплексному числу его модуль (неотрицательное действительное число), обладает свойствами (32.2) и (32.3).

Доказательство см. выше. □

Замечание 32.1. Уравнение $|z| = r$, где r — положительное действительное число, определяет на комплексной плоскости *окружность* радиуса r.

Вторым по важности понятием в евклидовой геометрии служит понятие *угла между векторами*. Рассмотрим (*ориентированный*) угол (см. рис. 32.1) от положительного направления действительной оси до вектора z (можно сказать иначе: угол от вектора $1 = (1, 0)$ до вектора z).

Как всегда в тригонометрии, этот угол определен не однозначно, а лишь *с точностью до целого кратного* 2π (т. е. с точностью до $2\pi k$, где $k \in \mathbb{Z}$), однако определение угла можно сделать однозначным, если потребовать, например, чтобы он принадлежал промежутку $(-\pi, \pi]$, задающему один полный оборот. (Иногда предпочитают другое требование: принадлежность угла промежутку $[0, 2\pi)$.)

В теорию комплексных чисел углы вводятся посредством следующего определения.

Определение 32.2. *Аргументом* ненулевого комплексного числа $z \in \mathbb{C}$ называется (любой) угол от положительного направления действительной оси до вектора z. Множество всех таких углов обозначается $\operatorname{Arg} z$. *Главным значением* аргумента комплексного числа

$z \in \mathbb{C}$ называется то из значений аргумента, которое принадлежит промежутку $(-\pi, \pi]$. Главное значение аргумента обозначается $\arg z$. (Аргумент числа 0 не определяется.)

Замечание 32.2. "Аргумент с большой буквы" (т. е. множество всех значений аргумента) можно выразить через "аргумент с маленькой буквы" формулой

$$\operatorname{Arg} z = \arg z + 2\pi k;\ k \in \mathbb{Z}. \tag{32.5}$$

На рис. 32.1 показан угол φ, являющийся главным значением аргумента числа z.

Замечание 32.3 (ностальгическое). Аналогичные формулы для *обратных тригонометрических функций* не так давно изучались в школьном курсе тригонометрии. Например, рассматривалась *многозначная* функция *арксинус* $y = \operatorname{Arcsin} x$ от действительного аргумента $x \in [-1, 1]$ и ее *главное значение* $\arcsin x \in [-\pi/2, \pi/2]$. Устанавливалось выражение

$$\operatorname{Arcsin} x = (-1)^n \arcsin x + \pi n;\ n \in \mathbb{Z}.$$

Однако буквально на глазах у автора настоящего пособия приходило в упадок преподавание тригонометрии и геометрии в средней школе. И вот уже простейшая и нагляднейшая задача решения (*косоугольных*) треугольников — представьте себе геодезиста с теодолитом! — исключается из обязательной программы. А дальше — пошло-поехало. Сегодня на математический факультет приходят выпускники, для которых серьезной проблемой является решение *прямоугольных* треугольников.

Если вдруг это является проблемой и для вас, не поленитесь — разберитесь с "школярской мудростью", не усвоенной вами в детстве. Без нее вам не стать ни математиком, ни настоящим компьютерщиком.

(Задачи из алгебры комплексных чисел, сводящиеся к тригонометрическим вычислениям, увы, очень плохо даются современным первокурсникам. Лет тридцать назад эти задачи входили в обязательную программу для школьников...)

Замечание 32.4. Уравнение $\arg z = \varphi$, где φ — действительное число, принадлежащее промежутку $(-\pi, \pi]$, определяет на комплексной плоскости *луч* такой, что угол от положительного направления

действительной оси до этого луча равен φ (число 0 исключается, поскольку для него аргумент не определен).

Замечание 32.5. Пара действительных чисел (φ, r), где φ — любое из значений аргумента для комплексного числа z, а $r = |z|$, есть не что иное, как *полярные координаты* точки на плоскости, являющейся концом вектора z. (Правда, в математическом анализе, для того чтобы "красивее получались розы", полярному радиусу r иногда разрешают принимать отрицательные значения. Мы такого допускать не будем.)

Решая прямоугольный треугольник (см. рис. 32.1), можем выразить действительную и мнимую части комплексного числа z через его модуль r и аргумент $\varphi \in \operatorname{Arg} z$:

$$a = r\cos\,\varphi;\ \ b = r\sin\,\varphi. \tag{32.6}$$

После этого можно данное комплексное число представить в виде

$$z = r(\cos\,\varphi + i\sin\,\varphi). \tag{32.7}$$

Определение 32.3. Запись (32.7) называется *тригонометрической формой* (ненулевого) комплексного числа $z \in \mathbb{C}$.

Замечание 32.6. Из сказанного выше ясно, что тригонометрическая форма для комплексного числа определена не однозначно, а лишь с точностью до добавления к аргументу угла $2\pi k$, где $k \in \mathbb{Z}$.

Выразим это обстоятельство иначе. Два комплексных числа z и w, где z задано формулой (32.7), а w — формулой

$$w = \rho(\cos\,\psi + i\sin\,\psi), \tag{32.8}$$

будут равны тогда и только тогда, когда выполнены два условия:

$$1)\ r = \rho;\quad 2)\ (\exists\ k \in \mathbb{Z})\ (\varphi = \psi + 2\pi k). \tag{32.9}$$

Обсудим задачу перехода от алгебраической формы задания комплексного числа $z = a + bi$ к тригонометрической форме (32.7). Для осуществления такого перехода необходимо

1) вычислить модуль r числа z, что совсем легко:

$$r = \sqrt{a^2 + b^2}; \tag{32.10}$$

2) найти какое-либо значение $\varphi \in \operatorname{Arg} z$; для это надо найти какое-либо решение тригонометрической системы [равносильной (32.6)]:

$$\begin{cases} \cos\varphi &= \dfrac{a}{r}\ ; \\ \sin\varphi &= \dfrac{b}{r}\ . \end{cases} \tag{32.11}$$

Обычно предпочтительным является выбор главного значения аргумента. Знаки косинуса и синуса, определяемые по правым частям уравнений (32.11) (или, что то же самое, знаки действительной и мнимой частей данного числа), позволяют выяснить, *в какой четверти* будет находиться искомый угол φ. После этого (если $a \neq 0$) систему (32.11) можно заменить одним уравнением

$$\operatorname{tg}\varphi = \frac{b}{a}\ , \tag{32.12}$$

которое, хотя и не вполне равносильно системе, но с учетом знания четверти, содержащей φ, позволяет найти этот угол однозначно:

$$\varphi = \operatorname{arctg}\frac{b}{a}\ , \tag{32.13}$$

если φ принадлежит первой или четвертой четверти; если угол находится во второй четверти, то к этому значению надо добавить π, а если — в третьей, то π надо вычесть.

Пример 32.1. Представим в тригонометрической форме комплексное число $z = -\sqrt{3} + i$. Модуль этого числа $r = 2$. Аргумент подлежит определению из уравнений

$$\begin{cases} \cos\varphi &= -\dfrac{\sqrt{3}}{2}\ ; \\ \sin\varphi &= \dfrac{1}{2}\ . \end{cases}$$

Действительная часть данного числа отрицательна, а мнимая — положительна, значит, угол $\varphi = \arg z$ находится во второй четверти, его можно вычислить с помощью арктангенса (как это излагалось выше):

$$\varphi = \pi + \operatorname{arctg}(-\frac{\sqrt{3}}{3}) = \pi - \frac{\pi}{6} = \frac{5\pi}{6}.$$

Но это не обязательно; зная четверть, можно получить тот же результат с помощью арккосинуса:

$$\varphi = \arccos(-\frac{\sqrt{3}}{2}) = \frac{5\pi}{6}.$$

Предостережем неопытных вычислителей от злоупотребления заученными в школе формулами для отыскания *всех решений* тригонометрического уравнения. Общее решение тригонометрического уравнения $\cos\varphi = -\sqrt{3}/2$ дается формулой

$$\varphi = \pm\arccos(-\frac{\sqrt{3}}{2}) + 2\pi k;\ k \in \mathbb{Z}.$$

Можно работать и по этой формуле, но тогда придется из всего бесконечного множества определяемых ею углов выбирать тот единственный, который принадлежит интервалу $(\pi/2,\, \pi)$. Нет никакой пользы от зубрежки громоздких тригонометрических формул, если вы не в состоянии, не решая уравнения, "нарисовать на глаз" искомый угол.

Осталось записать ответ:

$$z = 2(\cos\ \frac{5\pi}{6} + i\sin\ \frac{5\pi}{6}).$$

Пример 32.2. Решим аналогичную задачу для числа

$$z = 1 + \frac{\sqrt{3}}{2} + \frac{i}{2}.$$

По формулам (32.10), (32.11) с использованием техники вычислений с иррациональными выражениями, в частности избавляясь от иррациональности в знаменателе (вам следует выписать все выкладки подробно), получим:

$$r = \frac{\sqrt{2}}{2}(1+\sqrt{3});\ \cos\varphi = \frac{\sqrt{2}}{4}(1+\sqrt{3});\ \sin\varphi = \frac{\sqrt{2}}{4}(\sqrt{3}-1).$$

Как синус, так и косинус угла φ положительны, значит, этот угол находится в первой четверти. Можно вычислить синус и косинус двойного угла:

$$\sin 2\varphi = \frac{1}{2};\ \cos 2\varphi = \frac{\sqrt{3}}{2}.$$

Выясняется, во-первых, что угол 2φ также принадлежит первой четверти, а во-вторых, что он в точности равен $\pi/6$, откуда окончательно получаем: $\varphi = \pi/12$.

О т в е т:

$$z = \frac{\sqrt{2}}{2}(1+\sqrt{3})(\cos\frac{\pi}{12} + i\sin\frac{\pi}{12}).$$

Замечание 32.7. В последующих примерах (и в типовых задачах по изучаемой теме) нам понадобятся найденные выше значения синуса и косинуса (а также значение тангенса) для угла $\pi/12$. То же самое относится к значениям тригонометрических функций для угла $\pi/8$. Попробуйте проверить самостоятельно с помощью формул для синуса (косинуса, тангенса) половинного аргумента следующие формулы:

$$\cos\frac{\pi}{12} = \frac{\sqrt{2}}{4}(1+\sqrt{3});\ \sin\frac{\pi}{12} = \frac{\sqrt{2}}{4}(\sqrt{3}-1);\ \operatorname{tg}\frac{\pi}{12} = 2-\sqrt{3};\quad (32.14)$$

$$\cos\frac{\pi}{8} = \frac{\sqrt{2+\sqrt{2}}}{2};\ \sin\frac{\pi}{8} = \frac{\sqrt{2-\sqrt{2}}}{2};\ \operatorname{tg}\frac{\pi}{8} = \sqrt{2}-1. \qquad (32.15)$$

Занесите эти значения в свой "рабочий справочник" (шпаргалку), где вы храните $\sin(\pi/6)$ и т. п., или же научитесь быстро восстанавливать их вывод, когда в том возникнет необходимость.

Пример 32.3. В системе Maple предусмотрены команды abs и argument для вычисления модуля и главного значения аргумента комплексного числа, а также команда polar для перехода к "полярному представлению", которое, по сути, совпадает с тригонометрической формой, но выглядит иначе (см. образцы вычислений ниже).

```
> z := -sqrt(3) + I;
```

$$z := -\sqrt{3} + I$$

```
> zm := abs(z);  za := argument(z);  zp := polar(z);
```

$$zm := 2$$

$$za := \frac{5\pi}{6}$$

$$zp = \text{polar}\left(2, \frac{5\pi}{6}\right)$$

Обратный переход от полярного представления к алгебраической форме комплексного числа осуществляется уже известной командой evalc .

Перерешаем теперь с помощью Maple пример 32.2.

```
> w := 1 + sqrt(3) / 2 + I / 2;
```

$$w := \frac{\sqrt{3}}{2} + 1 + \frac{1}{2}I$$

```
> wp := simplify( polar(w) );
```

$$wp := \mathrm{polar}\left(\frac{\sqrt{2}\sqrt{3}}{2} + \frac{\sqrt{2}}{2}, \arctan\left(\frac{1}{2+\sqrt{3}}\right)\right)$$

Если бы мы не применили универсальную команду simplify (упростить), то в ответе получили бы "трехэтажные" формулы. Но и полученный ответ не совсем нас удовлетворяет: вручную было найдено точное значение аргумента $\arg w = \pi/12$. К сожалению, Maple знает несколько меньше тригонометрических формул, чем требуется знать выпускнику средней школы. Но зато он легко обучаем!

Мы могли бы продолжить исследование тем же приемом, что и при ручном счете, но с использованием Maple:

```
> wa := argument(w): simplify( expand ( sin( 2 * wa) ) );
```

$$\frac{1}{2}$$

(Команда expand "раскрывает" синус двойного угла по известной формуле, затем результат упрощается.)

Получается: $2\ wa = \arcsin(1/2) = \pi/6$, и следовательно, $wa = \pi/12$.

Можно пойти по другому пути — приближенно ("с плавающей точкой") вычислить (evalf) частное π/wa и узнать, во сколько раз найденный угол меньше π:

```
> evalf( Pi / wa);
```

12.00000000

Замечание 32.8. Компьютерные математические системы стараются (по возможности) "избегать" неопределенностей. Так, Maple считает, что аргумент числа нуль также является нулевым, и приписывает числу 0 полярное представление polar(0, 0).

Кроме того, имейте в виду, что лишь в России тангенс обозначается аббревиатурой tg; во всем остальном мире используют обозначение tan. (Выше вам встретилась обратная функция arctan.)

32.2. Умножение и деление комплексных чисел в тригонометрической форме. Формула Муавра. Тригонометрическая форма представления комплексных чисел очень удобна для вычисления произведений (и, следовательно, частных, а также целых степеней) комплексных чисел. Суммируем факты, относящиеся к данной теме, в следующем утверждении.

Предложение 32.2. *Пусть комплексные числа z и w представлены своими тригонометрическими формами (32.7) и (32.8). Тогда*

1) их произведение представляется в виде

$$z \cdot w = r\rho(\cos(\varphi + \psi) + i \sin(\varphi + \psi)); \tag{32.16}$$

2) число, сопряженное к z, представляется в виде

$$\overline{z} = r(\cos(-\varphi) + i \sin(-\varphi)); \tag{32.17}$$

3) число, обратное к z, представляется в виде

$$z^{-1} = \frac{1}{r}(\cos(-\varphi) + i \sin(-\varphi)); \tag{32.18}$$

4) частное w/z представляется в виде

$$\frac{w}{z} = \frac{\rho}{r}(\cos(\psi - \varphi) + i \sin(\psi - \varphi)); \tag{32.19}$$

5) для любого целого n степень z^n представляется в виде

$$z^n = r^n(\cos n\varphi + i \sin n\varphi). \tag{32.20}$$

Доказательство. 1. Первое утверждение получается непосредственным перемножением выражений (32.7) и (32.8) с последующим применением формул "косинус суммы" и "синус суммы" из школьной тригонометрии:

$$\begin{aligned} z \cdot w &= r(\cos\,\varphi + i \sin\,\varphi) \cdot \rho(\cos\,\psi + i \sin\,\psi) = \\ &= r\rho(\,(\cos\,\varphi \cos\,\psi - \sin\,\varphi \sin\,\psi) + i(\sin\,\varphi \cos\,\psi + \cos\,\varphi \sin\,\psi)) = \\ &= r\rho(\cos(\varphi + \psi) + i \sin(\varphi + \psi)). \end{aligned}$$

Доказанная выше формула может быть представлена как следующее правило: *при умножении комплексных чисел их модули перемножаются, а аргументы складываются.* Первый факт нам уже известен (см. предложение 32.1), а второй может быть выражен так:

$$[\varphi \in \operatorname{Arg} z] \wedge [\psi \in \operatorname{Arg} w] \Rightarrow [\varphi + \psi \in \operatorname{Arg}(z \cdot w)].$$

Правило умножения комплексных чисел иллюстрируется рисунком 32.3 в приложении 1.

Заметьте, что в случае, когда φ и ψ являются главными значениями аргументов, их сумма $\varphi + \psi$ не обязательно является главным значением аргумента, т. е. формула $\arg(z \cdot w) = \arg z + \arg w$ выполняться не обязана. Можно считать справедливым вариант этой формулы

$$\operatorname{Arg}(z \cdot w) = \operatorname{Arg} z + \operatorname{Arg} w,$$

использующий "аргументы с большой буквы" (надо только корректно определить сложение и равенство для *многозначных функций*).

2. Второе утверждение практически очевидно. Вектор $\overline{z} = (x, -y)$ симметричен вектору $z = (x, y)$ относительно действительной оси. Поэтому, если $\varphi \in \operatorname{Arg} z$, то $-\varphi \in \operatorname{Arg}(\overline{z})$, а модули у чисел z и $\overline{z}$ одинаковы.

3. Третье утверждение вытекает из второго с учетом выражения (31.24) обратного числа через сопряженное. В обозначениях данного параграфа указанная формула приобретает вид:

$$z^{-1} = \frac{1}{r^2}\overline{z}. \tag{32.21}$$

Из (32.21) следует, во-первых, что обратное и сопряженное число имеют одинаковые аргументы и, во-вторых, что справедлива формула (32.18). (См. рис. 32.4.)

4. Четвертое утверждение немедленно следует из первого и третьего и может быть выражено правилом: *при делении комплексных чисел их модули делятся, а аргументы вычитаются.*

5. Пятое утверждение именуется *формулой Муавра.* Для положительных n оно получается из первого утверждения простейшей индукцией. Нулевая степень любого элемента (в любом поле) по определению считается равной единице; поэтому при $n = 0$ доказываемое утверждение превращается в тождество $1 = 1$. При $n = -1$ формула (32.20) также справедлива [в силу (32.18)]. Общий случай отрицательной степени $n < 0$ сводится к первому случаю, поскольку отрицательные степени определяются формулой $z^n = (z^{-1})^{-n}$, где уже $-n > 0$. $\square$

Пример 32.4. Приведем несколько простых примеров вычислений над комплексными числами в тригонометрической форме.

Обратите внимание на то, что в простых случаях переход к тригонометрической форме осуществляется практически "устно".

$$z = -\sqrt{3} + i = 2(-\frac{\sqrt{3}}{2} + i\frac{1}{2}) = 2(\cos\ \frac{5\pi}{6} + i\sin\ \frac{5\pi}{6});$$

$$v = -1 - i = \sqrt{2}(-\frac{\sqrt{2}}{2} - i\frac{\sqrt{2}}{2}) = \sqrt{2}(\cos(-\frac{3\pi}{4}) + i\sin(-\frac{3\pi}{4}));$$

$$z \cdot v = 2\sqrt{2}(\cos(\frac{5\pi}{6} - \frac{3\pi}{4}) + i\sin(\frac{5\pi}{6} - \frac{3\pi}{4})) =$$
$$= 2\sqrt{2}(\cos\ \frac{\pi}{12} + i\sin\ \frac{\pi}{12}) = 2\sqrt{2}(\frac{\sqrt{2}}{4}(1+\sqrt{3}) + i\frac{\sqrt{2}}{4}(\sqrt{3}-1)) =$$
$$= (\sqrt{3}+1) + i(\sqrt{3}-1).$$

Разумеется, тот же ответ гораздо быстрее получился бы без тригонометрии, в алгебраической форме. Мощь тригонометрического метода начинает проявляться при возведении комплексных чисел в высокие степени. Здесь формула Муавра дает великолепные результаты.

$$w = 1 + \frac{\sqrt{3}}{2} + \frac{i}{2} = \frac{\sqrt{2}}{2}(1+\sqrt{3})(\cos\ \frac{\pi}{12} + i\sin\ \frac{\pi}{12});$$

$$w^6 = (\frac{\sqrt{2}}{2}(1+\sqrt{3}))^6(\cos(6\cdot\frac{\pi}{12}) + i\sin(6\cdot\frac{\pi}{12})) =$$
$$= (\frac{\sqrt{2}}{2}(1+\sqrt{3}))^6(\cos\ \frac{\pi}{2} + i\sin\ \frac{\pi}{2}) = \frac{1}{8}(1+\sqrt{3}))^6 i = \frac{i}{8}(4+2\sqrt{3}))^3 =$$
$$= (2+\sqrt{3})^3 i = (26+15\sqrt{3})i.$$

Замечание 32.9. (Для тех, кто помнит, что такое линейный оператор. Непомнящие могут заглянуть в § 15.)

В соответствии с векторной моделью для поля комплексных чисел (см. п. 31.3), представим $\mathbb{C}$ как действительную плоскость $\mathbb{R}^2$ и рассмотрим отображение умножения на фиксированное ненулевое комплексное число $z = a + bi = r(\cos\ \varphi + i\sin\ \varphi)$:

$$\mu_z : \mathbb{R}^2 \longrightarrow \mathbb{R}^2;\ \mu_z(w) = z \cdot w;\ w \in \mathbb{R}^2. \tag{32.22}$$

Из дистрибутивного закона $\boxed{5}$ и из замечания об умножении комплексного числа на действительное [см. (31.30)] следует, что отображение (32.22) является линейным (над полем $\mathbb{R}$).

Если переменный вектор (комплексное число w) также представлять в тригонометрической форме $w = c + di = \rho(\cos\psi + i\sin\psi)$, то (в соответствии с правилом из первого пункта предложения 32.2) действие линейного оператора можно будет описать как поворот произвольного вектора w на фиксированный угол φ с последующим растяжением (сжатием) в r раз.

Естественный базис в $\mathbb{C} = \mathbb{R}^2$ составляют векторы $1 = (1, 0)$ и $i = (0, 1)$. Под действием μ_z они перейдут соответственно в векторы

$$z \cdot 1 = z = (a, b) = (r\cos\varphi, r\sin\varphi)$$

и

$$z \cdot i = (-b, a) = (-r\sin\varphi, r\cos\varphi).$$

Стало быть, оператору (31.22) будет соответствовать (см. п. 15.2) матрица

$$M = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} r\cos\varphi & -r\sin\varphi \\ r\sin\varphi & r\cos\varphi \end{pmatrix}. \tag{32.23}$$

Эту матрицу можно представить в виде произведения двух (перестановочных между собой) матриц:

$$M = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix} \cdot \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} = B \cdot A,$$

что отражает отмеченный выше факт представимости оператора μ_z в виде композиции оператора поворота (см. пример 15.3; этому оператору соответствует матрица A) и оператора умножения на положительное действительное число (т. е. оператора растяжения или сжатия; ему отвечает матрица B).

Здесь проявляется некоторая особенность (по сравнению с п. 15.2): запись векторов *в строчку* приводит к тому, что действие линейного оператора выражается не левым умножением вектора-столбца на матрицу M, а *правым умножением вектора-строки на транспонированную матрицу* M^t:

$$w = (c,\ d) \overset{\mu_z}{\mapsto} w \cdot z = (c,\ d) \cdot M^t.$$

Кстати, транспонированная матрица

$$M^t = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \tag{32.24}$$

должна напомнить вам о *матричной модели* для поля $\mathbb{C}$ [см. п. 31.4; матрица (32.24) имеет вид (31.38); в частности, если фиксированным множителем z является мнимая единица i, то матрица M^t совпадает с матрицей J].

Замечание 32.10. В любом поле P множество $P^* = P \setminus \{0\}$ ненулевых элементов является группой по умножению (понятие *группы* вводилось в замечании 14.1; мы многократно к нему возвращались; вспомните, например, о группе перестановок, группе обратимых матриц и др.).

В случае поля $\mathbb{R}$ множество $\mathbb{Z}^*$, состоящее из всех действительных чисел, по модулю равных единице (таких чисел всего два: 1 и -1; см. п. 20.2), является подгруппой в $\mathbb{R}^*$.

В случае поля $\mathbb{C}$ множество

$$U = \{z \in \mathbb{C} : |z| = 1\},$$

состоящее из всех комплексных чисел, по модулю равных единице (это множество представляет собой единичную окружность на комплексной плоскости; см. замечание 32.1), является подгруппой в $\mathbb{C}^*$. (Почему?)

32.3. Извлечение корней из комплексных чисел в тригонометрической форме. Рассмотрим комплексное число w и натуральное число n. *Корнем степени* n из числа w называется комплексное число z, такое, что

$$z^n = w. \tag{32.25}$$

Извлечь корень степени n из числа w — это значит *найти все решения* уравнения (32.25). Для $n = 1$ эта задача тривиальна, для $n = 2$ она разобрана в п. 31.5 (см. задачу 3). Если $w = 0$, то уравнение (32.25) имеет единственное решение $z = 0$, поэтому далее мы будем рассматривать случай $w \neq 0$ и представлять данное число w и искомое число z в тригонометрической форме [см. формулы (32.8) и (32.7) соответственно]. Показатель n будем считать бóльшим единицы (для $n = 2$ будет получен второй способ решения задачи, как правило, менее удобный, чем разобранный в п. 31.5, но также приемлемый).

Предложение 32.3. *Пусть* $w = c + di = \rho(\cos\psi + i\sin\psi)$ — *ненулевое комплексное число;* $n \in \mathbb{N}$; $n \geqslant 2$.

Существует в точности n *различных значений корня степени* n *из комплексного числа* w, *т. е. уравнение (32.25) имеет* n *различных решений* z_k $(k = 0, ..., n-1)$. *Все они располагаются на окружности радиуса*

$$r = \sqrt[n]{\rho} \tag{32.26}$$

в вершинах правильного n*-угольника, положение которых задается углами*

$$\varphi_k = \frac{\psi + 2\pi k}{n} = \frac{\psi}{n} + k \cdot \frac{2\pi}{n};\ k = 0, ..., n-1. \tag{32.27}$$

Все решения (32.25) могут быть представлены формулой

$$\begin{aligned} z_k &= r(\cos\varphi_k + i\sin\varphi_k) = \\ &= \sqrt[n]{\rho}\left(\cos\frac{\psi + 2\pi k}{n} + i\sin\frac{\psi + 2\pi k}{n}\right);\ k = 0, ..., n-1. \end{aligned} \tag{32.28}$$

Доказательство. Будем искать решение уравнения (32.25) в виде $z = a + bi = r(\cos\varphi + i\sin\varphi)$. Подстановка этого выражения в (32.25) и применение формулы Муавра (32.20) дает:

$$r^n(\cos n\varphi + i\sin n\varphi) = \rho(\cos\psi + i\sin\psi). \tag{32.29}$$

В силу замечания 32.6, равенство (32.29) равносильно системе уравнений

$$\begin{cases} r^n = \rho\,; \\ n\varphi = \psi + 2\pi k \end{cases} \tag{32.30}$$

с неизвестными r и φ и с произвольным целым k.

Первое уравнение этой системы однозначно определяет значение (32.26) неизвестной r (корень берется в поле $\mathbb{R}$).

Из второго уравнения получаем выражение (32.27), где k — произвольное целое число. Однако легко заметить, что при изменении значения k на число, кратное n, к углу φ добавляется угол, кратный 2π, и следовательно, луч, определяемый этим углом, не изменяется. Значит, комплексные числа, определяемые формулой (32.28), повторяются с периодом n (т. е. $z_{k+n} = z_k$ для любого $k \in \mathbb{Z}$).

При этом углы φ_k при $k = 0, ..., n-1$, очевидно, определяют попарно различные лучи, разбивающие полный угол (равный 2π) на n равных частей. Следовательно, точки z_k при $k = 0, ..., n-1$ все располагаются на окружности радиуса (32.26), разбивают эту окружность на n равных частей и исчерпывают множество решений уравнения (32.25). □

Замечание 32.11. При $n = 2$ следует смириться с термином "двуугольник", понимая под этим фигуру, состоящую из двух диаметрально противоположных точек на окружности и соединяющего их диаметра.

При $n \geqslant 3$ возникает картинка, изображенная на рис. 32.5 в прил. 1. Разберитесь в ней внимательно.

Осознайте, что для отыскания всех значений корня n-й степени достаточно найти одно значение (например, z_0, аргумент которого находится по формуле $\varphi_0 = \psi/n$).

Если для w в качестве ψ выбрано *главное значение аргумента* ($\psi = \arg w \in (-\pi, \pi]$), то $\varphi_0 = \arg z_0$ и z_0 называется *главным значением корня степени* n из комплексного числа w.

Остальные значения корня определяются тригонометрическими средствами (могут применяться формулы для синуса и косинуса суммы и, в частности, *формулы приведения*) или же с использованием *соображений симметрии* (см. примеры ниже).

Пример 32.5. Извлечем кубический корень из числа

$$w = \sqrt{2}(1 + i),$$

т. е. решим в поле $\mathbb{C}$ уравнение $z^3 = w$.

Приведем данное число w к тригонометрической форме:

$$\rho = |z| = 2;\ \psi = \arg z = \operatorname{arctg} 1 = \frac{\pi}{4};\ w = 2\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right).$$

Уравнение имеет три решения z_k ($k = 0, 1, 2$). Все они лежат на окружности радиуса

$$r = \sqrt[n]{\rho} = \sqrt[3]{2}.$$

Главное значение (см. замечание выше) z_0 имеет аргумент

$$\varphi_0 = \frac{\psi}{n} = \frac{\pi}{12}$$

и, следовательно, может быть вычислено по формуле

$$z_0 = \sqrt[3]{2}\left(\cos\frac{\pi}{12} + i\sin\frac{\pi}{12}\right) \overset{(32.14)}{=\!=\!=} \\ = \frac{\sqrt[3]{2}\sqrt{2}}{4}((\sqrt{3}+1)+i(\sqrt{3}-1)) = \frac{1}{2\sqrt[6]{2}}((\sqrt{3}+1)+i(\sqrt{3}-1)).$$

Вся окружность разбивается точками z_k на три равные дуги величины $\Delta\varphi = 2\pi/3$. Аргумент для следующего значения корня находится по формуле

$$\varphi_1 = \varphi_0 + \Delta\varphi = \frac{\pi}{12} + \frac{2\pi}{3} = \frac{3\pi}{4},$$

и следовательно,

$$z_1 = \sqrt[3]{2}\left(\cos\frac{3\pi}{4} + i\sin\frac{3\pi}{4}\right) = \sqrt[3]{2}\left(-\frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}\right) = \frac{1}{\sqrt[6]{2}}(-1+i).$$

Последнее значение корня вычисляется так:

$$\varphi_2 = \varphi_1 + \Delta\varphi = \frac{17\pi}{12} = \frac{3\pi}{2} - \frac{\pi}{12};$$

$$z_2 = \sqrt[3]{2}\left(\cos\frac{17\pi}{12} + i\sin\frac{17\pi}{12}\right) = \sqrt[3]{2}\left(-\sin\frac{\pi}{12} - i\cos\frac{\pi}{12}\right) = \\ = -\frac{1}{2\sqrt[6]{2}}((\sqrt{3}-1)+i(\sqrt{3}+1)).$$

О т в е т:

$$\frac{1}{2\sqrt[6]{2}}((\sqrt{3}+1)+i(\sqrt{3}-1));\ \frac{1}{\sqrt[6]{2}}(-1+i);\ -\frac{1}{2\sqrt[6]{2}}((\sqrt{3}-1)+i(\sqrt{3}+1)).$$

Данное число w и все значения корня кубического из него изображены на рис. 32.6 в прил. 1.

Пример 32.6. Извлечем корень степени $n = 8$ из числа $w = -2$, т. е. решим уравнение $z^8 = -2$.

$$w = -2 = 2(\cos\pi + i\sin\pi);\ \rho = 2;\ \psi = \pi;$$

$$r = \sqrt[8]{2};\ \varphi_0 = \frac{\pi}{8};\ \Delta\varphi = \frac{\pi}{4};\ \varphi_1 = \frac{3\pi}{8};\ \ldots;\ \varphi_7 = \frac{15\pi}{8};$$

$$z_0 = \sqrt[8]{2}\left(\cos\frac{\pi}{8} + i\sin\frac{\pi}{8}\right) = \frac{\sqrt[8]{2}}{2}\left(\sqrt{2+\sqrt{2}} + i\sqrt{2-\sqrt{2}}\right).$$

Комплексное число z_1 симметрично числу z_0 относительно биссектрисы первого и третьего координатных углов. При такой симметрии происходит "транспозиция координат" (абсцисса симметричной точки равна ординате данной точки и наоборот). Следовательно,

$$z_1 = \frac{\sqrt[8]{2}}{2}\left(\sqrt{2-\sqrt{2}} + i\sqrt{2+\sqrt{2}}\right).$$

Точки z_2 и z_3 симметричны точкам z_1 и z_0 соответственно относительно мнимой оси; следовательно,

$$z_2 = \frac{\sqrt[8]{2}}{2}\left(-\sqrt{2-\sqrt{2}} + i\sqrt{2+\sqrt{2}}\right);$$
$$z_3 = \frac{\sqrt[8]{2}}{2}\left(-\sqrt{2+\sqrt{2}} + i\sqrt{2-\sqrt{2}}\right).$$

Точки z_4, z_5, z_6, z_7 симметричны точкам z_0, z_1, z_2, z_3 соответственно относительно начала координат; следовательно,

$$z_4 = -z_0;\ z_5 = -z_1;\ z_6 = -z_2;\ z_7 = -z_3.$$

О т в е т (в графическом виде) представлен на рис. 32.7.

Пример 32.7. Любопытно будет попытаться извлечь "тригонометрическим способом" корень квадратный из комплексного числа $w = -32 - 126i$ (см. "алгебраический способ" в примере 31.5; см. также замечание 32.11, в котором упоминались "двуугольники").

Находим модуль и аргумент данного числа:

$$\rho = |w| = 130;\ \psi = \arg w = \operatorname{arctg}\left(\frac{-126}{-32}\right) - \pi = \operatorname{arctg}\frac{63}{16} - \pi,$$

а также модуль и аргумент для главного значения z_0 квадратного корня из w:

$$r = \sqrt{\rho} = \sqrt{130};\ \varphi_0 = \frac{\psi}{2} = \frac{1}{2}\operatorname{arctg}\frac{63}{16} - \frac{\pi}{2} = \frac{u}{2} - \frac{\pi}{2},$$

где введен вспомогательный угол $u = \operatorname{arctg}(63/16) \in (0, \pi/2)$.

Вычислим далее косинус угла u, а также косинус и синус для угла φ_0:

$$\cos u = \frac{16}{\sqrt{16^2 + 63^2}} = \frac{16}{65};$$

$$\cos\varphi_0 = \sin\frac{u}{2} = \sqrt{\frac{1 - \cos u}{2}} = \sqrt{\frac{1 - \frac{16}{65}}{2}} = \frac{7}{\sqrt{130}};$$

$$\sin\varphi_0 = -\cos\frac{u}{2} = -\sqrt{\frac{1 + \cos u}{2}} = -\sqrt{\frac{1 + \frac{16}{65}}{2}} = -\frac{9}{\sqrt{130}}.$$

И наконец, найдем само число z_0:

$$z_0 = \sqrt{\rho}(\cos\varphi_0 + i\sin\varphi_0) = \sqrt{130}\left(\frac{7}{\sqrt{130}} - i\frac{9}{\sqrt{130}}\right) = 7 - 9i.$$

Результат сошелся с полученным в примере 31.5.

Замечание 32.12. Вспомним, что в п. 31.1 мы ввели мнимую единицу и комплексные числа с тем, чтобы "исправить недостатки" поля $\mathbb{R}$ (не извлекались квадратные корни из отрицательных действительных чисел). Однако результат расширения поля $\mathbb{R}$ намного превзошел наши ожидания: в поле $\mathbb{C}$ всегда извлекаются не только квадратные корни, но и корни произвольной натуральной степени n (причем для ненулевого числа $w \in \mathbb{C}$ оказывается ровно n различных значений корня n-й степени из w).

В § 35 мы убедимся в том, что результат введения поля комплексных чисел является еще более значимым: будет доказана так называемая *основная теорема алгебры*, утверждающая, что любое уравнение степени $n \geqslant 1$ с комплексными коэффициентами имеет хотя бы один комплексный корень.

Замечание 32.13. В системе Maple предусмотрена команда

```
> root [ n ] ( w );
```

вычисляющая главное значение корня заданной степени n из заданного комплексного числа w:

```
> root [3] ( sqrt( 2 ) * (1 + I) ); z[0] := evalc( % );
```

$$((1+I)\sqrt{2})^{\left(\frac{1}{3}\right)}$$

$$z_0 := 2^{\left(\frac{1}{3}\right)} \cos\frac{\pi}{12} + 2^{\left(\frac{1}{3}\right)} \sin\frac{\pi}{12}\, I$$

Можно ли заставить Maple точно вычислить косинус (синус, тангенс) половинного угла? Можно. Например так:

1) определим для $t \in (0, \pi/2)$ *функции пользователя*, вычисляющие тригонометрические функции угла t, представляя этот угол как половинный (по отношению к углу $2t$, значения тригонометрических функций для которого Maple знает):

```
> coshalf := t -> sqrt( (1 + cos( 2 * t) / 2);
> sinhalf := t -> sqrt( (1 - cos( 2 * t) / 2);
```

$$coshalf := t \longrightarrow \sqrt{\frac{1+\cos 2t}{2}}$$

$$sinhalf := t \longrightarrow \sqrt{\frac{1-\cos 2t}{2}}$$

2) теперь можно вычислить, скажем, косинус угла $\pi/12$:

```
> coshalf( Pi / 12);
```

$$\frac{\sqrt{6}}{4} + \frac{\sqrt{2}}{4},$$

что совпадает со значением, приведенным в (32.14);

3) но можно сразу применить "подстановку" subs, заменяя в выражении для z_0 функции cos и sin на coshalf и sinhalf соответственно:

```
> eval( subs({ cos=coshalf, sin=sinhalf }, z[0]) );
```

$$2^{\left(\frac{1}{3}\right)} \left(\frac{\sqrt{6}}{4} + \frac{\sqrt{2}}{4}\right) + 2^{\left(\frac{1}{3}\right)} \left(\frac{\sqrt{6}}{4} - \frac{\sqrt{2}}{4}\right) I\,,$$

что совпадает с выражением для z, полученным выше вручную.

Есть возможность найти сразу все значения корня, применив команду solve. Общий ответ (в данном примере) будет опять же содержать синусы и косинусы угла $\pi/12$. Можно попытаться собрать все решения в список, применить к списку команду evalc, а затем описанную выше замену, после чего упростить ответ. (Попробуйте сделать это сами; введите приводимые ниже команды и проследите за выводимыми результатами.)

```
> sol := solve(z ^ 3 = sqrt(2) * (1+I));
> map( evalc, [ sol ] );
> eval( subs({ cos=coshalf, sin=sinhalf }, %) );
> map( simplify, % );
```

Поэкспериментируйте аналогичным образом с уравнением из следующего примера ($z^8 = -2$).

Замечание 32.14. Все рисунки к настоящему параграфу построены с помощью графических средств системы Maple. В данном пособии вряд ли уместно рассказывать о Maple-графике. (Хотя для любознательного студента-компьютерщика самостоятельное освоение этого материала, несомненно, будет и интересным, и очень полезным.)

§ 33. Корни из единицы. Первообразные корни

33.1. Комплексные корни из единицы и их свойства. Применяя предложение 32.3 к комплексному числу $w = 1$, мы сразу получаем, что существует ровно n различных решений уравнения

$$z^n = 1 \tag{33.1}$$

(n значений корня n-й степени из 1).

Все z_k располагаются на единичной окружности, в вершинах правильного n-угольника.

Из общей формулы (32.27) получается ($\rho = 1$; $r = 1$; $\psi = 0$; $\varphi_0 = 0$) представление

$$z_k = \cos\,\frac{2\pi k}{n} + i\sin\,\frac{2\pi k}{n};\ \ k = 0, ..., n-1. \tag{33.2}$$

Главным значением корня n-й степени из 1 является $z_0 = 1$.

Рассмотрим корень из единицы с номером $k = 1$, т. е. z_1, и введем для него специальное обозначение:

$$\zeta_n = \cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n}. \tag{33.3}$$

Из формулы Муавра (32.20) следует, что для любого $k = 0, ..., n-1$ справедливо равенство

$$z_k = \zeta_n^k. \tag{33.4}$$

(Геометрическое представление чисел ζ_n^k для случая $n = 12$ см. на рис. 33.1 в прил. 1.)

Замечание 33.1. Можно рассматривать произвольные целые степени ζ_n^s ($s \in \mathbb{Z}$); они будут повторяться с периодом n:

$$\zeta_n^s = \zeta_n^r, \tag{33.5}$$

где r — остаток от деления целого числа s на натуральное число n.

В частности, $\zeta_n^s = 1$ тогда и только тогда, когда s делится на n.

Здесь мы вновь встречаемся с ситуацией, подобной той, что была разобрана в п. 18.2: там имела место периодическая повторяемость степеней перестановки $\varphi \in S_n$ и вводилось понятие порядка $o(\varphi)$ как наименьшего из натуральных показателей k, таких, что φ^k совпадает с единичной перестановкой.

По тому же принципу может быть определен *порядок* $o(g)$ элемента $g \in G$ в произвольной *группе* $(G, \cdot)$:

$$o(g) = \min\{k \in \mathbb{N} \,:\, g^k = 1\},$$

если такие показатели степени k существуют (в противном случае порядок считается бесконечным).

В настоящее время мы работаем в группе $G = \mathbb{C}^*$ (см. замечание 32.10). Для $z \in \mathbb{C}^*$, таких, что $|z| > 1$ (соответственно $|z| < 1$) формула Муавра влечет строгое возрастание (соответственно строгое убывание) модулей степеней z^k (при $k \to \infty$). Следовательно, ни одна из степеней такого числа z не может стать равной единице, и поэтому $|z| \neq 1$ влечет $o(z) = \infty$.

В то же время среди чисел, принадлежащих единичной окружности (подгруппе U группы $\mathbb{C}^*$), существуют элементы конечного порядка. Они (и только они) являются корнями натуральных степеней из единицы.

В частности, все степени ζ_n^k числа (33.3) при $k = 0, ..., n-1$ являются попарно различными, и следовательно,

$$o(\zeta_n) = n.$$

Можно доказать (это — интересное упражнение, см. указания в [2, гл. 5, § 1]), что число $z = (2+i)/(2-i)$ имеет единичный модуль и бесконечный порядок.

Замечание 33.2. По-видимому, необходимо предостеречь читателей: обозначение z_k не несет полной информации о числе; предварительно должна быть указана степень n. Так, z_1 при $n = 3$ и z_1 при $n = 7$ — это совершенно разные числа.

В то же время в обозначении ζ_n^k явно указывается как степень n, так и номер корня в списке всех корней из единицы степени n.

Переходим к формулировке основных свойств корней из единицы.

Предложение 33.1. *Пусть z_k $(k = 0, ..., n-1)$ — все корни степени n из единицы [определенные формулой (33.2)]. Тогда*

1) для любых $k, l \in \{0, ..., n-1\}$ справедливо равенство

$$z_k \cdot z_l = z_r, \tag{33.6}$$

где r — остаток от деления числа $k + l$ на n;

2) для любого $k \in \{0, ..., n-1\}$ справедливо

$$z_k^{-1} = z_{n-k}. \tag{33.7}$$

Доказательство. Согласно формулам (33.4), $z_k = \zeta_n^k$, $z_l = \zeta_n^l$. Следовательно, с учетом (33.5) получаем:

$$z_k \cdot z_l = \zeta_n^k \cdot \zeta_n^1 = \zeta_n^{k+l} = \zeta_n^r,$$

где r — остаток от деления $k + l$ на n.

Первое утверждение предложения доказано. Второе немедленно из него вытекает. □

Замечание 33.3. Из формул (33.4) и (33.5) следует также формула

$$z_k^s = z_r, \tag{33.8}$$

где $k \in \{0, 1, ...n-1\}$; $s \in \mathbb{Z}$; r — остаток от деления ks на n.

Замечание 33.4. Утверждения предложения 33.1 можно трактовать следующим образом: комплексные корни заданной степени n из единицы образуют *группу по умножению*. Выражаясь точнее, множество

$$C_n = \{\, 1,\, z_1,\, z_2,\, ...,\, z_{n-1} \,\}, \tag{33.9}$$

состоящее из всех корней n-й степени из 1, является подгруппой в группе U комплексных чисел, равных по модулю единице. (В самом деле, произведение двух элементов из C_n снова принадлежит C_n; число 1 принадлежит C_n; вместе с каждым своим элементом множество C_n содержит обратный элемент.)

У группы C_n имеется собственное имя — *циклическая группа порядка* n.

Пример 33.1. 1. Квадратные корни из 1 суть $1, -1$; циклическая группа порядка 2, составленная из этих чисел, есть не что иное, как неоднократно уже встречавшаяся нам группа $\mathbb{Z}^*$.

2. Кубические корни из 1 образуют группу

$$C_3 = \{\, 1,\, \zeta_3,\, \zeta_3^2 \,\}; \tag{33.10}$$

где

$$\zeta_3 = \cos\,\frac{2\pi}{3} + i\sin\,\frac{2\pi}{3} = \frac{-\sqrt{3}+i}{2};$$
$$\zeta_3^2 = \cos\,\frac{4\pi}{3} + i\sin\,\frac{4\pi}{3} = -\frac{\sqrt{3}+i}{2} = \overline{\zeta_3}. \tag{33.11}$$

3. Корни четвертой степени из 1 образуют группу

$$C_4 = \{\, 1,\, i,\, -1,\, -i \,\} \tag{33.12}$$

(здесь $z_1 = \zeta_4 = i$).

4. Столь же легко найти корни шестой, восьмой и двенадцатой степени из 1; проделайте это самостоятельно. Приведем ответ для C_{12}, полученный с помощью Maple:

```
> map( simplify, map( evalc, [ solve( z ^ 12 = 1 ) ] ) );
```

$$[\, -I,\ I,\ -1,\ 1,\ -\frac{1}{2}+\frac{1}{2}I\sqrt{3},\ \frac{1}{2}-\frac{1}{2}I\sqrt{3},\ -\frac{\sqrt{3}}{2}-\frac{1}{2}I,$$
$$\frac{\sqrt{3}}{2}+\frac{1}{2}I,\ -\frac{1}{2}-+\frac{1}{2}I\sqrt{3},\ \frac{1}{2}+\frac{1}{2}I\sqrt{3},\ -\frac{\sqrt{3}}{2}+\frac{1}{2}I,\ \frac{\sqrt{3}}{2}-\frac{1}{2}I\,]$$

На рис. 33.1 вы уже видели эти корни как вершины правильного 12-угольника. Обратите внимание на то, что список корней Maple упорядочил "по-своему", отнюдь не в естественном порядке z_k $(k = 0, ..., 11)$. (Разумеется, его можно "попросить" переупорядочить ответы так, как вам захочется.)

5. Интересным (с точки зрения геометрии) является случай $n = 5$.

Вычисления в тригонометрической форме немедленно дают пять значений корня пятой степени из единицы:

$$z_k = \cos\ \frac{2\pi k}{5} + i \sin\ \frac{2\pi k}{5};\ k = 0, 1, 2, 3, 4.$$

Замечательно то, что участвующие в этих формулах синусы и косинусы можно точно *вычислить в радикалах*. Для этого придется решить уравнение $z^5 - 1 = 0$ другим способом. Приведем лишь план решения, оставляя подробности в качестве упражнения для любознательных читателей. (Если вы — любознательный читатель, но решение не получается, подсмотрите в [9].)

Очевидный корень $z = 1$ дает разложение на множители

$$z^5 - 1 = (z - 1)(z^4 + z^3 + z^2 + z + 1).$$

Необходимо решить уравнение $z^4 + z^3 + z^2 + z + 1 = 0$. Нуль не является решением этого уравнения, поэтому можно (почленно) поделить уравнение на z^2. Получим (после перегруппировки):

$$(z^2 + \frac{1}{z^2}) + (z + \frac{1}{z}) + 1 = 0.$$

Произведем замену переменной:

$$z + \frac{1}{z} = t\ \ (\Rightarrow\ \ z^2 + \frac{1}{z^2} = t^2 - 2\,).$$

Для t получается квадратное уравнение $t^2 + t - 1 = 0$, решая которое находим два (действительных) корня: $t_{1,2} = (-1 \pm \sqrt{5})/2$.

Для отыскания z получится совокупность двух квадратных уравнений:

$$\left[\begin{array}{lllll} z^2 & - & t_1 z & + & 1 = 0; \\ z^2 & - & t_2 z & + & 1 = 0. \end{array}\right.$$

Решая эти уравнения в поле $\mathbb{C}$, получим совокупность ответов:

$$z_{1,2} = \frac{\sqrt{5} - 1}{4} \pm i\frac{\sqrt{10 + 2\sqrt{5}}}{4};\ z_{3,4} = -\frac{\sqrt{5} - 1}{4} \pm i\frac{\sqrt{10 - 2\sqrt{5}}}{4}.$$

Найдены все пять значений корня пятой степени из 1; в частности,

$$z_1 = \zeta_5 = \cos\frac{2\pi}{5} + i\sin\frac{2\pi}{5} = \frac{\sqrt{5}-1}{4} + i\frac{\sqrt{10+2\sqrt{5}}}{4},$$

откуда следуют выражения в радикалах:

$$\cos\frac{2\pi}{5} = \frac{\sqrt{5}-1}{4};\ \sin\frac{2\pi}{5} = \frac{\sqrt{10+2\sqrt{5}}}{4}\,. \tag{33.13}$$

[Кстати, выражения (33.13) обеспечивают возможность построения (циркулем и линейкой) правильного пятиугольника; см. рис. 33.2 в прил. 1.]

Вы можете сами убедиться в том, что Maple также знает описанное выше "точное" вычисление корней пятой степени из единицы.

33.2. Первообразные (примитивные) корни из единицы. Рассмотрим еще раз список корней степени $n = 12$ из единицы и рис. 33.1, их представляющий. Среди корней $z_0 = 1$, z_1, ... , z_{11} есть

— один корень, $z_0 = 1$, который является корнем из 1 *первой* (и вообще любой натуральной) степени;

— один корень, $z_6 = \zeta_{12}^6 = -1$, который является корнем из 1 *второй* (и вообще любой четной) степени; можно записать выражение: $\zeta_{12}^6 = \zeta_2$;

— два корня, z_4 и z_8, которые являются корнями из 1 *третьей* (и вообще любой делящейся на 3) степени; можно записать выражение: $z_4 = \zeta_{12}^4 = \zeta_3$;

— три корня, z_3, z_6 и z_9, которые являются корнями из 1 *четвертой* (и любой кратной четырем) степени; $z_3 = \zeta_{12}^3 = \zeta_4 = i$;

— четыре корня, z_2, z_4, z_8 и z_{10}, которые являются корнями из 1 *шестой* (и любой кратной шести) степени; $z_2 = \zeta_{12}^2 = \zeta_6$;

— четыре корня, $z_1 = \zeta_{12}$, $z_5 = \zeta_{12}^5$, $z_7 = \zeta_{12}^7$ и $z_{11} = \zeta_{12}^{11}$, которые *не* являются корнями из 1 никакой степени, меньшей 12 (на рисунке эти корни выделены).

[Последнее утверждение требует обоснования, которое можно было бы легко получить с помощью формулы Муавра и формулы (33.8), но мы не будем этого делать, поскольку указанный факт будет следовать из более общего утверждения (критерия), которое мы докажем ниже, в предложении 33.2.]

Определение 33.1. Корень степени n из единицы называется *первообразным* (или *примитивным*), если он не является корнем из единицы никакой степени, меньшей n.

Докажем *критерий первообразности* корня из единицы.

Предложение 33.2. *Пусть z_k — один из корней степени n из единицы, заданных формулой (33.2). Этот корень будет первообразным тогда и только тогда, когда его номер k взаимно прост со степенью n.*

Доказательство. 1. Пусть наибольший общий делитель

$$d = (k, n) > 1.$$

Тогда $k = k'd$ и $n = n'd$, причем $n' < n$. Кроме того, очевидно, $kn' = k'n$. Следовательно,

$$z_k^{n'} \overset{(33.4)}{=\!=} (z_1^k)^{n'} = z_1^{kn'} = z_1^{k'n} = (z_1^n)^{k'} = 1,$$

а это означает, что z_k является корнем степени $n' < n$ из 1, т. е. z_k не является первообразным корнем степени n из 1.

2. Пусть теперь $d = 1$. Докажем первообразность корня z_k. Предположим, что для некоторого натурального n' справедливо $z_k^{n'} = 1$. Вследствие (33.4), это равносильно равенству $z_1^{kn'} = 1$, которое, в свою очередь (см. замечание 33.1), равносильно факту делимости $n \,|\, kn'$. Но числа n и k взаимно просты, поэтому указанная делимость равносильна делимости $n \,|\, n'$, что уже влечет неравенство $n \leqslant n'$. Значит, z_k не может являться корнем из 1 степени, меньшей n. □

Замечание 33.5 (упражнение). Условие первообразности корня $z_k \in C_n$ равносильно любому из следующих условий: 1) степени z_k^s $(s = 0, ..., n-1)$ исчерпывают все множество (группу) C_n; 2) порядок $o(z_k)$ корня z_k (в группе C_n) равен n.

Пример 33.2. При любом n корень $z_1 = \zeta_n$ является первообразным. В случае *простой* степени n все корни $z_k = \zeta_n^k$, кроме $z_0 = 1$, являются первообразными.

Среди корней четвертой степени из 1 к первообразным относятся два: i и $-i$; среди корней шестой степени — тоже два (какие?); в группе C_{12} — четыре (вы их уже видели отмеченными на рис. 33.1).

Попробуйте найти все корни из единицы степени 24 и выделить среди них первообразные. Затем для какого-либо из них (не первого) найдите все степени (с нулевой по 23-ю).

33.3. Использование корней из единицы при вычислении корней из других комплексных чисел. Рассмотрим снова задачу извлечения корня степени n из ненулевого комплексного числа w, т. е. задачу отыскания всех решений уравнений (32.25).

Допустим, найдено какое-то одно из значений корня z^*. Это значение может (но не обязано) быть *главным*.

Чтобы найти все значения корня n-й степени из w, можно воспользоваться [см. формулы (33.3), (33.4) и (33.9)] набором корней из единицы:

$$C_n = \{1,\, \zeta_n,\, \zeta_n^2,\, \ldots,\, \zeta_n^{n-1}\}. \tag{33.14}$$

В самом деле, умножение комплексного числа z^* на ζ_n^k равносильно повороту вектора z^* в комплексной плоскости на угол $2\pi k/n$.

Таким образом, все множество корней степени n из w может быть описано следующим образом:

$$z^*\, C_n = \{z^*,\, z^*\,\zeta_n,\, z^*\,\zeta_n^2,\, \ldots,\, z^*\,\zeta_n^{n-1}\}. \tag{33.15}$$

§ 34. Показательная функция комплексного аргумента. Показательная форма комплексного числа

34.1. Основные понятия математического анализа функций комплексного переменного. Введение комплексных чисел приводит к значительным продвижениям и замечательным результатам не только в алгебре, но и в математическом анализе.

В элементарном математическом анализе *функций действительного переменного* простейшим объектом изучения служат функции $f : D \to \mathbb{R}$, определенные на подмножествах $D \subseteq \mathbb{R}$. В настоящем небольшом параграфе мы познакомимся с *функциями комплексного переменного*, т. е. с отображениями

$$f \,:\, D \longrightarrow \mathbb{C};\;\; z = x + yi \mapsto w = u + vi;\;\; z \in D;\;\; D \subseteq \mathbb{C}. \tag{34.1}$$

Функция (34.1) определяется двумя действительнозначными функциями от двух действительных переменных:

$$\begin{cases} u = g(x, y); \\ v = h(x, y). \end{cases} \tag{34.1a}$$

(Вам предстоит на втором курсе значительно подробнее познакомиться с удивительной наукой ТФКП — теорией функций комплексного переменного.)

Замечание 34.1. Несколько слов о графическом представлении функций комплексного переменного.

График функции действительного переменного $y = f(x)$ $(x \in D)$ представляет собой множество (кривую) $G_f = \{\, (x, f(x)) \,:\, x \in D \,\}$ на плоскости $\mathbb{R}^2$ с координатами (x, y).

Аналогично графиком функции (34.1) будет подмножество

$$G_f = \{\, (z, f(z)) \,:\, z \in D \,\}$$

в двумерном комплексном пространстве $\mathbb{C}^2$ с координатами (z, w).

Если это пространство "овеществить", т. е. превратить в действительное четырехмерное пространство $\mathbb{R}^4$ с координатами (x, y, u, v), то график G_f будет задаваться двумя уравнениями (34.1a). Можно сказать, что G_f является *комплексной кривой* в $\mathbb{C}^2$, или же *двумерной поверхностью* в $\mathbb{R}^4$.

Возможности "человеческой" графики в изображении четырехмерных объектов весьма ограничены. Функции (34.1a) можно представить графиками в двух различных трехмерных действительных пространствах, изоморфных $\mathbb{R}^3$: график G_g будет состоять из точек с координатами (x, y, u), где $u = g(x, y)$, а график G_h — из точек (x, y, v), где $v = h(x, y)$.

По другому пути пошли разработчики системы Maple: четвертое измерение v представляется *цветом*: поверхность G_f раскрашивается, подобно тому как земной рельеф показывается на физических картах. (Рекомендуем вам поэкспериментировать с командой complexplot из пакета plots.)

Частичную графическую информацию о функции (34.1) доставляет график $G_{|f|}$ модуля этой функции, представляющий собой двумерную поверхность в трехмерном пространстве, составленную из точек с координатами (x, y, ρ), где $\rho = |w| = \sqrt{g^2(x, y) + h^2(x, y)}$. Эта поверхность целиком располагается *над* горизонтальной плоскостью переменных (x, y); некоторые точки поверхности могут принадлежать этой плоскости; они будут соответствовать *корням* (*нулям*) данной функции.

Первым понятием, специфическим для математического анализа (не относящимся к сфере чистой алгебры), является понятие *предельного перехода* в поле $\mathbb{R}$.

Предельный переход (*покомпонентно*) переносится на действительные арифметические векторы. В частности, поскольку комплексные числа моделируются как двумерные действительные векторы, определена операция предельного перехода в поле $\mathbb{C}$.

Считается, что последовательность $z_k = x_k + y_k i$ *сходится* в поле $\mathbb{C}$ к числу $z_0 = x_0 + y_0 i$ (обозначение: $z_k \to z_0$), если одновременно $x_k \to x_0$ и $y_k \to y_0$ (в поле $\mathbb{R}$). Это равносильно стремлению к нулю (в поле $\mathbb{R}$) расстояния $|z_k - z_0|$. Число z_0 называется *пределом последовательности* z_k; используется обозначение $z_0 = \lim_{k\to\infty} z_k$; если предел последовательности существует, то он определен однозначно.

(Круговая) *окрестность* точки $z_0 \in \mathbb{C}$ — это открытый круг с центром в z_0 некоторого радиуса $\varepsilon > 0$. Данное выше определение сходимости последовательности можно выразить на *языке окрестностей*: любая окрестность точки z_0 содержит все члены последовательности, начиная с некоторого номера.

Обычным образом дается определение *ограниченной последовательности* комплексных чисел: последовательность z_k ограничена в $\mathbb{C}$, если найдется такое неотрицательное действительное число M, что для всех $k \in \mathbb{N}$ выполняется неравенство $|z_k| \leqslant M$. Из свойств модуля (см. предложение 32.1) легко выводится, что ограниченность последовательности $z_k = x_k + y_k i$ в $\mathbb{C}$ равносильна ограниченности двух последовательностей, x_k и y_k, в $\mathbb{R}$.

Всякая сходящаяся последовательность является ограниченной.

В комплексной области остается справедливой *теорема Больцано — Вейерштрасса*: из любой ограниченной последовательности комплексных чисел можно выбрать сходящуюся подпоследовательность.

В самом деле, из ограниченности z_k в $\mathbb{C}$ следует ограниченность x_k и y_k в $\mathbb{R}$ и теорема Больцано — Вейерштрасса (в действительной области) влечет существование сходящейся подпоследовательности x_{k_l}. Рассмотрим (также ограниченную) подпоследовательность y_{k_l}. Из нее можно выбрать сходящуюся подпоследовательность $y_{k_{l_m}}$, вместе с которой будут сходящимися подпоследовательность $x_{k_{l_m}}$ (в $\mathbb{R}$) и подпоследовательность $z_{k_{l_m}}$ (в $\mathbb{C}$).

Предел функции (34.1) при $z \to z_0$ определяется точно так же, как и предел для функций действительного переменного (на *языке окрестностей* или на хорошо вам известном *ε-δ-языке*.

Остаются справедливыми алгебраические свойства операции предельного перехода: предел суммы (разности, произведения, частного) двух функций (или последовательностей) равен сумме (разности, произведению, частному) пределов данных функций (последовательностей). (В случае частного требуется обычная оговорка: предел знаменателя должен быть отличен от нуля.)

Без всяких изменений в комплексную область переносится определение *непрерывной* функции: функция (34.1) непрерывна в точке $z_0 \in D$, если

$$\lim_{z \to z_0} f(z) = f(z_0),$$

т. е. для любого $\varepsilon > 0$ найдется $\delta > 0$ такое, что выполнение условий $z \in D$ и $|z - z_0| < \delta$ влечет $|f(z) - f(z_0)| < \varepsilon$.

Можно выразить непрерывность функции *по Гейне* (на *языке последовательностей*): функция $f(z)$ непрерывна в точке z_0 тогда и только тогда, когда для любой последовательности z_k из множества D

$$[\, z_k \to z_0 \,] \;\Rightarrow\; [\, f(z_k) \to f(z_0) \,].$$

Композиция непрерывных функций непрерывна.

Из свойств модуля комплексного числа следует, что функция "модуль", $z \mapsto |z|$, действующая из $\mathbb{C}$ в $\mathbb{R}$, непрерывна в каждой точке комплексной плоскости. Значит, непрерывность функции (34.1) в точке z_0 влечет непрерывность (в той же точке) функции $z \mapsto |f(z)|$, действующей из $\mathbb{C}$ в $\mathbb{R}$.

Многочлен степени n ($n \geqslant 0$) с комплексными коэффициентами понимается как функция

$$f(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n; \qquad a_k \in \mathbb{C}\ (k = 0, ..., n);\ a_0 \neq 0;\ z \in \mathbb{C}. \quad (34.2)$$

(При $n = 0$ многочлен является ненулевой константой; нулевая константа также считается многочленом, но для этого, нулевого многочлена степень не определятся.)

Многочлен (34.2) является непрерывной функцией в любой точке комплексной плоскости $\mathbb{C}$. (Это доказывается точно так же, как и в обычном математическом анализе для многочленов с действительными коэффициентами.)

Непрерывность функции (34.2), действующей из $\mathbb{C}$ в $\mathbb{C}$, влечет непрерывность функции "модуль многочлена": $z \mapsto |f(z)|$, которая действует из $\mathbb{C}$ в $\mathbb{R}$.

Замечание 34.2. На рис. 34.1 в прил. 1 представлен график модуля (см. замечание 34.1) для некоторого многочлена степени 4; видны точки, в которых график достигает горизонтальной координатной плоскости (они соответствуют четырем *корням многочлена*).

Сохраняется определение *производной* (в точке z_0) для функции комплексного переменного $w = f(z)$, заданной в некоторой окрестности z_0:

$$f'(z_0) = \lim_{z \to z_0} \frac{\Delta w}{\Delta z};\ \Delta z = z - z_0;\ \Delta w = f(z) - f(z_0). \tag{34.3}$$

Если существует производная $f'(z_0)$, то функция (34.1) называется *дифференцируемой* в точке z_0. Дифференцируемость функции в некоторой точке влечет ее непрерывность в этой точке. Остаются справедливыми основные свойства производных: производная суммы и произведения, производная сложной функции, производная многочлена и т. д.

34.2. Показательная функция (экспонента) комплексного переменного. Великий математик Леонард Эйлер предложил следующее определение *показательной функции* (*экспоненты*) от комплексного переменного:

$$e^z = e^x(\cos\, y + i \sin\, y);\ z = x + yi \in \mathbb{C}. \tag{34.4}$$

Определение не требует доказательства, однако желательной бывает *мотивировка*: откуда взялось такое определение?

Вы знакомы с показательной функцией действительного переменного по школьному курсу математики и по университетскому курсу математического анализа. Знакомство это вряд ли можно назвать глубоким, поскольку почти все, что относится к построению теории действительных чисел, в начальном курсе анализа остается "за кадром". Определение показательной функции для действительного переменного существенно опирается на (строгое) построение теории действительных чисел. Мы уже упоминали в начале настоящего пособия (см. замечание 2.1) о дисциплине "Числовые системы". Придется отложить все "строгости" до той поры, когда вы начнете изучать упомянутую науку, а пока будем основываться на поверхностном (интуитивном) представлении о поле $\mathbb{R}$ (и об элементарных функциях на нем).

Тем не менее мы распространяем [с помощью формулы (34.4)] определение показательной функции на более широкое поле $\mathbb{C}$.

Прежде всего, очевидно, что для действительных z (т. е. при $y = 0$) "новая" показательная функция совпадает со "старой знакомой": $e^z = e^x$; в частности, $e^0 = 1$.

Далее, функция (33.4) обладает характерным свойством показательной функции

$$e^{z_1+z_2} = e^{z_1}e^{z_2} \tag{34.5}$$

для произвольных комплексных чисел $z_1 = x_1 + iy_1$, $z_2 = x_2 + iy_2$.

В самом деле,

$$\begin{aligned} e^{z_1} \cdot e^{z_2} &= e^{x_1}(\cos\, y_1 + i \sin\, y_1) \cdot e^{x_2}(\cos\, y_2 + i \sin\, y_2) = \\ &= e^{x_1}e^{x_2} \cdot ((\cos y_1 \cos y_2 - \sin y_1 \sin y_2) + i(\sin y_1 \cos y_2 + \cos y_1 \sin y_2)) = \\ &= e^{x_1+x_2}(\cos(y_1 + y_2) + i \sin(y_1 + y_2)) = e^{(x_1+x_2)+i(y_1+y_2)} = e^{z_1+z_2}. \end{aligned}$$

(В цепочке преобразований использованы известные свойства показательной и тригонометрических функций в действительной области, а также правило умножения комплексных чисел.)

Разумность определения, данного Эйлером, получила достаточно убедительное обоснование. (Более подробные аргументы в пользу такого определения вы можете найти в учебнике [9].)

Можно обосновать определение Эйлера и другим способом. Для этого следует доказать, наверное, самую знаменитую в студенческой среде формулу:

$$(e^z)' = e^z. \tag{34.6}$$

Равенство (34.6), вместе с условием $e^0 = 1$, *однозначно* определяет функцию из $\mathbb{C}$ в $\mathbb{C}$, и для этой функции, уже как следствие, получается представление (34.4).

Замечание 34.3. Равенство (34.6) означает, что функция $w = e^z$ удовлетворяет *дифференциальному уравнению* $w' = w$. Теорию дифференциальных уравнений вам предстоит изучать на втором курсе в рамках большой и важной дисциплины, так и называемой "Дифференциальные уравнения".

Пример 34.1. Вычисляя по формуле (34.4) значение e^z при чисто мнимом значении аргумента $z = yi$, получим:

$$e^{yi} = \cos\, y + i \sin\, y; \tag{34.7}$$

в частности, при $y = \pi$ приходим к "мистической" формуле

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

содержащей четыре основные математические константы: 1, i, π, e.

Замечание 34.4. Отметим еще одно важное свойство комплексной экспоненты: все ее значения отличны от нуля. В самом деле, первый множитель e^x в формуле (34.4) есть положительное действительное число; второй же множитель есть комплексное число (34.7), которое могло бы обратиться в нуль лишь при равенстве нулю как синуса, так и косинуса числа y, что невозможно одновременно.

34.3. Формулы Эйлера. Вместе с формулой (34.7) рассмотрим вытекающее из нее выражение для e^{-yi}:

$$e^{-yi} = \cos y - i \sin y. \tag{34.7a}$$

Складывая формулы (34.7) и (34.7a), можем получить выражение для $\cos y$:

$$\cos y = \frac{e^{yi} + e^{-yi}}{2}. \tag{34.8}$$

Если же формулу (34.7a) вычесть из формулы (34.7), то можно получить выражение для $\sin y$:

$$\sin y = \frac{e^{yi} - e^{-yi}}{2i}. \tag{34.8a}$$

Формулы (34.8) и (34.8a), выражающие косинус и синус действительного числа y через показательную функцию (от чисто мнимого аргумента), носят название *формул Эйлера*.

Замечание 34.5. В формулах (34.7) и (34.7a) нет ничего удивительного или загадочного: "родственность" тригонометрических и показательной функций усматривается многими различными способами. Например, в курсе математического анализа вам предстоит изучать разложения этих функций в бесконечные суммы (*ряды Маклорена*):

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \frac{x^6}{6!} + \frac{x^7}{7!} + \dots;$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots;$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots.$$

Если в разложении для e^x положить сначала $x = yi$, а потом $x = -yi$, затем взять полусумму результатов, то получится представление для $\cos y$, в точности совпадающее с (34.8).

(Не сочтите проведенное выше "вычисление" доказательством. Понятие суммы бесконечного ряда, условия ее существования, правила оперирования с такими суммами — будут изучаться лишь на втором курсе. В "героический" период развития математического анализа первооткрыватели часто выполняли преобразования, мало заботясь об их законности. Обоснования были получены позже, уже не "героями", а "тружениками".)

По-видимому, вы знакомы (по курсу анализа) еще с одним типом функций, с так называемыми *гиперболическим косинусом* и *гиперболическим синусом*, которые определяются с помощью формул, очень напоминающих формулы Эйлера, но не выходящих за пределы поля действительных чисел:

$$\operatorname{ch} x = \frac{e^x + e^{-x}}{2};\ \operatorname{sh} x = \frac{e^x - e^{-x}}{2}.$$

(Эти функции активно используются в геометрии. Они имеют прямое отношение к давшей им имя гиперболе и, в частности, удовлетворяют так называемому *основному гиперболическому тождеству*: $\operatorname{ch}^2 x - \operatorname{sh}^2 x = 1$.)

Разложения гиперболических косинуса и синуса по формуле Маклорена отличаются от разложений для соответствующих тригонометрических функций лишь отсутствием чередования знаков.

34.4. Показательная форма комплексного числа. Рассмотрим (ненулевое) комплексное число $z = a + bi$, приведем его к тригонометрической форме (32.7), а затем воспользуемся формулой (34.7). Получим так называемую *показательную форму* комплексного числа:

$$z = re^{i\varphi};\ r = |z|;\ \varphi \in \operatorname{Arg} z. \tag{34.9}$$

Замечание 34.6. Показательная форма записи комплексных чисел (так же как и тригонометрическая) весьма удобна при выполнении алгебраических действий "второй и третьей ступени" (т. е. при умножении, делении, возведении в степень и извлечении корня) в поле $\mathbb{C}$. Ее дополнительным преимуществом (по сравнению с тригонометрической формой) является бóльшая лаконичность.

Пример 34.2. Главное значение корня n-й степени из комплексного числа $w = \rho e^{i\psi}$ находится по [вытекающей из (32.27)] формуле

$$\sqrt[n]{\rho e^{\psi i}} = \sqrt[n]{\rho}\, e^{\frac{\psi}{n}}. \tag{34.10}$$

Конкретно (см. пример 32.6):

$$\sqrt[8]{-2} = \sqrt[8]{2}\, e^{\frac{\pi}{8} i}.$$

34.5. Натуральный логарифм комплексного числа. Найти натуральный логарифм для комплексного числа w — это значит найти такое комплексное число z, что $e^z = w$.

Из замечания 34.4 следует, что для числа $w = 0$ логарифм не существует. Если же $w \neq 0$, то приведем это число к тригонометрической (или показательной) форме:

$$w = \rho(\cos\, \psi + i \sin\, \psi) = \rho e^{i\psi};\ \psi \in \operatorname{Arg} w;$$

найдем действительный натуральный логарифм $r = \ln \rho$ для положительного действительного числа ρ и заметим, что

$$e^{r+i\psi} = e^r\, e^{i\psi} = \rho e^{i\psi} = w.$$

Таким образом, ненулевое комплексное число w всегда имеет логарифм

$$z = r + i\,\psi;\ r = \ln \rho;\ \rho = |z|;\ \psi \in \operatorname{Arg} w. \tag{34.11}$$

Однако неоднозначность выбора аргумента ψ для данного числа w приводит к тому, что и логарифм определен не однозначно.

Главное значение логарифма обозначается обычной аббревиатурой ln и определяется по формуле

$$\ln\, w = \ln\, |w| + i \arg w. \tag{34.12}$$

Многозначная функция "логарифм" обозначается Ln и определяется формулой

$$\operatorname{Ln} w = \ln\, |w| + i \operatorname{Arg} w = \ln\, w + 2\pi i k;\ k \in \mathbb{Z}. \tag{34.13}$$

Пример 34.3. Вычислим все значения логарифма для "двух единиц" (действительной и мнимой):

$$\operatorname{Ln} 1 = \ln\, 1 + 2\pi i k = 2\pi i k;\ \operatorname{Ln} i = \ln\, 1 + \frac{\pi}{2} i + 2\pi i k = \pi i(\frac{1}{2} + 2k).$$

§ 35. Основная теорема алгебры

35.1. Корни многочленов (с комплексными коэффициентами). В математике термин "корень" употребляется в двух (близких, но различных) смыслах:

1) *корень* (= *нуль*) *функции* — это такое *значение* ее *аргумента*, при котором *значение функции* равно *нулю*;

2) *корень* (= *решение*) *уравнения* — это такое *значение неизвестной*, при котором *уравнение* обращается в *истинное равенство*.

Применительно к многочленам с комплексными коэффициентами вида (34.2) можно говорить

1) о корнях (= нулях) многочлена как о таких комплексных числах z_0, на которых многочлен $f(z)$, рассматриваемый как функция $f : \mathbb{C} \longrightarrow \mathbb{C}$, принимает нулевое значение: $f(z_0) = 0$;

2) о корнях (= решениях) *алгебраического уравнения* степени n вида $f(z) = 0$, т. е. таких комплексных числах z_0, при подстановке которых в уравнение (вместо неизвестной z) это уравнение обращается в истинное равенство.

Замечание 35.1. Увы, многозначность термина провоцирует путаницу в головах бедных первокурсников. Зачастую приходится слышать "определение": "Корень многочлена — это такое число, которое обращает его в истинное равенство". Но многочлен есть не "равенство", а функция...

Усугубляет путаницу еще и наличие третьего значения у термина "корень", с которым мы встречались, например, в § 32, говоря о *корнях заданной степени из заданного* (комплексного) *числа*.

Третье понимание термина "корень" может быть сведено к первым двум следующим образом. Комплексное число z_0 является корнем степени n из комплексного числа w тогда и только тогда, когда z_0 является корнем (= решением) уравнения $z^n = w$ или, что равносильно, корнем (= нулем) функции $z \mapsto z^n - w$.

Многочлены нулевой степени корней не имеют. Корнем нулевого многочлена (степень которого не определена) можно считать любое число. Очевидно, что всякий многочлен первой степени имеет корень, причем только один. Задача отыскания корней для многочлена второй степени с комплексными коэффициентами рассматривалась в п. 31.5 (задача 4); квадратичный многочлен в поле $\mathbb{C}$ всегда имеет два корня (различных или одинаковых).

В конце XVIII века Гауссом было получено первое строгое доказательство того факта, что всякий многочлен положительной степени

(с комплексными коэффициентами) имеет хотя бы один (комплексный) корень. (Мы уже упоминали сформулированную выше теорему в замечании 32.12.)

Замечание 35.2. История доказательства этой великой теоремы драматична и поучительна. Фактически основные идеи доказательства были ясны значительно раньше. Д'Аламбер, например, считал (в середине XVIII века), что доказательство у него есть, но "цензоры"-потомки признали его рассуждения недостаточно строгими.

(Заинтересованных читателей можно отослать к специальной математической литературе, а также к историко-математическим сочинениям и научно-популярным статьям. Укажем, например, на работу В. М. Тихомирова и В. В. Успенского "Десять доказательств основной теоремы алгебры" в сборнике "Математическое просвещение" (Сер. 3. 1997. Вып. 1. С. 50 — 70). Из этой статьи мы заимствуем кое-какую информацию исторического и методического плана.)

Есть еще один весьма примечательный аспект, связанный с основной теоремой алгебры: не существует чисто алгебраических ее доказательств. И это не случайно: построение поля $\mathbb{R}$ опирается, помимо алгебры, на неалгебраические (*топологические*) понятия и методы (вы с ними знакомитесь в курсе математического анализа). И хотя переход от $\mathbb{R}$ к $\mathbb{C}$ является уже чисто алгебраическим, не удается доказать основную теорему, относящуюся к полю комплексных чисел, минуя ключевые топологические факты, относящиеся к полю $\mathbb{R}$, и их следствия, относящиеся к полю $\mathbb{C}$. (В п. 34.1, посвященном функциям комплексного переменного, их пределам, непрерывности и т. п., мы уже сформулировали, и даже отчасти доказали, упомянутые факты.)

35.2. Основная теорема алгебры (формулировка, идеи и этапы доказательства)

Теорема 35.1 (основная теорема алгебры — ОТА). *Всякий многочлен*

$$f(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n \tag{35.1}$$

положительной степени ($n > 0$) *с комплексными коэффициентами* ($a_k \in \mathbb{C}$; $k = 0, ..., n$; $a_0 \neq 0$ *имеет хотя бы один комплексный корень, т. е. всегда найдется такое число* $z_0 \in \mathbb{C}$, *что*

$$f(z_0) = 0. \tag{35.2}$$

Доказательство (основные идеи). Прежде всего сведем изучение корней многочлена $f(z)$, как комплексной функции комплексного переменного z, к изучению корней (нулей) действительной функции комплексного переменного

$$\rho = |\, a_0 z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n|. \tag{35.3}$$

Как указывалось в замечании 34.2, точки $z_0 \in \mathbb{C}$, удовлетворяющие (34.2), представляются как точки, в которых достигается *наименьшее значение* (равное нулю) для действительной функции $\rho = |f(z)|$. В этих точках поверхность-график $G_{|f|}$ (расположенная в *верхнем полупространстве* $\rho \geqslant 0$ трехмерного пространства с координатами (x, y, ρ), где $x = \operatorname{Re} z$, $y = \operatorname{Im} z$) достигает горизонтальной плоскости $\rho = 0$. (Полюбуйтесь еще раз рисунком 34.1.)

После этого возникает естественный план.

Э т а п 1. Рассмотрим точную нижнюю грань (*инфинум*) значений функции $|f|$:

$$m = \inf_{z \in \mathbb{C}} |f(z)|; \tag{35.4}$$

неотрицательность рассматриваемой функции влечет существование и неотрицательность числа m.

Далее установим, что инфинум m *достигается* в некоторой точке z_0, т. е. $m = |f(z_0)|$ является наименьшим значением функции (35.3).

(Именно в этом пункте потомки обнаружили основной пробел в доказательстве ОТА, данном Д'Аламбером; дело в том, что во времена Д'Аламбера не была еще доказана теорема Больцано — Вейерштрасса; и даже не родились ни Больцано, ни Вейерштрасс.)

На этом этапе доказывается так называемая *лемма о возрастании модуля многочлена* (у нас это — лемма 1), а затем, в лемме 2, устанавливается достижимость инфинума.

Э т а п 2. Докажем *лемму Д'Аламбера* (у нас это — лемма 3), в которой утверждается, что если в некоторой точке z_0 значение многочлена $f(z_0)$ отлично от нуля (и следовательно, $|f(z_0)| > 0$), то найдется такая точка $z_1 \in \mathbb{C}$, что $|f(z_1)| < |f(z_0)|$.

Э т а п 3. Заметим, что наименьшее значение m, достижимость которого установлена в лемме 2, не может быть положительным, ибо это противоречило бы лемме 3.

Этим доказательство завершается: точка z_0, в которой достигается инфинум значений модуля данного многочлена, должна быть *корнем* этого многочлена.

35.3. Основная теорема алгебры (подробное доказательство). Прежде всего заметим, что не будет ничего плохого, если чтение этого пункта вы отложите до конца следующей главы, которая будет специально посвящена *теории многочленов*.

Замечание 35.3 (для служебного пользования). Даже при лекционном изложении часто удобнее отложить подробное доказательство теоремы 35.1 до самой последней лекции первого семестра: так она лучше запоминается, становится событием в студенческой жизни.

Приходится, однако, считаться и с нередкой "пропажей" лекций, в основном вызванной (превышающим всякую меру) изобилием праздников в календаре. Случается и так, что студенты остаются без рассказанного у доски доказательства. Тогда приходится обращаться к учебникам и пособиям.

Можно, конечно, не спешить, дождаться второго или третьего курса, чтобы в какой-либо из "продвинутых" наук (например, топологии или ТФКП) преподаватель после семестра или целого года лекций мог торжественно заявить: "Вот вам доказательство так называемой основной теоремы алгебры! Оно занимает всего две строчки!"

Как не вспомнить по этому поводу древний математический афоризм: "Доказательство любой теоремы может быть сведено в две строчки (если сделать строчки достаточно длинными)"!

Приводимый здесь вариант доказательства является некоторой компиляцией доказательств, содержащихся в учебниках [2, 3, 9].

Доказательство ОТА (подробности). Реализуем план, изложенный в предыдущем пункте.

Э т а п 1. Докажем первое вспомогательное утверждение.

Лемма 1 (о возрастании модуля многочлена). *Рассмотрим многочлен с комплексными коэффициентами $f(z)$ положительной степени. Если последовательность комплексных чисел z_k такова, что $|z_k| \to \infty$ (в поле $\mathbb{R}$), то и $|f(z_k)| \to \infty$ (в поле $\mathbb{R}$).*

Доказательство. Подчеркнем прежде всего, что в формулировке леммы речь идет о сходимости к бесконечности последовательностей *действительных* чисел. (Так, сходимость к ∞ первой из последовательностей означает, что ее члены $|z_k|$, начиная с некоторого номера, становятся больше любого наперед заданного положительного числа. Это равносильно сходимости к нулю последовательности $1/|z_k|$.)

Пользуясь свойствами модуля (32.2) (причем свойство "модуль суммы" будет применяться в обобщенном виде: для суммы нескольких слагаемых), выпишем следующую цепочку, состоящую из равенств и неравенств:

$$\begin{aligned}
|f(z_k)| &= \left|a_0 z_k^n + (a_1 z_k^{n-1} + \cdots + a_{n-1} z_k + a_n)\right| \geqslant \\
&\geqslant \left||a_0 z_k^n| - |a_1 z_k^{n-1} + \cdots + a_{n-1} z_k + a_n|\right| \geqslant \\
&\geqslant |a_0 z_k^n| - |a_1 z_k^{n-1} + \cdots + a_{n-1} z_k + a_n| \geqslant \\
&\geqslant |a_0 z_k^n| - (|a_1 z_k^{n-1}| + \cdots + |a_{n-1} z_k| + |a_n|) = \\
&= |a_0||z_k|^n - (|a_1||z_k|^{n-1} + \cdots + |a_{n-1}|z_k| + |a_n|) = \\
&= |z_k|^n \left[|a_0| - \left(\frac{|a_1|}{|z_k|} + \cdots + \frac{|a_{n-1}|}{|z_k|^{n-1}} + \frac{|a_n|}{|z_k|^n}\right)\right].
\end{aligned}$$

Исследуем поведение последнего из выражений в цепочке при $k \to \infty$. Сумма в круглых скобках содержит n дробей с постоянными (неотрицательными) числителями и со знаменателями, стремящимися к $+\infty$. Значит, каждая из дробей стремится к нулю и все выражение в круглых скобках также стремится к нулю. Следовательно, выражение в квадратных скобках стремится к положительной константе $|a_0|$. Перед квадратными скобками стоит множитель $|z_k|^n$, стремящийся к $+\infty$. Следовательно, и все (последнее в цепочке) выражение стремится к $+\infty$. А с ним вместе и $|f(z_k)| \to \infty$. $\square$

При описании первого этапа в предыдущем пункте уже введено в рассмотрение неотрицательное число (35.4), являющееся точной нижней гранью значений $|f(z)|$. Докажем второй вспомогательный результат: достижимость этого инфинума.

Лемма 2. *Пусть m — точная нижняя грань значений функции $|f(z)|$. Тогда существует такая точка $z_0 \in \mathbb{C}$, что $|f(z_0)| = m$, т. е.*

$$(\,\forall z \in \mathbb{C}\,)\ (\,|f(z_0)| \leqslant |f(z)|\,). \tag{35.5}$$

Доказательство. По свойствам точной нижней грани (загляните в ваши учебники по математическому анализу), найдется такая последовательность z_k (в поле $\mathbb{C}$), что

$$|f(z_k)| \to m \tag{35.6}$$

(в поле $\mathbb{R}$).

Докажем, что последовательность z_k ограничена (см. определение ограниченности в п. 34.1).

Предположим противное.

Неограниченность рассматриваемой последовательности должна означать, что для любого натурального l найдется член этой последовательности (с номером k_l, зависящим от l) такой, что $|z_{k_l}| > l$.

Так возникает подпоследовательность z_{k_l}, обладающая тем свойством, что $|z_{k_l}| \to \infty$. Выбросив из последовательности z_k "лишние члены", мы можем считать это свойство выполняющимся уже для исходной последовательности: $|z_k| \to \infty$.

В силу леммы 1, отсюда следует, что $|f(z_k)| \to \infty$, а это противоречит (35.6). Значит, предположение неверно и данная последовательность z_k является ограниченной.

Применим теорему Больцано — Вейерштрасса (см. п. 34.1). Эта теорема позволяет выделить из z_k сходящуюся подпоследовательность. Снова (чтобы не усложнять без нужды обозначения) будем считать, что лишние члены выброшены и что сходящейся является уже исходная последовательность, т. е. существует точка $z_0 \in \mathbb{C}$ такая, что $z_k \to z_0$.

В силу непрерывности модуля многочлена (т. е. функции $|f(z)|$; см. п. 34.1), отсюда будет следовать:

$$|f(z_k)| \to |f(z_0)|. \tag{35.7}$$

Утверждения (35.6) и (35.7) с учетом свойства единственности предела последовательности влекут равенство $|f(z_0)| = m$; таким образом, достижимость инфинума m доказана. Вместе с ней доказано и утверждение (35.5). □

Э т а п 2. Этот этап будет самым серьезным и продолжительным. Прежде всего обоснуем некоторую упрощающую замену переменных, которая будет использоваться при доказательстве леммы Д'Аламбера.

Рассмотрим многочлен (35.1) степени $n > 0$ и произвольную точку $z_0 \in \mathbb{C}$. Вместе с (35.1) рассмотрим функцию $g(z) = f(z + z_0)$, т. е.

$$g(z) = a_0(z + z_0)^n + a_1(z + z_0)^{n-1} + ... + a_{n-1}(z + z_0) + a_n. \tag{35.8}$$

Очевидно, функция $g(z)$ также является многочленом, той же степени n и с тем же старшим коэффициентом a_0.

[Чтобы в этом убедиться, достаточно раскрыть с помощью бинома Ньютона (31.26) выражение $(z + z_0)^n$:

$$(z+z_0)^n = z^n + C_n^1 z_0 z^{n-1} + ... + C_n^{n-1} z_0^{n-1} z + z_0^n;$$

это разложение начинается с члена z^n, который после раскрытия скобок окажется умноженным на a_0; ни в этом, ни в последующих младших членах n-я степень z не встретится.]

Новый многочлен $g(z)$ называется *сдвигом* многочлена $f(z)$; для него справедливо свойство $g(0) = f(z_0)$.

Дополнительно предположив, что значение $f(z_0) \neq 0$, т. е. что число z_0 не является корнем многочлена $f(z)$, мы, после сдвига, проведем деление на константу (так называемую *нормализацию*; см. ниже п. 36.5), т. е. рассмотрим многочлен

$$h(z) = \frac{g(z)}{f(z_0)} = \frac{f(z+z_0)}{f(z_0)}. \tag{35.9}$$

Этот многочлен также будет иметь степень n, и для него будет выполняться свойство: $h(0) = 1$. Другими словами, свободный член многочлена $h(z)$ будет равен единице. Вводя общее обозначение b_k $(k = 0, ..., n)$ для коэффициентов $h(z)$, получим представление:

$$h(z) = b_0 z^n + b_1 z^{n-1} + ... + b_{n-1} z + b_n, \tag{35.10}$$

где

$$b_n = 1;\ b_0 = \frac{a_0}{f(z_0)} \neq 0. \tag{35.11}$$

Теперь мы готовы к тому, чтобы сформулировать и доказать следующий замечательный результат.

Лемма 3 (лемма Д'Аламбера). *Пусть $f(z)$ — многочлен положительной степени с комплексными коэффициентами; $z_0 \in \mathbb{C}$. Если значение $f(z_0) \neq 0$, то найдется точка $z_1 \in \mathbb{C}$ такая, что*

$$|f(z_1)| < |f(z_0)|. \tag{35.12}$$

Доказательство. Вместе с данным многочленом (35.1) рассмотрим многочлен $h(z)$, заданный формулой (35.9). Если будет найдена точка $v \in \mathbb{C}$ такая, что

$$|h(v)| < 1, \tag{35.13}$$

то для точки $z_1 = z_0 + v$ будет выполняться условие (35.12). [Действительно, (35.13) влечет $|f(v + z_0)|/|f(z_0)| < 1$, что равносильно (35.12).]

Точку v, обеспечивающую выполнение (35.13), будем искать так: сначала специальным образом выберем направление, задаваемое комплексным числом (вектором) u, а затем подберем такой действительный множитнль $t \in (0, 1)$, чтобы комплексное число $v = tu$ удовлетворяло (35.13).

Обратимся к представлению (35.10). В нем наверняка отличны от нуля старший коэффициент и свободный член $b_n = 1$. Один или несколько подряд коэффициентов, предшествующих b_n, могут оказаться равными нулю. Пусть первым справа ненулевым коэффициентом в (35.10), отличным от свободного члена, будет коэффициент b_{n-p}, стоя́щий при z^p (здесь $p \in \{1, ..., n\}$).

Разложение (35.10) примет вид:

$$h(z) = b_0 z^n + b_z z^{n-1} + ... + b_{n-p} z^p + 1, \tag{35.14}$$

причем $b_{n-p} \neq 0$.

Подставим в формулу (35.14) $z = tu$, где u будет фиксированным (но пока неопределенным) ненулевым комплексным числом, а t будет *действительной* переменной. При изменении t от 0 до 1 точка z пробегает на комплексной плоскости отрезок от числа 0 до числа u.

Рассмотрим (комплексную) функцию $q(t) = h(tu)$ действительной переменной t, т. е. многочлен

$$q(t) = c_0 t^n + c_1 t^{n-1} + ... + c_{n-p-1} t^{p+1} + c_{n-p} t^p + 1 \tag{35.15}$$

с комплексными коэффициентами

$$c_k = b_k u^{n-k};\ k = 0, 1, ..., n - p. \tag{35.16}$$

Подберем теперь комплексное число u так, чтобы коэффициент $c_{n-p} = -1$. Для этого надо решить (относительно неизвестной u) уравнение $b_{n-p} u^p = -1$, или, что равносильно (т. к. $b_{n-p} \neq 0$), уравнение

$$u^p = -\frac{1}{b_{n-p}}. \tag{35.17}$$

Решить (35.17) — это значит извлечь корень степени p из (ненулевого) комплексного числа, находящегося в правой части уравнения.

В силу предложения 32.3, уравнение (35.17) имеет p решений. Выберем и зафиксируем одно (любое) из них. Формула (35.16) примет вид

$$q(t) = c_0 t^n + c_1 t^{n-1} + ... + c_{n-p-1} t^{p+1} - t^p + 1. \qquad (35.18)$$

Если $p = n$, то $q(t) = -t^n + 1$ и для всех $t \in (0, 1)$ получается $|q(t)| = q(t) < 1$.

Если $p < n$, то (35.18) можно переписать следующим образом:

$$q(t) = 1 - t^p + t^{p+1} \left(c_0 t^{n-p-1} + c_1 t^{n-p-2} + ... + c_{n-p-1} \right). \qquad (35.19)$$

Обозначим $r(t)$ многочлен (степени $n - p - 1$, с комплексными коэффициентами), стоя́щий в круглых скобках в формуле (35.19).

Оценим по модулю при $t \in (0, 1)$ многочлены $r(t)$ и $q(t)$.

Пусть $C = \max(|c_0|, |c_1|, ..., |c_{n-p-1}|)$. По свойствам модуля (с учетом того, что $|t| = t < 1$ и, следовательно, $|t^k| < 1$ для любой натуральной степени k) получим:

$$|r(t)| \leqslant |c_0||t|^{n-p-1} + |c_1||t|^{n-p-2} + ... + |c_{n-p-1}| < C(n-p). \qquad (35.20)$$

Обозначив константу [фигурирующую в (35.20)] $C(n - p) = A$, запишем окончательную оценку для $r(t)$:

$$|r(t)| < A; \ t \in (0, 1). \qquad (35.21)$$

Для $q(t)$ получим [с учетом (35.21) и того, что $1 - t^p > 0$]:

$$|q(t)| = |1 - t^p + t^{p+1} r(t)| \leqslant$$
$$\leqslant |1 - t^p| + t^{p+1}|r(t)| < 1 - t^p + At^{p+1} = 1 - t^t(1 - At). \qquad (35.22)$$

Если выбрать $t \in (0, \frac{1}{A})$, то из (35.22) будет следовать неравенство

$$|q(t)| < A; \ t \in (0, 1). \qquad (35.23)$$

(Вспомним, что и в особом случае $p = n$ получилось такое же неравенство.)

Пользуясь определением многочлена $q(t)$, заключаем, что если u удовлетворяет (35.17), а t принадлежит интервалу от 0 до $1/A$, то выполняется неравенство $|h(tu)| < 1$.

А это и есть (если положить $v = tu$) требуемое неравенство (35.13). Доказательство леммы завершено. □

Замечание 35.4. На самом деле доказано больше, чем утверждалось: точка v такая, что выполняется (35.13), найдется *в любой окрестности* точки 0. Отсюда легко вывести, что точку z_1 такую, что выполняется (35.12), можно найти *в любой окрестности* точки z_0.

Э т а п 3. Осталось самое простое: закончить доказательство ОТА. (Фактически это уже сделано в предыдущем пункте; здесь мы приведем чуть более подробное рассуждение.)

Пусть в точке z_0 достигается наименьшее значение функции $|f(z)|$ (такая точка существует в силу леммы 2). Значение $f(z_0)$ является нулевым, т. к. в противном случае, в силу леммы 3, число $|f(z_0)|$ не могло бы являться наименьшим значением функции $|f(z)|$.

Доказательство основной теоремы алгебры завершено. □

Глава 6

АЛГЕБРА МНОГОЧЛЕНОВ

§ 36. Векторная модель кольца многочленов (над полем)

36.1. Алгебраическое определение многочлена над полем. Мы уже работали с многочленами в § 34 — 35. Более того, понятие о многочленах дает уже школьный курс алгебры и начал анализа. К тому же многочлены играют ключевую роль во многих разделах университетского курса математического анализа.

Однако и в анализе, и в предыдущей главе нашего пособия мы рассматривали многочлены как *функции*. Их коэффициенты считались принадлежащими какому-либо *числовому полю* (чаще всего полю $\mathbb{R}$ или полю $\mathbb{C}$).

Теперь пришла пора изучить другую точку зрения на многочлены, чисто *алгебраическую*, более тонкую и богатую (но и более формальную). *Аналитическая* точка зрения не будет оставлена насовсем: мы разберемся с ситуациями, когда две названные точки зрения равносильны и когда аналитический подход перестает работать.

Коэффициенты многочленов будут предполагаться принадлежащими *произвольному* (не обязательно числовому) полю. (Можете перечитать в связи с этим п. 2.1 и, в частности, вспомнить о существовании *конечных* полей.)

Вместо выражения "многочлены с коэффициентами из поля P" принято употреблять термин "многочлены *над полем* P".

Естественным, однако, является еще более общий подход, когда рассматриваются многочлены *над коммутативным кольцом*. (Понятие кольца встречалось нам в пп. 14.1 и 15.5. Коммутативное кольцо отличается от поля лишь тем, что может не выполняться аксиома $\boxed{9}$ из п. 2.1. Скажем, множество $\mathbb{Z}$ целых чисел является коммутативным кольцом, но не полем.)

В основном в настоящей главе мы будем заниматься многочленами *от одной переменной* (которая, кстати, будет трактоваться совсем не так, как в математическом анализе, а более абстрактно, на "векторном языке").

Переходим к строгим определениям.

Пусть P — некоторое поле, n — неотрицательное целое число, x — произвольная *буква* (*формальная переменная*, которой пока не будет приписываться никакого алгебраического значения).

Определение 36.1. Многочленом (или *полиномом*) *степени* n над полем P, от переменной x, называется выражение вида

$$f(x) = f_0 + f_1 x + f_2 x^2 + ... + f_n x^n, \tag{36.1}$$

где f_k $(k = 0, ..., n)$ являются элементами поля P (они называются *коэффициентами* многочлена), причем $f_n \neq 0$ (этот коэффициент называется *старшим*).

Многочлены степени $n = 0$, т. е. *ненулевые* элементы поля P, называются *скалярными* (или *постоянными*).

Нулевой элемент поля P также считается скалярным многочленом (никакой степени ему не приписывается).

Множество всех многочленов над полем P, от переменной x, обозначается $P[x]$.

Замечание 36.1. Применяется также краткая запись многочленов с помощью знака суммирования:

$$f(x) = \sum_{k=0}^{n} f_k x^k; \ f_n \neq 0. \tag{36.1a}$$

В математике мало бывает дать определение *объекта*. Необходимо еще определить, когда два объекта (рассматриваемого класса) *равны* между собой.

Определение 36.2. Два многочлена $f(x)$ и $g(x)$ (над одним и тем же полем P и от одной и той же переменной x) называются равными (это обозначается $f(x) = g(x)$ или короче $f = g$), если

1) их степени одинаковы;

2) все соответствующие коэффициенты равны.

Если многочлен $f(x)$ задан формулой (36.1), а многочлен $g(x)$ формулой

$$g(x) = g_0 + g_1 x + g_2 x^2 + ... + g_m x^m; \ g_m \neq 0, \tag{36.2}$$

то условия 1 и 2 определения 36.2 будут выглядеть следующим образом: 1) $n = m$ и 2) для любого $k = 0, ..., n$: $f_k = g_k$.

Сразу заметим, что равенство многочленов (одинаковой степени n) очень похоже на равенство *арифметических векторов-строк*. Длина (количество координат) этих строк равняется $n + 1$ [ввиду наличия координаты (коэффициента) с нулевым номером].

Многочлену $f(x)$ будет соответствовать вектор-строка

$$\overline{f}^t = (\, f_0,\, f_1,\, f_2,\, ...,\, f_n\,).$$

(Напомним, что буква t является знаком транспонирования. Черточками мы помечаем векторы-столбцы, черточками и знаком транспонирования — векторы-строки.)

Возможно, вы обратили внимание на то, что в формуле (36.1) члены многочлена идут *в порядке возрастания степеней*, тогда как до сего момента мы записывали многочлены *в порядке убывания степеней*.

Это изменение формы записи обусловлено *векторным* подходом к многочленам. (Позже будут применяться оба порядка записи.)

Есть, однако, существенная трудность, связанная с необходимостью одновременного рассмотрения многочленов различных степеней. Соответствующие векторы будут при этом принадлежать различным арифметическим линейным пространствам; следовательно, их нельзя будет складывать.

Алгебраисты разрешают эту коллизию решительно и радикально: вводятся *бесконечные* (но *финитные*) векторы-строки, координаты которых образуют бесконечную последовательность (нумеруются от 0 до бесконечности), но лишь конечное число координат может быть отличным от нуля (последней по номеру ненулевой координатой будет старший коэффициент):

$$\overline{f}^t = (f_0, f_1, f_2, ..., f_n, 0, 0, ...);\ f_k \in P;\ k = 0, 1, ..., n;\ f_n \neq 0. \qquad (36.3)$$

Множество векторов вида (36.3) будем обозначать символом P_0^∞. Для элементов этого множества применяется также краткая запись:

$$\bar{f}^t = (f_k)_{k=0}^\infty;\ f_n \neq 0;\ (\forall k > n)\,(f_k = 0). \qquad (36.3а)$$

Обозначение P^∞ остается зарезервированным для множества всех (не обязательно финитных) последовательностей вида $(f_k)_{k=0}^\infty$ с элементами из поля P.

Множество P^∞ с покомпонентными сложением и умножением на скаляр является *линейным пространством* (над полем P) [этот термин означает, что на данном множестве определены две указанные линейные алгебраические операции и что они удовлетворяют условиям (i) — (viii)]; так получается бесконечномерный аналог арифметических линейных пространств P^n. (Мы не приводим здесь строгого определения понятия *бесконечномерного* линейного пространства. Подождите до следующего семестра. На интуитивном уровне вы должны чувствовать, о чем идет речь.)

Подмножество P_0^∞ будет, очевидно, *линейным подпространством* (см. определение и пояснения в п. 3.2) в пространстве P^∞. Это подпространство само является бесконечномерным. По определению равенства многочленов, существует естественная биекция между множествами $P[x]$ и $P_0^\infty : f(x) \mapsto \bar{f}^t$.

Краткой записи (36.1a) многочлена $f(x)$ можно придать такую форму:

$$f(x) = \sum_{k=0}^{\infty} f_k x^k;\ \ f_n \neq 0; (\forall k > n)\,(f_k = 0). \tag{36.1b}$$

Если в P_0^∞ выделить подмножество, состоящее из последовательностей $(f_k)_{k=0}^\infty$ таких, что все f_k с номерами $k > n$ (n фиксировано) обращаются в нуль, то получится уже *конечномерное* линейное пространство, *изоморфное* P^{n+1}. Векторам из этого подпространства отвечают многочлены, степень которых *не превышает* заданного числа n; множество таких многочленов будет обозначаться $P_n[x]$.

Замечание 36.2. А что при описанном подходе будет соответствовать произвольным (не обязательно финитным) последовательностям $(f_k)_{k=0}^\infty$ из P^∞? Здесь возникают так называемые *формальные степенные ряды*, т. е. *бесконечные суммы* вида

$$\sum_{k=0}^{\infty} f_k x^k = f_0 + f_1 x + f_2 x^2 + \ldots + f_k x^k + \ldots\ . \tag{36.4}$$

Множество таких рядов (находящееся в биективном соответствии с пространством P^∞) обозначается $P[[x]]$.

Имейте в виду, что подобным бесконечным суммам не приписывается никакого *значения*. Позже, в курсе математического анализа, вы будете заниматься *сходящимися* рядами — бесконечными суммами, для которых будет определено их (конечное) значение.

Формальные ряды вам тоже обязательно встретятся в важнейшей для компьютерщиков дисциплине "Дискретная математика".

С помощью векторного представления многочленов можно пересказать определение 36.2 равенства многочленов: два многочлена считаются равными ($f(x) = g(x)$ в множестве $P[x]$) тогда и только тогда, когда равны соответствующие векторы ($\bar{f}^t = \bar{g}^t$ в пространстве P_0^∞).

Уже самим определением каждому многочлену (кроме нулевого) сопоставляется неотрицательное целое число — его степень. Это число обозначается $\deg(f(x))$ или $\deg(f)$.

Так возникает функция "степень", являющаяся отображением из множества всех ненулевых многочленов над заданным полем в множество всех неотрицательных целых чисел:

$$\deg : P[x] \setminus \{0\} \longrightarrow \mathbb{N} \cup \{0\};$$
$$f(x) \mapsto \deg(f(x));\ f(x) \in P[x];\ f(x) \neq 0. \quad (36.5)$$

Замечание 36.3. Иногда функцию deg распространяют и на нулевой многочлен, полагая

$$\deg(0) = -\infty. \quad (36.6)$$

К мотивировке этого определения мы вернемся в следующих пунктах. Формула (36.6) является единственным разумным способом определения степени нулевого многочлена (таким, что остаются в силе "свойства степени"; см. ниже).

Пример 36.1. Рассмотрим "самое маленькое" из конечных полей — поле $\mathbb{F}_2$ (см. замечание 2.1). Это поле состоит из двух элементов: 0 и 1, действия над которыми выполняются по модулю 2. Перечислим все многочлены степени не выше трех над этим полем.

$$(\mathbb{F}_2)_3[x] = \{0, 1;\ x, x+1;$$
$$x^2, x^2+1, x^2+x, x^2+x+1;\ x^3, x^3+1, x^3+x, x^3+x+1,$$
$$x^3+x^2, x^3+x^2+1, x^3+x^2+x, x^3+x^2+x+1\}.$$

Разберитесь самостоятельно с принципом перечисления многочленов (над конечным полем). Сколько существует над рассматриваемым полем многочленов четвертой степени? степени n? Обратите также внимание на тот факт, что, хотя данное поле является конечным, множество *всех* многочленов над этим полем будет бесконечным.

36.2. Линейное пространство многочленов. Биекция между множеством многочленов $P[x]$ и линейным пространством P_0^∞ позволяет наделить $P[x]$ структурой линейного пространства: многочлены будут складываться и умножаться на скаляр по тому же принципу, что и векторы, т. е. покомпонентно.

Это будет означать, что при сложении многочленов $f(x)$ и $g(x)$ будут складываться коэффициенты при соответствующих степенях. Получится новый многочлен $h(x) = f(x) + g(x)$ с коэффициентами

$$h_k = f_k + g_k \ (k = 0, 1, 2, ...). \tag{36.7}$$

(Именно так вас учили складывать многочлены в школе.)

Пусть данные многочлены определены формулами (36.1) и (36.2). При $k > \max(n, m)$ коэффициенты как многочлена $f(x)$, так и многочлена $g(x)$, а значит, и многочлена $h(x)$ станут нулевыми. Если степени данных многочленов различны ($n \neq m$), то степень многочлена $h(x)$ будет равна $\max(n, m)$. Если $n = m$, то степень суммы может оказаться меньше наибольшей из степеней слагаемых (так получится, если старшие члены многочленов-слагаемых взаимно уничтожатся). Оба случая можно охватить следующим правилом.

П р а в и л о 1:

$$\deg(f(x) + g(x)) \leqslant \max(\deg(f(x)), \deg(g(x))), \tag{36.8}$$

причем, если $\deg(f(x)) \neq \deg(g(x))$, то неравенство (36.8) обращается в равенство.

Замечание 36.4. Нетрудно убедиться в том, что если степень нулевого многочлена определять в соответствии с формулой (36.6), то правило 1 сохраняет силу и тогда, когда какие-либо из участвующих в ней многочленов окажутся нулевыми. (Проследите внимательно за всеми возможными случаями.)

При умножении многочлена на скаляр все его коэффициенты умножаются на этот скаляр. Если скаляр является ненулевым, то после умножения на него степень многочлена останется неизменной.

Пример 36.2. Рассмотрим *поле классов вычетов* по модулю 5, обозначаемое $\mathbb{F}_5$ и состоящее из пяти элементов: 0, 1, 2, 3 и 4 (всех возможных остатков при делении на 5), которые складываются и умножаются по модулю 5.

[Тот факт, что остатки от деления на *простое* натуральное число p действительно образуют поле (обозначаемое $\mathbb{F}_p$), для математиков

доказывается в курсе теории чисел и в курсе алгебры (3-й семестр). Компьютерщикам придется — и настоятельно рекомендуется! — изучить этот материал самостоятельно.]

Рассмотрим два многочлена над полем $\mathbb{F}_5$:

$$f(x) = 4x^2 + x + 3;\ g(x) = x^2 + 2x + 2.$$

Выполним алгебраические действия сложения и вычитания над этими многочленами:

$$f(x) + g(x) = 3x;\ f(x) - g(x) = 3x^2 + 4x + 1.$$

Итак, множество $P[x]$ многочленов над полем P является линейным пространством над полем P, изоморфным линейному пространству P_0^∞. Подобно тому как в арифметическом линейном пространстве P^n существует естественный базис, состоящий из единичных векторов-строк $[\bar{e}_1^t, ..., \bar{e}_n^t]$, *естественный базис* $[\bar{e}_0^t, \bar{e}_1^t, ..., \bar{e}_k^t, ...]$ (уже бесконечный) можно обнаружить в пространстве P_0^∞. Здесь $\bar{e}_k^t = (0, ..., 0, 1, 0, 0, ...)$, где 1 стоит на k-м месте ($k = 0, 1, 2, ...$).

Всякая финитная вектор-строка, принадлежащая P_0^∞, однозначно представляется в виде (конечной) линейной комбинации базисных строк:

$$\bar{f}^t = \sum_{k=0}^{n} f_k \bar{e}_k^t. \tag{36.9}$$

Базисным строкам соответствуют в пространстве многочленов базисные многочлены (одночлены): $\bar{e}_0^t$ соответствует скаляр 1; $\bar{e}_1^t$ соответствует одночлен x; $\bar{e}_2^t$ соответствует x^2 и т. д. Формула (36.1) [или (36.1а)], по сути, представляет собой не что иное, как разложение многочлена $f(x)$ по базисным одночленам x^k ($k = 0, 1, 2, ...$).

Замечание 36.5. В какой-то мере это проясняет алгебраическую сущность переменной x, которая до сих пор никак не уточнялась. *Переменная* является *базисным вектором с номером 1* в естественном базисе пространства $P[x]$.

(Данное замечание подлежит дальнейшему уточнению в следующем пункте, в замечании 36.7, после изучения умножения многочленов.)

36.3. Кольцо многочленов. В школьной математике произведение $h(x) = f(x) \cdot g(x)$ многочленов (36.1) и (36.2) определялось следующим образом: надо записать произведение двух указанных сумм

(в скобках); раскрыть скобки; упростить все nm полученных одночленов, пользуясь правилом перемножения степеней $x^k \cdot x^l = x^{k+l}$; привести подобные члены и перегруппировать их в порядке степеней. Выразим это определение в терминах векторной модели для множества многочленов.

Пусть многочлен $f(x)$ степени n задан формулой (36.1b), а многочлен $g(x)$ степени m — аналогичной формулой:

$$g(x) = \sum_{l=0}^{\infty} g_l x^l;\ g_m \neq 0; (\forall l > m)\,(g_l = 0). \tag{36.2b}$$

Индекс суммирования в формуле (36.2b) выбран отличным от индекса суммирования в (36.1a). Это необходимо для корректной записи последующих формул.

Найдем теперь аналогичную запись для многочлена $h(x)$. Индекс суммирования снова придется обозначить новой буквой, скажем s. Степень x^s $(s \geqslant 0)$ в разложении для $h(x)$ может получиться как произведение $x^s = x^k \cdot x^l$, где k и l — любые неотрицательные целые числа, такие, что $k + l = s$. При каждом из указанных произведений будет получаться коэффициент, равный $f_k g_l$.

После группировки и вынесения x^s за скобки сформируется одночлен $h_s x^s$ с коэффициентом

$$h_s = f_0 g_s + f_1 g_{s-1} + \ldots + f_{s-1} g_1 + f_s g_0 = \sum_{\substack{k \geqslant 0,\, l \geqslant 0,\\ k+l=s}} f_k g_l. \tag{36.10}$$

В сумме (36.10) число слагаемых равно $s + 1$, однако некоторые из них могут обращаться в нуль: f_k (соответственно g_l) будет наверняка нулевым, если $k > n$ (соответственно $l > m$). Значит, h_s будет наверняка нулевым, если $s > n + m$. В то же время $h_{n+m} \neq 0$, поскольку

$$h_{n+m} = f_n g_m. \tag{36.11}$$

Скаляр (36.11) будет, таким образом, старшим коэффициентом многочлена $h(x)$, и мы приходим ко второму правилу, касающемуся функции deg .

П р а в и л о 2:

$$\deg(f(x) \cdot g(x)) = \deg(f(x)) + \deg(g(x)). \tag{36.12}$$

Замечание 36.6. В дополнение к замечанию 36.4, отметим, что формула (36.12) останется справедливой, если один или оба сомножителя окажутся нулевыми многочленами.

[Убедитесь в этом самостоятельно. Воспользуйтесь тем, что, как принято считать, $a + (-\infty) = -\infty$ для любого (конечного) числа a.]

Итак, теперь в множестве многочленов $P[x]$, наряду со *сложением* [определяемым по формуле (36.7)] и *умножением на скаляр*, задано [с помощью формулы (36.10)] третье алгебраическое действие — *умножение*.

Однако второе действие (умножение на скаляр) здесь сводится к третьему действию: умножить многочлен $f(x)$ на *скаляр* $c \in P$ — это то же самое, что умножить этот многочлен на *скалярный многочлен* $g(x) = c$. (Аналогичным образом дело обстояло, например, при изучении алгебры квадратных матриц; см. п. 14.1.)

Далее нам предстоит доказать, что два основных алгебраических действия в $P[x]$ (сложение и умножение) удовлетворяют аксиомам $\boxed{1} - \boxed{8}$.

Теорема 36.1. *Множество $P[x]$ многочленов над полем P с алгебраическими действиями сложения и умножения, заданными формулами (36.7) и (36.10), является коммутативным кольцом.*

Доказательство. Первые четыре аксиомы коммутативного кольца относятся к сложению и совпадают с первыми четырьмя аксиомами для линейных пространств. То, что $P[x]$ является линейным пространством, доказано в предыдущем пункте, поэтому эти аксиомы можно считать выполненными.

Из четырех оставшихся аксиом две совершенно очевидны: $\boxed{7}$ и $\boxed{8}$. Седьмая аксиома (коммутативность умножения) вытекает из *симметрии* формулы (36.10): если переставить сомножители $f(x)$ и $g(x)$, то это приведет лишь к перестановке слагаемых в сумме в правой части этой формулы, а также к перестановке сомножителей в этих слагаемых. В силу аксиом поля, выполняющихся в P, это никак не отразится на результате.

Восьмая аксиома утверждает существование единичного элемента. Роль такого элемента в $P[x]$ играет, очевидно, скалярный многочлен, равный 1.

Остается проверить две аксиомы. Одну из них, пятую, т. е. дистрибутивный закон, мы оставим в качестве упражнения для вдумчи-

вых читателей. Докажем аксиому $\boxed{6}$ (ассоциативность умножения):

$$(f(x)\cdot g(x))\cdot h(x) = f(x)\cdot(g(x)\cdot h(x)) \tag{36.13}$$

для любых трех многочленов $f(x), g(x), h(x) \in P[x]$.

Подсчитаем коэффициент с номером u ($u \geqslant 0$) для левой части формулы (36.13) (для краткости переменную будем всюду опускать):

$$[(f\cdot g)\cdot h]_u \overset{(36.10)}{=\!=}$$

$$= \sum_{\substack{v,s\geqslant 0\\ v+s=u}} (f\cdot g)_v \cdot h_s \overset{(36.10)}{=\!=} \sum_{\substack{v,s\geqslant 0\\ v+s=u}} \left(\left(\sum_{\substack{k,l\geqslant 0\\ k+l=v}} f_k\cdot g_l\right)\cdot h_s\right) =$$

$$= \sum_{\substack{v,s\geqslant 0\\ v+s=u}} \sum_{\substack{k,l\geqslant 0\\ k+l=v}} ((f_k\cdot g_l)\cdot h_s) = \sum_{\substack{k,l,s\geqslant 0,\\ k+l+s=u}} ((f_k\cdot g_l)\cdot h_s),$$

где на предпоследнем шаге мы внесли под знак внутренней суммы выражение, не зависящее от индексов суммирования в этой сумме, а на последнем шаге мы заменили "двойную" сумму "одинарной" (но имеющей уже три индекса суммирования). (Может быть, вам стоит освежить в памяти приемы обращения со знаками суммирования? См. п. 2.3.)

Теперь повторим все вычисления с правой частью (36.13):

$$[f\cdot(g\cdot h)]_u \overset{(36.10)}{=\!=}$$

$$= \sum_{\substack{k,t\geqslant 0\\ k+t=u}} f_k\cdot(g\cdot h)_t \overset{(36.10)}{=\!=} \sum_{\substack{k,t\geqslant 0\\ k+t=u}} \left(f_k\cdot\left(\sum_{\substack{l,s\geqslant 0\\ l+s=t}} g_l\cdot h_s\right)\right) =$$

$$= \sum_{\substack{k,t\geqslant 0\\ k+t=u}} \sum_{\substack{l,s\geqslant 0\\ l+s=t}} (f_k\cdot(g_l\cdot h_s)) = \sum_{\substack{k,l,s\geqslant 0,\\ k+l+s=u}} (f_k\cdot(g_l\cdot h_s)).$$

В поле P ассоциативность уже имеется, поэтому (для любых номеров k, l, s) справедливо равенство $(f_k\cdot g_l)\cdot h_s = f_k\cdot(g_l\cdot h_s)$. Значит, мы получили одинаковые результаты при вычислении левой и правой частей. Равенство (36.13) доказано.

Доказательство теоремы будет завершено, если вы не поленитесь и разберетесь с дистрибутивностью. □

Замечание 36.7 (продолжение замечания 36.5). Многочлен x вместе со всевозможными скалярами из поля P *порождает* $P[x]$ *как кольцо*. Это означает, что всякий многочлен $f(x) \in P[x]$ может быть получен из многочлена x и констант с помощью алгебраических действий сложения и умножения.

Пример 36.3. Maple легко выполняет любые действия с многочленами, причем не только над числовыми полями, но и, например, над полями классов вычетов (и вообще над произвольными конечными полями). Приведем (без комментариев) несколько характерных примеров.

Обратите внимание на полезные команды: expand — раскрытие скобок, collect — приведение подобных членов, sort — "сортировка" членов в порядке убывания степеней, а также на версии команд "с большой буквы", например Expand, принятые в *модулярной* арифметике.

```
> f := 2 * x^ 3 + 3 * x^ 2 - 4 * x + 1; g := -3 * x^ 2 + 2 * x + 5;
```

$$f := 2x^3 + 3x^2 - 4x + 1$$

$$g := -3x^2 + 2x + 5$$

```
> f * g; expand( % );
```

$$(2x^3 + 3x^2 - 4x)\,(-3x^2 + 2x + 5)$$

$$-6x^5 - 5x^4 + 28x^3 + 4x^2 - 18x + 5$$

```
> f := (1+3*I)*z^ 3 - 5*z^ 2 +(1+2*I);
  g := -(2+I)*z^ 2 + (1-I)*z + I;
```

$$f := (1 + 3I)z^3 - 5z^2 + 1 + 2I$$

$$g := -(2 + I)z^2 + (1 - I)z + I$$

```
> collect (f*g, z); sort( % );
```

$$-2 + I + (1 - 7I)z^5 + (14 + 7I)z^4 - (8 - 6I)z^3 - 10Iz^2 + (3 + I)z$$

$$(1 - 7I)z^5 + (14 + 7I)z^4 - (8 - 6I)z^3 - 10Iz^2 + (3 + I)z - 2 + I$$

```
> f := 4*x^ 2 + x + 3; g := x^ 2+2*x+2;
```

$$f := 4x^2 + x + 3$$

$$g := x^2 + 2x + 2$$

```
> f + g mod 5; f − g mod 5; f ∗ g mod 5;
  Expand( % ) mod 5; sort( % );
```

$$3x$$

$$3x^2 + 4x + 1$$

$$(4x^2 + x + 3)\,(x^2 + 2x + 2)$$

$$1 + 4x^4 + 4x^3 + 3x^2 + 3x$$

$$4x^4 + 4x^3 + 3x^2 + 3x + 1$$

Познакомимся также с несколькими служебными командами (вычисление степени многочлена и его коэффициентов):

```
> f := 2 ∗ x^ 3 + 3 ∗ x^ 2 − 4 ∗ x + 1:
> degree(f), [coeffs(f)], coeff(f, x, 1), coeff(f, x^ 4);
```

$$3, [1, -4, 3, 2], -4, 0$$

Замечание 36.8. Практически никакого труда не составляет обобщить результат теоремы 36.1 на случай многочленов над коммутативным кольцом (см. п. 36.1): если L — коммутативное кольцо, то множество $K = L[x]$ многочленов над L также является коммутативным кольцом.

Самым интересным применением понятия многочленов над кольцом является определение многочлена от двух (и более) переменных следующим образом: $f(x, y) \in P[x, y]$ понимается как многочлен от y с коэффициентами, принадлежащими $P[x]$; например:

$$4x^3y^4 - x^2y^3 + 2y^3 + xy^2 - 3xy + 5x^2 - 7y^2 + x - 3y + 2 =$$
$$= (4x^3)y^4 + (-x^2 + 2)y^3 + (x - 7)y^2 + (-3x - 3)y + (5x^2 + x + 2).$$

Иначе говоря, кольцо многочленов от двух (трех и т. д.) переменных может быть описано конструкцией:

$$P[x, y] = P[x][y];\ \ P[x, y, z] = P[x, y][z];\ \ ...$$

36.4. Группа обратимых элементов кольца многочленов. В п. 14.1 для любого кольца K определялись *обратимые* элементы (элемент $a \in K$ назывался обратимым, если для него существует такой элемент $b \in K$, что $a \cdot b = b \cdot a = 1$).

Легко доказать (в частном случае кольца квадратных матриц это сделано в предложении 14.1), что множество K^* всех обратимых элементов кольца K оказывается *группой* по умножению.

Если данное кольцо K коммутативно, то и группа K^* будет коммутативной. Например, нам неоднократно встречалась группа $\mathbb{Z}^*$ обратимых целых чисел, состоящая всего из двух элементов: 1 и -1. Другой пример: для поля P группа обратимых элементов состоит из всех элементов, кроме нуля, т. е. $P^* = P \setminus \{0\}$.

Рассмотрим теперь кольцо $K = P[x]$ многочленов над полем P.

Предложение 36.1. *Группа обратимых элементов кольца $P[x]$ многочленов над полем P состоит из всех ненулевых констант, т. е.*

$$(P[x])^* = P^* = P \setminus \{0\}. \tag{36.14}$$

Доказательство. Пусть $f(x)$ — обратимый многочлен. Существует такой многочлен $g(x)$, что

$$f(x)g(x) = 1. \tag{36.15}$$

Ни $f(x)$, ни $g(x)$ не являются нулевыми, и следовательно, для обоих многочленов определена степень (являющаяся неотрицательным целым числом). Степень правой части равна 0. По свойству степени (36.12) имеем:

$$\deg(f(x)) + \deg(g(x)) = 0,$$

откуда с необходимостью следует, что

$$\deg(f(x)) = \deg(g(x)) = 0.$$

Значит, как $f(x)$, так и $g(x)$ являются ненулевыми константами (f_0 и g_0 соответственно; причем $f_0 \cdot g_0 = 1$). □

Замечание 36.9 (продолжение замечания 36.8). Предложение 36.1 очевидным образом переносится на случай многочленов над коммутативным кольцом: многочлен $f(x) \in L[x]$ обратим тогда и только тогда, когда он является обратимой константой, т. е. $f(x) = f_0 \in L^*$.

36.5. Отношение ассоциированности в коммутативном кольце. Ассоциированные (пропорциональные) многочлены. Нормализованные многочлены

Определение 36.3. Два элемента a, b из коммутативного кольца K называются *ассоциированными*, если они *отличаются обратимым множителем*. Этот факт обозначается

$$a \sim b.$$

(Внимание: новый термин! *Ассоциированность* и *ассоциативность* — совершенно разные понятия.)

Отношение $\sim$ является *отношением эквивалентности*.

(Символ $\sim$ относится к числу наиболее "перегруженных" в математическом алфавите: он обозначает самые разнообразные отношения эквивалентности, которыми изобилует математическая наука.)

Нулевой элемент ассоциирован лишь с самим собой. Элементы, ассоциированные с единичным элементом, суть обратимые элементы и только они. Все кольцо K разбивается на попарно не пересекающиеся *классы ассоциированных элементов*.

В кольце $\mathbb{Z}$ два ненулевых элемента a, b ассоциированы тогда и только тогда, когда $b = \pm a$.

В кольце $K = P[x]$, где P — поле, два ненулевых элемента (многочлена) ассоциированы $[f(x) \sim g(x)]$ тогда и только тогда, когда они *пропорциональны*, т. е. отличаются ненулевым постоянным множителем $[g(x) = cf(x);\ c \in P^*]$.

Последняя формулировка практически не меняется с заменой поля коэффициентов P на коммутативное кольцо L; надо только учитывать, что в L могут существовать ненулевые необратимые элементы. В случае многочленов над полем в каждом классе пропорциональных многочленов найдется единственный многочлен со старшим коэффициентом, равным единице. Такие многочлены мы будем называть *нормализованными*.

Замечание 36.10. Последний термин не является общепринятым. Чаще такие многочлены называют *приведенными*. (Вспомните из школьного курса алгебры термин "приведенное квадратное уравнение".)

Однако термин "приведенный многочлен" очень плохо сочетается с похожим термином "приводимый многочлен" (который будет введен позже, он обозначает значительно более сложное и важное понятие). Во избежание путаницы мы остановимся на термине "нормализованный многочлен".

Всякий многочлен над полем можно *нормализовать*, поделив его на старший коэффициент. (Этот прием фактически уже использовался выше при доказательстве ОТА.)

36.6. Целостные коммутативные кольца. Целостность кольца многочленов

Определение 36.4. Коммутативное кольцо K называется *целостным*, если произведение двух любых ненулевых элементов этого кольца снова является ненулевым элементом.

Это равносильно следующей импликации:

$$(\,a, b \in K\,) \wedge (\,a \cdot b = 0\,) \Rightarrow (\,a = 0\,) \vee (\,b = 0\,). \tag{36.16}$$

Еще одно свойство равносильно свойству целостности. Оно называется *свойством сокращения* и формулируется следующим образом:

$$(\,a \cdot c = b \cdot c\,) \wedge (\,c \neq 0\,) \Rightarrow (\,a = b\,). \tag{36.16а}$$

[В самом деле, пусть выполнено (36.16). Равенство $a{\cdot}c = b{\cdot}c$ можно переписать в виде $(a-b)\cdot c = 0$, и тогда $c \neq 0$ с учетом (36.16) влечет $a = b$, и (36.16а) доказано. Обратно еще проще: пусть справедливо (36.16а); тогда из равенства $a \cdot b = 0 (= a \cdot 0)$, в предположении $a \neq 0$, будет следовать $b = 0$; свойство (36.16) доказано.]

Замечание 36.11. Не запутайтесь в словах! Нередко студенты считают, что "целостное кольцо" и "кольцо целых чисел" — это одно и то же. На самом же деле целостное кольцо — это такое коммутативное кольцо K, которое "похоже" на кольцо $\mathbb{Z}$ [в том плане, что в K выполняется свойство (36.16)].

Поле тривиальным образом является целостным кольцом. Нетривиальными примерами целостных колец являются всевозможные *подкольца в полях*. Ниже, в п. 46.4, будет объяснено обратное: всякое целостное кольцо можно *вложить* в некоторое поле.

Предложение 36.2. *Кольцо $P[x]$ многочленов над полем P является целостным.*

Доказательство. Пусть $f(x)$ и $g(x)$ — два ненулевых многочлена. Тогда, по свойству степени (36.12),

$$\deg(f(x)g(x)) = \deg(f(x)) + \deg(g(x)) \geqslant 0,$$

и следовательно, $f(x)g(x) \neq 0$. $\square$

Замечание 36.12. Рассмотрим кольцо многочленов $K = L[x]$ над целостным кольцом L. Для таких многочленов сохраняется свойство степени (36.12). (В самом деле, в доказательстве правила 2 ключевым звеном было наблюдение, состоящее в том, что произведение старших коэффициентов двух многочленов является ненулевым элементом, поскольку таковыми были сами старшие коэффициенты. Это наблюдение сохраняет свою силу и для случая многочленов с коэффициентами из целостного кольца.)

Значит, справедлив следующий вывод: если кольцо L является целостным, то и кольцо $L[x]$ является целостным.

В частности (см. замечание 36.8), целостными будут кольца многочленов от двух (трех и более) переменных (подробности — ниже).

Может быть, вам хочется увидеть пример *нецелостного* коммутативного кольца?

Рассмотрите кольцо классов вычетов по составному модулю, например, $\mathbb{Z}_6 = \{0, 1, 2, 3, 4, 5\}$. В этом кольце будет $2 \cdot 3 = 0$ (умножение производится по модулю 6).

Замечание 36.13. Мы уже говорили, что, наряду с записью многочленов $f(x) \in P[x]$ *в порядке возрастания степеней* [см. формулу (36.1)]:

$$f(x) = f_0 + f_1 x + ... + f_{n-1} x^{n-1} + f_n x^n; \\ f_k \in P \ (k = 0, 1, ..., n); \ f_n \neq 0, \quad (36.17)$$

применяется также запись *в порядке убывания степеней* [см. формулу (34.2)]:

$$f(x) = a_0 x^n + a_1 x^{n-1} + ... + a_{n-1} x + a_n; \\ a_k \in P \ (k = 0, 1, ..., n); \ a_0 \neq 0. \quad (36.18)$$

Связь между этими двумя формами записи осуществляется соотношениями: $a_k = f_{n-k}$ $(k = 0, 1, ..., n)$.

Первый стиль записи мы использовали в настоящем пункте при рассмотрении теоретических вопросов. В практических приложениях значительно удобнее применять второй стиль.

§ 37. Деление с остатком и отношение делимости в кольце многочленов (над полем)

37.1. Деление с остатком в кольце многочленов. В предисловии к настоящему пособию была дана краткая сводка информации о целочисленной арифметике. В частности, приводились сведения о делении с остатком целого числа на натуральное. Эту хорошо знакомую (еще со школы) операцию можно модифицировать так, чтобы было осуществимым деление с остатком любого целого числа на любое целое ненулевое.

А именно: для любых $a, b \in \mathbb{Z}$ $(a \neq 0)$ найдутся такие $q, r \in \mathbb{Z}$, что

$$(\, b = aq + r\,) \ \wedge \ (\, |r| < |a|\,). \tag{37.1}$$

[Подумайте, как свести приведенную здесь версию деления с остатком к более привычной версии, изложенной в предисловии. Еще один вопрос: однозначно ли определены числа q и r в (37.1)?]

Подчеркнем то обстоятельство, что в кольце $\mathbb{Z}$ за величиной остатка следит функция "модуль" $(x \mapsto |x|)$, действующая из $\mathbb{Z}$ в множество неотрицательных целых чисел.

Теперь обратимся к кольцу $P[x]$ многочленов над полем P. Сейчас мы докажем, что в этом кольце также можно "организовать" деление с остатком, причем роль следящей функции будет играть функция "степень" (36.5) $[\, f(x) \mapsto \deg(f(x))\,]$, определенная на множестве всех ненулевых многочленов и принимающая значения в множестве неотрицательных целых чисел.

Теорема 37.1. *Пусть P — поле; $f(x), g(x)$ — два многочлена с коэффициентами из P $[f(x) \neq 0]$. Тогда существуют (однозначно определенные) многочлены $q(x), h(x) \in P[x]$, такие, что*

$$g(x) = f(x)q(x) + h(x); \tag{37.2}$$

и

$$[\, h(x) = 0\,] \ \vee \ [\, \deg(h(x)) < \deg(f(x))\,]. \tag{37.3}$$

Доказательство. 1. Докажем сначала *возможность* деления с остатком.

Пусть многочлен $f(x)$ представлен (в порядке убывания степеней) формулой (36.18), а многочлен $g(x)$ задан аналогичной формулой

$$g(x) = b_0x^m + b_1x^{m-1} + ... + b_{m-1}x + b_m; \qquad b_l \in P \ (l = 0, 1, ..., m); \ b_0 \neq 0. \quad (37.4)$$

В частности, $\deg(f(x)) = n$, и если $g(x) \neq 0$, то $\deg(g(x)) = m \geqslant 0$.

Если $g(x) = 0$ или $m < n$, то формулы (37.2), (37.3) удовлетворяются, если в них положить неполное частное $q(x)$ равным нулю, а остаток $h(x)$ равным $g(x)$.

Если же $m \geqslant n$, то неполное частное будет многочленом степени $m - n$ и его старший член может быть задан формулой

$$q_1(x) = b_0a_0^{-1}x^{m-n}. \quad (37.5)$$

В самом деле, запись (37.5) корректна, поскольку старший коэффициент a_0 многочлена $f(x)$ отличен от нуля и, следовательно, обратим; кроме того, выражение

$$\begin{aligned} h_1(x) =& g(x) - f(x)q_1(x) = \\ =& b_0x^m + b_1x^{m-1} + ... + b_{m-1}x + b_m - \\ & - b_0a_0^{-1}x^{m-n}\left(a_0x^n + a_1x^{n-1} + ... + a_{n-1}x + a_n\right) = \\ =& (b_1 - b_0a_0^{-1}a_1)x^{m-1} + (...) \end{aligned}$$

является многочленом степени не выше $m - 1$ [старшие члены уничтожились; символом $(...)$ обозначен многочлен степени не выше $m - 2$; не исключается случай $h_1(x) = 0$].

Получено, таким образом, представление

$$g(x) = f(x)q_1(x) + h_1(x), \quad (37.6)$$

где либо $h_1(x) = 0$, либо $\deg(h_1(x)) < n$, либо $n \leqslant \deg(h_1(x)) < m$.

В первых двух случаях искомое представление (37.2) достигнуто [искомый остаток $h(x) = h_1(x)$]. В третьем случае деление надо продолжать, применяя к многочлену $h_1(x)$ тот же прием, что выше был применен к $g(x)$:

$$h_1(x) = f(x)q_2(x) + h_2(x), \quad (37.7)$$

где многочлен $h_2(x)$ либо нулевой, либо его степень меньше степени $h_1(x)$. Из (37.6) и (37.7) следует представление

$$g(x) = f(x)\,(q_1(x) + q_2(x)) + h_2(x).$$

Наметившийся процесс приводит к строгому убыванию степеней "промежуточных" остатков вплоть до того момента, когда на некотором шаге (с номером $s \leqslant m-n+1$) очередной остаток $h_s(x)$ станет нулевым или же его степень станет меньше n.

Окончательный результат будет иметь вид

$$g(x) = f(x)\,(q_1(x) + q_2(x) + ... + q_s(x)) + h_s(x),$$

что совпадает с (37.2), если положить $q(x) = q_1(x) + ... + q_s(x)$ и $h(x) = h_s(x)$.

2. Докажем теперь единственность определения неполного частного $q(x)$ и остатка $h(x)$ в представлении (37.2).

Пусть наряду с (37.2) имеется представление

$$g(x) = f(x)\tilde{q}(x) + \tilde{h}(x), \tag{37.2a}$$

где многочлен $\tilde{h}(x)$ либо нулевой, либо его степень меньше n.

Вычитая последнюю формулу из формулы (37.2) и перенося остатки в одну сторону равенства, получим

$$f(x)\,(q(x) - \tilde{q}(x)) = \tilde{h}(x) - h(x). \tag{37.8}$$

Докажем сначала равенство многочленов $q(x) = \tilde{q}(x)$. Если предположить противное, то многочлен $q(x) - \tilde{q}(x)$ окажется ненулевым, и следовательно, по свойству степени (36.12), левая часть равенства (37.8) будет многочленом степени не ниже n. В то же время оба остатка имеют степень меньшую n. Значит, по свойству (36.8), степень правой части также меньше n. Получается противоречие. Тем самым равенство неполных частных доказано. Равенство остатков $h(x) = \tilde{h}(x)$ теперь немедленно следует из (37.8). □

Пример 37.1. Рассмотрим многочлены $g(x) = 6x^4 - 2x^3 - 3x + 2$ и $f(x) = 2x^2 - x - 3$. (Их можно рассматривать над полем рациональных чисел $\mathbb{Q}$ или над любым более широким полем, например над $\mathbb{R}$.)

На практике деление многочленов с остатком оформляется, как в школе, в столбик. (Из соображений удобства набора мы перенесли пример такого оформления в приложение 2; см. табл. 37.1а.)

Вычисления дают неполное частное $q(x) = 3x^2 + \frac{1}{2}x + \frac{19}{4}$ и остаток $h(x) = \frac{13}{4}x + \frac{65}{4}$.

Замечание 37.1. Различие между ассоциированными (пропорциональными; см. замечание 36.12) многочленами зачастую является несущественным. (Так, например, пропорциональные многочлены имеют одинаковые *корни*; подробнее о корнях многочленов см. ниже.) Далее, из двух результатов деления с остатком, безусловно, бóльшую ценность имеет именно остаток, причем иногда достаточно бывает знать его лишь с точностью до пропорциональности.

Если делимое [многочлен $g(x)$] заменить на пропорциональный многочлен, сохраняя неизменным делитель $h(x)$, то и результаты деления с остатком [неполное частное $q(x)$ и остаток $h(x)$] заменятся на пропорциональные.

Более того, если на любом из шагов процесса деления (описанного в ходе доказательства теоремы 37.1) заменить получаемый на этом шаге "промежуточный" остаток на пропорциональный, то и "окончательный" остаток окажется пропорциональным своему истинному значению (хотя неполное частное может сильно исказиться).

Это обстоятельство можно использовать (при ручном счете) при делении с остатком одного многочлена с целыми коэффициентами на другой. Всегда можно добиться того, чтобы остаток также имел целые коэффициенты.

(В табл. 37.1б приведен второй вариант решения примера 37.1. Вычисления дают: $h(x) \sim 13x + 65$. Неполное частное оказалось существенно искажено.)

Пример 37.2. Попробуйте самостоятельно разобраться в синтаксисе Maple-функции rem. Ниже мы приводим решение (с помощью этой команды) примера 37.1.

```
> g := 6*x^4 - 2*x^3 - 3*x + 2:
> f := 2*x^2 - x - 3:
> h: = rem(g, f, x, 'q'); q;
```

$$h := \frac{65}{4} + \frac{13x}{4}$$

$$3x^2 + \frac{1}{2}x + \frac{19}{4}x$$

Есть еще команда quo, возвращающая в первую очередь неполное частное. Предусмотрены "модулярные версии" этих команд. (Для аналогичных операций в кольце $\mathbb{Z}$ действуют "i-версии": irem, iquo.)

Замечание 37.2. Можно ли определить деление с остатком в кольце многочленов $K = L[x]$ над коммутативным кольцом L?

Да, если делитель $f(x)$ имеет обратимый старший коэффициент ($a_0 \in L^*$). [Действительно, каждый шаг алгоритма, описанного в доказательстве теоремы 37.1, опирается лишь на существование элемента a_0^{-1}.] В частности, многочлены над кольцом $\mathbb{Z}$ можно поделить с остатком один на другой в том лишь случае, когда старшим коэффициентом делителя будет 1 или -1.

37.2. Отношение делимости в целостном кольце. Делимость (нацело) для многочленов. Для элементов a, b в произвольном коммутативном кольце K можно определить *отношение делимости* (*нацело*) $a|b$, равносильное существованию такого элемента $c \in K$, что $b = a \cdot c$.

В частности, утверждение $0|0$ является истинным (почему?), а при $b \neq 0$ утверждение $0|b$ ложно. Единичный элемент делит любой элемент. Если какой-то элемент делит любой элемент, то, в частности, он делит 1 и, следовательно, обратим. (Почему?)

Если кольцо K является целостным (см. п. 36.6), то при $a \neq 0$ факт делимости $a|b$ влечет *единственность* элемента c, фигурирующего в определении делимости. (Почему?)

В кольцах, "снабженных" операцией деления с остатком (таких, например, как кольцо целых чисел или кольцо многочленов над полем), факт делимости $a|b$, вообще говоря, не равносилен обращению в нуль остатка, получающегося при делении b на a (дело в том, что неполное частное и остаток могут определяться не однозначно).

Ниже формулируется предложение, в котором собраны основные свойства отношения делимости в целостном кольце.

Предложение 37.1. *Пусть K — целостное кольцо. Справедливы следующие свойства отношения делимости (для любых элементов $a, b, c \in K$):*

1) $a|a$ *(рефлексивность);*

2) $(a|b) \wedge (b|c) \Rightarrow (a|c)$ *(транзитивность);*

3) $(a|b) \wedge (b|a) \Leftrightarrow (a \sim b)$;

4a) $1|a$;

4b) $(a|1) \Leftrightarrow (a \sim 1) \Leftrightarrow (a \in K^*)$;

5a) $a|0$;

5b) $(0|a) \Leftrightarrow (a = 0)$;

6) $(a|b) \wedge (a' \sim a) \wedge (b' \sim b) \Rightarrow (a'|b')$;

7) $(a|b) \wedge (a|c) \Rightarrow (a \mid b + c,\ b - c,\ b \cdot c)$;

8a) $(a|b) \Rightarrow (a\cdot c\,|\,b\cdot c)$;
8b) $(a\cdot c\,|\,b\cdot c)\wedge(c\neq 0)\Rightarrow(a|b)$.

Доказательство. Некоторые из сформулированных свойств совершенно очевидны, некоторые — уже объяснены выше. Докажем свойства 3, 6 и 8b. Рассматривайте восстановление остальных доказательств как полезное и несложное упражнение.

Свойство 3. Если $a\sim b$, то существует обратимый элемент $u\in K$, такой, что $b=au$, откуда, домножая на u^{-1}, получим: $a=bu^{-1}$. Значит, справедливо как $a|b$, так и $b|a$.

Обратно, если $a|b$ и $b|a$, то существуют элементы $c,d\in K$, такие, что $b=ac$ и $a=bd$, откуда следует (пользуемся ассоциативным законом), что $a=a(cd)$ и $b=b(cd)$. Если хотя бы один из элементов a,b отличен от нуля, то, применяя свойство сокращения (36.16a), мы получим $1=cd$ (или $1=dc$). В силу коммутативности кольца, любое из этих равенств равносильно обратимости обоих элементов, c и d, и, следовательно, влечет ассоциированность элементов a и b. Если же $a=b=0$, то доказывать нечего: равенство элементов влечет их ассоциированность.

Заметим, что доказанное только что свойство является несколько непривычным: укоренилось (вынесенное из области *натуральных* чисел) убеждение, что если a делит b и b делит a, то $a=b$. Уже в кольце *целых* чисел это не так! (А как?)

Свойство 6. Имеем $b=ac$, $a'=au$ и $b'=bv$ для некоторых $c\in K$; $u,v\in K^*$. Значит, $b'=(ac)v=(a'u^{-1})cv=a'c'$, где $c'=cu^{-1}v$; делимость $a'|b'$ доказана.

Свойство 8b. Делимость $ac|bc$ влечет существование такого элемента $d\in K$, что $bc=acd$. Предположение $c\neq 0$ с учетом свойства сокращения влечет теперь $b=ad$, и следовательно, $a|b$. □

Замечание 37.3. Свойства, сформулированные и доказанные выше, понадобятся нам прежде всего применительно к кольцу многочленов над полем. Например: два многочлена взаимно делят друг друга тогда и только тогда, когда они ассоциированы (пропорциональны); ненулевые константы и только они делят все многочлены и т. д. В конкретных кольцах имеют место также некоторые специфические свойства отношения делимости. Например, в кольце целых чисел факт $a|b$ (для ненулевых a и b) влечет неравенство $|a|\leqslant|b|$, а в кольце многочленов $P[x]$ делимость $f(x)|g(x)$ [для ненулевых многочленов $f(x)$ и $g(x)$] влечет неравенство для их степеней: $\deg(f(x))\leqslant\deg(g(x))$.

Замечание 37.4. Система Maple располагает функциями-тестами, проверяющими, делится ли один многочлен на другой. Скажем, если в условиях примера 37.2 набрать:

> divide(g,f);

то система выдаст:

$$false$$

Значение 'true' получилось бы, если бы выполнялось $f|g$.

Посмотрим на "модулярную" версию команды с возвращением (в случае делимости нацело) частного q:

> Divide(x^ 3 + x^ 2 + 2*x + 3, x + 2, 'q') mod 5;

$$true$$

> q;

$$x^2 + 4x + 4$$

(Проверьте вычисления, помножив частное на делитель по модулю 5.)

§ 38. Алгоритм Евклида отыскания наибольшего общего делителя в кольце многочленов над полем

38.1. Понятие наибольшего общего делителя в целостном кольце. Условие Безу. Пусть K — целостное кольцо; a, b — два произвольных элемента этого кольца.

Определение 38.1. Наибольшим общим делителем (НОД) элементов a и b называется такой их общий делитель, который делится на любой общий делитель.

(Для случая кольца $\mathbb{Z}$ данное выше определение не совсем совпадает с приведенным в предисловии: там требовалось, чтобы НОД был натуральным числом. Если же отказаться от этого предположения, то окажется, что для двух целых чисел имеется не одно значение НОД, а два. Например, как 3, так и -3 являются значениями НОД для чисел 12 и 15. Именно такой подход оказывается правильным с точки зрения общей алгебры.)

Хотелось бы, чтобы вы осознали обманчивость слова *наибольший* (являющегося первой частью термина НОД): это слово употребляется не в смысле какой-либо "величины измерения" для элементов кольца K (такой величины в общем случае нет), а *в смысле отношения делимости*. Общий делитель является наибольшим общим делителем, если он *делится* на все общие делители.

Символом НОД(a, b) будем обозначать множество всех наибольших общих делителей для элементов $a, b \in K$. По определению,

$$d \in \text{НОД}(a, b) \tag{38.1}$$

равносильно выполнению двух условий:

$$1)\ (\,d|a\,) \wedge (\,d|b\,); \tag{38.2}$$

$$2)\ (\,d^\circ|a\,) \wedge (d^\circ|b\,) \Rightarrow (\,d^\circ|d\,). \tag{38.3}$$

Если $a|b$, то, очевидно, $a \in \text{НОД}(a, b)$, в частности, $a \in \text{НОД}(a, 0)$. Множество НОД$(0, 0)$ состоит лишь из нулевого элемента.

Из утверждения 6 предложения 37.1 следует, что всякий элемент, ассоциированный с наибольшим общим делителем элементов a и b, сам является таковым. (Докажите.) Верно и обратное: если два элемента, d и d°, оба принадлежат НОД(a, b), то они взаимно делят друг друга, и следовательно, по утверждению 3 указанного предложения, они ассоциированы: $d \sim d^\circ$.

Вывод: если множество НОД(a, b) не пусто, то оно является классом элементов, ассоциированных друг с другом.

К сожалению, далеко не во всяком целостном кольце произвольная пара элементов обладает наибольшим общим делителем. (Студенты-математики в третьем семестре будут проходить общую теорию делимости, где им встретятся примеры колец, в которых не любые два элемента имеют НОД.)

К счастью, в тех кольцах, с которыми мы будем иметь дело в данной главе, множество НОД(a, b) будет всегда непустым. Например, в кольце $K = \mathbb{Z}$: НОД$(12, 15) = \{3, -3\}$; в кольце многочленов $K = \mathbb{Q}[x]$ множество НОД$(x^2 - 1, x^2 + 2x + 1)$ будет состоять из всех многочленов, ассоциированных (пропорциональных) многочлену $x + 1$. (Подробности — ниже, в пп. 38.2 и 38.3.)

Часто в конкретных кольцах некоторым *каноническим* (= узаконенным) способом в каждом классе ассоциированных элементов

можно бывает выбрать "представителя". (Подробнее о таких "выборах" мы поговорим в пп. 44.3 и 45.1.) Например, в кольце целых чисел, в классе $\{a, -a\}$ $(a \neq 0)$, один из элементов будет положительным. Именно этот элемент будет приниматься в качестве *канонического представителя*.

Символом (a, b) будет обозначаться *положительное* (т. е. натуральное, как в предисловии) значение НОД для двух целых чисел a и b [если хотя бы одно из них отлично от нуля; в исключительном случае принимается: $(0, 0) = 0$].

Ниже, в теореме 38.1, мы докажем, что два любых многочлена $f(x), g(x) \in P[x]$ имеют непустое множество НОД$(f(x), g(x))$. В этом множестве однозначно определяется *нормализованный* (см. п. 36.4) представитель $d(x) = (f(x), g(x))$.

А сейчас мы докажем очень простое, но принципиальное

Предложение 38.1. *Пусть K — целостное кольцо; $a, b \in K$. Пусть элемент d является общим делителем элементов a и b, таким, что для него существует линейное представление (с коэффициентами из K) через данные элементы a и b:*

$$d = au + bv;\ u, v \in K. \tag{38.4}$$

Тогда $d \in$ НОД(a, b).

Доказательство. По предположению выполнено условие (38.2). Наличие линейного представления (38.4) с учетом свойств отношения делимости (см. предложение 37.1) позволяет заключить, что условие (38.3) также выполнено. □

После предложения 38.1 становится ясной принципиальная важность соотношения (38.4). В связи с этим дается

Определение 38.2. Будем называть *условием Безу* следующее требование для элементов целостного кольца K.

(Б) Для любых двух элементов $a, b \in K$ существует наибольший общий делитель $d \in$ НОД(a, b), допускающий линейное представление (38.4).

Замечание 38.1. Сообщим (без доказательства), что существуют такие целостные кольца, в которых любые два элемента имеют НОД, однако условие (Б) не выполнено.

Если же оно выполнено, то, поскольку два НОД для одних и тех же элементов ассоциированы, линейное представление будет иметь *любой* НОД (в условии говорилось, что — *некоторый*).

Отметим также некоторые общие свойства наибольших общих делителей.

1. НОД(a, b) не изменится (как множество), если данные элементы a и b заменить на ассоциированные. (Доказательство легко выводится из утверждения 6 предложения 37.1.)

2. Ассоциированность элементов a и b равносильна тому, что каждый из них является наибольшим общим делителем для обоих. (Доказательство очевидно.)

Замечание 38.2. НОД для нескольких элементов можно определить по тому же принципу, что и для двух элементов, а можно — *индуктивно*, последовательно определяя НОД для трех, четырех элементов и т. д.:

$$\text{НОД}(a, b, c) = \text{НОД}(\text{НОД}(a, b), c);$$
$$\text{НОД}(a, b, c, d) = \text{НОД}(\text{НОД}(a, b, c), d); \ldots$$

(Разумеется, надо задуматься и придумать доказательство равносильности этих двух подходов и заодно убедиться в том, что результат не изменится, если элементы $a, b, c, d, \ldots$ произвольным образом переставить.)

38.2. Существование НОД и алгоритм Евклида для его отыскания в кольце целых чисел. Пусть a и b — целые числа. Если хотя бы одно из них равно нулю, то, как объяснялось в предыдущем пункте, множество НОД(a, b) не пусто.

Пусть оба данных числа отличны от нуля. Задействуем "следящую функцию" (модуль; см. п. 37.1). Пусть, для определенности, $|b| \geqslant |a|$.

Поделим с остатком число b на число a:

$$b = aq_1 + r_1.$$

Если остаток $r_1 = 0$ (т. е. $a|b$), то процесс сразу останавливается: мы объявляем $d = a$; число d (как объяснялось в предыдущем пункте) будет наибольшим общим делителем данных чисел.

Если же $r_1 \neq 0$, то с учетом неравенства $|r_1| < |a|$ начнем вторую итерацию: поделим с остатком число a на число r_1:

$$a = r_1 q_2 + r_2.$$

Если $r_2 = 0$, то берем $d = r_1$ и останавливаем процесс.

Если $r_2 \neq 0$, то с учетом $|r_2| < |r_1|$ проводим следующую итерацию:

$$r_1 = r_2 q_3 + r_3;$$

и т. д.

Поскольку модуль каждого последующего (ненулевого) остатка строго меньше модуля предыдущего остатка, то описанный процесс не может бесконечно продолжаться; он завершится на некотором шаге, скажем, с номером k (т. е. r_k будет *последним ненулевым остатком*, следующий остаток r_{k+1} уже будет нулевым).

На выходе алгоритма (его до нашей эры изобрел для нас Евклид) мы получаем ненулевое число $d = r_k$.

В курсе теории чисел доказывается, что

— этот элемент является общим делителем для данных чисел;

— для него существует линейное представление (38.4).

В силу предложения 38.1, отсюда будет уже вытекать, что $d \in \text{НОД}(a, b)$.

Тем самым 1) устанавливается существование НОД для любых двух целых чисел; 2) указывается алгоритм его отыскания.

В предыдущем пункте уже объяснялось, что НОД (если он существует) определен однозначно, с точностью до ассоциированности. В данном случае оказывается, что НОД в $\mathbb{Z}$ определен *с точностью до знака*.

Мы не будем воспроизводить здесь доказательство сформулированных выше результатов. Вместо этого мы докажем в следующем пункте аналогичный факт для кольца многочленов над полем. Усвоив алгоритм Евклида в этом, более сложном случае, вы без труда сможете вернуться к кольцу $\mathbb{Z}$ и самостоятельно восстановить доказательство в более простом случае.

А пока займемся вопросами практического отыскания НОД и линейного представления для него в кольце целых чисел.

Пример 38.1. Ваши счастливые однокурсники-математики уже, с подробностями и упражнениями, изучили алгоритм Евклида в $\mathbb{Z}$ на практических занятиях по теории чисел. Здесь мы ограничимся лишь одним целочисленным примером: найти $\text{НОД}(60, 1008)$. Постарайтесь разобраться в "цепочке столбиков" в табл. 38.1 (прил. 2).

В рассматриваемом примере данные числа положительны. Из замечания 38.1 следует (но и без него понятно), что в случае кольца $\mathbb{Z}$ достаточно уметь находить НОД для положительных чисел. (При

смене знака у одного или у обоих данных чисел положительное значение НОД останется тем же.)

Опишем теперь способ отыскания коэффициентов u и v в линейном представлении (38.4). Имеем систему равенств:

$$\begin{cases} b & = & a \cdot q_1 & + & r_1 & ; \\ a & = & r_1 \cdot q_2 & + & r_2 & ; \\ r_1 & = & r_2 \cdot q_3 & & & , \end{cases}$$

где $d = r_2$.

Выразим из предпоследнего (второго) равенства d через a и r_1, а из первого равенства выразим r_1 через a и b. Второй результат подставим в первый; получим выражение d через a и b:

$$d = a - r_1 \cdot q_2 = a - (b - a \cdot q_1) \cdot q_2 = a(1 + q_1 q_2) + b(-q_2),$$

откуда получаем:

$$u = 1 + q_1 q_2 = 1 + 16 \cdot 1 = 17;\ v = -q_2 = -1.$$

Проверка: $60 \cdot 17 + 1008 \cdot (-1) = 1020 - 1008 = 12$.

В более сложных примерах цепочка вычислений будет длиннее. Если вам захочется привлечь Maple, то можно воспользоваться командой igcd (= 'greatest common divisor of integers') или ее "улучшенным" вариантом igcdex, дающим не только НОД, но и коэффициенты линейного представления:

```
> a := 60; b := 1008; d := igcdex( a, b,'u','v' );
    u; v; evalb( a*u + b*v = d );
```

$$a := 60$$
$$b := 1008$$
$$d := 12$$
$$17$$
$$-1$$
$$true$$

(Вы наверняка обратили внимание на булевозначную функцию evalb, возвращающую 'true' или 'false' в зависимости от того, истинным или ложным утверждением является ее аргумент, в данном случае — равенство.)

38.3. Существование НОД и алгоритм Евклида для его отыскания в кольце многочленов над полем. Опишем теперь алгоритм отыскания НОД применительно к кольцу $P[x]$ многочленов над полем P, также называемый алгоритмом Евклида. (Хотя, разумеется, древние греки многочленами не занимались, но математики по традиции сохраняют имя первооткрывателя исходной версии теоремы или алгоритма и за последующими обобщениями; см. п. 38.8, посвященный дальнейшему абстрагированию идеи Евклида.)

В работе алгоритма вычисления НОД для многочленов решающую роль снова будет играть алгоритм деления (многочленов) с остатком и "следящая функция" — степень (36.5).

Сразу сформулируем основной результат:

Теорема 38.1. *Для любых двух многочленов $f(x), g(x) \in P[x]$ существует наибольший общий делитель $d(x) \in \text{НОД}(f(x), g(x))$; он определен однозначно (с точностью до пропорциональности), в случае ненулевых $f(x)$ и $g(x)$ может быть получен с помощью алгоритма Евклида (как последний ненулевой остаток в процессе последовательного деления) и может быть представлен в виде линейной комбинации*

$$d(x) = f(x)u(x) + g(x)v(x) \tag{38.5}$$

данных многочленов $f(x)$ и $g(x)$ с коэффициентами-многочленами $u(x), v(x) \in P[x]$.

Доказательство. Рассмотрим многочлены $f(x), g(x) \in P[x]$. Случай, когда хотя бы один из них является нулевым, снова тривиален. [Контрольный вопрос: каким в этом случае будет линейное представление (38.5)?]

Для ненулевых $f(x)$ и $g(x)$ запускаем алгоритм последовательного деления с остатком, описанный в предыдущем пункте для чисел, а теперь применяемый для многочленов. Считаем для определенности, что

$$\deg(f(x)) = n \leqslant m = \deg(g(x)).$$

Не станем повторять начало рассуждения: оно переносится без всяких изменений. Начнем подробное доказательство с уже сфор-

мировавшейся в ходе работы алгоритма Евклида системы равенств

$$\begin{cases} g(x) = f(x)\cdot q_1(x) + h_1(x) ; \\ f(x) = r_1(x)\cdot q_2(x) + h_2(x) ; \\ h_1(x) = h_2(x)\cdot q_3(x) + h_3(x) ; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ h_{k-3}(x) = h_{k-2}(x)\cdot q_{k-1}(x) + h_{k-1}(x) ; \\ h_{k-2}(x) = h_{k-1}(x)\cdot q_k(x) + h_k(x) ; \\ h_{k-1}(x) = h_k(x)\cdot q_{k+1}(x) , \end{cases} \tag{38.6}$$

где

$$\deg(h_1(x)) < n;\ \deg(h_i(x)) < \deg(h_{i-1}(x));\ i = 2, ..., k$$

и следующий остаток уже нулевой: $h_{k+1}(x) = 0$.

Обозначим последний ненулевой остаток

$$d(x) = h_k(x). \tag{38.7}$$

Докажем, что многочлен (38.7) является наибольшим общим делителем для $f(x)$ и $g(x)$. В силу предложения 38.1, для этого достаточно доказать, что

1) $d(x)$ является общим делителем данных многочленов;

2) для $d(x)$ существует линейное представление (38.5).

Первое утверждение устанавливается по ходу просмотра соотношений (38.6), в которых вместо $h_k(x)$ подставлено $d(x)$, снизу вверх. Из нижнего равенства следует, что $d(x)|h_{k-1}(x)$.

Из предпоследнего равенства (и из свойств отношения делимости, см. предложение 37.1) будет теперь следовать, что $d(x)|h_{k-2}(x)$.

Продолжая движение снизу вверх, мы убеждаемся, что на $d(x)$ делятся все предшествующие остатки $h_{k-3}(x), ..., h_1(x)$ и, далее, многочлены $f(x)$ и $g(x)$.

Чтобы доказать второе утверждение, снова приступаем к просмотру формул (38.6), начиная с предпоследней и продвигаясь снизу вверх.

Формула

$$d(x) = 1\cdot h_{k-2}(x) + (-q_k(x))\cdot h_{k-1}(x)$$

является линейным представлением многочлена $d(x)$ через два предыдущих остатка с коэффициентами-многочленами.

Из предшествующей строки выражаем $h_{k-1}(x)$:

$$h_{k-1}(x) = 1\cdot h_{k-3}(x) + (-q_{k-1}(x))\cdot h_{k-2}(x).$$

Последний результат подставляем в выражение для $d(x)$; после перегруппировки получаем формулу

$$d(x) = (1 + q_k(x)q_{k-1}(x)) \cdot h_{k-2}(x) + (-q_k(x)) \cdot h_{k-3}(x),$$

являющуюся линейным представлением $d(x)$ через остатки с номерами $k-2$ и $k-3$. (Нечто подобное, но в другом кольце и в совсем простой ситуации, мы делали в примере 38.1.)

Поднимаясь все выше, вплоть до первой строки, мы в итоге получим линейное выражение (38.5) для $d(x)$ через $f(x)$ и $g(x)$.

Итак, $d(x) \in \text{НОД}(f(x), g(x))$.

В п. 38.1 объяснялось, что все множество $\text{НОД}(f(x), g(x))$ будет исчерпываться многочленами, пропорциональными $d(x)$. (В частности, всегда можно добиться нормализованности НОД.) □

Замечание 38.3. На любом шаге вычисления НОД можно заменять (см. замечание 37.1) получившийся остаток на пропорциональный многочлен. В частности, в качестве ответа всегда можно привести нормализованный НОД.

Пример 38.2. Найдем в кольце $\mathbb{Q}[x]$ НОД многочленов

$$f(x) = x^3 + x^2 - x - 1;\ g(x) = x^4 + x^3 - 3x^2 - 4x - 1.$$

Вычисления организуются точно так же, как и в кольце $\mathbb{Z}$ (в примере 38.1), но с упрощающими поправками, упомянутыми в замечании 38.3 (остатки заменяются на пропорциональные). Решение приведено в табл. 38.2 (прил. 2). Получается (нормализованный) НОД:

$$(f(x), g(x)) = x + 1.$$

Многочлены $u(x)$ и $v(x)$ для представления (38.5) можно было бы найти по тому же принципу, как в примере 38.1, но тогда следовало бы (поскольку нужны "неискаженные" неполные частные) "точно делить" в табл. 38.2, т. е. вести вычисления с дробными коэффициентами без замены остатков на пропорциональные. Ниже будет изложен другой, более интересный и важный прием отыскания этих многочленов – *метод неопределенных коэффициентов.*

38.4. Отыскание линейного представления для НОД в кольце многочленов методом неопределенных коэффициентов. Рассмотрим снова два (непропорциональных) многочлена $f(x)$ и $g(x)$ (степеней n и m соответственно; для определенности считаем

$n \leqslant m$). Согласно теореме 38.1, существует $d(x) \in \text{НОД}(f(x), g(x))$, а также многочлены $u(x)$ и $v(x)$ — коэффициенты линейного представления (38.5). Как станет ясно ниже, последние многочлены определены отнюдь не однозначно, однако, наложив ограничения на их степени, можно добиться однозначности в их выборе.

Первым этапом в исследовании будет следующее

Предложение 38.2. *Пусть многочлены $f(x), g(x) \in P[x]$ (степеней n и m соответственно) не пропорциональны.*

Тогда многочлены $u(x)$ и $v(x)$ в линейном представлении для $d(x) \in \text{НОД}(f(x), g(x))$ всегда можно выбрать так, чтобы

$$\deg(u(x)) \leqslant m^\circ - 1;\ \deg(v(x)) \leqslant n^\circ - 1, \tag{38.8}$$

где

$$m^\circ = m - k;\ n^\circ = n - k;\ k = \deg(d(x)). \tag{38.9}$$

Доказательство. Как сказано выше, перед формулировкой предложения, какие-то многочлены $u(x)$ и $v(x)$ (степеней, скажем, s и t соответственно), обеспечивающие выполнение равенства (38.5), найдутся.

По определению НОД, имеем:

$$f(x) = d(x)f^\circ(x);\ g(x) = d(x)g^\circ(x) \tag{38.10}$$

для некоторых многочленов $f^\circ(x)$ и $g^\circ(x)$ (степеней $n-k$ и $m-k$ соответственно).

По свойствам отношения делимости (см. предложение 37.1), обе части формулы (38.5) делятся на $d(x)$. Выполнив это деление, получим соотношение

$$1 = f^\circ(x)u(x) + g^\circ(x)v(x), \tag{38.11}$$

равносильное (38.5).

Поделим с остатком многочлен $v(x)$ степени t на многочлен $f^\circ(x)$ степени $n-k$;

$$v(x) = f^\circ(x)q(x) + v^\circ(x), \tag{38.12}$$

где остаток $v^\circ(x)$ имеет степень t° меньшую $n-k$ (т. е. $t^\circ \leqslant n-k-1$).

Подставив (38.12) в (38.11) и перегруппировав члены, получим:

$$1 = f^\circ(x)u^\circ(x) + g^\circ(x)v^\circ(x), \tag{38.11°}$$

где $u^\circ(x)$ — многочлен, степень которого мы обозначим s°.

Осталось доказать неравенство $s^\circ < m-k$ (или $s^\circ \leqslant m-k-1$).

Предположим противное, т. е. допустим, что $s^\circ \geqslant m-k$.

По свойствам степени, степень первого слагаемого в правой части формулы (38.11°) будет не меньше $(n-k)+(m-k) = n+m-2k$.

В то же время степень второго слагаемого будет строго меньше этого же числа.

Следовательно, степень всей правой части формулы (38.11°) совпадает со степенью первого слагаемого, т. е. не ниже $n+m-2k$.

Степень левой части равна нулю. Равенство $n+m-2k = 0$ (с учетом $k \leqslant n \leqslant m$) возможно, лишь если $k = n = m$.

Последнее же равенство означает, что вопреки предположению данные многочлены пропорциональны. [В самом деле, $d(x)$ делит $f(x)$ и при этом имеет с ним одинаковые степени; значит, частное имеет степень 0, т. е. является ненулевым скаляром, и поэтому $d(x) \sim f(x)$; аналогично $d(x) \sim g(x)$, и в итоге $f(x) \sim g(x)$.]

Таким образом, возможность представления (38.11°) доказана. Домножая эту формулу на $d(x)$, мы получаем линейное представление

$$d(x) = f(x)u^\circ(x) + g(x)v^\circ(x) \qquad (38.5^\circ)$$

с многочленами-коэффициентами, удовлетворяющими требуемым условиям. □

Замечание 38.4. Несколько позже, в замечании 38.8, после изучения свойств так называемых *взаимно простых* многочленов, мы сможем уточнить утверждение предложения 38.2: будет доказана единственность линейного представления для НОД с установленными в этом предложении ограничениями на степень многочленов $u(x)$ и $v(x)$.

На практике многочлены $u^\circ(x)$ и $v^\circ(x)$ можно искать с помощью метода неопределенных коэффициентов.

Чтобы не усложнять обозначения, опустим кружочки в обозначениях этих многочленов и будем считать, что уже в формуле (38.5) искомые многочлены $u(x)$ и $v(x)$ имеют степени $s \leqslant m-k-1$, $t \leqslant n-k-1$.

Вычисления несколько упростятся, если мы перейдем к равносильной формуле (38.11), заменив многочлены $f(x)$ и $g(x)$ на $f^\circ(x)$ и $g^\circ(x)$.

Рассматриваемые многочлены запишем в общем виде:

$$\begin{aligned} f^\circ(x) &= \alpha_0 x^{n-k} + \alpha_1 x^{n-k-1} + ... + \alpha_{n-k-1}x + \alpha_{n-k};\\ g^\circ(x) &= \beta_0 x^{m-k-1} + \beta_1 x^{m-k-1} + ... + \beta_{m-k-1}x + \beta_{m-k};\\ u(x) &= A_0 x^{m-k-1} + A_1 x^{m-k-2} + ... + A_{m-k-2}x + A_{m-k-1};\\ v(x) &= B_0 x^{n-k-1} + B_1 x^{n-k-2} + ... + B_{n-k-2}x + B_{n-k-1}, \end{aligned}$$

коэффициенты $A_0, A_1, ..., A_{m-k-1}$; $B_0, B_1, ..., B_{n-k-1}$ являются неопределенными (и подлежат определению), причем заранее не известно, равны ли на самом деле степени искомых многочленов числам s и t соответственно или окажутся меньше (т. е. коэффициенты A_0 и B_0 не предполагаются ненулевыми). Подставим эти выражения в формулу (38.11) и раскроем скобки, а затем, воспользовавшись определением равенства многочленов, приравняем коэффициенты при одинаковых степенях x в левой части (там они все нулевые, кроме свободного члена, равного 1) и в правой части.

Получим $n^\circ + m^\circ = n + m - 2k$ линейных уравнений относительно такого же количества неизвестных (неопределенных) коэффициентов.

Для ясности ниже мы проделаем все вышеописанное для случая $n^\circ = 3$, $m^\circ = 4$. Равенство (38.11) примет вид

$$\begin{aligned} 1 =&(\alpha_0 x^3 + \alpha_1 x^2 + \alpha_3 x + \alpha_4)(A_0 x^3 + A_1 x^2 + A_2 x + A_3) +\\ &+(\beta_0 x^4 + \beta_1 x^3 + \beta_2 x^2 + \beta_3 x + \beta_4)(B_0 x^2 + B_1 x + B_2). \end{aligned}$$

Раскрываем скобки, приводим подобные члены, приравниваем коэффициенты при x^6, ..., x^0:

$$\begin{array}{lccccccc} 0 = & \alpha_0 A_0 & & & & +\beta_0 B_0 & & \\ 0 = & \alpha_1 A_0 & +\alpha_0 A_1 & & & +\beta_1 B_0 & +\beta_0 B_1 & \\ 0 = & \alpha_2 A_0 & +\alpha_1 A_1 & +\alpha_0 A_2 & & +\beta_2 B_0 & +\beta_1 B_1 & +\beta_0 B_2 \\ 0 = & \alpha_3 A_0 & +\alpha_2 A_1 & +\alpha_1 A_2 & +\alpha_0 A_3 & +\beta_3 B_0 & +\beta_2 B_1 & +\beta_1 B_2 \\ 0 = & & \alpha_3 A_1 & +\alpha_2 A_2 & +\alpha_1 A_3 & +\beta_4 B_0 & +\beta_3 B_1 & +\beta_2 B_2 \\ 0 = & & & \alpha_3 A_2 & +\alpha_2 A_3 & & +\beta_4 B_1 & +\beta_3 B_2 \\ 1 = & & & & \alpha_3 A_3 & & & +\beta_4 B_2 \end{array}$$

Получилась система из семи линейных уравнений с семью неизвестными. Записываем ее в матричном виде:

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_0 & 0 & 0 & 0 & \beta_0 & 0 & 0 \\ \alpha_1 & \alpha_0 & 0 & 0 & \beta_1 & \beta_0 & 0 \\ \alpha_2 & \alpha_1 & \alpha_0 & 0 & \beta_2 & \beta_1 & \beta_0 \\ \alpha_3 & \alpha_2 & \alpha_1 & \alpha_0 & \beta_3 & \beta_2 & \beta_1 \\ 0 & \alpha_3 & \alpha_2 & \alpha_1 & \beta_4 & \beta_3 & \beta_2 \\ 0 & 0 & \alpha_3 & \alpha_2 & 0 & \beta_4 & \beta_3 \\ 0 & 0 & 0 & \alpha_3 & 0 & 0 & \beta_4 \end{pmatrix} \cdot \begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \\ B_0 \\ B_1 \\ B_2 \end{pmatrix}.$$

Замечание 38.5. Матрица системы (обозначим ее M) имеет очень любопытный вид. Проследите, как располагаются и сколько раз повторяются в матрице M коэффициенты многочленов $f^\circ(x)$ и $g^\circ(x)$.

Матрицы такого вида в дальнейшем еще встретятся нам. Определитель $\det(M)$ называется *результантом* рассматриваемых многочленов. Можно доказать, что результант отличен от нуля тогда и только тогда, когда многочлены *взаимно просты* (т. е. 1 является их НОД; подробности см. ниже, в замечании 38.8, а также в учебнике [9, гл. 11, § 3]).

Пример 38.3 (продолжение примера 38.2). Действуем по разобранной выше схеме. Сначала следует поделить данные многочлены на их наибольший делитель. (Внимание: делать это надо по схеме, показанной в табл. 37.1а, т. е. *точно*, без искажения частного, а остаток здесь равен нулю.)

Получим:

$$f^\circ(x) = x^2 - 1;\ g^\circ(x) = x^3 - 3x - 1.$$

Далее:

$$1 = (x^2 - 1)(Ax^2 + Bx + C) + (x^3 - 3x - 1)(Dx + E).$$

(В простых задачах удобнее неопределенные коэффициенты обозначать *различными* буквами.)

Раскрываем скобки:

$$\begin{aligned} 1 = \quad & Ax^4 + Bx^3 + Cx^2 \\ & \qquad\quad - Ax^2 - Bx - C \\ & + Dx^4 + Ex^3 \\ & \qquad\quad - 3Dx^2 - 3Ex \\ & \qquad\quad\quad\ - Dx - E\,. \end{aligned}$$

Система уравнений:

$$\begin{array}{lrrrrrrl}
(x^4): & 0 = & A & & & +D & & ; \\
(x^3): & 0 = & & B & & & +E & ; \\
(x^2): & 0 = & -A & & +C & -3D & & ; \\
(x^1): & 0 = & & -B & & -D & -3E & ; \\
(x^0): & 1 = & & & -C & & -E & .
\end{array}$$

Решать эту систему можно "по науке" (с использованием метода Жордана — Гаусса) или "на глаз", исключая легко исключаемые неизвестные и понижая тем самым количество остающихся неопределенными коэффициентов. В любом случае легко находится (единственное) решение:

$$A = \frac{2}{3};\ B = -\frac{1}{3};\ C = -\frac{4}{3};\ D = -\frac{2}{3};\ E = \frac{1}{3},$$

или

$$u(x) = \frac{1}{3}(2x^2 - x - 4);\ v(x) = \frac{1}{3}(-2x + 1).$$

Пример 38.4. Maple-команда для отыскания НОД двух многочленов отличается от аналогичной "целочисленной" команды отсутствием префикса 'i' (= integer):

```
> f := x^3 + x^2 - x - 1; g := x^4 + x^3 - 3*x^2 - 4*x - 1;
```

$$f := x^3 + x^2 - x - 1$$
$$g := x^4 + x^3 - 3x^2 - 4x - 1$$

```
> d := gcd(f, g);
```

$$d := 1 + x$$

Предусмотрена версия этой команды с четырьмя аргументами, вместо двух; она, помимо НОД, возвращает (по запросу) частные от деления данных многочленов на найденный НОД:

```
> gcd(f, g, 'f0', 'g0'): f0; g0;
```

$$x^2 - 1$$
$$x^3 - 3x - 1$$

Другая команда, gcdex (с пятью аргументами), требует указания переменной в рассматриваемых многочленах, а также резервирует

имена и (по запросу) выдает многочлены, являющиеся коэффициентами линейного представления НОД:

```
> d := gcdex(f, g, x, 'u', 'v'): u; v;
```

$$-\frac{4}{3}+\frac{2}{3}x^2-\frac{1}{3}x$$

$$\frac{1}{3}-\frac{2x}{3}$$

(Сравните с результатами ручного счета в примере 38.3.)

38.5. Взаимно простые элементы в целостном кольце (с условием Безу). Пусть K — целостное кольцо, в котором выполняется условие (Б) (см. п. 38.1).

Кольца $K=\mathbb{Z}$ и $K=P[x]$, которым мы уделяем в данной главе наибольшее внимание, как показано выше, этому условию удовлетворяют.

Определение 38.3. Элементы a и b кольца K называются *взаимно простыми*, если $1\in\text{НОД}(a,b)$.

Общими делителями двух взаимно простых элементов являются обратимые элементы кольца K и только они. В частности, в кольце $\mathbb{Z}$ числа a и b взаимно просты тогда и только тогда, когда $(a,b)=1$; и аналогично в кольце многочленов $f(x)$ и $g(x)$ взаимно просты тогда и только тогда, когда $(f(x),g(x))=1$. (Для случая многочленов над полем их взаимная простота равносильна тому, что общими делителями для них являются лишь ненулевые константы.)

Предложение 38.3. *Элементы a и b целостного кольца K, удовлетворяющего условию* (Б)*, взаимно просты тогда и только тогда, когда существуют такие $u,v\in K$, что*

$$au+bv=1. \tag{38.13}$$

Доказательство. Если данные элементы a и b взаимно просты, то $1\in\text{НОД}(a,b)$. В силу условия (Б), будет существовать линейное представление (38.13).

Обратно, если имеет место (38.13), то всякий общий делитель элементов a и b будет делителем единицы, т. е. обратимым элементом. Следовательно, единичный элемент будет принадлежать $\text{НОД}(a,b)$. $\square$

Предложение 38.4. *Предположим, кольцо K удовлетворяет условию* (Б). *Пусть элемент $d \in K$ является общим делителем для ненулевых элементов $a, b \in K$ и*

$$a^\circ = \frac{a}{d};\ b^\circ = \frac{b}{d}.$$

Тогда d является наибольшим общим делителем для a и b в том и только том случае, если элементы a° и b° взаимно просты.

Доказательство. Если $d \in \text{НОД}(a, b)$, то, по условию (Б), найдутся такие элементы u и v, что выполняется равенство (38.4). Сокращая это равенство на d, получаем равенство

$$a^\circ u + b^\circ v = 1, \tag{38.14}$$

что, в силу предложения 38.3, влечет взаимную простоту a° и b°.

Обратно, если a° и b° взаимно просты, то, в силу предложения 38.3, найдутся элементы u и v такие, что справедливо (38.14). Домножая это равенство на d, мы приходим к (38.4).

Предполагалось, что d является *общим* делителем данных элементов. Наличие линейного представления (38.4) позволяет, в силу предложения 38.1, утверждать, что d является *наибольшим* общим делителем для a и b. □

Замечание 38.6. С операцией *сокращения на НОД* (применительно к многочленам) мы уже встречались при доказательстве предложения 38.2. [Причем с некоторым опережением были введены "обозначения с кружочками" и получалось равенство типа (38.14).]

Взаимная простота многочленов $f(x), g(x) \in P[x]$ равносильна наличию линейного представления

$$1 = f(x)u(x) + g(x)v(x) \tag{38.11a}$$

с некоторыми $u(x), v(x) \in P[x]$.

Факт $d(x) \in \text{НОД}(f(x), g(x))$ равносилен взаимной простоте частных $f^\circ(x)$ и $g^\circ(x)$ [см. (38.10)], т. е. соотношению (38.12).

Предложение 38.5. *Пусть кольцо K обладает свойством* (Б).

1. Если элемент a делит произведение элементов bc и при этом элементы a и b взаимно просты, то a делит c.

2. Если элемент a взаимно прост с каждым из элементов b и c, то он взаимно прост и с их произведением.

Доказательство. 1. Взаимная простота элементов a и b влечет, в силу предложения 38.3, существование линейного представления (38.13). Домножая это равенство на c, мы получаем: $c = a(uc)+(bc)v$, откуда, очевидно, следует, что $a|c$.

2. Взаимная простота a с элементом b влечет существование представления (38.13), а взаимная простота a и c влечет наличие аналогичного представления $a\tilde{u} + c\tilde{v} = 1$.

Перемножая последнюю формулу с (38.13) и подходящим образом группируя члены, мы получим:

$$a(a\tilde{u}\tilde{v} + cu\tilde{v} + bv\tilde{u}) + bc(v\tilde{v}) = 1.$$

В силу предложения 38.3, последнее равенство влечет взаимную простоту элементов a и bc. □

Замечание 38.7. Индукция (по количеству сомножителей) позволяет легко обобщить предложение 38.5 на случай нескольких сомножителей.

Если каждый из элементов a_i $(i = 1, ..., s)$ кольца K взаимно прост с каждым из элементов b_j $(j = 1, ..., t)$, то произведения $a_1...a_s$ и $b_1...b_t$ также будут взаимно простыми.

38.6. Взаимно простые многочлены. Теперь мы от вопросов делимости в *абстрактных* кольцах снова возвращаемся к рассмотрению *конкретного* кольца $K = P[x]$.

Предложение 38.6. *Многочлен $f(x)$ [степени n] и многочлен $g(x)$ [степени m] над полем P взаимно просты тогда и только тогда, когда не существует ненулевых многочленов $u(x)$ [степени меньшей m] и $v(x)$ [степени меньшей n], таких, что выполняется соотношение*

$$f(x)u(x) + g(x)v(x) = 0. \tag{38.15}$$

Доказательство. 1. Равенство (38.15) влечет, очевидно, следующие факты делимости: $f(x)|g(x)v(x)$ и $g(x)|f(x)u(x)$. Если многочлены $f(x)$ и $g(x)$ являются взаимно простыми, то отсюда, в силу первого утверждения предложения 38.5, следует, что $f(x)|v(x)$ и $g(x)|u(x)$. Последние два факта влекут (см. замечание 37.3) неравенства для степеней $\deg(u(x)) \geqslant m$ и $\deg(v(x)) \geqslant n$, противоречащие условиям предложения.

Значит, таких ненулевых многочленов $u(x)$ и $v(x)$, которые удовлетворяли бы уравнению (38.15), а также указанным в формулировке предложения ограничениям на степени, не существует.

2. Если, напротив, многочлены $f(x)$ и $g(x)$ не взаимно просты, то они имеют отличный от константы наибольший общий делитель $d(x)$. Сократим данные многочлены на НОД, т. е. рассмотрим многочлены $f^\circ(x) = f(x)/d(x)$ и $g^\circ(x) = g(x)/d(x)$.

Положив $u(x) = g^\circ(x)$ и $v(x) = -f^\circ(x)$, мы получим ненулевые многочлены, удовлетворяющие (38.15), причем $\deg(u(x)) < m$ и $\deg(v(x)) < n$, поскольку $\deg(d(x)) \geqslant 1$. □

Замечание 38.8. Теперь мы можем завершить обсуждение вопросов, связанных с методом неопределенных коэффициентов (см. предыдущий пункт и, в частности, замечание 38.4).

Представление (38.5) для $d(x) \in \text{НОД}(f(x), g(x))$, с ограничениями (38.8) на степени многочленов $u(x)$ и $v(x)$, определено однозначно.

В самом деле, пусть, наряду с (38.5), имеется еще одно представление

$$d(x) = f(x)\tilde{u}(x) + g(x)\tilde{v}(x),$$

причем степени как $u(x)$, так и $\tilde{u}(x)$ меньше $m - k$, а степени как $v(x)$, так и $\tilde{v}(x)$ меньше $n - k$, где $n = \deg(f(x))$, $m = \deg(g(x))$, $k = \deg(d(x))$.

Перейдем посредством деления на $d(x)$ в последнем равенстве и равенстве (38.5) к взаимно простым многочленам $f^\circ(x)$ и $g^\circ(x)$, а затем вычтем одно равенство из другого. Получим:

$$0 = f^\circ(x)[u(x) - \tilde{u}(x)] + g^\circ(x)[v(x) - \tilde{v}(x)]. \tag{38.16}$$

Равенство (38.16) имеет тот же вид, что и (38.15), причем коэффициенты при многочленах $f^\circ(x)$ и $g^\circ(x)$ (выражения в двух квадратных скобках) являются многочленами, степени которых меньше $m - k$ (для первого выражения) и меньше $n - k$ (для второго).

Согласно предложению 38.6, в силу взаимной простоты $f^\circ(x)$ и $g^\circ(x)$, оба многочлена в квадратных скобках должны быть нулевыми. Этим доказаны равенства $\tilde{u}(x) = u(x)$ и $\tilde{v}(x) = v(x)$, т. е. совпадение двух рассматриваемых линейных представлений для НОД.

Можно рассуждать иначе, с помощью *результанта*.

В замечании 38.5 вы (правда, лишь в частном случае) видели матрицу M, которая отвечает с.л.у., позволяющей определить неопределенные коэффициенты многочленов $u(x)$ и $v(x)$, задающих линейное

представление НОД и удовлетворяющих ограничениям на степень (38.8). Уравнение (38.16) с неизвестными многочленами $u(x)$ и $v(x)$ является однородным аналогом уравнения (38.11); в его левой части фигурирует та же самая квадратная матрица M порядка $n^\circ + m^\circ$.

Факт наличия у однородной с.л.у. с матрицей M лишь нулевого решения равносилен, как известно (см. предложения 6.1 и 28.2), тому, что $\det(M)$, т. е. результант многочленов $f^\circ(x)$ и $g^\circ(x)$, отличен от нуля.

Попутно мы убеждаемся в справедливости упомянутого в замечании 38.5 утверждения: два многочлена взаимно просты тогда и только тогда, когда их результант не равен нулю.

38.7. Наименьшее общее кратное. Связь НОК и НОД. В этом пункте мы снова предположим, что кольцо K удовлетворяет условию (Б) из пункта 38.1.

Наряду с понятием НОД для двух (и нескольких) элементов в целостном кольце K, можно рассматривать *двойственное* понятие *наименьшего общего кратного* (НОК).

(Запись факта делимости $a|b$ $(a, b \in K)$ может быть прочитана трояко: *a делит b*; *b делится* на a или *b кратно a*. Очень популярный в математике термин *двойственность* в данном случае будет означать, что в определении 38.1 слово "делит" всюду будет заменено словом "делится" или "кратно".)

Определение 38.4. *Наименьшим общим кратным элементов a* и b кольца K называется такое их общее кратное q, которое делит любое общее кратное q°.

Множество всех наименьших кратных для элементов a и b обозначается $\mathrm{НОК}(a, b)$.

Элемент $q \in \mathrm{НОК}(a, b)$ тогда и только тогда, когда выполняются условия:

$$1)\ (\,a|q\,) \wedge (\,b|q\,); \tag{38.17}$$

$$2)\ (\,a|q^\circ\,) \wedge (b|q^\circ\,) \Rightarrow (\,q|q^\circ\,). \tag{38.18}$$

Так же как в случае с НОД, множество $\mathrm{НОК}(a, b)$ может оказаться пустым. Но если оно не пусто, то представляет из себя класс ассоциированных друг с другом элементов. Для любого элемента a, очевидно, $\mathrm{НОК}(a, 0) = \{0\}$.

В конкретных кольцах $K = \mathbb{Z}$ и $K = P[x]$, где P — поле, в каждом из (не сводящихся к нулю) множеств НОК может быть указан

канонический представитель (положительное целое число или нормализованный многочлен соответственно). Это выделенное значение НОК будем обозначать квадратными скобками.

Например,

$$q(x) = [f(x), g(x)] \in \text{НОК}(f(x), g(x)) \tag{38.19}$$

будет нормализованным наименьшим общим кратным для двух многочленов.

В кольце целых чисел наименьшее общее кратное $[a, b]$ всегда существует и (для ненулевых a и b) оно, с точностью до знака, совпадает с частным $ab/(a, b)$.

Аналогичное утверждение (с заменой словосочетания "с точностью до знака" словосочетанием "с точностью до пропорциональности") справедливо для кольца многочленов над полем.

Оба эти утверждения будут частными случаями общего факта, который мы установим в предложении 38.8. Но сначала мы докажем вспомогательное

Предложение 38.7. *Пусть целостное кольцо K удовлетворяет условию* (Б). *Рассмотрим элементы $a, b, c \in K$ ($c \neq 0$). Тогда*

$$1)\ d \in \text{НОД}(a, b) \Leftrightarrow cd \in \text{НОД}(ca, cb); \tag{38.20}$$

$$2)\ q \in \text{НОК}(a, b) \Leftrightarrow cq \in \text{НОК}(ca, cb). \tag{38.21}$$

Доказательство. 1. Пусть $d \in \text{НОД}(a, b)$. Это означает, в силу предположения (Б), что d является общим делителем для a и b, допускающим линейное представление (38.4).

Во-первых, отсюда следует, что cd является общим делителем для ca и cb (здесь мы используем свойство 8a отношения делимости; см. предложение 37.1), а во-вторых, получается линейное представление $cd = (ca)u + (cb)v$, из которого, в силу предложения 38.1, можно заключить: $cd \in \text{НОД}(ca, cb)$.

Обратное рассуждение: факт $cd \in \text{НОД}(ca, cb)$ влечет $cd|ca$, $cd|cb$ и $cd = (ca)u + (cb)v$ для некоторых $u, v \in K$. В силу свойства делимости 8b (с учетом целостности кольца K), отсюда следует: $d|a$, $d|b$ и $d = au + bv$, или $d \in \text{НОД}(a, b)$.

2. Пусть $q \in \text{НО}K(a,b)$. Тогда прежде всего $a|q$ и $b|q$, откуда следует, что $ca|cq$ и $cb|cq$, т. е. cq является общим кратным для элементов ca и cb.

Пусть $\tilde{q}$ также является общим кратным для ca и cb. Это означает, что найдутся $s,t \in K$, такие, что $\tilde{q} = cas$ и $\tilde{q} = cbt$. Следовательно, $cas = cbt$, откуда, в силу целостности кольца, вытекает $as = bt$.

Кроме того, тот факт, что cbt кратно ca, влечет (см. свойство 8b), что bt кратно a. Получается, что элемент bt кратен не только b (в силу своего вида), но и a. Следовательно, он, по определению НОК, кратен q, т. е. найдется $r \in K$, такое, что $bt = rq$.

Отсюда следует, что элемент $\tilde{q} = cbt = crq$ кратен сq. Значит, $cq \in \text{НОК}(ca,cb)$.

Докажем обратное утверждение. Пусть $cq \in \text{НОК}(ca,cb)$, т. е. прежде всего $ca|cq$ и $cb|cq$, откуда (по уже не раз упоминавшемуся свойству 8b) следует, что $a|q$ и $b|q$, т. е. q является общим кратным для a и b.

Пусть q° также является общим кратным для a и b. Это означает, что найдутся $s,t \in K$, такие, что $q^\circ = as$ и $q^\circ = bt$. Но тогда, домножая на c, мы получим $cq^\circ = cas$ и $cq^\circ = cbt$, т. е. cq° является общим кратным для ca и cb. По определению НОК, элемент cq° кратен cq, и следовательно, q° кратен q. Значит, $q \in \text{НОК}(a,b)$. □

Замечание 38.9. Как определение НОК, так и оба утверждения предложения 38.7 переносятся на случай нескольких элементов.

Установим теперь существование НОК и его связь с НОД.

Предложение 38.8. *При условии, что кольцо K удовлетворяет предположению* (Б), *справедливы следующие утверждения:*

1) для взаимно простых элементов a и b кольца K их произведение является для них наименьшим общим кратным:

$$ab \in \text{НОК}(a,b); \tag{38.22}$$

2) для любых ненулевых элементов a и b

$$d \in \text{НОД}(a,b) \Rightarrow q = \frac{ab}{d} \in \text{НОК}(a,b); \tag{38.23}$$

3) для любых a и b множество $\text{НОК}(a,b)$ *не пусто.*

Доказательство. 1. Для любых элементов $a,b \in K$ их произведение ab является их общим кратным. Предположим, что элементы

a и b являются взаимно простыми и докажем, что в этом случае $ab \in \text{НОК}(a, b)$.

Пусть q° — любое общее кратное для a и b, т. е. $q^\circ = as$ и $q^\circ = bt$ для некоторых $s, t \in K$. Следовательно, $as = bt$ и $a|bt$. В силу взаимной простоты a и b, последний факт влечет (см. первое утверждение предложения 38.5) делимость $a|t$, т. е. существование такого элемента $t' \in K$, что $t = at'$. Значит, $q^\circ = bt = (ab)t'$ кратно ab.

2. Докажем импликацию (38.23) с помощью предложения 38.4 (и используемых там обозначений).

Пусть $d \in \text{НОД}(a, b)$ и a°, b° — частные от деления данных элементов на d. По предложению 38.4, элементы a° и b° взаимно просты. По доказанному выше первому пункту настоящего предложения, мы будем иметь:

$$a^\circ\, b^\circ = \frac{ab}{d^2} \in \text{НОК}(a^\circ,\, b^\circ). \tag{38.24}$$

Используя эквивалентность (38.20) (см. предложение 38.7), получим из (38.24):

$$\frac{ab}{d} = d \cdot \frac{ab}{d^2} \in \text{НОК}(da^\circ, db^\circ) = \text{НОК}(a, b),$$

что и требовалось.

3. Третье утверждение немедленно следует из второго и из предположения (Б). □

Пример 38.5. Приведем небольшую Maple-сессию по вычислению НОК в кольцах целых чисел и многочленов. Используемые функции: ilcm и lcm (= least common multiple).

```
> a:=60: b:=1008: d:=igcd(a, b); q:=ilcm(a, b); evalb(q = a*b/d);
```

$$d := 12$$

$$q := 5040$$

$$true$$

```
> f := x^3 + x^2 - x - 1: g := x^4 + x^3 - 3*x^2 - 4*x - 1:
  h := gcd(f, g); q := expand( lcm(f, g) );
  evalb( q = normal( expand( f*g )/h ) );
```

$$h := x + 1$$

$$q := x^6 + x^5 - 4x^4 - 5x^3 + 2x^2 + 4x + 1$$

$$true$$

Обратите внимание на то, что в случае многочленов понадобились команды раскрытия скобок и нормализации (сокращения) дроби.

38.8. Понятие о евклидовых кольцах. Доказательство того факта, что кольцо $P[x]$ многочленов над полем P удовлетворяет условию Безу, и аналогичное доказательство для кольца целых чисел основывались на применении *алгоритма Евклида*, работа которого, в свою очередь, опирается на существование в указанных кольцах "следящей функции", позволявшей определить в кольце *деление с остатком*. (Мы уже говорили об этом в пп. 37.1, 38.2 и 38.3.)

Обобщением указанных конструкций служит следующее

Определение 38.5. Целостное кольцо K называется *евклидовым*, если на множестве его ненулевых элементов задана *следящая функция*

$$\nu \,:\, K \setminus \{0\} \longrightarrow \mathbb{N} \cup \{0\} \tag{38.25}$$

и для любых двух элементов $a, b \in K$ $(a \neq 0)$ существуют такие элементы $q, r \in K$, что

$$(\, b = aq + r \,) \wedge [\, (\, r = 0 \,) \vee (\, \nu(r) < \nu(a) \,) \,]; \tag{38.26}$$

кроме того, должно выполняться *условие монотонности*

$$(\forall a, b \in K \setminus \{0\})\ [\, (\, a \,|\, b \,) \Rightarrow (\, \nu(a) \leqslant \nu(b) \,) \,]. \tag{38.27}$$

В наших основных примерах: 1) следящая функция "модуль" определена на всем кольце $\mathbb{Z}$; 2) функция "степень" — не на всем кольце $P[x]$, а лишь на ненулевых многочленах; 3) монотонность этих функций отмечалась в замечании 37.3.

Предложение 38.9. *Любое евклидово кольцо удовлетворяет условию Безу.*

Доказательство состоит в организации работы алгоритма Евклида и ничем (кроме большей абстрактности) не отличается от доказательства теоремы 38.1. □

Пример 38.6. Еще одним, очень популярным примером евклидова кольца является кольцо *целых гауссовых чисел*

$$\mathbb{Z}[i] = \{\, z = x + yi \in \mathbb{C} \,:\, x, y \in \mathbb{Z} \,\} \tag{38.29}$$

со следящей функцией — "норма":

$$N : \mathbb{Z}[i] \to \mathbb{N} \cup \{0\};\ N(z) = |z|^2 = x^2 + y^2. \tag{38.30}$$

Интересные и важные подробности об этом кольце можно найти в указанных в списке литературы учебниках и задачниках. (Не забывайте о том, что сборник задач — это не только средство для "озадачивания" студентов, но и важный источник дополнительной информации об изучаемом материале.)

§ 39. Многочлены и полиномиальные функции. Корни многочленов. Теорема Безу

39.1. Полиномиальные функции. Равенство многочленов и равенство полиномиальных функций. Рассмотрим кольцо $P[x]$ многочленов над полем P.

Произвольному многочлену $f(x) \in P[x]$, заданному формулой (36.1), — мы снова возвращаемся к записи многочленов *по возрастанию степеней* — и произвольному элементу $c \in P$ можно сопоставить новый элемент поля P, обозначаемый $\mathrm{f}(c)$, который называется *значением* многочлена $f(x)$ в точке c и определяется формулой

$$\mathrm{f}(c) = f_0 + f_1 c + f_2 c^2 + ... + f_n c^n. \tag{39.1}$$

Так возникает *отображение вычисления* из кольца многочленов в поле коэффициентов:

$$\nu_c : P[x] \to P;\ f(x) \mapsto \mathrm{f}(c);\ f(x) \in P[x]. \tag{39.2}$$

Это отображение зависит от элемента $c \in P$ как от параметра.

Считая, что c пробегает все поле P, мы получаем функцию (отображение) f, действующую из P в P и заданную правилом

$$\mathrm{f} : P \to P;\ c \mapsto \mathrm{f}(c) = \sum_{k=0}^{n} f_k c^k;\ c \in P. \tag{39.3}$$

Определение 39.1. Функция, заданная формулой (39.3), называется *полиномиальной функцией*, отвечающей многочлену (36.1).

Замечание 39.1. Обратите внимание на "легкое" видоизменение шрифта в обозначении значений многочлена и полиномиальной функции. Мы хотим подчеркнуть, что имеем дело с новым математическим объектом, связанным с многочленом, но не тождественным ему. (Впоследствии мы о таких тонкостях будем забывать, но пока условимся многочлены обозначать принятым при наборе формул так называемым *математическим курсивом*, а соответствующие полиномиальные функции — прямым шрифтом.)

Термины *значение полиномиальной функции* (в некоторой точке) и *значение многочлена* (в той же точке) мы различать не будем. Уже в следующем пункте, допуская "вольность речи" и "вольность в обозначениях", мы будем говорить о значении многочлена в точке c и писать $f(c)$, не заботясь о шрифтах.

Вообще, не взирая на репутацию "строжайшей" из наук, математика изобилует подобными отклонениями от строгой терминологии. Иначе математические тексты было бы очень трудно читать. Пассаж "вольность речи" введен в оборот группой французских математиков, работавшей под псевдонимом Н. Бурбаки (чем и прославившей не очень популярного французского генерала). В прошлом веке Бурбаки буквально потрясли всю математику, затеяв ее ревизию от самых оснований и до наиболее "продвинутых" ветвей. Многотомный тракт Бурбаки под названием "Элементы математики" вызвал и продолжает вызывать как неумеренные восторги, так и яростное отторжение. (Это информация для тех, кто думает, что математика — "спокойная" и "мирная" наука.)

Множество *всех функций* из поля P в себя представляет собой коммутативное кольцо, обозначим его $\mathcal{F}(P)$, с *поточечными* алгебраическими действиями сложения и умножения. Для двух функций $\mathrm{f}, \mathrm{g} \in \mathcal{F}(P)$ их сумма и произведение определяются формулами

$$(\mathrm{f}+\mathrm{g})(c) = \mathrm{f}(c) + \mathrm{g}(c);\ (\mathrm{f}\cdot\mathrm{g})(c) = \mathrm{f}(c)\cdot\mathrm{g}(c) \tag{39.4}$$

для любого $c \in P$.

(Выполнение аксиом коммутативного кольца для $\mathcal{F}(P)$ с очевидностью следует из того факта, что при любом фиксированном c они выполняются, поскольку справедливы в поле коэффициентов P.)

Однако свойством целостности (36.16) кольцо функций не обладает. (Рассмотрите две функции f и g из $\mathbb{R}$ в $\mathbb{R}$; пусть первая из них

равна нулю при $c \leqslant 0$ и совпадает с тождественной функцией при $c > 0$, а вторая совпадает с тождественной при $c < 0$ и равна нулю при $c \geqslant 0$; произведение fg этих функций есть нулевая функция.)

Вообще говоря, не всякая функция является полиномиальной. (Для случая поля действительных чисел это должно быть вам хорошо известно из курса анализа. Однако ситуация оказывается совершенно противоположной для конечных полей; см. ниже пример 39.1.) Множество всех полиномиальных функций из поля P в себя обозначим $\mathcal{F}_{poly}(P)$. Ниже мы покажем, что это подмножество является на самом деле *подкольцом* в $\mathcal{F}(P)$.

По определению 39.1, имеется отображение

$$\nu : P[x] \to \mathcal{F}(P);\ f(x) \mapsto \mathrm{f};\ f(x) \in P[x], \tag{39.5}$$

сопоставляющее многочлену порождаемую этим многочленом полиномиальную функцию. Множество $\mathcal{F}_{poly}(P) \subseteq \mathcal{F}(P)$, по определению, является образом отображения (39.5).

Убедимся теперь, что отображения ν_c (при любом фиксированном $c \in P$) и ν согласованы с алгебраическими операциями (другими словами, являются *гомоморфизмами колец*).

Проверку согласованности проведем для умножения (для сложения она значительно проще и будет по силам читателям).

Коэффициенты для произведения $h(x) = f(x)g(x)$ двух многочленов (36.1) и (36.2) задаются формулой (36.10):

$$h_s = \sum_{\substack{0 \leqslant k \leqslant n \\ 0 \leqslant l \leqslant m \\ k+l=s}} f_k g_l.$$

[Здесь мы записали эту формулу чуть иначе, с учетом того факта, что коэффициенты f_k (при $k > n$) и коэффициенты g_l (при $l > m$) обращаются в нуль.]

Поэтому значение многочлена $h(x)$ на элементе $c \in P$ определится формулой

$$\mathrm{h}(c) = \sum_{s=0}^{n+m} \left(\sum_{\substack{0 \leqslant k \leqslant n \\ 0 \leqslant l \leqslant m \\ k+l=s}} f_k g_l \right) \mathrm{c}^s. \tag{39.6}$$

Теперь перемножим значения в точке c для многочленов $f(x)$ и $g(x)$ (умножение будет производиться в поле P с использованием аксиомы дистрибутивности):

$$\mathrm{f}(c)\mathrm{g}(c) = \left(\sum_{k=0}^{n} f_k c^k\right)\left(\sum_{l=0}^{m} g_l c^l\right) = \sum_{s=0}^{n+m}\left(\sum_{\substack{0\leqslant k\leqslant n\\ 0\leqslant l\leqslant m\\ k+l=s}} f_k g_l\right)\mathrm{c}^s. \qquad (39.7)$$

Правые части (39.6) и (39.7) совпадают, следовательно,

$$\mathrm{h}(c) = \mathrm{f}(c)\mathrm{g}(c). \qquad (39.8)$$

Формула (39.8) выражает согласованность отображения ν_c с умножением. Поскольку она имеет место для любого $c \in P$, то и отображение ν оказывается согласованным с умножением: произведению многочленов $f(x)g(x)$ отвечает произведение fg соответствующих полиномиальных функций.

Итак, всякое отображение вычисления (39.1) является гомоморфизмом кольца многочленов в поле коэффициентов. (Легко видеть, что оно является *эпиморфизмом*: достаточно применить его к постоянным многочленам.)

Отображение (39.5) является гомоморфизмом (вообще говоря, не эпиморфизмом) кольца многочленов в кольцо функций (*на* подкольцо полиномиальных функций).

Образ $\mathcal{F}_{poly}(P)$ гомоморфизма колец ν является подкольцом, т. е. устойчив относительно взятия сумм, разностей и произведений, в кольце $\mathcal{F}(P)$.

Сейчас мы убедимся, что, вообще говоря, гомоморфизм (39.5) не является *мономорфизмом*, т. е. инъективность для этого отображения может не иметь места. Выражаясь конкретнее, мы установим, что различным многочленам могут отвечать одинаковые полиномиальные функции.

На самом деле это совершенно ясно из общих "перечислительных" соображений. Если поле P является *конечным* (имеет, скажем, q элементов), то существует в точности q^q различных функций из P в P (любой из q элементов поля может отображаться опять же в любой из q элементов). В то же время многочленов над любым полем P всегда бесконечно много.

(Просмотрите еще раз пример 36.1; сейчас он получит продолжение и развитие.)

Пример 39.1. Поле $P = \mathbb{F}_2$ содержит два элемента, и поэтому существуют всего четыре функции из этого поля в себя:

— постоянная функция, тождественно равная нулю: $0 \mapsto 0$; $1 \mapsto 0$;

— еще одна константа, равная единице: $0 \mapsto 1$; $1 \mapsto 1$;

— тождественная функция ε: $0 \mapsto 0$; $1 \mapsto 1$;

— транспозиция элементов поля τ: $0 \mapsto 1$; $1 \mapsto 0$.

Совершенно очевидно, что все эти четыре функции являются полиномиальными и соответствуют четырем первым многочленам 0, 1, x, $1 + x$ из списка примера 36.1. Разберемся с четырьмя многочленами второй степени x^2, $1 + x^2$, $x + x^2$, $1 + x + x^2$. Первый из них отвечает функции ε, второй — функции τ, третий — нулевой константе, четвертый — единичной. (Выясните, каким функциям соответствуют многочлены третьей степени.)

Замечание 39.2. Подводя итог, констатируем, что аналитическая точка зрения на многочлены "грубее", а алгебраическая "тоньше". Различным многочленам могут соответствовать одинаковые функции.

Однако в случае бесконечного поля коэффициентов ситуация меняется: ниже, в п. 39.4, будет доказано, что над *бесконечным* полем равенство многочленов равносильно равенству соответствующих полиномиальных функций, т. е. гомоморфизм (39.5) является мономорфизмом.

Замечание 39.3. Понятие полиномиальной функции может быть обобщено в двух направлениях.

Во-первых, многочлен f с коэффициентами f_k, принадлежащими полю P, определяет отображение [обозначаемое так же, как и (39.3)]

$$\mathrm{f} : P' \to P';\ c \mapsto \mathrm{f}(c) = \sum_{k=0}^{n} f_k c^k;\ c \in P' \qquad (39.3')$$

любого поля P', являющегося *расширением* поля P. (Скажем, многочлену с действительными коэффициентами соответствует полиномиальная функция из $\mathbb{C}$ в $\mathbb{C}$.)

Во-вторых, без всяких изменений эта конструкция переносится на многочлены с коэффициентами из коммутативного кольца: всякий многочлен $f(x) \in L[x]$ порождает функции $\mathrm{f} : L' \to L'$ (здесь L — коммутативное кольцо коэффициентов, а L' — произвольное его расширение).

Все сказанное о равенстве многочленов и равенстве соответствующих полиномиальных функций сохраняет силу и в более общей ситуации многочленов над кольцом.

39.2. Корни многочленов и теорема Безу

Определение 39.2. *Корнем многочлена* $f(x) \in P[x]$ называется *корень* соответствующей *полиномиальной функции* $\mathrm{f} : P \to P$, т. е. такой элемент $c \in P$, что значение

$$\mathrm{f}(c) = 0. \tag{39.9}$$

Замечание 39.4. В соответствии с "вольным духом" замечания 39.1, в дальнейшем мы, не заботясь о различии шрифтов, будем писать $f(c) = 0$ и вполне по-школьному говорить, что корень многочлена — это такой элемент, принадлежащий полю коэффициентов, при подстановке которого в многочлен (вместо переменной) получается значение, равное нулю. Более того, мы будем рассматривать корни многочлена не только в поле коэффициентов P, но и в его *расширениях* (см. замечание 39.3).

Для многочленов над коммутативным кольцом также можно изучать корни (как в кольце коэффициентов, так и в любом расширении этого кольца). Например, многочлен $f(x) = x^2 + 1$ можно считать заданным над полем рациональных чисел $\mathbb{Q}$ или над кольцом целых чисел $\mathbb{Z}$. Но ни в $\mathbb{Z}$, ни в $\mathbb{Q}$, ни даже в $\mathbb{R}$ этот многочлен корней не имеет, тогда как в $\mathbb{C}$ у него существуют два корня $\pm i$.

Кстати, этот многочлен можно (при надлежащей договоренности о том, что 0 и 1 суть классы вычетов по модулю 2) рассматривать и над $\mathbb{F}_2$. В этом поле он будет иметь один корень, равный 1 (точнее говоря, два одинаковых корня, поскольку $x^2 + 1 \equiv (x+1)^2 \mod 2$). (Попробуйте найти корни того же многочлена в поле $\mathbb{F}_3$.)

Ни в каком поле (кольце) не имеют корней многочлены нулевой степени (ненулевые константы). Напротив, корнями нулевого многочлена являются все элементы рассматриваемого поля (кольца).

Пусть $f(x)$ есть многочлен степени $n \geqslant 1$ над полем P. Его корни оказываются связанными с его делителями степени 1.

Точнее, справедлива следующая (простая, но знаменитая)

Теорема 39.1 (теорема Безу). *1. Остаток от деления многочлена $f(x)$ положительной степени n над полем P на многочлен первой степени (двучлен) $x - c$, где $c \in P$, равен значению $f(c)$ многочлена $f(x)$ в точке c.*

2. Элемент c является корнем многочлена $f(x)$ тогда и только тогда, когда двучлен $x - c$ делит данный многочлен:

$$x - c \,|\, f(x). \tag{39.10}$$

Доказательство. 1. Согласно теореме 37.1, многочлен $f(x)$ можно поделить с остатком на двучлен $x - c$, т. е. представить его в виде

$$f(x) = (x - c)q(x) + h(x), \tag{39.11}$$

где неполное частное $q(x)$ является многочленом степени $n - 1$, а остаток $h(x)$ либо равен нулю, либо является многочленом степени меньшей, чем степень делителя, т. е. многочленом нулевой степени. Оба случая охватываются формулировкой: остаток $h(x) = h_0$ является скалярным многочленом (константой).

Итак, имеем равенство многочленов:

$$f(x) = (x - c)q(x) + h_0. \tag{39.11a}$$

Применим к этому равенству гомоморфизм вычисления ν_c (попросту говоря, подставим в обе части этого равенства $x = c$). Законность такого действия объяснялась в предыдущем пункте: отображение ν_c согласовано с алгебраическими действиями сложения и умножения.

Поскольку значение двучлена $x - c$ в точке c равно, очевидно, нулю, то значение всей правой части равенства (39.11a) равно константе h_0. Тем самым доказано равенство $h(x) = h_0 = f(c)$, и формула (39.11a) приобретает вид

$$f(x) = (x - c)q(x) + f(c). \tag{39.12}$$

2. Второе утверждение теоремы (именно его обычно называют теоремой Безу) очевидным образом вытекает из первого. В самом деле, тот факт, что c является корнем $f(x)$, означает, по определению 39.1, что значение $f(c) = 0$, а это, в силу первого утверждения теоремы, равносильно обращению в нуль остатка $h(x)$, т. е. *делимости нацело* (39.9).

В этом случае многочлен $f(x)$ *разлагается на множители*:

$$f(x) = (x - c)q(x), \tag{39.13}$$

один из которых является линейным двучленом, а другой — многочленом степени $n - 1$. □

Замечание 39.5. Теорема Безу без всяких оговорок переносится на случай многочленов над коммутативным (целостным) кольцом.

Дело в том, что делитель $x-c$ является *нормализованным* многочленом; его старший коэффициент (равный 1) *обратим*. Отсюда вытекает возможность деления с остатком на $x-c$ над любым кольцом коэффициентов (см. замечание 37.2). А для доказательства теоремы Безу больше ничего и не надо.

39.3. Оценка количества (различных) корней многочлена. Ненулевой многочлен (над любым полем или целостным кольцом) имеет лишь конечное множество корней: их количество не может превышать степени многочлена. (Это верно и для многочленов нулевой степени: они не имеют ни одного корня.)

Точнее, справедливо следующее

Предложение 39.1. *Пусть попарно различные элементы*

$$c_1, c_2, ..., c_s$$

поля P являются корнями многочлена $f(x) \in P[x]$ степени n.

Тогда $s \leqslant n$ и имеет место разложение на множители

$$f(x) = (x-c_1)(x-c_2)...(x-c_s)g(x), \tag{39.14}$$

где $g(x)$ — многочлен степени $n-s$.

Доказательство. По теореме Безу, тот факт, что c_1 является корнем $f(x)$, позволяет разложить [см. (39.9)] многочлен $f(x)$ на множители:

$$f(x) = (x-c_1)f_1(x), \tag{39.15}$$

где многочлен $f_1(x)$ имеет степень $n-1 \geqslant 0$.

Корень c_2, отличный от первого, при подстановке в (39.15) дает:

$$0 = (c_2-c_1)f_1(c_2). \tag{39.16}$$

Разность c_2-c_1 является ненулевым элементом поля P; поэтому (39.16) влечет $f_1(c_2)=0$, а это означает, что c_2 является корнем $f_1(x)$. По теореме Безу, многочлен $f_1(x)$ имеет разложение, аналогичное (39.15):

$$f_1(x) = (x-c_2)f_2(x), \tag{39.17}$$

где $\deg(f_2(x)) = n-2 \geqslant 0$.

Для исходного многочлена, в силу (39.17), в (39.15) будем иметь:

$$f(x) = (x - c_1)(x - c_2)f_2(x). \tag{39.18}$$

Подставляя третий корень c_3 в (39.18), получим:

$$0 = (c_3 - c_1)(c_3 - c_2)f_2(c_3). \tag{39.19}$$

Снова, сокращая на ненулевой скаляр, убеждаемся в том, что c_3 является корнем многочлена $f_2(x)$; и т. д.

В конце концов мы придем к разложению, содержащему s скобок (линейных двучленов), соответствующих s корням, и многочлен $f_s(x)$, который мы переобозначим в $g(x)$. Степень $n - s$ этого многочлена обязана быть неотрицательным числом, следовательно, должно выполняться неравенство $s \leqslant n$. □

Замечание 39.6. Если $s = n$, то многочлен $g(x)$ должен иметь нулевую степень, т. е. будет скаляром $g(x) = g_0$.

Мысленно раскройте скобки и приравняйте коэффициенты при старшей степени x^n. Вы убедитесь в том, что скаляр g_0 будет совпадать со старшим коэффициентом многочлена $f(x)$.

Предположим, что данный многочлен записан по убыванию степеней [см. (36.18)] и его старший коэффициент обозначается a_0. Тогда разложение (38.14) примет вид:

$$f(x) = a_0(x - c_1)(x - c_2)...(x - c_n). \tag{39.20}$$

39.4. Многочлены и полиномиальные функции над бесконечным полем. Рассмотрим кольцо $P[x]$ многочленов над *бесконечным* полем P и кольцо полиномиальных функций $\mathcal{F}_{poly}$. Равным многочленам по построению отвечают равные функции. Как подчеркивалось в п. 39.1, обратное, вообще говоря, не верно.

Обратное становится верным, если поле коэффициентов P является бесконечным. Точнее, справедливо следующее

Предложение 39.2. *Пусть поле P бесконечно. Тогда гомоморфизм колец (39.5) является мономорфизмом, т. е. различным многочленам отвечают различные функции.*

Доказательство. Пусть многочленам $f(x), g(x) \in P[x]$ соответствуют функции $\mathrm{f}, \mathrm{g} \in \mathcal{F}_{poly}$. Предположим $\mathrm{f} = \mathrm{g}$, т. е. для любого элемента $c \in P$ справедливо $\mathrm{f}(c) = \mathrm{g}(c)$. Рассмотрим многочлен-разность $h(x) = f(x) - g(x)$. В силу согласованности алгебраических

действий над многочленами и полиномиальными функциями, этому многочлену отвечает функция-разность $\mathrm{h} = \mathrm{f} - \mathrm{g}$, тождественно равная нулю: $\mathrm{h}(c) = \mathrm{f}(c) - \mathrm{g}(c) = 0$ $(c \in P)$.

Другими словами, всякий элемент поля P является корнем многочлена $h(x)$. Следовательно, $h(x)$ имеет бесконечно много попарно различных корней. Для ненулевого многочлена это невозможно (в силу предложения 39.1). Значит, многочлен $h(x)$ нулевой, и следовательно, $f(x)$ и $g(x)$ являются равными многочленами. □

Замечание 39.7. Предложения 39.1 и 39.2 сохраняют силу для случая, когда коэффициенты многочленов принадлежат не полю, а целостному кольцу.

§ 40. Кратность корня. Оценка суммы кратностей корней. Алгебраически замкнутые поля. Разложимость многочленов на линейные множители. Теорема Виета

40.1. Понятие кратности корня многочлена. То, что корни бывают кратными, вы знаете с детства. Например, вам не один раз повторяли, что если дискриминант равен нулю, то квадратное уравнение "имеет два одинаковых корня" или "имеет корень кратности два". Но без теоремы Безу (в "обычных" курсах школьной математики о ней обычно не упоминают, хотя в задачах повышенной трудности она необходима) строгое понятие кратности корня не ввести.

(Известно анекдотическое определение кратного корня: "Это, когда подставляешь число в многочлен — получается нуль, опять подставляешь — опять получается нуль и т. д.")

Ниже приводится серьезное и строгое определение.

Пусть $f(x)$ — многочлен положительной степени n над полем P; c — корень этого многочлена, принадлежащий P. По теореме Безу (теореме 39.1), двучлен $x - c$ делит $f(x)$. Не исключено, однако, что некоторая, более высокая степень $(x - c)^k$ (где $k > 1$) также делит $f(x)$. В связи с этим возникает

Определение 40.1. *Кратностью* корня c для многочлена $f(x)$ называется *наивысшая* (натуральная) степень m, такая, что

$$(x-c)^m | f(x), \tag{40.1}$$

т. е.

$$f(x) = (x-c)^m g(x) \tag{40.2}$$

для некоторого многочлена $g(x)$ степени $n-m$, причем c не является корнем $g(x)$.

Замечание 40.1. По определению, кратность корня не превышает его степени: $m \leqslant n$.

Корни кратности 1 обычно называются *простыми* или *однократными*. В некоторых случаях первый вариант названия оказывается не очень удачным ввиду возможной путаницы с другим понятием "простоты" (в смысле *неразложимости*; см. ниже п. 44.2).

Ради единообразия допускается словоупотребление "корень кратности нуль", означающее "не корень". С определением 40.1 это согласуется, поскольку $(x-c)^0 = 1$.

40.2. Оценка суммы кратностей корней многочлена. Как известно из предложения 39.1, количество корней многочлена $f(x)$ не превышает его степени n. Пусть

$$c_1,\ c_2,\ \ldots,\ c_s \tag{40.3}$$

— список всех попарно различных корней многочлена $f(x)$, принадлежащих P.

Каждому из этих корней приписывается натуральное число — его кратность; так получается список натуральных чисел

$$m_1,\ m_2,\ \ldots,\ m_s. \tag{40.4}$$

Рассмотрим сумму кратностей всех корней:

$$m' = \sum_{k=1}^{s} m_k. \tag{40.5}$$

(Многочлен может вообще не иметь корней, тогда принято считать, что $m' = 0$.)

Предложение 40.1. *Сумма кратностей (40.4) всех (попарно различных) корней (40.3) многочлена не превышает его степени:*

$$m' \leqslant n. \tag{40.6}$$

Имеет место разложение на множители:

$$f(x) = (x - c_1)^{m_1}(x - c_2)^{m_2}...(x - c_s)^{m_s} g(x), \tag{40.7}$$

где $g(x)$ — многочлен степени $n - m'$, не имеющий корней в поле P.

Доказательство. Следующее рассуждение в основном повторяет схему доказательства предложения 39.1, но добавляются трудности, связанные с учетом кратностей.

Корень c_1 кратности m_1 дает [по типу (40.2)] разложение

$$f(x) = (x - c_1)^{m_1} g_1(x), \tag{40.8}$$

где $g_1(x)$ — многочлен степени $n - m_1$ и $g_1(c_1) \neq 0$.

Совершенно аналогично тому, как это было сделано в доказательстве предложения 39.1, устанавливается, что корень c_2 многочлена $f(x)$ является в то же время и корнем многочлена $g_1(x)$.

Пусть m_2' — кратность c_2 как корня $g_1(x)$. Ясно, что $m_2' \leqslant m_2$, поскольку делимость $(x-c_2)^{m_2'} | g_1(x)$ влечет (в силу свойств отношения делимости; см. предложение 37.1) делимость $(x - c_2)^{m_2'} | f(x)$.

Если имеет место строгое неравенство $m_2' < m_2$, то, представив многочлены $f(x)$ и $g_1(x)$ в виде произведений

$$f(x) = (x - c_2)^{m_2} f_2(x);\ g_1(x) = (x - c_2)^{m_2'} g_2(x),$$

где $\deg(f_2(x)) = n - m_2$; $\deg(g_2(x)) = n - m_1 - m_2'$; $f_2(c_2) \neq 0$; $g_2(c_2) \neq 0$, мы придем к формуле

$$(x - c_2)^{m_2} f_2(x) = (x - c_1)^{m_1}(x - c_2)^{m_2'} g_2(x),$$

которую можно будет (в силу целостности кольца многочленов) сократить на $(x - c_2)^{m_2'}$ и получить равенство

$$(x - c_2)^{m_2 - m_2'} f_2(x) = (x - c_1)^{m_1} g_2(x),$$

приводящее к противоречию, т. к. при подстановке в него c_2 вместо x слева получится нуль, а справа — не нуль.

Значит, $m_2' = m_2$, и можно записать разложение на множители

$$f(x) = (x - c_1)^{m_1}(x - c_2)^{m_2} g_2(x), \tag{40.9}$$

где $\deg(g_2(x)) = n - m_1 - m_2$; $g_2(c_1) \neq 0$; $g_2(c_2) \neq 0$.

Далее можно (вполне аналогичным рассуждением) установить, что c_3 будет корнем многочлена $g_2(x)$, причем такой же кратности m_3, каковую этот корень имел в многочлене $f(x)$ и т. д.

Перебрав все корни (40.3), мы придем к разложению (40.7), с многочленом $g(x) = g_s(x)$ степени $n - m'$, которая обязана быть неотрицательной, что приводит к неравенству (40.4). □

Замечание 40.2. Вместо термина "сумма кратностей корней", допускается использование (менее строгого, менее лаконичного, но более выразительного) термина "количество корней с учетом их кратностей".

Доказанное выше предложение представляет собой оценку этого количества, уточняющую ранее полученное (в предложении 39.1) неравенство для количества попарно различных корней: $s \leqslant n$.

Замечание 40.3. Предложение 40.1 остается в силе для многочленов с коэффициентами из целостного кольца L. (В самом деле, в этом случае целостным будет и кольцо многочленов $L[x]$, и следовательно, все сокращения, на которых основано проведенное выше доказательство, останутся законными.)

40.3. Алгебраически замкнутые поля

Определение 40.2. Поле P называется *алгебраически замкнутым*, если всякий многочлен положительной степени с коэффициентами из поля P имеет хотя бы один корень в этом поле.

Поле $\mathbb{Q}$ не является алгебраически замкнутым: уже квадратичный многочлен $x^2 - 2$ не имеет рациональных корней. (Это — известный из школьного курса алгебры факт, выражаемый иначе утверждением "число $\sqrt{2}$ *иррационально*". Умеете ли вы доказывать этот факт? Если нет, то у вас — очень существенный пробел в знаниях о природе, и вам необходимо его срочно ликвидировать, хотя бы порывшись в школьных учебниках.)

В поле $\mathbb{R}$ указанный выше многочлен имеет два корня и следующее разложение на линейные множители: $x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$.

Но и поле $\mathbb{R}$ не является алгебраически замкнутым. (Мы об этом хорошо знаем: иначе не понадобилась бы гл. 5, в которой отсутствие

корней у многочлена x^2+1 побудило нас расширить поле $\mathbb{R}$, добавив к нему мнимую единицу.)

Основная теорема алгебры (теорема 35.1) утверждает в точности следующее:

п о л е $\mathbb{C}$ а л г е б р а и ч е с к и з а м к н у т о .

40.4. Разложение многочлена над алгебраически замкнутым полем на линейные множители. Рассмотрим многочлен, записанный в виде (36.18):

$$f(x) = a_0x^n + a_1x^{n-1} + ... + a_{n-1}x + a_n.$$

Предложение 40.2. *Всякий многочлен (36.18) положительной степени n над алгебраически замкнутым полем P имеет n корней (с учетом их кратностей) и разлагается в произведение n линейных сомножителей:*

$$f(x) = a_0(x-c_1)^{m_1}(x-c_2)^{m_2}...(x-c_s)^{m_s}, \qquad (40.10)$$

где $c_1, c_2, ..., c_s$ — все его (попарно различные) корни; $m_1, m_2, ..., m_s$ — кратности этих корней; $m_1 + m_2 + ... + m_s = n$.

Доказательство. По определению 40.2, список корней многочлена $f(x)$ не пуст. Согласно предложению 40.1, этот многочлен разлагается на множители по формуле (40.7).

Фигурирующий в указанной формуле многочлен $g(x)$ не имеет корней в поле P, следовательно, в силу алгебраической замкнутости P, степень этого многочлена может быть только нулевой, т. е. $m' = n$ и $g(x)$ является ненулевым скаляром: $g(x) = g_0$.

Рассуждая в точности так же, как и в замечании 39.6, мы придем к выводу, что $g_0 = a_0$. Требуемое разложение (40.10) получено. □

Пример 40.1. Разложим (элементарными средствами, с помощью группировки) многочлен

$$f(x) = x^5 - x^4 + 2x^3 - 2x^2 - 8x + 8$$

на линейные множители:

$$\begin{aligned} f(x) &= x^4(x-1) + 2x^2(x-1) - 8(x-1) = \\ &= (x-1)(x^4 + 2x^2 - 8) = (x-1)(x^4 + 4x^2 - 2x^2 - 8) = \\ &= (x-1)(x^2(x^2+4) - 2(x^2+4)) = \\ &= (x-1)(x^2-2)(x^2+4) \text{ [в поле } \mathbb{Q}]; \\ &= (x-1)(x-\sqrt{2})(x+\sqrt{2})(x^2+4) \text{ [в поле } \mathbb{R}]; \\ &= (x-1)(x-\sqrt{2})(x+\sqrt{2})(x-2i)(x+2i) \text{ [в поле } \mathbb{C}]. \end{aligned}$$

Замечание 40.4. В курсе общей алгебры доказывается, что всякое поле P можно *вложить* в алгебраически замкнутое поле P', являющееся *алгебраически замкнутым расширением* данного поля P. Среди алгебраически замкнутых расширений P' существует в определенном смысле "наименьшее" поле $\overline{P}$, которое *вкладывается* в любое алгебраически замкнутое расширение данного поля. Поле $\overline{P}$ называется *алгебраическим замыканием* поля P.

(Точный смысл предыдущих высказываний требует довольно пространных разъяснений, которые были бы здесь не уместны и не своевременны. При построении *поля комплексных чисел*, являющегося *алгебраическим замыканием поля действительных чисел*, мы уже говорили о "минимальности" расширения $\mathbb{R} \subset \mathbb{C}$, но в несколько ином смысле; см. замечание 31.3.)

Только что был упомянут пример алгебраического замыкания: $\overline{\mathbb{R}} = \mathbb{C}$. Еще один важнейший пример — алгебраическое замыкание поля рациональных чисел: $\overline{\mathbb{Q}} = \mathbb{A}$. Поле $\mathbb{A}$ называется полем *алгебраических чисел*; оно состоит из всех комплексных чисел, являющихся корнями многочленов с рациональными коэффициентами (или, что достаточно, многочленов с целыми коэффициентами). (Такими числами являются, скажем: $\sqrt{2}$ — корень многочлена $x^2 - 2$; $(\sqrt[8]{2})(\sqrt{2+\sqrt{2}} + i\sqrt{2-\sqrt{2}})/2$ — корень многочлена $x^8 + 2$ и т. п.)

Всякий многочлен положительной степени с коэффициентами из произвольного поля P можно разложить на линейные множители [по формуле (40.10)] над полем $\overline{P}$.

40.5. Вычисление корней многочленов и разложение многочленов на множители средствами системы Maple. Мы уже встречались выше с отдельными приемами вычисления корней многочленов и разложения многочленов на множители в системе Maple. Сейчас мы несколько более систематично изложим сведения об этих Maple-средствах.

Пример 40.2. Введем следующий многочлен:

```
> g := 4*x^4 + 12*x^3 + 5*x^2 - 16*x + 5;
```

$$g := 4x^4 + 12x^3 + 5x^2 - 16x + 5$$

Применим функцию roots (= корни), по умолчанию настроенную на отыскание корней в наименьшем из полей, содержащих коэффициенты заданного многочлена (в данном случае этим полем будет поле рациональных чисел), и сразу же — расширенную версию этой

функции с указанием на *расширение* поля $\mathbb{Q}$ с помощью добавления мнимой единицы (напомним, что в Maple она обозначается заглавной буквой I):

> roots(g); roots(g, I);

Результаты будут выданы в форме списков пар вида [*корень, кратность*]:

$$\left[\left[\frac{1}{2},\ 2\right]\right]$$

$$\left[[-2+I,\ 1],\ [-2-I,\ 1],\ \left[\frac{1}{2},\ 2\right]\right]$$

Используем теперь другую команду, factors (= множители), тоже в двух версиях. Эта команда возвращает списки вида

[a_0, [список пар вида [двучлен $x - c_i$, натуральное число m_i]]],

где a_0 — *старший коэффициент* многочлена, c_i — его *корни*, m_i — их *кратности*; во внутреннем списке могут присутствовать также *неприводимые* (т. е. не разложимые на множители меньшей степени) многочлены, тоже с кратностями (в частности, у этих многочленов *нет корней* в рассматриваемом поле).

> factors(g); factors(g, I);

$$\left[4,\ \left[\left[x-\frac{1}{2},\ 2\right],\ [x^2+4x+5,\ 1]\right]\right]$$

$$\left[4,\ \left[\left[x-\frac{1}{2},\ 2\right],\ [x+2-I,\ 1],\ [x+2+I,\ 1]\right]\right]$$

И наконец, применим команду factor (= разложить на множители), возвращающую результат в привычной алгебраической записи:

> factor(g); factor(g, I);

$$(2x-1)^2(x^2+4x+5)$$

$$(x+2-I)(x+2+I)(2x-1)^2$$

Пример 40.3. Повторим вычисления, проделанные в предыдущем примере, применительно к многочлену $f(x)$, рассмотренному в "ручном" примере 40.1. Здесь к полю рациональных чисел придется добавить не только мнимую единицу, но и квадратный корень из двух. [На Maple-языке это оформляется указанием *множества* с (заключаемыми в фигурные скобки) элементами I и $\sqrt{2}$.]

```
> f:= x^5 - x^4 + 2*x^3 - 2*x^2 - 8*x + 8:
> roots( f ); roots( f, sqrt(2) ); roots( f, I ); roots( f, { sqrt(2), I } );
```

$$[[1,\ 1]]$$

$$[[-\sqrt{2},\ 1]],\ [\sqrt{2},\ 1],\ [1,\ 1]]$$

$$[[2I,\ 1],\ [-2I,\ 1],\ [1,\ 1]]$$

$$[[2I,\ 1],\ [-2I,\ 1],\ [-\sqrt{2},\ 1]],\ [\sqrt{2},\ 1],\ [1,\ 1]]$$

```
> factors(f); factors(f, sqrt(2)); factors(f, I); factors(f, { sqrt(2), I });
```

$$[1,\ [[x^2-2,\ 1],\ [x^2+4,\ 1],\ [x-1,\ 1]]]$$

$$[1,\ [[x-\sqrt{2},\ 1],\ [x^2+4,\ 1],\ [x+\sqrt{2},\ 1],\ [x-1,\ 1]]]$$

$$[1,\ [[x-2I,\ 1],\ [x+2I,\ 1],\ [x^2-2,\ 1],\ [x-1,\ 1]]]$$

$$[1,\ [[x-2I,\ 1],\ [[x-\sqrt{2},\ 1],\ [x+2I,\ 1],\ [x+\sqrt{2},\ 1],\ [[x-1,\ 1]]]$$

```
> factor(f); factor(f, sqrt(2)); factor(f, I); factor(f, { sqrt(2), I });
```

$$(x-1)(x^2+4)(x^2-2)$$

$$(x^2+4)(x+\sqrt{2})(x-\sqrt{2})(x-1)$$

$$(x^2-2)(x-2I)(x+2I)(x-1)$$

$$(x-2I)(x+2I)(x+\sqrt{2})(x-\sqrt{2})(x-1)$$

40.6. Теорема Виета. Рассмотрим многочлен $f(x)$ вида (36.18), степени $n > 0$, с коэффициентами из некоторого поля P, *разложимый* над этим полем *на линейные множители* по формуле типа (40.10). (В случае *алгебраической замкнутости* поля P указанным свойством, в силу предложения 40.2, будет обладать *любой* многочлен положительной степени.)

Многочлен, разложимый на линейные множители, будет иметь ровно n корней (с учетом их кратностей). Сформируем n-элементный список корней

$$c_1,\ c_2,\ ...,\ c_n\,, \tag{40.11}$$

в котором каждый корень *повторяется* столько раз, какова его *кратность*.

Разложение (40.10) также представим в "несгруппированном" виде:

$$a_0x^n + a_1x^{n-1} + ... + a_{n-1}x + a_n = a_0(x-c_1)(x-c_2)...(x-c_n). \tag{40.12}$$

Введем далее в рассмотрение следующие суммы:

— сумму σ_1 всех n элементов списка (40.11);

— сумму σ_2 всех возможных произведений, в которых берутся по два (различных) сомножителя из (40.11); количество таких произведений будет равняться C_n^2;

— и так далее;

— сумму σ_{n-1} всех $C_n^{n-1} = n$ возможных произведений по $n-1$ (различных) сомножителей из списка (40.11);

— единственное возможное произведение σ_n всех элементов этого списка.

В формульной записи определенные выше суммы выглядят так:

$$\left\{\begin{array}{rcc} \sigma_1 & = & c_1 + c_2 + ... + c_n\,; \\ \sigma_2 & = & c_1c_2 + c_1c_3 + ... + c_1c_n + c_2c_3 + ... + c_{n-1}c_n\,; \\ \ldots\ldots & \ldots\ldots & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \sigma_{n-1} & = & c_1c_2...c_{n-1} + c_1c_2...c_{n-2}c_n + ... + c_2c_3...c_n\,; \\ \sigma_n & = & c_1c_2...c_n\,. \end{array}\right. \tag{40.13}$$

Общим видом сумм σ_k $(k = 1, ..., n)$ является

$$\sigma_k = \sum_{1\leqslant j_1<j_2<...<j_k\leqslant n} c_{j_1}c_{j_2}...c_{j_k}\,, \tag{40.14}$$

где количество слагаемых равно числу сочетаний C_n^k.

Замечание 40.5. Суммы (40.13) являются *многочленами от n переменных $c_1, c_2, ..., c_n$*. Эти многочлены обладают свойством *симметрии*: при любой перестановке элементов списка (40.11) значение каждого из них остается неизменным. Для σ_k используется название *элементарные симметрические многочлены.*

(Строгое определение многочленов от нескольких переменных дается ниже, в § 48; симметрические многочлены, в том числе элементарные, изучаются в § 49.)

Теорема 40.1 (теорема Виета). *Пусть многочлен $f(x) \in P[x]$ положительной степени n разлагается над полем P на линейные множители и пусть (40.11) есть список корней этого многочлена. Коэффициенты многочлена $f(x)$ связаны со значениями сумм (40.13) формулами*

$$\frac{a_k}{a_0} = (-1)^k \sigma_k, \tag{40.15}$$

где $k = 1, ..., n$.

Доказательство. Обе части формулы (40.12) поделим на a_0 и в правой части полученной формулы раскроем скобки. При этом получится 2^n слагаемых, каждое из которых будет произведением n сомножителей, причем некоторые из этих сомножителей будут равняться x, а остальные будут иметь вид $(-c_j)$.

Приведем подобные в полученной сумме, т. е. сгруппируем ее по степеням x.

Один из членов будет равен x^n (из всех n скобок берется x).

Количество членов, содержащих x^{n-k} $(k = 1, ..., n)$, будет равняться C_n^{n-k} $(= C_n^k)$ — столькими способами из n скобок можно выбрать те $n - k$ скобок, из которых берется x. Все эти члены будут иметь общий знак $(-1)^k$, поскольку из оставшихся k скобок берутся члены вида $(-c_j)$. Если вынести из этих членов за скобку $(-1)^k x^{n-k}$, то в скобке останется не что иное, как сумма σ_k [см. (40.14)].

Таким образом, коэффициентом при x^{n-k} в правой части будет $(-1)^k \sigma_k$; в левой части (после деления на старший коэффициент) коэффициент при x^{n-k} будет равняться a_k/a_0.

Приравнивая коэффициенты при степенях x, начиная с $(n-1)$-й и заканчивая нулевой, получим формулы Виета (40.15). □

Замечание 40.6. Вы, конечно, знаете формулы Виета для квадратного трехчлена. Для многочлена третьей степени

$$f(x) = a_0 x^3 + a_1 x^2 + a_2 x + a_3$$

эти формулы будут выглядеть следующим образом:

$$\begin{cases} \dfrac{a_1}{a_0} &= -(c_1+c_2+c_3)\,; \\ \dfrac{a_2}{a_0} &= c_1c_2+c_1c_3+c_2c_3\,; \\ \dfrac{a_3}{a_0} &= -c_1c_2c_3\,. \end{cases}$$

В качестве упражнения попробуйте выписать формулы Виета для многочлена четвертой степени.

Замечание 40.7. Формулы Виета позволяют однозначно (с точностью до пропорциональности) восстановить многочлен по его корням (в случае, когда их количество равно степени многочлена). Нормализованный многочлен восстанавливается однозначно: $a_0 = 1$; $a_k = (-1)^k\sigma_k$ $(k = 1, ..., n)$.

В задачниках по алгебре есть задачи типа: "Найти многочлен наименьшей степени, имеющий своими корнями числа 1, 2 (кратности не ниже 2), 3, -1 (кратности не ниже 3).

Ответ (нормализованный) дается следующей формулой:

$$f(x) = (x-1)(x-2)^2(x-3)(x+1)^3.$$

Надо только раскрыть скобки и привести подобные.

Но можно решать эту задачу "по Виету": составить список корней

$$[1, 2, 2, 3, -1, -1, -1];$$

вычислить суммы σ_j и найти коэффициенты a_j:

$$a_1 = -(1+2+2+3+(-1)+(-1)+(-1)) = -5$$

и т. д.; суммы для $k = 2, ..., 6$ будут получаться довольно громоздкими.

Может помочь Maple (для него первый способ решения реализовать проще).

```
> expand( (x-1) * (x-2) ^ 2 * (x-3) * (x+1) ^ 3 );
```

$$x^7 - 5x^6 + 2x^5 + 18x^4 - 11x^3 - 25x^2 + 8x + 12$$

§ 41. Схема Горнера

41.1. Схема Горнера для вычисления значений многочленов. Рассмотрим задачу практического вычисления значения $f(c)$ многочлена $f(x)$ с коэффициентами из поля P на элементе $c \in P$. (Напомним, что более корректным было бы говорить о значении полиномиальной функции f(c), но мы условились в замечании 39.1 пока не усложнять себе жизнь такими тонкостями.)

При высоких степенях многочленов задача вычисления значений может оказаться весьма трудоемкой. (Хочется обратиться к будущим программистам: знаете ли вы, сколько специфических трудностей таит в себе "простейшая" задача вычисления степени числа? Если нет, то откройте удивительную книгу П. Нодена и К. Китте "Алгебраическая алгоритмика"; М.: Мир, 1999.)

Так называемая схема Горнера является алгоритмом, в значительной степени упрощающим вычисление значений многочлена. Этот алгоритм основан на первом утверждении теоремы Безу [см. формулу (39.12), согласно которой $f(c)$ есть остаток от деления многочлена $f(x)$ на двучлен $x - c$].

Будем искать неполное частное $q(x)$ *методом неопределенных коэффициентов* (для других целей применявшимся уже в п. 5).

Пусть многочлен $f(x)$ задан формулой (36.18), а многочлен $q(x)$ — формулой

$$q(x) = b_0x^{n-1} + b_1x^{n-2} + ... + b_{n-2}x + b_{n-1}, \tag{41.1}$$

с неопределенными коэффициентами $b_0, b_1, ..., b_{n-2}, b_{n-1}$.

Раскроем скобки в равенстве

$$\begin{aligned} a_0x^n + a_1x^{n-1} + ... + a_{n-1}x + a_n = \\ = (x - c)(b_0x^{n-1} + b_1x^{n-2} + ... + b_{n-2}x + b_{n-1}) + f(c). \end{aligned} \tag{41.2}$$

Получим:

$$\begin{aligned} a_0x^n + \ a_1x^{n-1} + \ a_2x^{n-2} + ... + \ a_{n-2}x^2 + \ a_{n-1}x + \ a_n = \\ = b_0x^n + \ b_1x^{n-1} + \ b_2x^{n-2} + ... + \ b_{n-2}x^2 + \ b_{n-1}x - \\ - cb_0x^{n-1} - cb_1x^{n-2} - ... - cb_{n-3}x^2 - \ cb_{n-2}x \ - cb_{n-1} + f(c). \end{aligned}$$

Приравняем коэффициенты при одинаковых степенях x:

$$\begin{cases} (x^n) & : a_0 & = b_0 & & ; \\ (x^{n-1}) & : a_1 & = b_1 & - cb_0 & ; \\ (x^{n-2}) & : a_2 & = b_2 & - cb_1 & ; \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots \\ (x^2) & : a_{n-2} & = b_{n-2} & - cb_{n-3} & ; \\ (x^1) & : a_{n-1} & = b_{n-1} & - cb_{n-2} & ; \\ (x^0) & : a_n & = f(c) & - cb_{n-1} & . \end{cases} \tag{41.3}$$

Систему (41.3) легко разрешить относительно неопределенных коэффициентов b_k $(k = 0, ..., n-1)$ и искомого значения $f(c)$:

$$\begin{cases} b_0 & = a_0 & & ; \\ b_1 & = a_1 & + cb_0 & ; \\ b_2 & = a_2 & + cb_1 & ; \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots \\ b_{n-2} & = a_{n-2} & + cb_{n-3} & ; \\ b_{n-1} & = a_{n-1} & + cb_{n-2} & ; \\ f(c) & = a_n & + cb_{n-1} & . \end{cases} \tag{41.4}$$

Соотношения (41.4) относятся к типу *рекуррентных*, характерных тем, что при отыскании очередного (по номеру) значения используются ранее найденные значения с меньшими номерами (см. § 30а). В данном случае сразу определяется b_0; при отыскании b_k используется ранее найденное значение b_{k-1} $(k = 1, ..., n-1)$; при вычислении $f(c)$ — значение b_{n-1}.

Удобно поместить данные и результаты вычислений по формулам (41.4) в таблицу, которая и называется *схемой Горнера* (см. табл. 41.1а в прил. 2).

Замечание 41.1. Заметим, что схема Горнера дает не только значение многочлена $f(c)$, являющееся остатком от деления данного многочлена на $x - c$, но и неполное частное (41.1).

Кроме того, получив нулевой остаток, мы можем констатировать, что элемент c является корнем многочлена $f(x)$.

Пример 41.1. Дан многочлен

$$f(x) = 2x^4 - (1 - 2i)x^3 - (2 + 3i)x - 4$$

над полем $\mathbb{C}$. Требуется поделить с остатком $f(x)$ на двучлен $x-c$, где $c=1+i$; в частности вычислить значение $f(c)$.

Р е ш е н и е см. в табл. 41.1б.

О т в е т:

$$f(x)=$$
$$=(\,2x^3+(1+4i)x^2+(-3+5i)x-(10+i)\,)(\,x-(1+i)\,)+(-13-11i);$$
$$f(1+i)=-13-11i.$$

41.2. Разложение многочлена по степеням $x-c$ (формула Тейлора). Выполним деление с остатком многочлена $f(x)$ степени $n>0$ на двучлен $x-c$ [см. (41.2)]; результат, имея в виду предстоящее повторение (итерирование) процесса, представим следующим образом:

$$f(x)=q_0(x)(x-c)+h_0;\ \deg(q_0(x))=n-1;\ h_0=f(c), \qquad (41.5)$$

где неполное частное $q_0(x)$ является многочленом степени $n-1$.

Если $n-1>0$, то поделим с остатком на $x-c$ многочлен $q_0(x)$:

$$q_0(x)=q_1(x)(x-c)+h_1;\ \deg(q_1(x))=n-2;\ h_1=q_0(c). \qquad (41.6)$$

Подставим (41.6) в (41.5) и сгруппируем члены по степеням $x-c$:

$$f(x)=q_1(x)(x-c)^2+h_1(x-c)+h_0. \qquad (41.7)$$

Если $n-2>0$, то поделим с остатком на $x-c$ многочлен $q_1(x)$; результат подставим в (41.7) и т. д.

На шаге с номером $n-1$ неполное частное $q_{n-2}(x)$ будет многочленом первой степени и получится разложение

$$f(x)=q_{n-2}(x)(x-c)^{n-1}+h_{n-2}(x-c)^{n-2}+\ldots+h_2(x-c)^2+h_1(x-c)+h_0.$$

На следующем шаге, с номером n, неполное частное станет скаляром, который мы обозначим h_n, и получится, уже в окончательном виде, формула

$$f(x)=h_n(x-c)^n+h_{n-1}(x-c)^{n-1}+$$
$$+h_{n-2}(x-c)^{n-2}+\ldots+h_2(x-c)^2+h_1(x-c)+h_0, \qquad (41.8)$$

называемая *разложением* (или, несколько жаргонно, *переразложением*) $f(x)$ по степеням $x-c$.

Замечание 41.2. Правая часть этой формулы является многочленом степени n от новой переменной $x-c$. Коэффициенты этого многочлена получились занумерованными в порядке возрастания степеней. Свободный член h_0 равен значению $f(c)$. Старший коэффициент h_n, как нетрудно понять, раскрыв степени по формуле бинома Ньютона (ср. с замечанием 39.6), равен старшему коэффициенту a_0 исходного многочлена $f(x)$.

Формулу (41.8) называют также *формулой Тейлора для многочлена $f(x)$ в точке c.*

Замечание 41.3. Формула с таким названием является одной из самых знаменитых и важнейших формул математического анализа. В анализе она записывается не только для многочленов, но и — с дополнительным (остаточным) членом — для произвольных *функций* от *действительной* или *комплексной* переменной, достаточное число раз *дифференцируемых* в окрестности точки c; *коэффициенты Тейлора* h_k $(k=0,\dots,n)$ выражаются через производные $f^{(k)}(c)$.

В алгебре изучаются многочлены на полях произвольной природы, без привлечения понятия предельного перехода (являющегося в анализе необходимой предпосылкой для определения производной). Тем не менее производные (для многочленов) можно определить и в алгебре, "своими" методами, не использующими предельного перехода. Этим мы займемся ниже, в § 47.

Выражения h_k через производные также будут нами получены (см. п. 47.3) чисто алгебраическими средствами.

А сейчас мы изучим метод вычисления коэффициентов формулы Тейлора с помощью схемы Горнера.

Осуществляя по Горнеру каждое из делений на $x-c$, мы придем к "ступенчатой" схеме (см. табл. 41.2а в прил. 2). К схеме добавлена "для красоты" еще одна, одноклеточная строка, которую можно понимать как изображение деления с остатком многочлена нулевой степени $q_n(x)=h_n$ на $x-c$: неполное частное будет нулевым, а остаток будет совпадать с h_n; именно этот элемент занимает последнюю клеточку таблицы.

Обратите внимание на то, что во всем столбце старших коэффициентов будет фигурировать один и тот же элемент a_0.

Искомые коэффициенты

$$h_0,\ h_1,\ h_2,\ ...,\ h_{n-2},\ h_{n-1},\ h_n$$

(остатки) будут последними элементами в строках таблицы, начиная со второй.

Пример 41.2. Дан многочлен

$$f(x) = 8x^6 - 44x^5 + 50x^4 + 55x^3 - 55x^2 - 56x - 12$$

над полем $\mathbb{Q}$. Требуется переразложить его по степеням $x - 2$.

Р е ш е н и е см. в табл. 41.2б (прил. 2).

О т в е т:

$$f(x) = -125(x-2)^2 - 25(x-2)^3 + 90(x-2)^4 + 52(x-2)^5 + 8(x-2)^6.$$

41.3. Определение кратности корня многочлена с помощью схемы Горнера. Посмотрите еще раз на ответ примера 41.2 и на диагональ в табл. 41.2б, содержащую коэффициенты формулы Тейлора.

Из ответа ясно, что $c = 2$ является корнем многочлена $f(x)$, причем этот многочлен, очевидно, делится на $(x-2)^2$ и в частном получается многочлен, который уже не делится на $x - 2$. Значит, кратность $c = 2$ как корня $f(x)$ равна 2.

Из таблицы усматривается следующее правило, справедливое в общем случае.

П р а в и л о: Кратность корня c многочлена $f(x)$ равна m тогда и только тогда, когда в табл. 41.2а, на диагонали, содержащей остатки, имеется в начале m нулей подряд: $h_0 = h_1 = ... = h_{m-1} = 0$, а следующий остаток отличен от нуля: $h_m \neq 0$.

Пример 41.3. Решим типовую "задачу с параметрами" на схему Горнера.

З а д а ч а: Выяснить, при каких значениях параметров A и B многочлен $f(x) = Ax^4 + Bx^3 + 1$ делится на $(x-1)^2$.

Р е ш е н и е: Задачу можно переформулировать следующим образом: при каких значениях A и B число $c = 1$ является корнем многочлена $f(x)$, кратности не ниже 2?

Выпишем ступенчатую схему Горнера с тремя ступеньками (см. табл. 41.3). Кратность корня будет не ниже двух тогда и только

тогда, когда $A+B+1=0$ и $4A+3B=0$. Этими условиями параметры однозначно определены: $A=3$, $B=-4$.

А третью ступеньку мы изобразили для того, чтобы узнать, каково будет точное значение кратности: не будет ли оно больше 2?

(Здесь автору хотелось бы продемонстрировать, как может проявиться одно из самых ценных свойств студента — любознательность. Получив и решив задачу, хороший студент задумывается:

— нельзя ли извлечь из условия дополнительную информацию?

— нельзя ли утверждение задачи усилить (получить из условия больше, чем предписывается заданием)?

— может быть, условие задачи можно "ослабить" (получить тот же результат при более слабых ограничениях)?

— а что если на используемом материале придумать более интересную задачу?

— и так далее, и тому подобное.)

Третий остаток $6A+4B=2\neq 0$, следовательно, кратность в этом случае в точности равна двум.

О т в е т : $(x-1)^2|f(x)$ тогда и только тогда, когда $A=3$, $B=-4$; $(x-1)^3$ не делит $f(x)$ ни при каких значениях A и B.

§ 42. Рациональные корни многочленов с рациональными коэффициентами

42.1. Многочлены с рациональными коэффициентами и многочлены с целыми коэффициентами. Рассмотрим многочлен (36.18), положительной степени n, с коэффициентами из поля рациональных чисел $\mathbb{Q}$:

$$a_i=\frac{p_i}{q_i};\ p_i,q_i\in\mathbb{Z};q_i\neq 0\ (i=0,...,n). \tag{42.1}$$

Представление рационального числа в виде дроби $a=p/q$ (где $p,q\in\mathbb{Z}$; $q\neq 0$) определено не однозначно: дробь p'/q' $(q'\neq 0)$ определяет то же самое число a тогда и только тогда, когда $pq'=qp'$ (в кольце $\mathbb{Z}$). Дробь всегда можно выбрать *нормализованной*, т. е. *несократимой*, с положительным знаменателем.

С другой стороны, набор дробей (42.1) можно привести к общему

(натуральному) знаменателю $q = [q_0, q_1, ..., q_n]$:

$$a_i = \frac{p'_i}{q};\ p'_i \in \mathbb{Z}\ (i = 0, ..., n);\ q \in \mathbb{N}. \tag{42.2}$$

Для целых чисел p'_i $(i = 0, ..., n)$ можно найти (натуральный) наибольший общий делитель $p' = (p'_0, p'_1, ..., p'_n)$ и представить эти числа в виде

$$p'_i = p'b_i;\ b_i \in \mathbb{Z}\ (i = 0, ..., n). \tag{42.3}$$

Образовавшийся общий множитель p'/q можно вынести за скобки; в скобках останется многочлен $g(x)$ с целыми *взаимно простыми* коэффициентами:

$$f(x) = \frac{p'}{q}g(x); \tag{42.4}$$

$$g(x) = b_0x^n + b_1x_{n-1} + ... + b_{n-1}x + b_n;$$
$$b_i \in \mathbb{Z}\ (i = 0, ..., n);\ b_0 \neq 0;\ (b_0, b_1, ..., b_n) = 1. \tag{42.5}$$

Данный многочлен $f(x) \in \mathbb{Q}[x]$ оказывается пропорциональным (с рациональным коэффициентом пропорциональности) многочлену $g(x) \in \mathbb{Z}[x]$ с взаимно простыми коэффициентами.

Корнями исходного многочлена $f(x)$ будут корни многочлена $g(x)$ и только они.

Поэтому рассматриваемая далее задача отыскания *рациональных корней* для многочлена с рациональными коэффициентами может быть сведена к аналогичной задаче для многочлена с целыми (и даже взаимно простыми) коэффициентами.

42.2. Рациональные корни многочлена с целыми коэффициентами. Рассмотрим многочлен вида (36.18), положительной степени n, с целыми коэффициентами:

$$f(x) = a_0x^n + a_1x^{n-1} + ... + a_{n-1}x + a_n;$$
$$a_i \in \mathbb{Z}\ (i = 0, ..., n);\ a_0 \neq 0. \tag{42.6}$$

Если свободный член $a_n = 0$, то многочлен $f(x)$ будет делиться на x (что равносильно наличию нулевого корня). Вынося за скобку наибольшую степень x, на которую делится данный многочлен, мы получим в скобках новый многочлен, с ненулевым свободным членом, который подлежит дальнейшему исследованию.

В связи с этим наложим с самого начала ограничение $a_n \neq 0$.

Кроме того, исследование лишь облегчится, если вынести за скобку наибольший общий делитель коэффициентов данного многочлена. Поэтому мы будем изначально считать коэффициенты многочлена взаимно простыми. (Для формулировки следующего предложения это не существенно, но при практических вычислениях по алгоритму следующего пункта предполагаемая взаимная простота коэффициентов может заметно *сократить перебор* и ускорить работу.)

Рациональные корни у многочлена (42.6) вообще могут отсутствовать, но если они есть, то их легко найти все с помощью простого перебора. Следующее предложение является совершенно элементарным и "школьным", его обязательно изучают в математических классах и с ним непременно знакомятся школьники, ориентирующиеся на поступление в "серьезные" вузы.

Предложение 42.1. *Рассмотрим многочлен (42.6), с целыми коэффициентами, положительной степени, с ненулевым свободным членом, а также несократимую дробь $c = s/t \in \mathbb{Q}$. Если число c является корнем данного многочлена, то*

1) числитель s делит свободный член a_n;

2) знаменатель t делит старший коэффициент a_0;

3) для любого целого u число $s - ut$ делит значение $f(u)$.

Доказательство. 1. Пусть $f(c) = 0$, т. е. выполняется равенство

$$a_0 \left(\frac{s}{t}\right)^n + a_1 \left(\frac{s}{t}\right)^{n-1} + ... + a_{n-1}\frac{s}{t} + a_n = 0. \tag{42.7}$$

Домножая обе части (42.7) на t^n, мы получим следующее равенство в кольце $\mathbb{Z}$:

$$a_0 s^n + a_1 s^{n-1} t + ... + a_{n-1} s t^{n-1} + a_n t^n = 0. \tag{42.8}$$

Все члены в левой части (42.8), кроме последнего, содержат множитель s; значит, и последний член $a_n t^n$ должен делиться на s. По предположению числа s и t являются взаимно простыми. Это влечет, во-первых, взаимную простоту s и t^n и, во-вторых, факт делимости на s числа a_n (использованы применительно к кольцу $\mathbb{Z}$ свойства отношения делимости и свойства взаимно простых элементов из предложения 38.5).

2. Рассуждая совершенно аналогично, мы получим сначала $t | a_0 s^n$ и затем $t | a_0$.

3. Произведем в многочлене $f(x)$ замену переменной $x = y + u$ (сдвиг). Получим новый многочлен

$$\begin{aligned} h(y) = f(x) = f(y+u) = \\ = a_0(y+u)^n + a_1(y+u)^{n-1} + ... + a_{n-1}(y+u) + a_n = \\ = v_0 y^n + v_1 y^{n-1} + ... + v_{n-1} y + v_n, \end{aligned}$$

где $v_i \in \mathbb{Z}$ $(i = 0, ..., n)$; $v_0 = a_0$; $v_n = h(0) = f(u)$.

Число $c \in \mathbb{Q}$ будет корнем многочлена $f(x)$ тогда и только тогда, когда число $c - u$ будет корнем $h(y)$. [В самом деле, $h(c-u) = f(c)$.]

В частности, если дробь $c = \frac{s}{t}$ несократима и является корнем $f(x)$, то выражение $c - u = \frac{s-ut}{t}$ также является несократимой дробью и будет корнем многочлена $h(y)$.

По уже доказанному в первом пункте доказательства, числитель этой дроби $s - ut$ должен делить свободный член $v_n = f(u)$. □

Следствие. *Если многочлен (42.6) является нормализованным, т. е. $a_0 = 1$, то всякий его рациональный корень является целым.*

Доказательство. Если $t|1$, то $t = \pm 1$ и число $c \in \mathbb{Z}$. □

42.3. Алгоритм отыскания всех рациональных корней для многочлена с целыми коэффициентами. В силу предложения 42.1, задача отыскания рациональных корней для многочлена (42.6), с целыми коэффициентами и ненулевым свободным членом, сводится к задаче перебора всех (можно считать, что — нормализованных) дробей $c = s/t$, таких, что $s \in \mathcal{S}$ и $t \in \mathcal{T}$, где введены обозначения: $\mathcal{S}$ — множество всех делителей свободного члена a_n; $\mathcal{T}$ — множество всех *положительных* делителей старшего коэффициента a_0.

Множество всех "подозрительных" дробей (повторяющиеся элементы, разумеется, исключаются) будет обозначаться $\mathcal{C}$.

Проверку каждой из дробей $c \in \mathcal{C}$ следует оформлять с помощью *схемы Горнера*, одновременно определяя *кратность* обнаруженного корня. Понадобится всего одна (ступенчатая) схема Горнера (см. ниже пример 42.1).

Отыскание каждого очередного корня уменьшает степень подлежащего дальнейшему исследованию многочлена на столько единиц, какова кратность найденного корня.

Некоторые пробы будут давать отрицательный результат. Соответствующие строки подлежат удалению (вычеркиванию) из ступенчатой схемы Горнера. (Следующую строку мы будем просчитывать так, как будто предыдущей, вычеркнутой нет.)

В связи с тем что список "подозрительных" дробей может оказаться довольно длинным, при ручном счете бывает уместным сократить перебор с помощью некоторых "тестов". Опишем два из них (наиболее употребительные).

Прежде всего заметим, что множество $\mathcal{C}$ для многочлена $f(x)$ обязательно содержит числа ± 1. Именно с проверки этих чисел мы рекомендуем начинать работу.

Допустим, что число 1 является корнем $f(x)$ кратности k, а число -1 — корнем кратности l. (Не исключается случай, когда одна или обе из этих кратностей обращаются в нуль.)

Тогда можно записать разложение на множители

$$f(x) = (x-1)^k (x+1)^l g(x), \tag{42.9}$$

где многочлен

$$g(x) = b_0 x^m + b_1 x^{m-1} + ... + b_{m-1} x + b_m \tag{42.10}$$

имеет степень $m = n - k - l$ и числа ± 1 не являются его корнями, т. е. $g(1) \neq 0$ и $g(-1) \neq 0$.

Отбор остальных "подозрительных" дробей, т. е. формирование множеств

$$\mathcal{S} = \{s \in \mathbb{Z} : s | b_m\},\ \mathcal{T} = \{t \in \mathbb{Z} : (t | b_0) \wedge (t > 0)\} \tag{42.11}$$

и

$$\mathcal{C} = \{\frac{s}{t} \in \mathbb{Q} : s \in \mathcal{S},\, t \in \mathcal{T}\}, \tag{42.12}$$

следует производить уже для многочлена $g(x)$, причем числа ± 1 из множества $\mathcal{C}$ можно сразу исключить.

(Напомним, что повторений в множестве $\mathcal{C}$ быть не должно и что все включенные в него дроби должны быть нормализованными. Целые числа, принадлежащие $\mathcal{C}$, мы также будем представлять как дроби с единичным знаменателем.)

Для дальнейшего понадобятся значения $g(1)$ и $g(-1)$, также определяемые по Горнеру.

Два обещанных теста таковы:

(test 1) если $c = s/t$ является корнем $g(x)$, то $t - s | g(1)$;

(test 2) если $c = s/t$ является корнем $g(x)$, то $t + s | g(-1)$.

Оба этих теста основываются на третьем утверждении предложения 42.1 (применяемом к многочлену $g(x)$ при $u = \pm 1$).

Пример 42.1. Рассмотрим многочлен с целыми коэффициентами:

$$f(x) = 12x^7 - 4x^6 - 17x^5 - 18x^4 + 37x^3 - 14x + 4.$$

Найдем все рациональные корни этого многочлена.

На "подготовительном этапе" проверим по Горнеру, являются ли корнями этого многочлена числа 1 и -1. (И если да, то — какой кратности.) Вычисления представлены в табл. 42.1а (прил. 2).

Оказалось, что 1 является корнем кратности 2, а -1 корнем не является. Получено разложение на множители

$$f(x) = (x-1)^2 g(x),$$

где многочлен

$$g(x) = 12x^5 + 20x^4 + 11x^3 - 16x^2 - 6x + 4$$

уже не имеет числа ± 1 своими корнями.

Попутно определены значения многочлена:

$$g(1) = 25;\ g(-1) = -9.$$

Формируем множество $\mathcal{S}$ всех целых делителей свободного члена $b_5 = 4$ многочлена $g(x)$ и множество $\mathcal{T}$ всех натуральных делителей старшего коэффициента $b_0 = 12$ этого многочлена:

$$\mathcal{S} = \{-4, -2, -1, 1, 2, 4\};\ \mathcal{T} = \{1, 2, 3, 4, 6, 12\}.$$

Формируем множество $\mathcal{C}$, состоящее из всевозможных дробей s/t, таких, что $s \in \mathcal{S}$, $t \in \mathcal{T}$ (целые числа записываем как дроби с единичным знаменателем):

$$\mathcal{C} = \left\{ \frac{1}{1}; \frac{-1}{1}; \frac{2}{1}; \frac{-2}{1}; \frac{4}{1}; \frac{-4}{1}; \frac{1}{2}; \frac{-1}{2}; \right.$$
$$\left. \frac{1}{3}; \frac{-1}{3}; \frac{2}{3}; \frac{-2}{3}; \frac{4}{3}; \frac{-4}{3}; \frac{1}{4}; \frac{-1}{4}; \frac{1}{6}; \frac{-1}{6}; \frac{1}{12}; \frac{-1}{12} \right\}.$$

Выбрасываем из множества $\mathcal{C}$ дроби, равные ± 1 (поскольку по построению $g(x)$ они не являются корнями этого многочлена).

К оставшимся дробям применяем **(test 1)** и **(test 2)**: разность $s - t$ должна делить число $g(1) = 25$, а сумма $s + t$ должна делить

$g(-1) = -9$. (Например, число $4/3$ проходит первый тест, т. к. $4-3 = 1$ делит 25, но не проходит второй, т. к. $4+3 = 7$ не делит -9.)

Оба теста выдерживают дроби из множества

$$\mathcal{C}' = \left\{2; -4; \frac{1}{2}; -\frac{2}{3}; -\frac{1}{4}; \right\}$$

(здесь мы уже используем обычную запись целых чисел).

Каждое число из множества $\mathcal{C}'$ мы проверяем по Горнеру. Некоторые строки в таблице "срабатывают": они заканчиваются нулями, что свидетельствует о том, что проверяемое число является корнем (кратность корня равняется количеству "заработанных" этим корнем нулей). Такие строки мы выделяем (тонируем). Остальные, "не сработавшие" строки мы (мысленно) удаляем: считаем их при переходе к следующей строке не существующими.

Например, в табл. 42.1б после двух неудачных попыток "сработала" строка, соответствующая числу $1/2$. "Сработала" и следующая строка; это свидетельствует о том, что $1/2$ является корнем $g(x)$ кратности не ниже 2. Следующая попытка оказалась неудачной. Значит, кратность корня $1/2$ в точности равна 2.

Дальнейшему исследованию подлежит последнее из полных частных, полученное как результат деления $g(x)$ на $(x - 1/2)^2$ и фигурирующее (в виде вектора коэффициентов) в пятой (тонированной) строке.

Игнорируя неудачную шестую строку, мы подвергаем испытанию дробь $-2/3$, которая оказывается корнем кратности 1.

(Два последних испытания, восьмая и девятая строки, в принципе, можно было не делать, поскольку в седьмой строке получилось полное частное $12(x^2 + 2x + 2)$, являющееся квадратным трехчленом, корни которого можно найти по формуле. В данном случае дискриминант квадратного трехчлена оказывается отрицательным, следовательно, не только рациональных, но и действительных корней больше не будет.)

Замечание 42.1. Maple легко находит рациональные корни для многочленов с целыми коэффициентами. Мы уже указывали в примере 40.2, что, по умолчанию, система настроена так, что корни ищутся в наименьшем из полей, содержащем коэффициенты многочлена. В случае коэффициентов из $\mathbb{Z}$ таким полем является $\mathbb{Q}$, поэтому команда roots (без "дополнительных просьб") возвращает лишь рациональные корни для многочлена с целыми коэффициентами.

§ 43. Многочлены с действительными коэффициентами и их разложение на линейные и квадратичные множители

43.1. Сопряженные многочлены для многочленов с комплексными коэффициентами. Рассмотрим многочлен $f(x)$ над полем $\mathbb{C}$, заданный формулой (36.18). К каждому из его коэффициентов можно применить операцию *комплексного сопряжения* (см. п. 31.3); при этом получится новый многочлен (такой же степени), называемый *сопряженным* данному:

$$\overline{f}(x) = \overline{a_0}\, x^n + \overline{a_1}\, x^{n-1} + \ \ldots\ + \overline{a_{n-1}}\, x + \overline{a_n}\,. \tag{43.1}$$

Установим связь между *значениями* (см. замечание 39.1) многочленов (36.18) и (43.1) в произвольной точке $c \in \mathbb{C}$.

Предложение 43.1. *1. Для любого комплексного числа c значения сопряженных друг другу многочленов связаны формулой*

$$\overline{f}(\overline{c}) = \overline{f(c)}\,. \tag{43.2}$$

2. Число $c \in \mathbb{C}$ является корнем (кратности k) для многочлена $f(x)$ тогда и только тогда, когда число $\overline{c}$ является корнем (такой же кратности) для многочлена $\overline{f}(x)$.

Доказательство. 1. Первое утверждение предложения очевидным образом вытекает из свойств операции комплексного сопряжения (см. предложение 31.1):

$$\overline{a_0c^n + a_1c^{n-1} + \ldots + a_{n-1}c + a_n} = \\ = \overline{a_0}\cdot\overline{c}^{\,n} + \overline{a_1}\cdot\overline{c}^{\,n-1} + \ldots + \overline{a_{n-1}}\cdot\overline{c} + \overline{a_n}\,.$$

2. Из формулы (43.2) немедленно следует, что c является корнем $f(x)$ тогда и только тогда, когда $\overline{c}$ является корнем $\overline{f}(x)$. Остается доказать совпадение кратностей этих корней.

То, что кратность c как корня $f(x)$ равняется k, означает, что имеет место разложение на множители

$$f(x) = (x-c)^k\, g(x), \tag{43.3}$$

где $g(c) \neq 0$.

Посмотрим на равенство (43.3) как на равенство *полиномиальных функций* (справедливое для произвольных значений $x \in \mathbb{C}$) и применим к обеим его частям операцию сопряжения, снова пользуясь предложением 31.1:

$$\overline{f(x)} = (\overline{x} - \overline{c})^k \, \overline{g(x)}.$$

Применим теперь (дважды) формулу (43.2):

$$\overline{f}(\overline{x}) = (\overline{x} - \overline{c})^k \, \overline{g}(\overline{x}).$$

Произведя в последней формуле замену переменной $y = \overline{x}$ (переменная y, так же как x, пробегает все $\mathbb{C}$), мы получим равенство полиномиальных функций

$$(\forall y \in \mathbb{C})[\, \overline{f}(y) = (y - \overline{c})^k \, \overline{g}(y) \,].$$

Бесконечность поля $\mathbb{C}$ гарантирует (в силу предложения 39.2) равносильность равенства многочленов и равенства соответствующих полиномиальных функций. Поэтому справедливо следующее равенство многочленов (мы снова возвращаемся к переменной x):

$$\overline{f}(x) = (x - \overline{c})^k \, \overline{g}(x). \tag{43.4}$$

Кроме того, имеет место неравенство

$$\overline{g}(\overline{c}) \overset{(43.2)}{=\!=} \overline{g(c)} \neq 0,$$

приводящее нас к выводу, что число $\overline{c}$ является корнем многочлена $\overline{g}(x)$ кратности, в точности равной k. $\square$

43.2. Комплексные корни для многочленов с действительными коэффициентами. Рассмотрим многочлен $f(x)$ вида (36.18), положительной степени n, с коэффициентами из поля действительных чисел $\mathbb{R}$.

Многочлен $f(x)$ можно рассматривать и над более широким полем $\mathbb{C}$ (являющимся алгебраическим замыканием поля $\mathbb{R}$; см. замечание 40.4). В поле $\mathbb{C}$ этот многочлен имеет ровно n корней (с учетом кратностей) и разлагается на линейные множители по формуле (40.10).

Среди комплексных корней многочлена $f(x)$ могут присутствовать действительные; занесем их в список

$$c_1, \, c_2, \, ..., \, c_s, \tag{43.5}$$

где $0 \leqslant s \leqslant n$, и сформируем список соответствующих кратностей:

$$m_1,\ m_2,\ ...,\ m_s. \tag{43.5a}$$

Остальные $n - s$ корней будут *комплексными, не действительными*. Сейчас мы докажем, что число $n - s$ является *четным*, по той причине, что вместе с каждым комплексным корнем многочлен с действительными коэффициентами имеет *сопряженный* комплексный корень, причем такой же кратности. [Для случая действительных корней это утверждение ничего не дает, поскольку является тавтологичным, а для случая *недействительных* корней приводит к выводу о том, что такие корни встречаются *парами* вида (*корень, сопряженный корень*).]

Предложение 43.2. *Если комплексное число c является корнем (кратности k) многочлена $f(x)$ с действительными коэффициентами, то и сопряженное комплексное число $\overline{c}$ является корнем этого многочлена (такой же кратности).*

Доказательство немедленно следует из второго утверждения предложения 43.1 с учетом того, что многочлен с действительными коэффициентами, рассматриваемый над полем комплексных чисел, является, очевидно, "самосопряженным". □

Следствие. *Всякий многочлен нечетной степени с действительными коэффициентами имеет хотя бы один действительный корень.*

Доказательство очевидно. □

Продолжим рассмотрение совокупности корней многочлена $f(x)$.

Обозначим (сгруппированные попарно) недействительные корни этого многочлена

$$w_1,\ \overline{w_1},\ w_2,\ \overline{w_2},\ ...,\ w_t,\ \overline{w_t} \tag{43.6}$$

и рассмотрим соответствующий список кратностей

$$l_1,\ l_1,\ l_2,\ l_2,\ ...,\ l_t,\ l_t. \tag{43.6a}$$

Объединим списки (43.5) и (43.6), а также соответствующие списки кратностей. Должно выполняться равенство

$$m_1 + m_2 + ... + m_s + 2(l_1 + l_2 + ... + l_t) = n. \tag{43.7}$$

43.3. Разложение многочлена с действительными коэффициентами на линейные и квадратичные множители. Разложим теперь многочлен $f(x) \in \mathbb{R}[x]$ над полем $\mathbb{C}$ на линейные множители по формуле (40.10), используя описанную в предыдущем пункте группировку корней:

$$f(x) = a_0 \cdot (x - c_1)^{m_1} \cdot \ldots \cdot (x - c_s)^{m_s} \cdot \\ \cdot [(x - w_1)(x - \overline{w_1})]^{l_1} \cdot \ldots \cdot [(x - w_t)(x - \overline{w_t})]^{l_t}. \quad (43.8)$$

Представим далее комплексные числа w_k $(k = 1, ..., t)$ и числа, им сопряженные, в алгебраической форме, а затем раскроем в формуле (43.8) круглые скобки внутри квадратных:

$$w_k = \alpha_k + \beta_k i;\ \overline{w_k} = \alpha_k - \beta_k i;\ \alpha_k, \beta_k \in \mathbb{R};\ \beta_k \neq 0; \quad (43.9)$$

$$(x - w_k)(x - \overline{w_k}) = x^2 - (w_k + \overline{w_k})x + w_k\overline{w_k} = \\ = x^2 - 2\alpha_k x + \alpha_k^2 + \beta_k^2 = x^2 + p_k x + q_k, \quad (43.10)$$

где введены обозначения

$$p_k = 2\alpha_k = 2\mathrm{Re}(w_k);\ q_k = \alpha_k^2 + \beta_k^2 = |w_k|^2. \quad (43.11)$$

Выражения (43.10) представляют собой квадратные трехчлены с действительными коэффициентами, которые по самому построению не имеют действительных корней и, следовательно, имеют отрицательные дискриминанты. (Последний вывод, впрочем, можно получить и непосредственно: $D_k = -4\beta_k^2 < 0$, поскольку $\beta_k \neq 0$.)

Таким образом, доказано следующее

Предложение 43.3. *Рассмотрим многочлен $f(x) \in \mathbb{R}[x]$ вида (36.18), положительной степени n, и список его действительных и недействительных корней*

$$c_1,\ c_2,\ ...,\ c_s;\ w_1,\ \overline{w_1},\ w_2,\ \overline{w_2},\ ...,\ w_t,\ \overline{w_t} \quad (43.12)$$

вместе со списком соответствующих кратностей

$$m_1,\ m_2,\ ...,\ m_s;\ l_1,\ l_1,\ l_2,\ l_2,\ ...,\ l_t,\ l_t. \quad (43.12a)$$

Тогда имеет место разложение многочлена $f(x)$ на линейные и квадратичные (с отрицательным дискриминантом) множители (над полем $\mathbb{R}$):

$$f(x) = a_0 \cdot (x - c_1)^{m_1} \cdot \ldots \cdot (x - c_s)^{m_s} \cdot \\ \cdot (x^2 + p_1 x + q_1)^{l_1} \cdot \ldots \cdot (x^2 + p_t x + q_t)^{l_t}, \quad (43.13)$$

в котором

— каждому действительному корню c_j $(j = 1, \ldots, s)$ отвечает линейный множитель $x - c_j$, входящий в разложение в степени, равной кратности m_j этого корня;

— каждой паре сопряженных корней w_k, $\overline{w_k}$ $(k = 1, \ldots, t)$ отвечает квадратичный множитель $x^2 + p_k x + q_k$, где $p_k = 2Re(w_k)$ и $q_k = \alpha_k^2 + \beta_k^2 = |w_k|^2$; этот множитель имеет отрицательный дискриминант, не разлагается над полем $\mathbb{R}$ на линейные множители и входит в разложение для $f(x)$ в степени, равной общей кратности l_k корней w_k и $\overline{w_k}$.

Доказательство изложено выше, перед формулировкой предложения. □

43.4. Примеры разложения многочленов над полем действительных чисел. Общий метод разложения на множители для многочлена с действительными коэффициентами состоит в том, чтобы разложить этот многочлен (над полем комплексных чисел) на линейные множители, затем сгруппировать множители, отвечающие комплексно сопряженным корням, так, что получатся уже квадратичные (но зато действительные) множители.

В некоторых примерах, однако, возможно применение (искусственных) "чисто действительных приемов", приводящих к разложению над полем $\mathbb{R}$, без использования разложения над $\mathbb{C}$.

Пример 43.1. Разложим на линейные множители над полем $\mathbb{C}$ и на линейные и квадратичные множители над полем $\mathbb{R}$ многочлен $f(x) = x^4 + 4$.

Чтобы найти комплексные корни этого многочлена, необходимо извлечь в поле $\mathbb{C}$ корень четвертой степени из числа -2. Сделать это можно с помощью перехода к тригонометрической форме записи комплексных чисел [см. (более сложный) образец в примере 32.6; пусть вас только не смущает тот факт, что комплексная неизвестная обозначается здесь не z, а x].

Корнями будут числа: $1+i$, $-1+i$, $-1-i$ и $1-i$. Разложение на линейные множители над $\mathbb{C}$ (с группировкой множителей, отвечающих комплексно сопряженным корням) будет иметь вид

$$f(x) = [(x-(1+i))(x-(1-i))] \cdot [(x-(-1+i))(x-(-1-i))].$$

Перемножая круглые скобки внутри квадратных [или пользуясь готовыми формулами (43.11)], получим разложение многочлена $f(x)$ на квадратичные (с отрицательным дискриминантом) множители над $\mathbb{R}$:

$$f(x) = (x^2 - 2x + 2)(x^2 + 2x + 2).$$

Это же самое разложение можно получить (элементарным) искусственным приемом "прибавить и отнять":

$$f(x) = x^4 + 4 = (x^4 + 4x^2 + 4) - 4x^2 = \\ = (x^2+2)^2 - (2x)^2 = (x^2 - 2x + 2)(x^2 + 2x + 2).$$

Пример 43.2. В "совсем школьном" примере $f(x) = x^3 - 1$ мы получим разложение над $\mathbb{R}$ с помощью формулы "разность кубов":

$$f(x) = (x-1)(x^2 + x + 1).$$

Разложение над $\mathbb{C}$ будет иметь вид

$$f(x) = (x-1)(x-(-\frac{1}{2} + \frac{\sqrt{3}}{2}i))(x-(-\frac{1}{2} - \frac{\sqrt{3}}{2}i)).$$

Пример 43.3. Еще один пример, связанный с извлечением корней из комплексных чисел: $f(x) = x^8 + 1$. Главное значение корня восьмой степени из -1 будет иметь аргумент $\pi/8$; понадобятся формулы (32.15) для синуса и косинуса этого угла.

Разложение над $\mathbb{C}$:

$$f(x) = (x-(\frac{\sqrt{2+\sqrt{2}}}{2} + i\frac{\sqrt{2-\sqrt{2}}}{2})) \cdot (x-(\frac{\sqrt{2-\sqrt{2}}}{2} + i\frac{\sqrt{2+\sqrt{2}}}{2})) \cdot \\ \cdot (x-(-\frac{\sqrt{2-\sqrt{2}}}{2} + i\frac{\sqrt{2+\sqrt{2}}}{2})) \cdot (x-(-\frac{\sqrt{2+\sqrt{2}}}{2} + i\frac{\sqrt{2-\sqrt{2}}}{2})) \cdot \\ \cdot (x-(-\frac{\sqrt{2+\sqrt{2}}}{2} - i\frac{\sqrt{2-\sqrt{2}}}{2})) \cdot (x-(-\frac{\sqrt{2-\sqrt{2}}}{2} - i\frac{\sqrt{2+\sqrt{2}}}{2})) \cdot \\ \cdot (x-(\frac{\sqrt{2-\sqrt{2}}}{2} - i\frac{\sqrt{2+\sqrt{2}}}{2})) \cdot (x-(\frac{\sqrt{2+\sqrt{2}}}{2} - i\frac{\sqrt{2-\sqrt{2}}}{2})).$$

Разложение над $\mathbb{R}$:

$$f(x) = (x^2 - \sqrt{2-\sqrt{2}}\,x + 1)\cdot(x^2 + \sqrt{2-\sqrt{2}}\,x + 1)\cdot \\ \cdot(x^2 - \sqrt{2+\sqrt{2}}\,x + 1)\cdot(x^2 + \sqrt{2+\sqrt{2}}\,x + 1).$$

То же разложение элементарными средствами получается так:

$$f(x) = x^8 + 1 = (x^4+1)^2 - 2x^2 = \\ = (x^4 - \sqrt{2}\,x^2 + 1)(x^4 + \sqrt{2}\,x^2 + 1) = \\ [(x^2+1)^2 - (2+\sqrt{2})x^2]\cdot[(x^2+1)^2 - (2-\sqrt{2})x^2] = \\ = (x^2 - \sqrt{2+\sqrt{2}}\,x + 1)\cdot(x^2 + \sqrt{2+\sqrt{2}}\,x + 1)\cdot \\ \cdot(x^2 - \sqrt{2-\sqrt{2}}\,x + 1)\cdot(x^2 + \sqrt{2-\sqrt{2}}\,x + 1).$$

Замечание 43.1. Проследите (это будет поучительным) за реакцией Maple на команды:

```
> factor( x ^ 8 + 1 ); factor( x ^ 8 + 1, sqrt(2) );
  factor( x ^ 8 + 1, I); factor( x ^ 8 + 1, sqrt( 2 + sqrt(2) ) );
```

§ 44. Неразложимые элементы в целостном кольце. Простые элементы. Неприводимые многочлены

44.1. Понятие неразложимого элемента в целостном кольце. В текущей главе основным предметом нашего изучения являются кольца многочленов. Однако некоторые факты теории таких колец становятся понятнее в общем контексте *теории делимости в целостных кольцах.* (Например, свойства кольца $P[x]$ многочленов над полем P очень похожи на свойства кольца целых чисел $\mathbb{Z}$.)

В связи с этим в ряде пунктов (см. 36.4 — 36.6, 37.2, 38.1, 38.5, 38.7, 38.8) теория делимости излагалась в большей общности, чем это необходимо для изучения только многочленов.

Напомним, в частности, что в п. 38.1 фигурировало *условие Безу* для целостных колец (состоящее в том, что для любых двух элементов кольца существует НОД, допускающий линейное представление), а в п. 38.8 был определен класс *евклидовых колец* (в которых возможна работа алгоритма Евклида).

В настоящем и последующих параграфах указанная "абстрактная линия" будет развиваться дальше; мы обнаружим, что сходство арифметики целых чисел и алгебры многочленов оказывается еще более глубоким: в кольцах многочленов возможно разложение на так называемые *неприводимые* множители, аналогичное разложению целых чисел на *простые* множители.

Пусть K — целостное кольцо (см. п. 36.6), K^* — группа его обратимых элементов (см. п. 36.4). На кольце K заданы отношения ассоциированности (см. п. 36.5) и делимости (см. п. 37.2).

Определение 44.1. *Неразложимым* называется такой элемент $a \in K$, который

1) необратим ($a \notin K^*$) и отличен от нуля;

2) делится лишь на обратимые элементы и на элементы, ассоциированные ему, т. е.

$$(\, b \,|\, a \,) \Rightarrow (\, b \in K^* \,) \vee (\, b \sim a \,). \tag{44.1}$$

Замечание 44.1. Две возможности, фигурирующие в условии 2 определения 44.1, являются (в силу условия 1) взаимоисключающими. Условие 2 можно пересказать иначе: если элемент a разлагается в произведение двух сомножителей ($a = b \cdot c$), то либо элемент b обратим (и тогда a и c ассоциированы), либо c обратим (и тогда a и b ассоциированы). В формульной записи:

$$(\, a = b \cdot c \,) \Rightarrow [\, (\, b \in K^* \,) \wedge (\, c \sim a \,) \,] \vee [\, (\, c \in K^* \,) \wedge (\, b \sim a \,) \,]. \tag{44.2}$$

Для необратимого и ненулевого элемента a *тривиальными* называются: 1) обратимые делители; 2) делители, ассоциированные с a. Пользуясь этим термином, условие 2 можно выразить еще одним способом: *элемент a имеет только тривиальные делители.*

Термин *разложимый* элемент также обычно относят к ненулевым и необратимым элементам (т. е. обратимые элементы и нуль не относятся ни к разложимым, ни к неразложимым). Свойство разложимости ненулевого необратимого элемента (т. е. отрицание условия

2) можно выразить так: *элемент разлагается в произведение двух необратимых элементов.*

Заметим, что в целостном кольце вообще может не быть неразложимых элементов. (Простейший пример — поле, оно является целостным кольцом, но все его элементы, кроме нулевого, обратимы.)

В следующем предложении приводятся общие свойства неразложимых элементов.

Предложение 44.1. *1. Элемент, ассоциированный с неразложимым элементом, сам является неразложимым.*

2. Если один неразложимый элемент делит другой неразложимый, то эти элементы ассоциированы.

Доказательство. 1. Пусть элемент a является неразложимым и $a' \sim a$, т. е. $a' = au$, где u обратим. Тогда, очевидно, элемент a', так же как и a, является необратимым и ненулевым.

Если бы a' оказался разложимым, то нашлись бы необратимые элементы $b, c \in K$, такие, что $a' = bc$. А значит, элемент $a = a'u^{-1} = b(cu^{-1})$ являлся бы произведением необратимых элементов и, следовательно, был бы разложимым, что противоречит предположению.

2. Второе утверждение непосредственно следует из определения неразложимого элемента. □

44.2. Понятие простого элемента в целостном кольце

Определение 44.2. Элемент $a \in K$ называется *простым*, если

1) он необратим и отличен от нуля;

2) из того факта, что a делит произведение двух элементов $b \cdot c$ вытекает, что a делит хотя бы один из сомножителей, т. е. справедлива импликация:

$$(a \mid b \cdot c) \Rightarrow (a \mid b) \vee (a \mid c). \tag{44.3}$$

Замечание 44.2. Свойство 2 из определения простых элементов по индукции легко обобщается на любое количество сомножителей.

В качестве упражнения, пользуясь свойствами отношения делимости (см. предложение 37.1), вы легко докажете утверждение, аналогичное первому пункту предложения 44.1: элемент, ассоциированный простому, сам прост.

Простота есть, вообще говоря, свойство более сильное, чем неразложимость, но в некоторых случаях эти свойства равносильны. Точнее, справедливо следующее

Предложение 44.2. *Пусть K — целостное кольцо; $a \in K$. Тогда*
1) если a прост, то a неразложим;
2) при дополнительном условии (Б) *(см. п. 38.1) верно обратное утверждение: неразложимость влечет простоту (и таким образом, эти свойства равносильны).*

Доказательство. 1. Пусть элемент a прост и пусть $a = bc$. Тогда, поскольку $a|a$, имеет место делимость $a|bc$. По определению простого элемента, из этого следует, что $a|b$ или $a|c$.

С другой стороны, как b, так и c делят a. По свойству 3 отношения делимости (см. предложение 37.1), получается, что $a \sim b$ или $a \sim c$, и следовательно (см. замечание 44.1), элемент a неразложим.

2. При дополнительном условии (Б) докажем обратное утверждение. Пусть элемент a неразложим. Докажем, что он прост. Предположим $a|bc$. В силу (Б), существует наибольший общий делитель $d \in \text{НОД}(a, b)$, допускающий (для некоторых $u, v \in K$) линейное представление $d = au + bv$. Поскольку $d|a$ и a неразложим, то либо $d \in K^*$, либо $d \sim a$.

Рассмотрим первый случай. Обратимость элемента эквивалентна его ассоциированности единичному элементу. Значит, во-первых, $1 \in \text{НОД}(a, b)$, и во-вторых, из линейного представления $d = au + bv$ следует линейное представление для единичного элемента $1 = au' + bv'$ (где $u' = ud^{-1}$, $v' = vd^{-1}$), из которого, по предложению 38.3, кстати тоже опирающемуся на условие (Б), следует взаимная простота элементов a и b. По предложению 38.5, также использующему (Б), делимость $a|bc$, при условии взаимной простоты a и b, влечет делимость $a|c$.

Переходим ко второму случаю. Если $d \sim a$, то, в частности, $a|d$; с учетом делимости $d|b$ получаем делимость $a|b$.

Простота элемента a доказана. □

Замечание 44.3. Как показано в § 38, в кольце целых чисел $\mathbb{Z}$ и в кольце $P[x]$ многочленов над полем P (и вообще в любом евклидовом кольце) условие (Б) выполнено. Значит, в силу только что доказанного утверждения, в этих кольцах неразложимость элемента равносильна его простоте.

Неразложимые (= простые) элементы в кольце $\mathbb{Z}$ вам хорошо знакомы с детства; это — числа: $\pm 2, \pm 3, \pm 5, \ldots$ Число $p \in \mathbb{Z}$ является простым элементом в $\mathbb{Z}$, если оно делится лишь на обратимые элементы ± 1 и на ассоциированные себе числа $\pm p$. Неразложимым (= простым) элементам в кольце многочленов мы посвятим п. 44.4.

44.3. Канонические неразложимые (простые) элементы. Поскольку неразложимые (простые) элементы встречаются целыми классами, содержащими элементы, ассоциированные друг другу, то было бы удобным в каждом из таких классов выбрать ("назначить") *канонического* ("узаконенного") представителя. В некоторых кольцах имеется естественный способ такого "назначения". (Мы уже кратко говорили об этом в п. 36.5.)

Например, в кольце $\mathbb{Z}$ *каноническими неразложимыми* элементами принято считать *положительные простые* числа. О кольце многочленов подробнее будет сказано в следующем пункте. В общем же случае приходится выбирать канонических представителей (сплошная демократия!) по произволу.

В дальнейшем мы иногда будем использовать аббревиатуру *к.н.* для термина "канонический неразложимый" (элемент, представитель или множитель).

44.4. Неприводимые многочлены. В этом пункте мы будем рассматривать кольцо $K = P[x]$ многочленов над полем P. Это кольцо (как и кольцо целых чисел) является евклидовым (см. п. 38.8); на его (ненулевых) элементах задана функция "степень", следящая за делением (с остатком или нацело) и, в частности, обладающая свойством монотонности.

Введем важнейшее в теории многочленов

Определение 44.3. Многочлен $f(x) \in P[x]$, положительной степени, называется *неприводимым*, если он не разлагается в произведение многочленов меньшей степени.

Непосредственно из определения 44.3 следует, что (над любым полем) многочлены первой степени неприводимы. (В самом деле, меньшей, чем первая, является лишь нулевая степень, а произведение двух многочленов нулевой степени также имеет нулевую степень.)

Предложение 44.3. *Неприводимые многочлены и только они являются неразложимыми (простыми) элементами кольца* $P[x]$.

Доказательство. Обратимыми элементами в кольце $P[x]$ являются ненулевые константы, т. е. многочлены нулевой степени. Ненулевыми необратимыми элементами будут, таким образом, многочлены положительной степени.

Многочлен $f(x)$ степени $n > 0$ является неразложимым элементом в $P[x]$ тогда и только тогда, когда его нельзя разложить в про-

изведение двух многочленов положительной степени (степень этих многочленов автоматически должна быть меньшей n).

Равносильность неразложимости и неприводимости доказана.

Остается заметить, что, согласно теореме 38.1, кольцо $P[x]$ удовлетворяет условию (Б), и следовательно, по предложению 44.2, в этом кольце неразложимость элемента равносильна его простоте. □

Замечание 44.4. Многочлен, ассоциированный (пропорциональный) неприводимому, сам является таковым. В качестве канонического представителя (см. п. 44.3) от класса пропорциональных неприводимых многочленов принято выбирать *нормализованный* неприводимый многочлен.

Замечание 44.5. Основным предметом изучения в данной главе являются многочлены *над полем*. Однако многие факты, доказанные для таких многочленов, сохраняют силу для многочленов *над целостным кольцом*.

К тому же в некоторых вопросах, касающихся многочленов над полями, без рассмотрения многочленов над кольцами не обойтись. (Например, при индуктивном определении многочленов от нескольких переменных или — см. § 46 — при рассмотрении многочленов над полем рациональных чисел.)

В связи с этим на протяжении всей главы мы тянули "тонкую нить замечаний", посвященных изучению кольца $K = L[x]$ многочленов над целостным кольцом L (см. замечания 36.8, 36.9, 36.12, 37.2, 39.3, 39.7, 40.3).

Продолжением этой нити служит и настоящее замечание.

Определение 44.3 неприводимых многочленов без всяких изменений переносится на случай многочленов над целостным кольцом L. Однако описание неразложимых элементов в кольце $L[x]$ несколько усложняется. Это обусловлено наличием необратимых многочленов нулевой степени (констант).

Рассмотрим, скажем, над кольцом $L = \mathbb{Z}$ многочлен

$$f(x) = 2x^2 + 2.$$

Он не представляется в виде произведения многочленов меньшей степени (над $\mathbb{Z}$, над $\mathbb{Q}$ и даже над $\mathbb{R}$), следовательно, он является неприводимым (над кольцом $\mathbb{Z}$ и над указанными полями).

Однако он не является неразложимым элементом кольца $\mathbb{Z}[x]$, поскольку может быть представлен в виде произведения двух необратимых элементов:

$$f(x) = 2 \cdot (x^2 + 1).$$

В кольце $L[x]$ неразложимыми элементами будут, во-первых, неразложимые скаляры (элементы кольца коэффициентов L) и, во-вторых, так называемые *примитивные* неприводимые многочлены, не имеющие общего для всех членов необратимого скалярного множителя (т. е. такие, для которых единица $1 \in L$ является наибольшим общим делителем всех коэффициентов).

Более подробное изложение теории многочленов над кольцами можно найти, скажем, в учебниках [1, 5, 7, 9]. Начала этой теории (в основном для многочленов над кольцом $\mathbb{Z}$) будут изложены в § 46, при рассмотрении многочленов над полем $\mathbb{Q}$.

44.5. Неприводимые многочлены над алгебраически замкнутыми полями и над полем действительных чисел

Предложение 44.4. *1. Над алгебраически замкнутым полем P (в частности, над полем $P = \mathbb{C}$) неприводимыми являются лишь многочлены первой степени; все нормализованные неприводимые многочлены имеют вид $x - c$, где $c \in P$.*

2. Над полем $P = \mathbb{R}$ неприводимыми являются все многочлены первой степени и многочлены второй степени с отрицательным дискриминантом; нормализованными неприводимыми многочленами являются многочлены вида $x - c$ ($c \in \mathbb{R}$) и вида $x^2 + px + q$ ($p, q \in \mathbb{R}$; $p^2 - 4q < 0$).

Доказательство. 1. Многочлены первой степени, как выше отмечалось, неприводимы над любым полем; докажем, что в случае алгебраически замкнутого поля других неприводимых многочленов нет. В силу предложения 40.10, всякий многочлен, степень которого больше единицы, разлагается на линейные множители (в количестве, равном степени) и, следовательно, не может быть неприводимым.

2. В случае $P = \mathbb{R}$ применяется предложение 43.3. Многочлены первой степени неприводимы. Многочлены второй степени, имеющие неотрицательный дискриминант, приводимы (разлагаются над $\mathbb{R}$ на два линейных множителя). Многочлены второй степени с отрицательным дискриминантом являются неприводимыми над $\mathbb{R}$, поскольку, будь они приводимыми, они разлагались бы на множители первой степени и имели бы, следовательно, действительные корни, что невозможно в силу отрицательности дискриминанта.

3. Утверждения о виде нормализованных неприводимых многочленов очевидны, многочлены такого вида уже фигурировали в упомянутых выше предложениях 40.10 и 43.3. □

Замечание 44.6. Советуем вам поупражняться с Maple-тестами неприводимости; введите, например, команды:

```
> f := x ^ 5 + x + 1; irreduc( f ); factor( f );
  irreduc( f, real ); factor( f, real );
  irreduc( f, complex ); factor( f, complex );
```

и внимательно проследите за результатами. (Над какими полями произведены эти вычисления?)

§ 45. Факториальные кольца. Факториальность кольца целых чисел (основная теорема арифметики) и факториальность колец многочленов (над полем)

45.1. Определение факториального кольца. Рассмотрим целостное кольцо K. Сейчас будет введено важнейшее из понятий теории делимости — понятие *факториального кольца.* Мы говорили выше, что целостные кольца "похожи" на кольцо целых чисел $\mathbb{Z}$. Факториальные кольца "очень похожи" на $\mathbb{Z}$: в этих кольцах будет справедливо утверждение, аналогичное *основной теореме арифметики* (ОТАр).

Самое ОТАр студенты-математики изучают в курсе теории чисел. Эта теорема утверждает, что любое целое число (отличное от 0, 1, -1) разлагается на простые множители, причем однозначно, с точностью до порядка сомножителей и их знака. (Например: $12 = 2 \cdot 2 \cdot 3 = (-2) \cdot (-3) \cdot 2$.)

В настоящем параграфе мы получим наряду с другими результатами и эту теорему (как частный случай некоторого общего утверждения; см. ниже теоремы 45.1 и 45.2).

Определение 45.1. Целостное кольцо K называется *факториальным*, если выполнены следующие два условия:

(Ф.1) всякий ненулевой и необратимый элемент $a \in K$ представляется в виде *произведения неразложимых элементов*

$$a = p_1 \cdot p_2 \cdot \ldots \cdot p_k; \tag{45.1}$$

(Ф.2) разложение (45.1) определено *однозначно, с точностью до порядка сомножителей и их ассоциированности*, т. е. если помимо

(45.1) имеется еще одно разложение на неразложимые элементы

$$a = q_1 \cdot q_2 \cdot \ldots \cdot q_l, \tag{45.2}$$

то, во-первых, эти разложения имеют одинаковое количество сомножителей ($k = l$) и, во-вторых, существует такая перестановка номеров

$$\sigma = \begin{pmatrix} 1 & 2 & \ldots & k \\ \sigma(1) & \sigma(2) & \ldots & \sigma(k) \end{pmatrix} \in S_k,$$

что для любого $i = 1, \ldots, k$ элементы p_i и $q_{\sigma(i)}$ будут ассоциированы ($p_i \sim q_{\sigma(i)}$).

Данное выше сложное определение заслуживает подробного обсуждения и комментариев. Прежде всего, не исключаются "произведения из одного сомножителя", что не слишком согласуется с языковы́ми нормами, но является обычной практикой в математических текстах.

Далее, среди сомножителей в (45.1) вполне могут быть *повторяющиеся*, и их можно будет *сгруппировать*, вводя соответствующие *кратности*.

Однако в действительности дело с группировкой обстоит несколько сложнее. Надо будет объединять в одну группу не только совпадающие, но и ассоциированные друг другу сомножители.

Пусть, скажем, помимо сомножителя p_1, в разложении имеются еще два сомножителя, ассоциированные p_1. Поскольку, в силу коммутативного закона для умножения, их номера не играют существенной роли, то пусть это будут $p_2 = u_2 p_1$ и $p_3 = u_3 p_1$, где $u_1, u_2 \in K^*$, т. е. обратимы. Группировка этих трех сомножителей даст $u_1 p_1^3$, где элемент $u_1 = u_2 u_3 \in K^*$. (Здесь мы наблюдаем появление *обратимого* множителя перед степенью неразложимого элемента.)

Берем следующий сомножитель, уже не ассоциированный с ранее рассмотренными, и производим группировку для него; при этом появляется новый обратимый множитель; и т. д. "Набежавшие" обратимые множители можно собрать в один $u \in K^*$ и расположить его в начале разложения.

Все обнаруженные (*попарно не ассоциированные*) неразложимые сомножители перенумеруем:

$$p_1, p_2, \ldots, p_s; \tag{45.3}$$

каждый из них будет иметь свою кратность:

$$m_1, m_2, \ldots, m_s; \tag{45.4}$$

разложение (45.1) примет сгруппированный вид

$$a = up_1^{m_1} p_2^{m_2} \dots p_s^{m_s};\ u \in K^*, \tag{45.5}$$

где сумма кратностей равна исходному количеству неразложимых сомножителей:

$$m_1 + m_2 + \dots + m_s = k.$$

Формула (45.5) "лучше" исходной формулы (45.1) еще и тем, что она годится для записи обратимых элементов: в такой записи будет $k = 0$ неразложимых множителей, т. е. "разложение" сведется к тавтологии $a = u \in K^*$. (Таким образом, только нулевой элемент окажется "не охваченным" формулой разложения.)

Есть еще одна возможность "унификации" разложений: в каждом классе ассоциированных элементов следует выбирать канонического представителя (см. п. 44.3), чтобы в формуле (45.5) присутствовали только *канонические неразложимые* множители. В этом случае мы будем называть формулу (45.5) *каноническим разложением* для элемента $a \in K$.

По построению каноническое разложение для элемента a будет определено однозначно, с точностью лишь до порядка сомножителей. (В самом деле, из факториальности кольца следует, что попарно не ассоциированные к.н. элементы p_i определены однозначно, вместе со своими кратностями m_i, а это уже влечет однозначность определения обратимого множителя u.)

Подведем итог в форме следующего утверждения.

Предложение 45.1. *Всякий ненулевой элемент a факториального кольца K однозначно, с точностью до порядка сомножителей, представляется своим каноническим разложением вида (45.5).*

Доказательство см. выше. □

Добавим, что выбор канонических представитетелей в классах ассоциированности неразложимых элементов определяет выбор таких представителей во всех (отличных от нулевого) классах. А именно: представителем класса элементов, ассоциированных с элементом, заданным разложением (45.5), назначается элемент

$$a' = p_1^{m_1} p_2^{m_2} \dots p_s^{m_s} \tag{45.5a}$$

с единичным обратимым множителем ($u = 1$). В частности, каноническим представителем класса обратимых элементов будет служить единица $1 \in K$.

45.2. Свойства факториальных колец. Изучим теперь некоторые свойства факториальных колец, связанные с делимостью их элементов.

Ненулевые элементы будем представлять их каноническими разложениями. Для двух (или любого конечного числа) различных элементов кольца списки [вида (45.3)] к.н. множителей в их разложениях будут, вообще говоря, различными, но эти списки можно "подогнать" под один список (в него следует включить те к.н. элементы, которые входят в разложение хотя бы одного из рассматриваемых элементов).

При этом придется разрешить некоторым кратностям принимать нулевые значения (что будет свидетельствовать о том, что в разложении данного элемента данный неразложимый множитель фактически отсутствует).

Пусть после подгонки списков к.н. множителей канонические разложения для ненулевых элементов $a, b \in K$ имеют вид (45.5) и

$$b = vp_1^{l_1} p_2^{l_2} \dots p_s^{l_s};\ v \in K^*, \tag{45.6}$$

где, возможно, некоторые из кратностей обращаются в нуль. Тогда легко получить каноническое разложение для произведения ab. (В нем будет фигурировать обратимый множитель, равный uv, а кратности неприводимых множителей *сложатся*.)

Переходим теперь к изучению делимости элементов и к задаче отыскания НОД и НОК в факториальных кольцах.

Предложение 45.2. *Пусть ненулевые элементы a и b факториального кольца K заданы своими каноническими разложениями (45.5) и (45.6). Тогда справедливы следующие утверждения.*

(1) Элемент a делит элемент b тогда и только тогда, когда для любого номера $i = 1, \dots, s$ кратность m_i не превосходит кратность l_i, или в формульной записи:

$$(\, a \,|\, b \,) \Leftrightarrow (\forall i = 1, \dots, s)(m_i \leqslant l_i). \tag{45.7}$$

(2) Любые два элемента $a, b \in K$ имеют как наибольший общий делитель $d \in \text{НОД}(a, b)$, так и наименьшее общее кратное $q \in \text{НОК}(a, b)$, которые (для ненулевых элементов) могут быть найдены по формулам

$$\begin{aligned} d &= p_1^{k_1} p_2^{k_2} \dots p_s^{k_s};\ k_i = \min(m_i, l_i);\ i = 1, \dots, s; \\ q &= p_1^{n_1} p_2^{n_2} \dots p_s^{n_s};\ n_i = \max(m_i, l_i);\ i = 1, \dots, s. \end{aligned} \tag{45.8}$$

(3) В факториальном кольце неразложимость элемента влечет его простоту.

Доказательство. 1. Если неравенство для кратностей, фигурирующее в (45.7), выполнено, то, очевидно, $a|b$, причем частное b/a будет иметь каноническое разложение с обратимым множителем u/v и кратностями $l_i - m_i$.

Обратно, если $a|b$, то найдется элемент $c \in K$, такой, что $b = ac$; он также может быть представлен своим каноническим разложением.

Мы заранее не знаем, какие к.н. элементы будут присутствовать в разложении для c. Предположим, что разложение для c начинается обратимым множителем w, что элементы списка (45.3) входят в это разложение с кратностями r_i $(i = 1, ..., s)$, возможно нулевыми, и что помимо этих элементов в разложении присутствуют другие к.н. элементы p_i (с кратностями r_i), где $i = s+1, ..., s+t$. При перемножении элементов a и c обратимые множители перемножатся, а кратности к.н. элементов сложатся.

Равенство $b = ac$ приведет, в силу единственности канонического разложения (см. предложение 45.1), к системе равенств

$$\begin{cases} v = uw; \\ l_i = m_i + r_i \qquad (i = 1, ..., s); \\ 0 = r_i \qquad (i = s+1, ..., s+t), \end{cases}$$

из которой, вследствие неотрицательности целых чисел r_i, получаются неравенства $m_i \leqslant l_i$ $(i = 1, ..., s)$, кроме того, становится ясным, что элемент c не может иметь в своем разложении "новых" к.н. элементов.

2. Рассмотрим два элемента a, b и элемент d, определяемый первой из формул (45.8). Из (45.7) немедленно следует, что d является общим делителем для a и b.

Рассмотрим еще один элемент, d°, являющийся общим делителем для a и b. Очевидно, его каноническое разложение не может содержать "новых" к.н. элементов, не входящих в формулы (45.5) и (45.6), ибо это противоречило бы первому утверждению настоящего предложения.

Согласно тому же утверждению, кратности m_i° множителей p_i в разложении элемента d° должны удовлетворять двум неравенствам: $m_i^\circ \leqslant m_i$ и $m_i^\circ \leqslant l_i$, а значит, и неравенству $m_i^\circ \leqslant \min(m_i, l_i) = k_i$, откуда уже будет следовать делимость $d^\circ|d$. Тем самым доказано, что $d \in$ НОД(a, b). Утверждение, относящееся к НОК, остается в качестве упражнения.

3. Прежде чем начинать доказательство третьего утверждения, заметим, что идея воспользоваться предложением 44.2 (неразложимость равносильна простоте — при выполнении условия Безу) не проходит, поскольку факториальное кольцо не обязательно удовлетворяет условию (Б). (Контрпример будет приведен в следующем параграфе; см. п. 46.6.)

Пусть p — неразложимый элемент факториального кольца K. Докажем, что p прост.

Если p делит произведение двух элементов $b \cdot c$, то найдется элемент $s \in K$, такой, что $p \cdot s = b \cdot c$. Разложив элементы b, c, s на неразложимые p_i, q_j, r_k $(i = 1, ..., n;\ j = 1, ..., m;\ k = 1, ..., l)$, получим равенство

$$p \cdot s_1 \cdot ... \cdot s_l = p_1 \cdot ... \cdot p_n \cdot q_1 \cdot ... \cdot q_m,$$

которое можно рассматривать как факт наличия двух разложений одного и того же элемента на неразложимые множители.

Условие (Ф.2) влечет, во-первых, равенство $1 + s = n + m$, а во-вторых, попарную (при надлежащей перенумерации) ассоциированность неразложимых множителей в левой и правой частях.

В частности, элемент p ассоциирован либо с одним из p_i (и тогда он делит a), либо с одним из q_j (и тогда он делит c).

Этим доказана простота элемента p. □

45.3. Достаточные условия факториальности кольца. Факториальность евклидовых колец. Теперь мы займемся выводом некоторых условий на целостное кольцо K, выполнение которых гарантирует его факториальность.

Таких условий будет два. Одно из них нам уже многократно встречалось, начиная с п. 38.1. Это — *условие Безу*, здесь мы приведем его в формульной записи:

$$\text{(Б)}\quad (\forall a, b \in K)\,(\exists d, u, v \in K)\,[\,(\,d \in \text{НОД}(a, b)\,) \wedge (\,d = au + bv\,)\,].$$

Второе условие мы сформулируем ниже и будем называть его *условием Нетер* [для ссылок будет использоваться обозначение (Н)]:

(Н) в кольце K не существует такой бесконечной последовательности элементов $\{a_k\}_{k=0}^{\infty}$, в которой каждый последующий элемент a_{k+1} является *нетривиальным делителем* (см. замечание 44.1) предыдущего элемента a_k, т. е.

$$a_{k+1} \,|\, a_k;\ a_{k+1} \notin K^*;\ a_{k+1} \not\sim a_k\ (k = 0, 1, 2, ...). \tag{45.9}$$

Замечание 45.1. Оба приведенных выше условия, (Б) и (Н), допускают более "научную" формулировку, в терминах так называемых *идеалов* (и в частности, *главных идеалов*) в кольцах. (С такой терминологией вы встретитесь, если обратитесь к уже упомянутым учебникам [1, 3, 5, 7].)

Мы предпочли отложить введение "идеалистических абстракций" и, следуя духу замечательного сочинения П. Нодена и Л. Китте (см. ссылку в начале п. 41.1), привести более элементарные формулировки. (Кстати, в указанной книге вы сможете найти также мотивировки используемых названий: свойство Безу, свойство Нетер. Отметим только, что наше условие (Н) является некоторой *ослабленной* формой широко употребительного в теории колец условия *нетеровости*, честь открытия которого принадлежит Эмми Нетер, удивительной женщине, имевшей редкую профессию "математик", обучавшей абстрактной алгебре молодых советских ученых в конце двадцатых годов прошлого века.)

Предложение 45.3. *Евклидовы кольца удовлетворяют условию Нетер.*

Доказательство. Пусть K — евклидово кольцо (см. определение 38.5) со следящей функцией ν. Из свойства монотонности этой функции вытекает, что если (ненулевые) элементы $a, b \in K$ ассоциированы, то $\nu(a) = \nu(b)$. Следовательно, если a является нетривиальным делителем b (и в частности, $a \not\sim b$), то будет иметь место строгое неравенство $\nu(a) < \nu(b)$.

Отсюда и из неотрицательности значений ν можно заключить, что не существует бесконечных последовательностей нетривиальных делителей вида (45.9). $\square$

Теорема 45.1. *1. Если в целостном кольце K выполнено условие (Н), то в нем выполнено условие (Ф.1).*

2. Выполнение условия (Б) влечет выполнение (Ф.2).

3. Кольцо, в котором выполняются оба условия, (Б) и (Н), является факториальным.

4. Всякое евклидово кольцо факториально.

Доказательство. 1. Пусть выполнено условие Нетер. Докажем, что всякий ненулевой и необратимый элемент кольца K разлагается в произведение неразложимых элементов.

Предположим противное, т. е. предположим, что существует элемент $a_0 \in K$, ненулевой, необратимый и обладающий тем "плохим"

свойством, что его нельзя разложить в произведение неразложимых элементов.

В частности, сам элемент a_0 не является неразложимым. Значит, он разлагается в произведение двух своих *нетривиальных* делителей, скажем, $a_0 = a_1 \cdot b_1$.

Если бы оба этих делителя допускали разложение на неразложимые, то такое разложение допускал бы и исходный элемент a_0, что противоречит предположению. Поэтому хотя бы один из элементов a_1 или b_1, так же как и a_0, обладает "плохим" свойством. Будем для определенности считать, что таким элементом является a_1.

Сделан первый шаг в построении последовательности $\{a_k\}$: элемент a_1 является нетривиальным делителем элемента a_0 и, что самое главное, обладает тем же свойством, что и a_0: не разлагается в произведение неразложимых элементов.

Рассуждение можно повторить и найти нетривиальный делитель a_2 элемента a_1, снова обладающий "плохим" свойством; и т. д. Мы приходим, таким образом, к бесконечной последовательности элементов кольца, удовлетворяющей (45.9), что противоречит условию (Н). Значит, наше допущение не верно и в кольце выполняется (Ф.1).

2. Докажем теперь импликацию (Б) $\Rightarrow$ (Ф.2). Пусть (ненулевой и необратимый) элемент $a \in K$ имеет два разложения, (45.1) и (45.2), на неразложимые множители.

Пусть для определенности количество k сомножителей в первом разложении не превышает количества l сомножителей во втором разложении. Рассмотрим равенство

$$p_1 \cdot p_2 \cdot \ldots \cdot p_k = q_1 \cdot q_2 \cdot \ldots \cdot q_k \cdot q_{k+1} \cdot \ldots \cdot q_l. \qquad (45.10)$$

Элементы p_i $(i = 1, \ldots, k)$ и q_j $(j = 1, \ldots, l)$ являются неразложимыми и, следовательно (в силу условия (Б); см. второе утверждение предложения 44.2), простыми.

Элемент p_1 делит правую часть равенства (45.10), и следовательно (по определению 44.2; см. также замечание 44.2), p_1 делит некоторый из сомножителей q_j правой части.

Оба этих элемента неразложимы, поэтому, в силу второй части предложения 44.1, они ассоциированы.

Наличие коммутативного закона для умножения позволяет как угодно переставлять (перенумеровывать) сомножители. Сделаем обнаруженный выше сомножитель q_j первым и будем далее считать, что $p_1 \sim q_1$, т. е. $q_1 = p_1 u_1$, где $u_1 \in K^*$.

Подставим полученное выражение в (45.9) и сократим (пользуясь целостностью кольца K) это равенство на p_1; получим:

$$p_2 \cdot \ldots \cdot p_k = u_1 \cdot q_2 \cdot \ldots \cdot q_k \cdot q_{k+1} \cdot \ldots \cdot q_l. \tag{45.10a}$$

Рассуждая совершенно аналогично, мы найдем в правой части (45.10a) элемент, ассоциированный с p_2. Этим элементом не может быть обратимый множитель u_1; им будет один из оставшихся неразложимых множителей.

Сделаем (с помощью перенумерации) номер этого множителя равным 2 и представим его в виде $q_2 = p_2 u_2$ ($u_2 \in K^*$); затем произведем очередное сокращение; и т. д.

После k сокращений мы получим равенство

$$1 = u \cdot q_{k+1} \cdot \ldots \cdot q_l;\ u = u_1 u_2 \ldots u_k \in K^*, \tag{45.11}$$

которое в случае $k < l$ приведет к противоречию, поскольку в правой части будет присутствовать хотя бы один неразложимый множитель, а левая часть равняется единице.

Значит, 1) $k = l$ и 2) неразложимые множители в правой части (45.10), после подходящей перенумерации, оказываются ассоциированными соответствующим множителям в левой части.

3. Третье утверждение теоремы следует из первых двух.

4. В евклидовом кольце выполняется как условие (Б), согласно предложению 38.9, так и условие (Н), согласно предложению 45.3. Таким образом, евклидовость кольца влечет его факториальность. □

45.4. Основная теорема арифметики

Теорема 45.2 (основная теорема арифметики, ОТАр). *1. Кольцо целых чисел факториально, т. е. всякое целое число, отличное от* 0, 1, -1, *разлагается на простые множители, причем такое разложение определено однозначно, с точностью до порядка сомножителей и их знака.*

2. Для любого $a \in \mathbb{Z} \setminus \{0\}$ однозначно, с точностью до порядка сомножителей, определено каноническое разложение

$$a = u p_1^{m_1} p_2^{m_2} \ldots p_s^{m_s}, \tag{45.12}$$

где $u = \pm 1$; p_i — положительные простые числа ($i = 1, \ldots, s$; $s \geqslant 0$).

Доказательство. Первое утверждение теоремы следует из четвертого утверждения теоремы 45.1, а второе — из предложения 45.1. □

45.5. Факториальность кольца многочленов над полем

Теорема 45.3. *1. Кольцо $P[x]$ многочленов над полем P факториально, т. е. всякий многочлен положительной степени разлагается на неприводимые множители, причем такое разложение определено однозначно, с точностью до порядка сомножителей и их пропорциональности.*

2. Для любого $f(x) \in P[x] \setminus \{0\}$ однозначно, с точностью до порядка сомножителей, определено каноническое разложение

$$f(x) = a_0 p_1(x)^{m_1} p_2(x)^{m_2} \dots p_s(x)^{m_s}, \tag{45.13}$$

где a_0 есть старший коэффициент многочлена $f(x)$; $p_i(x)$ — нормализованные неприводимые многочлены над полем P ($i = 1, \dots, s$; $s \geqslant 0$).

Доказательство. Следует повторить ту же фразу, что приведена в качестве доказательства предыдущей теоремы. (При переходе к другому кольцу меняются только описания обратимых, неразложимых и ассоциированных элементов.) □

§ 46. Факториальность кольца многочленов над факториальным кольцом. Неприводимые многочлены над полем $\mathbb{Q}$ и над кольцом $\mathbb{Z}$

46.1. Содержание многочлена над факториальным кольцом. Примитивные многочлены. В настоящем параграфе мы будем изучать кольцо $K = L[x]$ многочленов над целостным кольцом L. Мы уже обращались к этому вопросу в замечании 44.5, в котором (пока без строгого обоснования) были описаны неразложимые элементы в кольце $L[x]$.

Сейчас мы начнем с того, что констатируем: можно ожидать, что в кольце многочленов $L[x]$ получится "хорошая" теория делимости, только если такая теория уже имеется в кольце коэффициентов L.

В связи с этим мы с самого начала параграфа предполагаем, что кольцо L является *факториальным* (см. определение 45.1).

Далее, в предыдущих параграфах мы свободно переходили от одной формы записи многочленов (в порядке возрастания степеней) к другой (в порядке убывания степеней). Первая форма более удобна

в теоретических вопросах, а вторая — в практических вычислениях. (По крайней мере, там она более привычна.)

И сейчас нам удобнее будет обратиться к форме записи многочленов в виде бесконечных, но финитных сумм по возрастанию степеней [см. (36.1b)]:

$$f(x) = \sum_{k=0}^{\infty} f_k x^k;\ \ f_k \in L\ (k \geqslant 0);\ \ f_n \neq 0; (\forall k > n)\,(f_k = 0). \qquad (46.1)$$

Определение 46.1. *Содержанием* ненулевого многочлена (46.1) называется наибольший общий делитель коэффициентов этого многочлена. Используется обозначение

$$\operatorname{cont}(f) = \text{НОД}(f_0, f_1, ..., f_n). \qquad (46.2)$$

Замечание 46.1. Как всякий НОД, содержание определено лишь с точностью до ассоциированности (является множеством, классом ассоциированных друг другу элементов). Поэтому для элементов этого класса мы используем запись '$d \in \operatorname{cont}(f)$', вместо '$d = \operatorname{cont}(f)$'.

В конкретных кольцах, в которых возможен канонический выбор представителей в классах ассоциированных элементов, можно договориться о таком выборе из класса $\operatorname{cont}(f)$.

Например, для $K = \mathbb{Z}$ мы можем считать содержание не множеством, а положительным целым числом, т. е. полагать

$$\operatorname{cont}(f) = (\, a_0,\, a_1,\, ...,\, a_n\,). \qquad (46.2a)$$

При умножении многочлена на скаляр его содержание умножается на тот же скаляр. Точнее, для любого скаляра $c \in L$ $(c \neq 0)$:

$$d \in \operatorname{cont}(f) \Rightarrow cd \in \operatorname{cont}(cf).$$

[Это вытекает из свойства НОД, выражаемого формулой (38.22), с учетом замечания 38.9.]

Вспомним, что обратимыми многочленами над кольцом L служат обратимые скаляры (элементы L^*), а ассоциированными будут многочлены, отличающиеся лишь обратимым скалярным множителем. После этого становится очевидным, что ассоциированные многочлены имеют одинаковые (как множества) содержания.

Определение 46.2. Многочлен $f(x) \in L[x]$ называется *примитивным*, если его содержание $\mathrm{cont}(f)$ представляет собой класс обратимых элементов кольца L, т. е. если коэффициенты многочлена $f(x)$ являются взаимно простыми, или, что равносильно, если $1 \in \mathrm{cont}(f)$].

Замечание 46.2. Ясно, что многочлен, ассоциированный с примитивным, сам является таковым.

Очевидно также, что всякий многочлен можно представить в виде

$$f(x) = d \cdot f^\circ(x);\ d \in \mathrm{cont}(f), \tag{46.3}$$

где многочлен

$$f^\circ(x) = b_0 x^n + b_1 x^{n-1} + \ldots + b_{n-1} x + b_n \tag{46.4}$$

с коэффициентами $b_k = a_k/d$ $(k = 0, \ldots, n)$ является примитивным.

И обратно, из наличия представления

$$f(x) = \mathrm{c} \cdot g(x);\ \mathrm{c} \in L, \tag{46.5}$$

где многочлен $g(x)$ примитивен, следует [в силу того же свойства (38.22)], что

$$\mathrm{c} \in \mathrm{cont}(f). \tag{46.6}$$

Не только содержание d, но и примитивный многочлен $f^\circ(x)$ в формуле (46.3) определяется с точностью до ассоциированности.

Другими словами, если для двух примитивных многочленов, $g(x)$ и $g_1(x)$, над кольцом L и двух скаляров, c и c_1, из кольца L справедливо равенство

$$f(x) = \mathrm{c} \cdot g(x) = \mathrm{c}_1 \cdot g_1(x), \tag{46.5a}$$

то как скаляры, так и примитивные многочлены ассоциированы:

$$\mathrm{c} \sim \mathrm{c}_1 \text{ (в кольце } L) \text{ и } g(x) \sim g_1(x) \text{ (в кольце } L[x]).$$

В самом деле, как c, так и c_1 должны принадлежать $\mathrm{cont}(f)$ и, следовательно, должны быть ассоциированы, т. е. найдется обратимый элемент $u \in L^*$, такой, что $c_1 = cu$. Сокращая равенство (46.5a) на $c \neq 0$, получим $g_1(x) = ug(x)$.

46.2. Неразложимые элементы в кольце многочленов над факториальным кольцом. Вспомним (см. замечание 44.5), что *неприводимость над кольцом* L определяется по тому же принципу, что и над полем: $f(x)$ будет *неприводимым над* L, если его нельзя представить в виде произведения двух многочленов положительной степени с коэффициентами из L. (Такая неприводимость не равносильна *неразложимости* $f(x)$ как элемента $L[x]$.)

В следующем предложении описываются все неразложимые элементы в кольце $L[x]$.

Предложение 46.1. *Многочлен $f(x)$ с коэффициентами из факториального кольца L является неразложимым элементом кольца $L[x]$ тогда и только тогда, когда*

— либо $f(x) = f_0$ является скаляром, неразложимым в кольце L;

— либо $\deg(f(x)) = n > 0$ и $f(x)$ является примитивным неприводимым многочленом над L.

Доказательство. В кольце многочленов $L[x]$ имеются ненулевые необратимые элементы двух типов: необратимые скаляры и многочлены положительной степени.

Необратимые скаляры будут неразложимы как многочлены тогда и только тогда, когда они будут неразложимы в кольце коэффициентов. [Это с очевидностью вытекает из правила "степень произведения": см. формулу (36.12) и замечание 36.12.]

Для неразложимости многочлена положительной степени необходима (но не достаточна) его неприводимость.

Если неприводимый многочлен имеет содержание, являющееся необратимым элементом кольца L, то формула (46.3) будет свидетельствовать о разложимости этого многочлена.

Если же многочлен $f(x)$ является неприводимым и примитивным, то он неразложим. В самом деле, если бы он представлялся в виде произведения двух необратимых элементов в кольце $L[x]$, то один из этих элементов обязательно был бы необратимым скаляром (противное противоречило бы неприводимости данного многочлена). А это значит, что получилось бы разложение (46.5), из которого, в силу замечания 46.2, вытекало бы утверждение (46.6), противоречащее (поскольку скаляр c необратим) примитивности $f(x)$. □

46.3. Лемма Гаусса. Мы в очередной раз встречаемся с Карлом Фридрихом Гауссом. Сейчас будет доказано предложение (по традиции именуемое *леммой Гаусса*), выражающее ключевое для теории

делимости свойство примитивных многочленов и связанное с ним свойство содержания.

Предложение 46.2. *1. Произведение двух примитивных многочленов (над факториальным кольцом L) также является примитивным многочленом.*

2. Для любых двух многочленов $f(x), g(x) \in L[x]$ справедливо утверждение:

$$(\, d \in \mathrm{cont}(f)\,) \wedge (\, c \in \mathrm{cont}(g)\,) \Rightarrow (\, d \cdot c \in \mathrm{cont}(f \cdot g)\,). \tag{46.7}$$

Доказательство. 1. Пусть многочлен $f(x)$ степени n, заданный формулой (46.1), и многочлен $g(x)$ степени m, заданный формулой

$$g(x) = \sum_{l=0}^{\infty} g_l x^l;\ g_l \in L\ (l \geqslant 0);\ g_m \neq 0; (\forall l > m)\,(g_l = 0), \tag{46.8}$$

являются примитивными.

Рассмотрим многочлен-произведение $h(x) = f(x)g(x)$, который также представим аналогичной формулой

$$h(x) = \sum_{s=0}^{\infty} h_s x^s;\ h_{n+m} \neq 0; (\forall s > n+m)\,(h_s = 0), \tag{46.9}$$

где коэффициенты h_s $(s \geqslant 0)$ определяются [в соответствии с (36.10)] формулами

$$h_s = f_0 g_s + f_1 g_{s-1} + \ldots + f_{s-1} g_1 + f_s g_0. \tag{46.10}$$

(Напомним, что если номер какого-либо коэффициента превышает степень многочлена, то этот коэффициент считается равным нулю.) Надо доказать, что многочлен (46.9) является примитивным.

Предположим противное. Тогда содержание $\mathrm{cont}(f)$ является необратимым и ненулевым элементом кольца L. В силу факториальности L, этот элемент может быть разложен в произведение неразложимых элементов. Возьмем любой из этих элементов, обозначим его p.

В факториальном кольце всякий неразложимый элемент является простым (см. третье утверждение предложения 45.2), поэтому элемент p будет являться простым делителем всех коэффициентов h_s.

В то же время из примитивности многочлена $f(x)$ следует, что p не является делителем всех коэффициентов f_k. Выберем первый по номеру коэффициент, скажем f_{k_0} $(0 \leqslant k_0 \leqslant n)$, который не делится на p; все предыдущие коэффициенты f_k $(k = 0, ..., k_0 - 1)$ будут делиться на p. Аналогичным образом для многочлена $g(x)$ можно будет выбрать коэффициент g_{l_0} $(0 \leqslant l_0 \leqslant m)$ такой, что $p \,|\, g_l$ $(l = 0, ..., l_0 - 1)$ и $p \nmid g_{l_0}$.

Рассмотрим теперь коэффициент h_s при $s = k_0 + l_0$. Распишем для такого s формулу (46.10) несколько подробнее:

$$h_{k_0+l_0} = f_0 g_{k_0+l_0} + f_1 g_{k_0+l_0-1} + ... + f_{k_0-1} g_{l_0+1} + \\ + \boxed{f_{k_0} g_{l_0}} + f_{k_0+1} g_{l_0-1} + ... + f_{k_0+l_0-1} g_1 + f_{k_0+l_0} g_0. \quad (46.11)$$

По предположению левая часть формулы (46.11) делится на p.

В правой части этой формулы, в силу описанного выше выбора номеров k_0 и l_0, все члены, кроме выделенного, тоже делятся на p.

По свойствам отношения делимости, выделенный член также обязан делиться на p. Следовательно, в силу простоты p, хотя бы один из коэффициентов, f_{k_0} или g_{l_0}, должен делиться на p, что противоречит их построению. Первое утверждение предложения доказано.

2. Второе утверждение получается следующим образом. Вынесем из каждого многочлена его содержание [по типу формулы (46.3)]:

$$f(x) = df^\circ(x), \;\; g(x) = cg^\circ(x);$$

затем перемножим эти формулы; получим:

$$h(x) = (\, dc \,)\, h^\circ(x), \quad (46.12)$$

где обозначено

$$h^\circ(x) = f^\circ(x) g^\circ(x). \quad (46.13)$$

По (уже доказанному) первому утверждению, многочлен (46.13) является примитивным.

В силу замечания 46.2, формула (46.12) влечет тот факт, что скаляр dc принадлежит $\mathrm{cont}(fg)$. Таким образом, импликация (46.7) справедлива. □

Замечание 46.3. Многие математики любят писать учебники нарочито небрежно, выражаясь порой нестрого, но рассчитывая на

самостоятельность и здравый смысл читателей. Разумеется, они правы.

Если определить содержание многочлена как *некоторое значение НОД для коэффициентов многочлена*, а затем [вместо импликации (46.7)] написать формулу

$$\operatorname{cont}(fg) = \operatorname{cont}(f)\operatorname{cont}(g), \tag{46.7a}$$

то понимающий читатель сообразит, что здесь — вольность речи: "равенство" понимается "с точностью до ассоциированности" (или как равенство классов ассоциированных элементов).

46.4. Поле частных целостного кольца. В этом пункте нам предстоит познакомиться с одной важной конструкцией из общей алгебры — вложением целостного кольца в его *поле частных* (или, как еще говорят, *поле отношений*).

Конструкция эта весьма простая, и можно сказать, что основные ее идеи вам известны с той счастливой поры школьного детства, когда вас впервые знакомили с *обыкновенными дробями* и алгебраическими действиями над ними.

Однако строгое и полное изложение этих идей является весьма пространным и изобилует тонкими деталями и кропотливыми проверками. Мы считаем пока несвоевременным погружение на такую глубину в абстракции общей алгебры. (Математикам предстоит изучать эти вопросы в третьем семестре.) Поэтому последующее изложение будет беглым и не вполне строгим.

Те, кого такой стиль не устроит, легко смогут найти серьезное изложение данного вопроса (например, в учебниках [2] или [3]).

Другим, которым пока чужды понятия более абстрактные, чем числа, предлагается, начиная с этого пункта, считать, что рассматривается кольцо $K = \mathbb{Z}$, которое вкладывается в поле $F = \mathbb{Q}$.

Пусть K — целостное кольцо (см. определение 36.4). В замечании 36.11 мы уже отмечали, что всякое *подкольцо в поле* является целостным кольцом, и обещали объяснить обратное: как вложить целостное кольцо в поле. Фактически будет сделано большее: мы построим *минимальное* из таких полей, в которые вкладывается K.

Поле частных F для кольца K будет состоять "из дробей" вида $\frac{a}{b}$, где $a, b \in K$ $(b \neq 0)$, понимаемых просто как *пары* элементов из данного кольца, но "равенство пар" будет не обычным "равенством векторов", оно будет определяться следующим образом:

$$\left(\frac{a}{b} = \frac{c}{d}\right) \Leftrightarrow (\,ad = bc\,).$$

(При строгом изложении здесь следовало бы говорить не о равенстве, а об эквивалентности пар, рассматривать классы эквивалентности... Но как раз этого мы обещали не делать.)

Вполне школьным образом определяются алгебраические действия над дробями:

$$\frac{a}{b}+\frac{c}{d}=\frac{ad+bc}{bd};\ \frac{a}{b}\cdot\frac{c}{d}=\frac{ac}{bd}.$$

(Необходима проверка *корректности* этих определений.)

Далее (без особых проблем) проверяются все девять аксиом поля и описывается *вложение* кольца K в поле F: элементу $a\in K$ сопоставляется дробь $\frac{a}{1}$.

Затем устанавливается свойство *минимальности* построенного поля F, состоящее в том, что всякое вложение кольца K в произвольное поле F' *продолжается* до вложения поля F в F', и в заключение доказывается, что такое минимальное поле, в которое вкладывается данное целостное кольцо, определено *однозначно*, с точностью до изоморфизма.

Этим завершается процесс конструирования поля частных.

Замечание 46. 4. Для кольца $\mathbb{Z}$ поле частных совпадает с полем $\mathbb{Q}$; для кольца $\mathbb{Z}[i]$ целых гауссовых чисел (см. пример 38.5) — с полем $\mathbb{Q}[i]$ рациональных гауссовых чисел [вида $z=\frac{a+bi}{c+di}$ $(a,b,c,d\in\mathbb{Z};$ $c^2+d^2\neq 0)$, что легко сводится к виду $z=u+vi$ $(u,v\in\mathbb{Q})$].

Большой интерес представляет поле частных для кольца $K=P[x]$ многочленов над полем P. В этом случае поле частных, обозначаемое $F=P(x)$, будет состоять из *рациональных дробей* вида $\frac{f(x)}{g(x)}$, где числитель и знаменатель являются многочленами.

46.5. Сохранение неприводимости многочленов над факториальным кольцом при переходе к полю частных. Всякий многочлен $f(x)$ с коэффициентами из целостного кольца L можно рассматривать также над полем частных F, соответствующим кольцу L.

Далее всюду предполагается, что L — факториальное кольцо.

Рассмотрим многочлен

$$f(x)=a_0x^n+a_1x^{n-1}+...+a_{n-1}x+a_n;$$
$$a_i\in F\ (i=0,...,n);\ a_0\neq 0 \quad (46.14)$$

с коэффициентами из поля F. (Мы возвращаемся к записи многочленов по убыванию степеней.)

Коэффициенты $f(x)$ представляются дробями, описанными в предыдущем пункте:

$$a_i = \frac{p_i}{q_i};\ p_i, q_i \in L;\ q_i \neq 0\ (i = 0, ..., n); p_0 \neq 0. \tag{46.15}$$

Факториальность кольца L позволяет привести эти дроби к общему знаменателю

$$q \in \text{НОК}(q_0, q_1, ..., q_n), \tag{46.16}$$

т. е. представить коэффициенты в виде

$$a_i = \frac{p'_i}{q};\ p'_i = \frac{q}{q_i} \in L. \tag{46.17}$$

Общий скалярный множитель $1/q$ вынесем за скобку; в скобке останется многочлен той же степени n, что и исходный, с коэффициентами из кольца L:

$$f(x) = \frac{1}{q}(p'_0 x^n + p'_1 x^{n-1} + ... + p'_{n-1} x + p'_n). \tag{46.18}$$

Из многочлена в скобках можно [по принципу формулы (46.3)] вынести его *содержание*

$$p \in \text{НОД}(p'_0, p'_1, ..., p'_{n-1}, p'_n),$$

после чего мы придем к представлению исходного многочлена в виде

$$f(x) = \frac{p}{q} f^\circ(x), \tag{46.19}$$

где многочлен $f^\circ(x) \in L[x]$, с коэффициентами p'_i/p $(i = 0, ..., n)$, является примитивным.

[Теперь самое время сообразить, что проведенные выше выкладки фактически повторяют (в более общей ситуации) преобразования, производившиеся в начале п. 42.1 для случая $L = \mathbb{Z}$; $F = \mathbb{Q}$. Сравните формулы (46.19) и (42.4); различаются лишь обозначения.]

Формула (46.19) означает, в частности, что над полем F многочлен $f(x)$ (с коэффициентами из F) ассоциирован с примитивным многочленом $f^\circ(x)$ (с коэффициентами из L).

Следовательно, вопрос о неприводимости над полем F многочлена с коэффициентами из F сводится к вопросу о неприводимости (тоже над F) многочлена с коэффициентами из кольца L.

Поскольку ясно, что всякий многочлен над кольцом L можно рассматривать над (содержащим это кольцо) полем частных F, то возникает проблема: выяснить, как связана неприводимость над L для многочлена $f(x) \in L[x]$ с его неприводимостью над F.

Ответ дает следующее

Предложение 46.3. *Многочлен $f(x)$ над факториальным кольцом L неприводим над этим кольцом тогда и только тогда, когда он неприводим над полем частных F кольца L.*

Доказательство. В одну сторону утверждение предложения очевидно: если многочлен $f(x) \in L[x]$ неприводим над F, то он неприводим и над L.

Докажем обратное утверждение. Пусть многочлен $f(x)$ неприводим над кольцом L. Предположим, что над полем F он окажется приводимым. Тогда будет иметь место разложение $f(x)$ в произведение двух многочленов $g(x)$ и $h(x)$, положительных степеней, с коэффициентами из F:

$$f(x) = g(x)h(x).$$

Многочлен $f(x) \in L[x]$ мы представим в виде (46.3), а многочлены $g(x)$ и $h(x)$ — в виде (46.19):

$$d\, f^\circ(x) = \frac{s}{t} g^\circ(x) \cdot \frac{r}{v} h^\circ(x),$$

где d, s, t, r, v — ненулевые элементы из кольца L; $f^\circ(x)$, $g^\circ(x)$ и $h^\circ(x)$ — примитивные многочлены над L.

Домножая последнюю формулу на произведение tv, получим равенство

$$(dtv) \cdot f^\circ(x) = (sr) \cdot (\, g^\circ(x)h^\circ(x)\,). \tag{46.20}$$

В правой части (46.20) фигурирует произведение примитивных многочленов, само являющееся, в силу леммы Гаусса (см. предложение 46.2), примитивным многочленом.

Мы оказываемся в условиях замечания 46.2 [см. формулу (46.5а)] и можем сделать вывод:

$$f^\circ(x) \sim g^\circ(x)h^\circ(x). \tag{46.21}$$

Формула (46.21) означает, что многочлен $f^\circ(x)$ ассоциирован (над L) с приводимым (над L) многочленом и, следовательно, сам приводим. Значит, и исходный многочлен $f(x)$ приводим над L. Противоречие. В другую сторону предложение также доказано. □

46.6. Факториальность кольца многочленов над факториальным кольцом. Сейчас мы докажем теорему, существенно обобщающую теорему 45.3: окажутся факториальными не только кольца многочленов *над полями*, но и кольца многочленов *над факториальными кольцами коэффициентов*.

Теорема 46.1. *Кольцо $L[x]$ многочленов над любым факториальным кольцом L факториально.*

Доказательство. Итак, в кольце коэффициентов L предполагаются выполненными условия (Ф.1) и (Ф.2) [см. п. 45.1]. Докажем, что эти условия выполняются также и в $L[x]$.

1. Убедимся в выполнении (Ф.1). Рассмотрим ненулевой и необратимый элемент $f(x) \in L[x]$. Доказательство проведем индукцией по степени $n = \deg(f(x))$.

Если $n = 0$, т. е. $f(x) = f_0$ есть ненулевой и необратимый скаляр, то он разлагается (в кольце L) в произведение неразложимых (в L и, следовательно, в $L[x]$) скаляров.

Пусть теперь $n > 0$ и для всех многочленов степени меньшей n существует разложение на неразложимые множители.

Рассмотрим многочлен $f(x)$ степени n. Представим этот многочлен формулой (46.3), выделив в нем его содержание $d \in \mathrm{cont}(f)$ и примитивный многочлен $f^\circ(x)$.

Если многочлен $f^\circ(x)$ является неприводимым над L, то он будет (по предложению 46.1) неразложимым элементом в $L[x]$. Тогда, разлагая d на неразложиимые скаляры, мы получим разложение $f(x)$ на неразложимые элементы.

Если же многочлен $f^\circ(x)$ приводим, то он разлагается в произведение двух многочленов, степень каждого из которых положительна и меньше n. По предположению индукции, каждый из этих многочленов допускает разложение на неразложимые множители. Кроме того, содержание d также разложимо на неразложимые скаляры. В итоге получится разложение на неразложимые множители для $f(x)$.

Свойство (Ф.1) доказано.

2. Переходим к доказательству (Ф.2). Пусть имеются два разложения на неразложимые элементы для одного и того же многочлена $f(x) \in L[x]$:

$$f(x) = a_1 a_2 ... a_k p_1(x) p_2(x) ... p_s(x) = b_1 b_2 ... b_l q_1(x) q_2(x) ... q_t(x), \quad (46.22)$$

где участвуют неразложимые скаляры

$$a_1, a_2, ..., a_k;\ b_1, b_2, ..., b_l \in L$$

и неприводимые и примитивные многочлены положительной степени

$$p_1(x), p_2(x), ..., p_s(x);\ q_1(x), q_2(x), ..., q_t(x) \in L[x].$$

В силу леммы Гаусса, произведения примитивных многочленов

$$g(x) = p_1(x)p_2(x)...p_s(x);\ h(x) = q_1(x)q_2(x)...q_t(x)$$

являются примитивными многочленами.

В силу замечания 46.2, равенство

$$(a_1a_2..a_k)g(x) = (b_1b_2...b_l)h(x)$$

влечет ассоциированности

$$a_1a_2..a_k \sim b_1b_2...b_l \tag{46.23}$$

(в кольце L) и

$$g(x) = p_1(x)p_2(x)...p_s(x) \sim q_1(x)q_2(x)...q_t(x) = h(x) \tag{46.24}$$

(в кольце $L[x]$).

Ассоциированность (46.23) равносильна равенству

$$a_1a_2...a_k = vb_1b_2...b_l;\ v \in L^*, \tag{46.25a}$$

в котором обратимый множитель v можно отнести, например, к b_1, что не повлияет на неразложимость этого скаляра.

В силу факториальности L, равенство (46.25a) влечет, во-первых, равенство $k = l$ и, во-вторых, попарную ассоциированность $a_i \sim b_i$ $(i = 1, ..., k)$, после подходящей перенумерации множителей.

Ассоциированность (46.24) равносильна равенству

$$p_1(x)p_2(x)...p_s(x) = uq_1(x)q_2(x)...q_t(x);\ u \in L^*, \tag{46.25b}$$

в котором обратимый множитель u можно отнести, например, к первому из неприводимых множителей в правой части, что не повлияет на его неприводимость.

Но, в силу предложения 46.3, неприводимые над L многочлены останутся неприводимыми над полем частных F, соответствующим кольцу L.

Равенство (46.25b) можно рассматривать в кольце $F[x]$ многочленов над полем F. Из теоремы 45.3 известно, что это кольцо факториально. Поэтому указанное равенство влечет, в силу (Ф.2), что, во-первых, $s = t$, и, во-вторых, после подходящей перенумерации,

множители левой части будут ассоциированы соответствующим множителям правой части.

Но будем внимательны, эти ассоциированности имеют место уже над F, т. е. найдутся такие (ненулевые) элементы $r_i, v_i \in L$ (где $i = 1, ..., s$), что

$$p_i(x) = \frac{r_i}{v_i}\, q_i(x).$$

Последнее равенство можно, домножив на v_i, привести к виду

$$v_i\, p_i(x) = r_i\, q_i(x). \tag{46.26}$$

Снова оказываемся (в силу примитивности участвующих в этом равенстве многочленов) в условиях замечания 46.2 и приходим к выводу

$$p_i(x) \sim q_i(x);\ i = 1, ..., s. \tag{46.27}$$

Это уже будут ассоциированности над L, именно те, которые требовалось установить. Доказательство свойства (Ф.2) завершено.

Теорема доказана полностью. □

Пример 46.1. Применяя теорему 46.1 в случае кольца $L = \mathbb{Z}$, мы приходим к выводу, что кольцо $\mathbb{Z}[x]$ многочленов с целыми коэффициентами факториально.

Будет интересным заметить, что это кольцо *не удовлетворяет* условию Безу (Б): хотя в этом кольце (как во всяком факториальном) любые два элемента имеют НОД, линейное представление для НОД, вообще говоря, отсутствует.

В самом деле, элементы 2 и x, очевидно, взаимно просты. (Содержание многочлена x равно 1.) Однако единицу нельзя представить в виде

$$1 = 2u(x) + xv(x).$$

Действительно, свободный член в правой части этого равенства должен быть четным, а в левой части нет ничего, кроме свободного члена, равного 1.

Пример 46.2. Вторым важнейшим следствием теоремы 46.1 будет факториальность колец многочленов от нескольких переменных (над факториальным кольцом коэффициентов), которые определяются индуктивно (см. замечание 36.8):

$$L[x_1, x_2] = L[x_1][x_2]; ...;\ L[x_1, ..., x_n] = L[x_1, ..., x_{n-1}][x_n]. \tag{46.28}$$

Как и в предыдущем примере, будут получаться кольца, не удовлетворяющие условию Безу (даже в случае, когда кольцо коэффициентов L является полем).

Скажем, в кольце $\mathbb{R}[x, y]$ многочленов от двух переменных x, y (над полем $\mathbb{R}$) многочлены x и y взаимно просты, но линейное представление

$$1 = xu(x, y) + yv(x, y)$$

невозможно ни при каких $u(x, y), v(x, y) \in \mathbb{R}[x, y]$. (Почему?)

46.7. Признак Эйзенштейна неприводимости многочлена над факториальным кольцом. Из п. 46.5 нам известно, что изучение неприводимых многочленов над полем частных F факториального кольца L сводится к изучению (примитивных) неприводимых многочленов над L (что может оказаться более простой задачей).

Далее мы докажем самый популярный из признаков неприводимости для многочленов над факториальным кольцом — признак Эйзенштейна. (В своей исходной версии он, разумеется, относился к конкретному кольцу $L = \mathbb{Z}$.) Снова нам удобнее будет воспользоваться формой записи многочленов по возрастанию степеней.

Предложение 46.4 (признак Эйзенштейна). *Рассмотрим многочлен*

$$f(x) = f_0 + f_1 x + ... + f_{n-1} x^{n-1} + f_n x^n, \tag{46.29}$$

положительной степени n над факториальным кольцом L.

Если для некоторого неразложимого (= простого) элемента $p \in L$ старший коэффициент f_n не делится на p, а все остальные коэффициенты f_i $(i = 1, ..., n)$ делятся на p, причем свободный член f_0 не делится на p^2, то многочлен $f(x)$ является неприводимым над кольцом L.

Доказательство. Наше последующее рассуждение аналогично тому, которое применялось при доказательстве предложения 46.2.

Предположим, что многочлен $f(x)$, вопреки утверждению предложения, оказался приводимым, т. е. для него существует разложение в произведение двух многочленов положительной степени:

$$f(x) = g(x)h(x); \tag{46.30}$$

где

$$g(x) = g_0 + g_1 x + ... + g_{m-1} x^{m-1} + g_m x^m; \tag{46.31}$$

$$h(x) = h_0 + h_1 x + ... + h_{l-1} x^{l-1} + h_l x^l; \qquad (46.32)$$

и $g_j, h_k \in L$ $(j = 0, ..., m;\ k = 0, ..., l)$; $g_m, h_l \neq 0$; $m, l > 0$; $m + l = n$.

Приравняем свободные члены в формуле (46.30): $f_0 = g_0 h_0$.

Поскольку $p | f_0$ и элемент p прост, то p делит либо g_0, либо h_0 (оба этих коэффициента одновременно не могут делиться на p, поскольку f_0 не делится на p^2).

Пусть для определенности $p | g_0$ (и тогда $p \nmid h_0$).

Приравняем теперь старшие коэффициенты в (46.30): $f_n = g_m h_l$.

Ясно, что ни g_m, ни h_l не делятся на p (иначе на p делился бы f_n).

Таким образом, среди коэффициентов многочлена (46.31) имеются как делящиеся на p, так и не делящиеся. Рассмотрим коэффициент g_{j_0} с номером j_0 $(0 < j_0 < m)$, наименьшим из таких номеров j, что $p \nmid g_j$ (т. е. при всех $j \in \{0, ..., j_0 - 1\}$ имеем: $p | g_j$).

Приравняем теперь в (46.30) коэффициенты при x^{j_0}:

$$f_{j_0} = g_0 h_{j_0} + g_1 h_{j_0 - 1} + ... + g_{j_0 - 1} h_1 + \boxed{g_{j_0} h_0}. \qquad (46.33)$$

С одной стороны, т. к. $j_0 < m < n$, коэффициент f_{j_0} делится на p.

С другой стороны, в правой части (46.33) все слагаемые, кроме (выделенного) последнего, делятся на p. Выделенное же слагаемое не может делиться на p, поскольку элемент p прост и не делит ни g_{j_0}, ни h_0. Значит, вся правая часть (46.33) не делится на p.

Полученное противоречие доказывает неприводимость исходного многочлена $f(x)$. □

Во всех последующих примерах будут рассматриваться многочлены над кольцом целых чисел (и записываться они будут по убыванию степеней).

Пример 46.3. Многочлен

$$f(x) = 3x^7 - 4x^5 + 2x^4 - 6x^3 - 8x^2 - 4x - 2$$

неприводим над кольцом $\mathbb{Z}$, т. к. его старший коэффициент не делится на простое число $p = 2$; все остальные коэффициенты на это число делятся, но свободный член не делится на $p^2 = 4$.

Пример 46.4. Двучлен

$$f(x) = x^n - pq,$$

где p — простое число, а q — любое целое число, не делящееся на p, является, очевидно, неприводимым.

Замечание 46.5. Рассмотрев последний пример, можем сделать следующее наблюдение.

Над полем $\mathbb{C}$ (и вообще над любым алгебраически замкнутым полем) неприводимыми являются лишь многочлены первой степени. Над полем $\mathbb{R}$ к числу неприводимых добавляются квадратичные многочлены (с отрицательным дискриминантом).

А вот над кольцом $\mathbb{Z}$ (и, следовательно, над полем $\mathbb{Q}$) существуют неприводимые многочлены *любой* положительной степени.

(Только на первый взгляд кажется парадоксальным, что кольцо многочленов над более "простым" полем устроено сложнее, чем над более "сложным". Именно затем и расширялось ("усложнялось") самое простое из числовых полей, поле рациональных чисел, чтобы "упростилось" строение кольца многочленов.)

Пример 46.5. Рассмотрим теперь так называемые *многочлены деления круга*:

$$f(x) = x^{p-1} + x^{p-2} + ... + x + 1, \tag{46.34}$$

где p — простое натуральное число.

[Термин возник ввиду наличия тесной связи (см. ниже) многочлена (46.34) с многочленом $x^p - 1$, комплексные корни которого делят на p равных частей единичную окружность (см. § 33).]

Непосредственно применить к многочлену (46.34) признак Эйзенштейна нельзя, поскольку все его коэффициенты равны 1 и ни о каких простых делителях речи идти не может.

Заметим, однако, что этот многочлен можно "свернуть" по формуле суммы геометрической прогрессии:

$$f(x) = \frac{x^p - 1}{x - 1}. \tag{46.35}$$

[Причем ни о какой О.Д.З. (кошмар, характерный для всех "школярских" задач!) здесь беспокоиться не придется: многочлены рассматриваются не как функции, а как формальные суммы (или векторы); многочлен $x-1$ делит многочлен x^p-1 (хотя бы в силу теоремы Безу); правая часть (46.35) есть соответствующее частное.]

Далее необходимо сделать такое наблюдение: при "сдвиге" аргумента (в данном случае — на произвольное целое число; фактически понадобится сдвиг на единицу) свойство приводимости (неприводимости) многочлена сохраняется. Точнее, многочлен $f(x)$ является приводимым (неприводимым) тогда и только тогда, когда приводим (неприводим) многочлен

$$g(x) = f(x+1). \tag{46.36}$$

(Попробуйте сами провести совсем простое доказательное рассуждение.)

Для "сдвинутого" многочлена (46.36) получим (используя формулу бинома Ньютона):

$$\begin{aligned} g(x) &= \frac{(x+1)^p - 1}{x} = \\ &= \frac{1}{x}(x^p + C_p^1 x^{p-1} + C_p^2 x^{p-2} + ... + C_p^{p-2} x^2 + C_p^{p-1} x) = \\ &= x^{p-1} + C_p^1 x^{p-2} + C_p^2 x^{p-3} + ... + C_p^{p-2} x + C_p^{p-1}. \end{aligned}$$

В окончательном выражении старший коэффициент, будучи единицей, не делится на p.

Докажем, что все остальные коэффициенты C_p^k $(k = 1, ..., p-1)$ делятся на p. В самом деле, число

$$C_p^k = \frac{p \cdot (p-1) \cdot ... \cdot (p-k+1)}{1 \cdot 2 \cdot ... \cdot k} \tag{46.37}$$

является натуральным (это — число сочетаний из p элементов по k), т. е. числитель дроби в правой части (46.37) делится нацело на знаменатель.

Как числитель, так и знаменатель содержат k сомножителей (где $1 \leqslant k \leqslant p-1$), причем в числителе присутствует множитель p. Все сомножители в знаменателе меньше p, и следовательно, ни один из них не делится на p. Простота p влечет теперь тот факт, что весь знаменатель не делится на p. Значит, рассматриваемая дробь не может быть сокращена на p и получающееся (после всех возможных сокращений) целое число C_p^k будет делиться на p.

Остается проверить, что свободный член не делится на p^2. Но это так, поскольку $C_p^{p-1} = C_p^1 = p$.

По признаку Эйзенштейна, многочлен $g(x)$ и с ним вместе $f(x)$ — неприводимы.

Необходимо добавить, что при *составном* показателе n многочлен

$$f(x) = x^{n-1} + x^{n-2} + ... + x + 1$$

является *приводимым*. Об этом свидетельствует формула

$$\frac{x^{kl}-1}{x-1} = \frac{x^{kl}-1}{x^k-1} \cdot \frac{x^k-1}{x-1}.$$

Замечание 46.6. Maple прекрасно справляется с *тестированием* многочленов (с целыми или рациональными коэффициентами) на неприводимость, а также с задачей *разложения* на неприводимые множители. Примеры уже рассматривались (см. пп. 43.4 и 44.5).

Познакомиться с алгоритмами *факторизации многочленов* (являющимися важнейшей частью всех систем компьютерной алгебры) можно, например, по превосходной книге В. В. Прасолова "Многочлены" (М.: МЦНМО, 2003).

§ 47. Дифференцирование в кольце многочленов. Отделение кратных множителей

47.1. Понятие характеристики поля. В настоящем параграфе мы попытаемся развить в кольце $P[x]$ многочленов над полем P *дифференциальное исчисление*. Но предварительно нам предстоит познакомиться с важнейшим для теории полей понятием *характеристики* поля.

Рассмотрим *аддитивную группу* $(P;+)$ поля P. Элемент $1 \in P$ в этой группе может иметь

— либо *конечный порядок* m, если сумма m штук (но не меньше) единиц равна нулю:

$$\underbrace{1+1+...+1}_{m \text{ раз}} = m \cdot 1 = 0,$$

где использовано "мультипликативное" обозначение $m \cdot 1$ для этой суммы;

— либо *бесконечный порядок*, когда все такие суммы отличны от нуля (и, как следствие, все попарно различны).

(Понятие порядка элемента, для случая *мультипликативных групп* обсуждалось в п. 18.2 (порядок перестановки) и 33.1 (порядок ненулевого комплексного числа). Здесь, в аддитивной ситуации, роль нейтрального элемента играет не единица, а нуль и вместо операции *возведения в целую степень* действует операция *умножения на целое число*.)

В общем случае порядок элемента группы может быть произвольным натуральным числом. Ниже мы убедимся, что в рассматриваемой ситуации, для аддитивной группы поля, порядок элемента 1 (если он конечен) является *простым* числом.

В самом деле, наличие в поле *двух* алгебраических действий, сложения и умножения, связанных дистрибутивным законом, позволяет в случае *составного* числа m ($m = k \cdot l$; $1 < k, l < m$) провести следующую выкладку:

$$0 = m \cdot 1 = \underbrace{1 + 1 + ... + 1}_{m \text{ раз}} =$$

$$= (\underbrace{1 + 1 + ... + 1}_{k \text{ раз}}) \cdot (\underbrace{1 + 1 + ... + 1}_{l \text{ раз}}) = (k \cdot 1) \cdot (l \cdot 1) \Rightarrow$$

$$\Rightarrow (\, k \cdot 1 = 0 \,) \vee (\, l \cdot 1 = 0 \,)$$

— и получить противоречие с определением порядка.

Итак, аддитивный порядок полевой единицы либо равен простому числу, либо бесконечен.

Беря произвольный ненулевой элемент поля $a \in P \setminus \{0\}$, мы легко убеждаемся, что он имеет такой же порядок, что и $1 \in P$.

(Это следует из формулы

$$\underbrace{a + a + ... + a}_{m \text{ раз}} = a \cdot (\underbrace{1 + 1 + ... + 1}_{m \text{ раз}}),$$

которую можно выразить "мультипликативно":

$$m \cdot a = a \cdot (m \cdot 1);\ m \in \mathbb{N};\ a, 1 \in P;$$

подробности рассуждения восстановите сами.)

Определение 47.1. *Характеристикой* поля P называется
— либо (простое) натуральное число p, если $p \cdot 1 = 0$ в поле P,
— либо *нуль*, если все элементы $m \cdot 1$ ($m \in \mathbb{N}$) отличны от нуля.
Для характеристики поля используется обозначение $\mathrm{char}(P)$.

Замечание 47.1. Вот вам еще пример ситуации, когда в математической терминологии традиция побеждает логику: было бы, разумеется, более естественным *поля нулевой характеристики* именовать *полями бесконечной характеристики*.

Пример 47.1. Для всех числовых полей ($P = \mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ и др.) $\mathrm{char}(P) = 0$. Для поля $\mathbb{F}_p$ классов вычетов по простому модулю p характеристика $\mathrm{char}(\mathbb{F}_p) = p$.

47.2. Понятие (формальной) производной от многочлена. Определим теперь отображение кольца многочленов $P[x]$ в себя

$$' : P[x] \longrightarrow P[x];\ f(x) \mapsto f'(x);\ f(x) \in P[x], \tag{47.1}$$

называемое *дифференцированием* или взятием *производной* и удовлетворяющее следующим условиям:

1) (*линейность*) для любых многочленов $f(x), g(x) \in P[x]$ и любых скаляров $\lambda, \mu \in P$ справедливо

$$(\lambda f(x) + \mu g(x))' = \lambda f'(x) + \mu g'(x); \tag{47.2}$$

2) (*свойство Лейбница,* или *правило дифференцирования произведения*) для любых многочленов $f(x), g(x) \in P[x]$ справедливо

$$(\, f(x)g(x)\,)' = f'(x)g(x) + f(x)g'(x); \tag{47.3}$$

3) отображение (47.1) переводит единичный скаляр в нуль:

$$(1)' = 0; \tag{47.4}$$

4) отображение (47.1) переводит многочлен x в единицу:

$$(x)' = 1. \tag{47.5}$$

При формулировке следующего предложения будет использоваться запись многочлена по возрастанию степеней.

Предложение 47.1. *Отображение (47.1), удовлетворяющее условиям (47.2) — (47.5), существует, определено однозначно и сопоставляет многочлену степени n*

$$f(x) = f_0 + f_1 x + f_2 x^2 + \ldots + f_{n-1}x^{n-1} + f_n x^n; \qquad f_i \in P\ (i = 0, \ldots, n);\ f_n \neq 0, \tag{47.6}$$

многочлен

$$f'(x) = f_1 + 2f_2x^2 + ... + (n-1)f_{n-1}x^{n-2} + nf_nx^{n-1} \qquad (47.7)$$

степени, не превосходящей $n-1$.

Доказательство. 1. Докажем сначала единственность искомого оператора дифференцирования, т. е. докажем, что если отображение (47.1), обладающее требуемыми свойствами, существует, то лишь одно. Формулой (47.4) это отображение задано (имеет нулевое значение) на единичном скаляре. Отсюда немедленно следует, что оно имеет нулевое значение на любом скаляре $\lambda \in P$:

$$(\lambda)' = (\lambda \cdot 1)' = \lambda \cdot (1)' = 0.$$

Формулой (47.5) отображение (47.1) задано на одночлене x. Свойства (47.4) достаточно, чтобы (однозначно) задать это отображение на всевозможных одночленах x^m $(m \geqslant 0)$.

В самом деле, докажем по индукции формулу

$$(x^m)' = mx^{m-1} : \ m = 1, 2, 3, ... \qquad (47.8)$$

Сделаем мы это после небольшой "беседы" с читателем. Беседа будет состоять из двух замечаний. (Доказательство пока прерывается.)

Замечание 47.2. Автору в очередной раз приходится ловить себя (и своих коллег) на неаккуратности обозначений и формулировок.

Натуральное число m, вообще говоря, не обязано содержаться в поле P. Более корректной была бы следующая запись правой части формулы (47.8): $(m \cdot 1)x^{m-1}$, где произведение полевой единицы 1 на натуральное число m понимается как сумма m единиц (см. п. 47.1).

Однако обычно пишут именно так, как в (47.8). А то, что сказано выше,"имеется в виду". Если бы математики всегда соблюдали предписанные ими же самими строгости, то они не смогли бы работать. Приходится утешать себя тем, что в случае необходимости полная строгость всегда может быть наведена.

В связи с этим можно припомнить неосторожный пассаж из трактата Н. Бурбаки (об этом великом сочинении мы упоминали в замечании 39.1; сейчас речь идет о книге первой — "Теория множеств"; М.: Мир, 1965). Автор трактата пишет о строгости как о некотором "горизонте", чем дает повод редактору русского перевода съязвить:

дескать, сравнение оказывается более точным, чем того хотелось бы автору.

Замечание 47.3. И снова о формуле (47.8). На этот раз понадобится понятие характеристики поля, введенное выше.

Если $\operatorname{char}(P) = 0$, то производная от одночлена степени m является одночленом степени $m-1$.

Если же $\operatorname{char}(P) = p$ — простое число, то $p \cdot 1 = 0$, и поэтому $(x^p)' = 0$, т. е. бывают многочлены положительной степени с нулевой производной.

Теперь мы возвращаемся к прерванному доказательству предложения, и конкретно к доказательству (вызвавшей столь пространные обсуждения) формулы (47.8).

При $m = 1$ эта формула справедлива (если, конечно, принять обычное соглашение: $x^0 = 1$).

Если предположить, что она справедлива при некотором $m = k$, то при $m = k+1$ получим:

$$(x^{k+1})' = (x^k \cdot x)' \overset{(47.3)}{=\!=} (x^k)'x + x^k \cdot (x)' = \\ = (kx^{k-1}) \cdot x + x^k \cdot 1 = kx^k + x^k = (k+1)x^k,$$

т. е. (47.8) оказывается справедливой и при $m = k+1$. Формула доказана.

Одночлены x^m образуют базис в (бесконечномерном) линейном пространстве многочленов. Задание линейного отображения на базисе однозначно определяет это отображение на всем пространстве. (Это ничего, что оно бесконечномерно: все равно всякий многочлен однозначно представляется в виде линейной комбинации конечного числа одночленов.)

Продолжая по линейности отображение $'$ с одночленов на произвольные многочлены (47.6), мы получим формулу (47.7).

2. Имея готовую формулу (47.7), можно обратиться к доказательству существования оператора дифференцирования.

Это доказательство сводится теперь к проверке того, что отображение, заданное указанной формулой, удовлетворяет всем требованиям (47.2) — (47.5).

Выполнение требований (47.4) и (47.5) очевидно.

Линейность (47.2) тоже практически очевидна (в силу того что сложение многочленов и умножение многочлена на скаляр производятся "покомпонентно"):

$$
\begin{aligned}
&(f(x)+g(x))' = \\
&= \big((f_0+g_0) + (f_1+g_1)x + (f_2+g_2)x^2 + \ldots + (f_N+g_N)x^N \big)' = \\
&\qquad = (f_1+g_1) + 2(f_2+g_2)x + \ldots + N(f_N+g_N)x^{N-1} = \\
&= (f_1 + 2f_2x + \ldots + Nf_Nx^{N-1}) + (g_1 + 2g_2x + \ldots + Ng_Nx^{N-1}) = \\
&\qquad\qquad = f'(x) + g'(x),
\end{aligned}
$$

где, чтобы не задумываться о (возможно, различных) степенях многочленов, выбрано $N = \max(\deg(f(x)), \deg(g(x)))$.

Остается проверить свойство Лейбница (47.3). Сначала докажем частный случай этой формулы для одночленов x^k и x^l, где $k, l \geqslant 0$:

$$
(x^k \cdot x^l)' = (x^k)'x^l + x^k(x^l)'. \tag{47.9}
$$

Если $k, l \geqslant 1$, то доказательством служит следующая выкладка:

$$
\begin{aligned}
(x^k \cdot x^l)' = (x^{k+l})' = (k+l)x^{k+l-1} = \\
= kx^{k-1}x^l + lx^{l-1}x^k = (x^k)'x^l + x^k(x^l)'.
\end{aligned}
$$

[При "честной" записи вычислений в (47.9) в качестве коэффициентов должны фигурировать выражения вида $k \cdot 1$ и т. п. Обратите также внимание на использование ассоциативного закона для сложения, но в "завуалированном" виде, больше похожем на дистрибутивный закон: $(k+l)\cdot 1 = k\cdot 1 + l\cdot 1$.]

Если же k или l обращаются в нуль, то доказательство (47.9) совсем тривиально.

Теперь, пользуясь свойством линейности оператора $'$ и справедливостью правила дифференцирования произведения одночленов, можно доказать это правило в общем случае, для произвольных многочленов $f(x)$ и $g(x)$ степеней n и m соответственно:

$$
\big(f(x)\cdot g(x) \big)' = \left(\left(\sum_{k=0}^{n} f_k x^k \right) \cdot \left(\sum_{l=0}^{m} g_l x^l \right) \right)' =
$$

$$= \left(\sum_{k=0}^{n}\sum_{l=0}^{m}(f_k g_l)(x^k x^l)\right)' \overset{(47.2)}{=\!=} \sum_{k=0}^{n}\sum_{l=0}^{m}(f_k g_l)(x^k x^l)' \overset{(47.9)}{=\!=}$$

$$= \sum_{k=0}^{n}\sum_{l=0}^{m}(f_k g_l)((x^k)' x^l + x^k (x^l)') =$$

$$= \sum_{k=0}^{n}\sum_{l=0}^{m}(f_k (x^k)')(g_l x^l) + \sum_{k=0}^{n}\sum_{l=0}^{m}(f_k x^k)(g_l (x^l)') =$$

$$= \sum_{k=0}^{n} f_k (x^k)' \sum_{l=0}^{m} g_l x^l + \sum_{k=0}^{n} f_k x^k \sum_{l=0}^{m} g_l (x^l)' \overset{(47.2)}{=\!=} f'(x)g(x) + f(x)g'(x).$$

Формула (47.3) доказана.

[По ходу преобразований использованы правила обращения с суммами (и двойными суммами), к которым вы должны бы уже привыкнуть. Например, при переходе от четвертой строки к пятой мы (в первой из двойных сумм) вынесли из-под внутренней суммы (по индексу l) выражение, зависящее лишь от индекса k; после этого внутренняя сумма оказывается равной $g(x)$; мы выносим этот множитель из-под знака суммы по k (вправо); оставшаяся сумма равна $f'(x)$. Не исключено, однако, что, испугавшись двойных сумм, вы не различили за их "частоколом" простую идею: *произведение* многочленов *билинейно* (линейно по каждому из двух сомножителей); *дифференцирование линейно*; формулу "производная произведения" достаточно доказывать для *произведений базисных векторов* (одночленов).]

В заключение упомянем свойство степени:

$$\deg(f'(x)) \leqslant \deg(f(x)) - 1 \tag{47.10}$$

для любого многочлена $f(x) \in P[x]$ положительной степени. (Причем, как показано в замечании 47.3, неравенство может быть строгим.) $\square$

Замечание 47.4. Отметим несколько простых свойств оператора дифференцирования многочленов, непосредственно следующих из определяющих свойств (47.2) — (47.5).

1. Формула Лейбница для нескольких сомножителей

$$(f_1(x)f_2(x)...f_m(x))' = f_1'(x)f_2(x)...f_m(x) + \\ + f_1(x)f_2'(x)...f_m(x) + ... + f_1(x)f_2(x)...f_m'(x) \tag{47.11}$$

получается из (47.3) индукцией по m.

2. Частным случаем (47.11) является формула для производной от степени многочлена:

$$(\,(f(x))^m\,)' = m(f(x))^{m-1}f'(x);\ m \in \mathbb{N}. \tag{47.12}$$

3. Из (47.12), в свою очередь, получается:

$$(\,(x-c)^m\,)' = m(x-c)^{m-1};\ c \in P;\ m \in \mathbb{N}. \tag{47.13}$$

47.3. Высшие производные и формула Тейлора для многочленов. Вторая производная от многочлена определяется как производная от первой производной. Общее определение дается по индукции:

$$f^{(0)}(x) = f(x);\ f^{(1)}(x) = f'(x);$$
$$f^{(k+1)}(x) = (\,f^{(k)}(x)\,)';\ k = 1, 2, \ldots \tag{47.14}$$

Также по индукции доказывается *формула Лейбница для высших производных* произведения, очень похожая на формулу бинома Ньютона [см. (31.26)]:

$$(\,f(x)g(x)\,)^{(m)} = \sum_{k=0}^{m} C_m^k f^{(k)}(x)g^{(m-k)}(x);\ m \in \mathbb{N}. \tag{47.15}$$

Итерируя формулу (47.13), получим представление для высших производных:

$$(\,(x-c)^m\,)^{(k)} = m(m-1)\ldots(m-k+1)(x-c)^{m-k};$$
$$c \in P;\ m, k \in \mathbb{N};\ k \leqslant m;\ (\,(x-c)^m\,)^{(m+1)} = 0. \tag{47.16}$$

А теперь мы повторно обратимся к формуле Тейлора (см. п. 41.2), представляющей из себя "переразложение" многочлена

$$f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n;$$
$$a_k \in P\ (k = 0, \ldots, n);\ a_0 \neq 0; \tag{47.17}$$

по степеням двучлена $x - c$. Это переразложение получилось у нас в виде формулы (41.8), в которой коэффициенты h_k в правой части (*коэффициенты Тейлора*) были занумерованы по возрастанию степеней:

$$f(x) = h_0 + h_1(x-c) + h_2(x-c)^2 + h_3(x-c)^3 + ...+ \\ + h_{n-1}(x-c)^{n-1} + h_n(x-c)^n;\ h_n = a_0. \quad (47.18)$$

Продифференцируем (47.18), пользуясь формулой (47.13):

$$f'(x) = h_1 + 2h_2(x-c) + 3h_3(x-c)^2 + ...+ \\ + (n-1)h_{n-1}(x-c)^{n-2} + nh_n(x-c)^{n-1}. \quad (47.18')$$

Повторное дифференцирование даст:

$$f''(x) = 2h_2 + (3\cdot 2)h_3(x-c) + ...+ \\ + (n-1)(n-2)h_{n-1}(x-c)^{n-3} + n(n-1)h_n(x-c)^{n-2}; \quad (47.18'')$$

k-кратное $(k \leqslant n)$:

$$f^{(k)}(x) = k!h_k + (\,(k+1)k\cdot ...\cdot 2\,)h_{k+1}(x-c) + ...+ \\ + n(n-1)...(n-k+1)h_n(x-c)^{n-k}; \quad (47.18^{(k)})$$

в частности, при $k = n$ получится скаляр:

$$f^{(n)}(x) = n!h_n. \quad (47.18^{(n)})$$

Подставляя в формулы (47.18) — $(47.18^{(n)})$ значение $x = c$, получим:

$$f^{(k)}(c) = k!h_k;\ k = 0, 1, 2, ..., n. \quad (47.19)$$

Из формул (47.19) можно выразить "кому что надо": если, скажем, коэффициенты Тейлора найдены с помощью *схемы Горнера*, то формулы (47.19) позволяют вычислить значения всех производных данного многочлена в точке $c \in P$; можно, наоборот, придать формуле Тейлора "аналитический вид", выразив ее коэффициенты через производные:

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}(x-c)^k. \quad (47.20)$$

Замечание 47.5. Именно в записи (47.20) формула Тейлора (для многочленов) встретится вам в курсе математического анализа. В этой науке по Тейлору разлагаются не только многочлены, но и другие, достаточное число раз дифференцируемые функции; для них разложение должно содержать дополнительный (*остаточный*) член.

47.4. Кратность корня многочлена и значения производных. В п. 41.3 было выведено правило: элемент $c \in P$ является корнем кратности m $(1 \leqslant m \leqslant n)$ для многочлена $f(x) \in P[x]$ положительной степени n тогда и только тогда, когда коэффициенты Тейлора h_k обращаются в нуль при $k = 0, ..., m-1$, а коэффициент h_m отличен от нуля. Теперь это правило можно переформулировать следующим образом.

Предложение 47.2. *Элемент $c \in P$ является корнем кратности $m \geqslant 1$ для многочлена $f(x) \in P[x]$ тогда и только тогда, когда*

$$f(c) = f'(c) = ... = f^{(m-1)}(c) = 0;\ f^{(m)}(c) \neq 0. \tag{47.21}$$

Доказательство немедленно следует из формул (47.19). □

Замечание 47.6. В частности, *однократные* корни многочлена (они еще — очень неудачно — называются *простыми*; не путайте с простотой в смысле теории делимости) характеризуются тем, что они не являются корнями (первой) производной.

47.5. Уменьшение кратности неприводимого множителя многочлена при дифференцировании. Обратимся теперь к разложению (45.13) произвольного многочлена $f(x)$ положительной степени (над полем P) на неприводимые множители и выясним, как связано разложение на неприводимые множители для производной $f'(x)$ с разложением для данного многочлена.

Предложение 47.3. *Предположим, что поле P имеет нулевую характеристику. Тогда*

1) если многочлен $p(x)$ является неприводимым делителем кратности $m \geqslant 1$ для многочлена $f(x)$, то $p(x)$ является делителем кратности $m-1$ для многочлена $f'(x)$;

2) в частности, если элемент $c \in P$ является корнем кратности $m \geqslant 1$ для многочлена $f(x) \in P[x]$, то он является корнем кратности $m-1$ для производной $f'(x)$ этого многочлена.

Доказательство. 1. Начнем доказательство с напоминания (см. замечание 40.1) о том, что, как было условлено, "корень кратности 0" — это "не корень". Аналогичная условность предполагается и применительно к неприводимым делителям: если $(p(x))^m | f(x)$, а $(p(x))^{m+1} \nmid f(x)$, то $p(x)$ считается делителем кратности m; "делитель кратности 0" — это "не делитель".

По условию, имеет место представление

$$f(x) = (p(x))^m g(x), \tag{47.22}$$

где $p(x) \nmid g(x)$, что, в силу неприводимости $p(x)$, равносильно взаимной простоте многочленов $g(x)$ и $p(x)$ [наличие у этих многочленов общего делителя положительной степени противоречило бы неприводимости $p(x)$].

Продифференцируем равенство (47.22), используя свойства производной (47.3) и (47.12):

$$f'(x) = m(p(x))^{m-1} g(x) + (p(x))^m g'(x). \tag{47.23}$$

Если $m = 1$, то $p(x)$ делит второе слагаемое в (47.23) и не делит первое. Следовательно, $p(x)$ не делит $f'(x)$.

Если $m > 1$, то приходится использовать тот факт, что

$$\operatorname{char}(P) = 0.$$

Равенство нулю характеристики поля обеспечивает нам выполнение неравенства $m \cdot 1 \neq 0$ [без этого предположения первое слагаемое в (47.23) могло бы оказаться нулевым и дальнейшее рассуждение не проходило бы].

Очевидно, что правая часть (47.23) делится на $(p(x))^{m-1}$. Но она не делится на $(p(x))^m$. (Иначе, по свойствам отношения делимости, многочлен $p(x)$ делил бы $g(x)$, а это противоречит предположению.)

Значит, $(p(x))^m \nmid f'(x)$. Первое утверждение доказано.

2. Второе утверждение является частным случаем первого, поскольку многочлены первой степени $x - c$ являются неприводимыми и, по определению, кратность корня c (для некоторого многочлена) есть не что иное, как кратность неприводимого делителя $x - c$ (для того же многочлена). □

47.6. Отделение кратных неприводимых множителей (кратных корней). Рассмотрим так называемую задачу *отделения кратных неприводимых множителей* многочлена (*кратных корней* многочлена). [Не очень вразумительный термин "отделение", возможно, является следствием неудачного перевода.]

Задача состоит в следующем: по заданному многочлену $f(x)$ положительной степени найти другой многочлен $g(x)$, который имел бы все *те же неприводимые множители* в разложении типа (45.13),

что и $f(x)$, но лишь *однократными* (отыскание самих неприводимых множителей при этом производиться не должно!).

В случае неприводимых множителей вида $x - c$ говорят об *отделении кратных корней.*

Если данное поле является *алгебраически замкнутым*, то обе версии задачи равносильны.

Замечание 47.7. Можно привести следующую мотивировку для постановки задачи отделения кратных корней. Численные методы приближенного отыскания корней многочленов, как правило, хорошо работают, лишь если эти корни однократные. Поэтому перед применением численных методов многочлен следует "подготовить", избавившись от кратностей, но не потеряв ни одного из корней.

Решение поставленной выше задачи отделения достигается применением производной; его описанию посвящено следующее

Предложение 47.4. *Пусть многочлен $f(x) \in P[x]$ степени n над полем P характеристики нуль представлен своим разложением на (нормализованные) неприводимые множители:*

$$f(x) = a_0 p_1(x)^{m_1} p_2(x)^{m_2} \dots p_s(x)^{m_s}. \tag{47.24}$$

Пусть

$$d(x) = (\, f(x),\, f'(x)\,) \tag{47.25}$$

есть (нормализованный) НОД данного многочлена и его производной.

Тогда $d(x)$ представляется разложением

$$d(x) = p_1(x)^{m_1-1} p_2(x)^{m_2-1} \dots p_s(x)^{m_s-1}, \tag{47.26}$$

а многочлен

$$g(x) = \frac{f(x)}{d(x)} \tag{47.27}$$

решает задачу отделения кратных неприводимых множителей, т. е. имеет разложение

$$g(x) = a_0 p_1(x) p_2(x) \dots p_s(x). \tag{47.28}$$

Доказательство. Согласно предложению 47.3 (в котором используется условие обращения в нуль характеристики поля), многочлен

$f'(x)$ имеет в своем разложении неприводимые множители $p_k(x)$, где $k = 1, ..., s$, с кратностями $m_k - 1$. Возможно, некоторые из этих кратностей обращаются в нуль, и тогда соответствующие множители в разложении для производной будут просто отсутствовать. Кроме того, в разложение для $f'(x)$ могут входить другие неприводимые многочлены, не участвующие в разложении (47.24).

Вследствие этого, НОД $d(x)$ будет иметь разложение (47.25), а многочлен $g(x)$ — разложение (47.27). □

Пример 47.2 (с использованием Maple, без комментариев).

```
> f := x ^ 24 + 7* x^ 23 + 24*x^ 22 + 56*x^ 21 + 92*x^ 20
  + 81*x^ 19 -61* x^ 18 -400*x^ 17 - 924*x^ 16 - 1456*x^ 15
  - 1633*x^ 14 - 1083*x^ 13 + 392*x^ 12 + 2732*x^ 11
  + 5472*x^ 10 + 7819*x^ 9 + 9121*x^ 8 + 9120*x^ 7 + 7884*x^ 6
  + 5856*x^ 5 + 3732*x^ 4 + 2016*x^ 3 + 864*x^ 2 + 252*x + 36 ;
```

$$f := x^{24} + 7x^{23} + 24x^{22} + 56x^{21} + 92x^{20} + 81x^{19} - 61x^{18} - 400x^{17} - 924x^{16} - 1456x^{15} - 1633x^{14} - 1083x^{13} + 392x^{12} + 2732x^{11} + 5472x^{10} + 7819x^9 + 9121x^8 + 9120x^7 + 7884x^6 + 5856x^5 + 3732x^4 + 2016x^3 + 864x^2 + 252x + 36$$

```
> factor( f );
```

$$(x^4 + x^3 + x^2 + x + 1)^3 (x^4 - 6)^2 (x + 1)^4$$

```
> d := gcd( f, diff( f, x ) ); factor( d );
```

$$d := x^{15} + 5x^{14} + 12x^{13} + 20x^{12} + 22x^{11} + 4x^{10} - 38x^9 - 92x^8 - 148x^7 - 192x^6 - 199x^5 - 167x^4 - 120x^3 - 71x^2 - 30x - 6$$

$$(x^4 - 6)(x^4 + x^3 + x^2 + x + 1)^2 (x + 1)^3$$

```
> divide( f, d, 'q' ); g := q; factor( g );
```

$$true$$

$$g := x^9 + 2x^8 + 2x^7 + 2x^6 - 4x^5 - 11x^4 - 12x^3 - 12x^2 - 12x - 6$$

$$(x + 1)(x^4 + x^3 + x^2 + x + 1)(x^4 - 6)$$

Пример 47.3. Рассмотрим поле $\mathbb{F}_p$ классов вычетов по простому модулю p и многочлен $f(x) = x^p - x$. Тогда, как объяснялось в замечании 47.3, $f'(x) = -1$. Производная является ненулевым скаляром ($-1 = p-1 \in \mathbb{F}_p$) и поэтому не имеет корней. Следовательно, исходный многочлен имеет лишь однократные корни.

§ 48. Первоначальные понятия теории многочленов от нескольких переменных

48.1. Мультииндексы и их лексикографическое упорядочение. В первом семестре мы лишь познакомимся с очень большой и сложной теорией многочленов от нескольких переменных. Более подробное изложение можно будет при желании найти в учебниках, указанных в списке литературы.

В замечании 36.8 и примере 46.2 был намечен индуктивный подход к определению многочленов от нескольких переменных и даже была доказана *факториальность* кольца $L[x_1, x_2, ..., x_n]$ многочленов от n переменных с коэффициентами из факториального кольца L. В частности, являются факториальными кольца многочленов от нескольких переменных над полем.

Но здесь мы изложим иной подход к построению теории "многопеременных" (калька с английского "multivariant") многочленов, основанный на применении так называемых *мультииндексов*.

Пусть n — натуральное число, большее единицы (оно будет "количеством переменных").

Определение 48.1. *Мультииндексами* длины n будем называть n-мерные векторы с неотрицательными целыми компонентами вида

$$I = (i_1\, i_2 \,...\, i_n);\ i_k \in \mathbb{Z};\ i_k \geqslant 0\ (k = 1, ..., n). \tag{48.1}$$

(Запятые без особой необходимости не ставятся.)

Замечание 48.1. Мультииндексы длины n можно считать элементами прямой степени

$$\mathbb{Z}_+^n = \underbrace{\mathbb{Z}_+ \times \mathbb{Z}_+ \times ... \times \mathbb{Z}_+}_{n \text{ раз}}, \tag{48.2}$$

т. е. прямого произведения n экземпляров множества $\mathbb{Z}_+$ неотрицательных целых чисел.

Мультииндексы из $\mathbb{Z}_+^n$ можно складывать (по принципу сложения векторов — покомпонентно); среди них имеется нулевой элемент — мультииндекс $(0\,0\,...\,0)$.

Определим отношение порядка на множестве $\mathbb{Z}_+^n$.

Определение 48.2. Пусть $I, J \in \mathbb{Z}^n$ — два мультиииндекса:

$$I = (i_1\, i_2\, ...\, i_n);\ \ J = (j_1\, j_2\, ...\, j_n).$$

Будем говорить, что I *раньше* J (J *позже* I) и фиксировать этот факт с помощью "искривленного" знака неравенства $I \prec J$ (соответственно $J \succ I$), если найдется номер m $(1 \leqslant m \leqslant n)$ такой, что все компоненты мультииндексов I и J с номерами меньшими m (если таковые имеются) одинаковы, а компоненты с номером m связаны неравенством $i_m < j_m$. (Иногда этот факт выражают другими словами: первая из ненулевых разностей $i_s - j_s$ $(s = 1, ..., n)$ отрицательна.)

Будем использовать также обозначения $\preccurlyeq$ ("раньше или равен") и $\succcurlyeq$ ("позже или равен") для "нестрогого варианта" рассматриваемых отношений.

Предложение 48.1. *1.* (Трихотомия.) *Для любых двух мультииндексов $I, J \in \mathbb{Z}_+^n$ выполняется одно и только одно из трех соотношений:*

$$(I \prec J) \vee (I = J) \vee (I \succ J).$$

2. (Транзитивность.) *Для любых трех мультииндексов $I, J, K \in \mathbb{Z}_+^n$ справедливо*

$$(I \prec J) \wedge (J \prec K) \Rightarrow (I \prec K).$$

Доказательство. 1. Если $I \neq J$, то среди разностей $i_s - j_s$ (где $s = 1, ..., n$) найдутся ненулевые. Пусть первой из них будет $i_m - j_m$. Если эта разность отрицательна, то будет выполняться $I \prec J$; если положительна, то наоборот: $I \succ J$.

2. Неравенство $I \prec J$ означает, что среди разностей $i_s - j_s$ первая ненулевая $i_m - j_m$ является отрицательной; аналогично неравенство $J \prec K$ означает, что среди разностей $j_s - k_s$ первая ненулевая разность $j_l - k_l < 0$.

Если $l < m$, то первой ненулевой разностью среди $i_s - k_s$ будет

$$i_l - k_l = j_l - k_l < 0.$$

Если $l = m$, то снова первой ненулевой разностью среди $i_s - k_s$ будет $i_l - k_l$, причем

$$i_l - k_l = (i_l - j_l) + (j_l - k_l) = (i_m - j_m) + (i_l - k_l) < 0.$$

Если $l > m$, то первой ненулевой среди $i_s - k_s$ будет

$$i_m - k_m = i_m - j_m < 0.$$

Во всех трех случаях приходим к выводу: $I \prec K$. $\square$

Замечание 48.2. Доказанное выше предложение 48.1 позволяет утверждать, что отношение "раньше" является отношением *линейного порядка* на множестве $\mathbb{Z}_+^n$.

Этот порядок принято называть *лексикографическим*.

Отметим еще, что нулевой мультииндекс является *наименьшим* ("самым ранним") элементом в линейно упорядоченном множестве $(\mathbb{Z}_+^n, \preccurlyeq)$.

Пример 48.1. В множестве $\mathbb{Z}_+^4$ мультииндексов длины 4:

$$(7\,0\,0\,0) \succ (5\,9\,8\,1) \succ (5\,9\,6\,8) \succ (5\,9\,6\,1) \succ (3\,2\,2\,0) \succ (3\,2\,0\,4).$$

Определение 48.3. Сумма всех компонент мультииндекса (48.1) называется его *нормой*. Используется обозначение

$$|I| = i_1 + i_2 + \ldots + i_n. \tag{48.3}$$

Норма мультииндекса есть неотрицательное целое число (положительное, если мультииндекс ненулевой). Норма суммы мультииндексов равна сумме норм.

Заметьте, что более ранний мультииндекс может иметь бóльшую норму.

Среди мультииндексов выделяются наши "старые знакомые" — единичные векторы

$$E_1 = (1\,0\ldots0),\ E_2 = (0\,1\ldots0),\ \ldots,\ E_n = (0\,0\ldots1); \tag{48.4}$$

они имеют единичную норму и упорядочены следующим образом:

$$E_1 \succ E_2 \succ \ldots \succ E_n. \tag{48.5}$$

Замечание 48.3. Поясним происхождение термина лексикографический (= словарный) порядок. Этот термин уже встречался нам: в примере 16.1 мы лексикографически упорядочивали перестановки степени n. В качестве "букв" фигурировали номера от 1 до n; в качестве "слов" — нижние строки в двустрочной записи перестановок. Все "слова" имели длину n, и все "буквы" в них были попарно различны. "Буквы" естественным образом упорядочены (как числа). Мы считали, что перестановка с нижней строкой $(i_1\, i_2 \ldots i_n)$ идет раньше перестановки с нижней строкой $(j_1\, j_2 \ldots j_n)$, если для некоторой позиции s справедливо:

— первые $s-1$ соответствующих "букв" в двух "словах" одинаковы: $i_1 = j_1; \ldots; i_{s-1} = j_{s-1}$;

— "буква" с номером s из первого "слова" идет в "алфавите" раньше, нежели "буква" с номером s из второго "слова", т. е. выполняется неравенство $i_s < j_s$.

Именно так располагаются слова в обычных словарях. [Самым ранним "словом" будет $(1\,2 \ldots n)$, самым поздним — $(n\ n-1\ \ldots 1)$.]

Теперь у нас работает иная версия лексикографического порядка: хотя "слова" по-прежнему имеют фиксированную длину n, но "алфавит" уже бесконечен: он состоит из всех неотрицательных целых чисел и естественным образом упорядочен. Отношение "раньше" определяется точно так же, как и для перестановок, но, в силу бесконечности "алфавита", отсутствует "последняя буква" и среди "слов" нет самого позднего.

В следующем предложении будет установлена согласованность лексикографического порядка для мультииндексов с их сложением.

Предложение 48.2. *1. Для любых мультииндексов $I, J, K \in \mathbb{Z}_+^n$ справедливо:*

$$(\, I \prec J \,) \Rightarrow (\, I + K \prec J + K \,); \tag{48.6}$$

2. Для любых четырех мультииндексов $I, J, K, L \in \mathbb{Z}_+^n$ справедливо:

$$(\, I \prec J \,) \wedge (\, K \prec L \,) \Rightarrow (\, I + K \prec J + L \,). \tag{48.7}$$

Доказательство. 1. Если первые $s-1$ номеров в I совпадают с соответствующими номерами в J и $i_s < j_s$, то первые $s-1$ номеров

в $I+K$ будут совпадать с соответствующими номерами в $J+K$ и будет выполняться неравенство $i_s+k_s<j_s+k_s$. Этим доказывается первое утверждение.

2. Второе утверждение получается с помощью первого "в три приема": из двух импликаций $(I \prec J) \Rightarrow (I+K \prec J+K)$ и $(K \prec L) \Rightarrow (J+K \prec J+L)$ следует третья — (48.7). □

Замечание 48.4. Лексикографический порядок не является единственным разумным порядком на множестве мультииндексов. Во многих отношениях более предпочтительным является d-*лексикографический* порядок $\overset{\mathrm{d}}{\prec}$, с предварительной группировкой мультииндексов *по возрастанию нормы,* с последующим лексикографическим упорядочением мультииндексов внутри каждой группы.

Убедитесь в том, что d-лексикографический порядок, так же как и описанный ранее "чисто лексикографический", согласован со сложением.

Важнейшее свойство лексикографического порядка в множестве мультииндексов выражает следующее

Предложение 48.3. *Не существует бесконечной строго убывающей последовательности мультииндексов (фиксированной длины) вида*

$$I^{(1)} \succ I^{(2)} \succ I^{(3)} \succ \ldots \tag{48.8}$$

Доказательство. Обратите с самого начала внимание на систему нумерации. Номер мультииндекса, члена последовательности $\{I^{(k)}\}_{k=1}^{\infty}$, ставится сверху в скобках; номер компоненты мультииндекса ставится снизу.

Первые компоненты мультииндексов (48.8) образуют последовательность неотрицательных целых чисел $\{i_1^{(k)}\}_{k=1}^{\infty}$. Согласно определению лексикографического порядка, эта последовательность является *невозрастающей,* и следовательно, она неизбежно *стабилизируется*, т. е., начиная с некоторого номера k_1, все числа $i_1^{(k)}$ $(k \geqslant k_1)$ одинаковы.

Теперь надо смотреть на вторые компоненты: $\{i_2^{(k)}\}_{k=k_1}^{\infty}$. В силу убывания последовательности мультииндексов (48.8), эта последовательность неотрицательных целых чисел невозрастает и, как следствие, стабилизируется, начиная с некоторого номера $k_2 \geqslant k_1$; и т. д.

Начиная с некоторого номера k_n ($k_n \geqslant ... \geqslant k_2 \geqslant k_1$), стабилизируются все последовательности компонент $\{i_s^{(k)}\}_{k=k_n}^{\infty}$ ($s = 1, ..., n$).

Значит, стабилизируется и последовательность мультииндексов $\{I^{(k)}\}_{k=k_n}^{\infty}$, что противоречит ее строгому убыванию.

Тем самым доказано, что последовательностей вида (48.8) не существует. □

Замечание 48.5. Предложение 48.3 остается в силе для d-лексикографического порядка (см. замечание 48.4).

[Обратим ваше внимание на еще одно, очень интересное сочинение, посвященное компьютерной алгебре: Д. Кокс, Дж. Литтл, Д. О'Ши. "Идеалы, многообразия и алгоритмы" (М.: Мир, 2000). В этой книге, в частности, подробно обсуждаются различные способы упорядочения мультииндексов, похожие на лексикографический, но в некоторых отношениях более предпочтительные для организации вычислений.]

48.2. Многочлены от нескольких переменных. Лексикографическое упорядочение одночленов. Дадим теперь строгое определение множества $P[x_1, x_2, ..., x_n]$ многочленов от n переменных $x_1, x_2, ..., x_n$, над полем P. (Точнее, это будет целая серия определений, объединенных в одно.)

Определение 48.4. *Одночленом* от переменных из списка

$$x = [\, x_1, x_2, ..., x_n], \tag{48.9}$$

соответствующим мультииндексу

$$I = (\, i_1\, i_2 \ldots i_n) \in \mathbb{Z}_+^n$$

называется выражение, обозначаемое и определяемое следующим равенством:

$$ax^I = ax^{i_1} x^{i_2} \ldots x^{i_n}, \tag{48.10}$$

где коэффициент $a \in P$.

Два одночлена называются *подобными*, если они отвечают одному и тому же мультииндексу.

Мультииндекс I называется *мультистепенью* одночлена (48.10) с ненулевым коэффициентом a; *степенью* одночлена (48.10) называется норма $|I|$ мультииндекса I (иногда говорят о *тотальной* степени одночлена "по совокупности переменных").

(Можно определить также степень одночлена отдельно по каждой из участвующих в нем переменных. А если некоторая переменная в рассматриваемый одночлен не входит, то его степень по этой переменной считается равной нулю.)

Многочленом от переменных (48.9) называется *формальная сумма*

$$f(x_1, x_2, ..., x_n) = \sum_{I \in \mathbb{Z}_+^n} f_I x^I, \tag{48.11}$$

бесконечная, но *финитная*, т. е. такая, что лишь конечное число коэффициентов $f_I \in P$ отлично от нуля.

[Сумма (48.11) не содержит подобных членов. Можно использовать сокращенное обозначение $f(x)$, понимая под x список (48.9).]

Два многочлена, $f(x)$ и $g(x)$, от n переменных, называются *равными*, если они равны "покоэффициентно", т. е. если для любого мультииндекса $I \in \mathbb{Z}_+^n$ справедливо $f_I = g_I$.

Степенью (или *тотальной* степенью) многочлена (48.11) называется максимальная из степеней, входящих в него (ненулевых) одночленов. Обозначение:

$$\deg(\, f(x_1, x_2, ..., x_n)\,) = \max\{\, |I| \,:\, I \in \mathbb{Z}_+^n \wedge f_I \neq 0 \,\}. \tag{48.12}$$

Приступаем к описанию алгебраических действий в множестве многочленов от нескольких переменных. Описание это будет не совсем строгим, обзорным, вполне "школьным" по духу. (Ведь многочлены от многих переменных фактически вводятся в школьном курсе алгебры.)

Обоснование корректности приводимых ниже определений (и тем более проверку выполнения аксиом кольца) мы здесь приводить не будем. (Для случая многочленов от одной переменной более или менее строгое обоснование излагалось в § 36.)

Сложение двух многочленов вида (48.11) осуществляется по "векторному принципу": складываются все соответствующие коэффициенты. Умножение двух одночленов вида (48.10) производится по правилу:

$$(ax^I) \cdot (bx^J) = ab\, x^{I+J} = ab\, x_1^{i_1+j_1} x_2^{i_2+j_2} ... \, x_n^{i_n+j_n}. \tag{48.13}$$

Замечание 48.6. Обратите внимание на использование сложения мультииндексов — "мультистепеней" одночленов. "Настоящие" (*тотальные*) степени тоже складываются.

Кроме того, в формуле (48.13) заложена возможность перестановки неизвестных: запись x_2x_1 считается столь же допустимой, как и запись x_1x_2, но в окончательном виде рекомендуется размещать переменные в произведениях в установленном для них порядке.

Процедура *умножения многочленов* такова:

— перемножить две конечные суммы (раскрыть скобки; здесь "закладывается" будущий дистрибутивный закон);

— каждое из полученных произведений "упростить" по правилу умножения одночленов (48.13);

— привести подобные одночлены.

Замечание 48.7. Итак, вам предлагается "поверить", что с описанными выше алгебраическими действиями сложения и умножения множество $P[x_1, x_2, ..., x_n]$ является коммутативным кольцом.

Один из способов не поверить, а понять — убедиться в совпадении описанных действий с теми, которые возникают в кольце многочленов при его *индуктивном* построении [см. (46.28)].

При индуктивном построении тот факт, что получается *кольцо*, не требует отдельного доказательства. Более того, как мы уже знаем (см. пример 46.2), это кольцо оказывается *факториальным*. (Причем не только в случае поля коэффициентов, но и в случае, когда коэффициенты берутся из факториального кольца.)

По-видимому, очевидными являются *свойства степени* многочленов от нескольких переменных: 1) при сложении степень суммы не превосходит максимальной из степеней слагаемых; 2) при умножении степень произведения равняется сумме степеней сомножителей.

Для многочленов от одной переменной существуют две естественные записи: *по возрастанию степеней* и *по убыванию степеней*. Для многочленов от двух и более переменных может существовать много одночленов-слагаемых одинаковой степени и возникает проблема выбора "естественного порядка" записи членов в многочлене.

Эта проблема решается с помощью предварительного выбора порядка в множестве переменных.

Порядок в множестве переменных устанавливается либо их нумерацией по типу (48.9), либо, если переменные обозначаются все различными буквами, явным указанием порядка (скажем, алфавитного) в множестве используемых букв. (Второй способ ничем принципиально не отличается от первого: фактически, каждой из букв-переменных присваивается свой порядковый номер.)

Пусть переменные уже упорядочены по типу (48.9). Определим *иерархию* одночленов от этих переменных *по высоте*.

Определение 48.5. Рассмотрим два (ненулевых) одночлена ax^I и bx^J. Будем считать, что ax^I *ниже* bx^J (а bx^J *выше* ax^I), если мультииндекс I *раньше* мультииндекса J (J *позже* I).

Для отношений "ниже" ("выше") среди одночленов будем использовать те же символы, что и для отношений "раньше" ("позже") среди мультииндексов. Таким образом,

$$(\, ax^I \prec bx^J \,) \Leftrightarrow (\, I \prec J \,) \tag{48.14}$$

независимо от значений (ненулевых) коэффициентов a и b.

Замечание 48.8. Согласованность лексикографического порядка для мультииндексов с их сложением влечет (см. замечание 48.6) согласованность лексикографического порядка для одночленов с их умножением:

$$(\, ax^I \prec bx^J \,) \Rightarrow (\, ax^I \cdot cx^K \prec bx^J \cdot cx^K \,).$$

Все сказанное о лексикографическом порядке переносится на случай d-лексикографического порядка.

Сформулируем далее утверждение, относящееся к лексикографическому порядку для одночленов, вытекающее из аналогичного утверждения для мультииндексов (см. предложение 48.3).

Предложение 48.4. *Не существует бесконечной последовательности одночленов от переменных (48.9), такой, что каждый следующий одночлен ниже предыдущего.*

Доказательство немедленно следует из предложения 48.3 и определения 48.5. (Данное предложение также сохраняет силу для порядка $\overset{\mathrm{d}}{\prec}$.) $\square$

Замечание 48.9. После выбора лексикографического порядка среди одночленов $f_I x^I$ ($f_I \neq 0$), входящих в многочлен $f(x)$ (напомним: предполагается, что многочлен не содержит различных подобных членов; именно поэтому все его (ненулевые) члены оказываются линейно упорядоченными по высоте), снова, как и для многочленов от одной переменной, возникают две естественные возможности записи членов многочлена: либо в порядке возрастания "высоты", либо в противоположном порядке.

Второй способ (начиная с высшего члена) является более распространенным. Для высшего члена будем использовать обозначение h.t.(f) (от "higher term").

Вот пример записи многочленов в порядке убывания высоты членов:

$$f(x_1,x_2,x_3)=3x_1^5x_3-2x_1^4x_2^8x_3^9-7x_1x_2+x_2^5+x_2x_3;\ \text{h.t.}(f)=3x_1^5x_3.$$

При такой, *чисто лексикографической* записи степень одночленов не играет решающей роли: высший член может иметь меньшую степень, нежели следующие за ним.

Можно (считая, что степень "главнее" лексикографии) применить d-*лексикографический* порядок для мультииндексов и одночленов (см. замечания 48.5 и 48.8), при котором члены разбиваются на группы в порядке убывания *тотальных степеней*, а внутри каждой из групп наводится лексикографический порядок (по убыванию высоты).

В предыдущем примере d-лексикографическая запись будет иметь вид

$$f(x_1,x_2,x_3)=-2x_1^4x_2^8x_3^9+3x_1^5x_3+x_2^5-7x_1x_2+x_2x_3.$$

Пример 48.2. Упорядочение членов весьма существенно в теории (для организации индуктивных доказательств) и в практических вычислениях. Не случайно в системе Maple предусмотрена команда sort, допускающая реализацию самых различных методов "сортировки" членов в многочлене от нескольких переменных.

В том числе, по умолчанию, заложен d-лексикографический порядок (его имя в системе Maple — 'tdeg', от "total degree"); возможно также применение чисто лексикографического ('plex', от "pure lexicographic") порядка. (Сам Maple "добровольно" члены не сортирует, поскольку это может помешать замыслам пользователя.)

Применение команды sort предполагает явное упорядочение переменных (например, в виде списка).

```
> p := 3*x[1]^5*x[2]^9 + 4*x[1]^6*x[2] - 2*x[1]*x[2]* x[3]^5
   - 11*x[2]^4*x[3] + 5*x[1]^2*x[3] + 2*x[1]*x[2]*x[3]^{10}
  - 23*x[1] + 4*x[2] - 8*x[3] + x[3]^{11} - x[1]*x[2]^{67}
  + x[1]*x[3]^{10};
```

$$p := 3x_1^5x_2^9 + 4x_1^6x_2 - 2x_1x_2x_3^5 - 11x_2^4x_3 + 5x_1^2x_3 + 2x_1x_2x_3^{10} - 23x_1 + 4x_2 - 8x_3 + x_3^{11} - x_1x_2^{67} + x_1x_3^{10}$$

> sort(p, [x[1], x[2], x[3]], tdeg);

$$-x_1x_2^{67} + 3x_1^5x_2^9 + 2x_1x_2x_3^{10} + x_1x_3^{10} + x_3^{11} + 4x_1^6x_2 - 2x_1x_2x_3^5 - 11x_2^4x_3 + 5x_1^2x_3 - 23x_1 + 4x_2 - 8x_3$$

> sort(p, [x[1], x[2], x[3]], plex);

$$4x_1^6x_2 + 3x_1^5x_2^9 + 5x_1^2x_3 - x_1x_2^{67} + 2x_1x_2x_3^{10} - 2x_1x_2x_3^5 + x_1x_3^{10} - 23x_1 - 11x_2^4x_3 + 4x_2 + x_3^{11} - 8x_3$$

48.3. Лемма о высшем члене произведения многочленов. В заголовок подпункта вынесено традиционное название следующего утверждения.

Предложение 48.5. *Высшим членом произведения многочленов является произведение высших членов сомножителей.*

Доказательство. Пусть $f_{I_0}x^{I_0}$ — высший член многочлена $f(x)$, а $g_{J_0}x^{J_0}$ — высший член $g(x)$.

Это означает, что для любых (ненулевых) членов f_Ix^I и g_Jx^J справедливо: $I_0 \succcurlyeq I$ и $J_0 \succcurlyeq J$. Если хотя бы один из этих членов отличен от высшего (в своем многочлене), то хотя бы одно из ("искривленных") неравенств для мультииндексов является строгим. По предложению 48.2, отсюда следует (строгое) неравенство $I_0 + J_0 \succ I + J$.

А значит, справедливо (строгое) неравенство для одночленов

$$f_{I_0}g_{J_0}x^{I_0+J_0} \succ f_Ig_Jx^{I+J},$$

что собственно и утверждалось.

(Правда, среди одночленов $f_Ig_Jx^{I+J}$ могут быть подобные, но при приведении подобных "мультистепени" $I + J$ сохраняются, либо же

соответствующий член совсем пропадает (если коэффициенты взаимно гасятся). Так что членов более высоких, чем $f_{I_0}g_{J_0}x^{I_0+J_0}$, не появится.) $\square$

Замечание 48.10. Справедлива также "d-версия" доказанного выше предложения. (Отдельного доказательства не потребуется, если вы разобрались с предыдущими замечаниями на тему порядка $\overset{\mathrm{d}}{\prec}$.)

48.4. Однородные многочлены (формы)

Определение 48.6. Многочлен от n переменных вида (48.11) называется *однородным* многочленом (или *формой*) степени m, если все его (ненулевые) члены имеют одинаковую степень, равную m.

Всякий многочлен, очевидно, представляется в виде суммы форм. Для форм лексикографический порядок и d-лексикографический порядок их членов не отличаются.

Формы нулевой степени — это просто ненулевые константы.

Общим видом формы первой степени (или *линейной формы*) является

$$f(x) = a_1x_1 + a_2x_2 + ... + a_nx_n, \tag{48.15}$$

где $a_i = f_{E_i}$; $E_i = (0 \dots 1 \dots 0)$ (единица на i-м месте).

Форма второй степени (*квадратичная форма*) представляется в виде

$$\begin{aligned} f(x) = a_{11}x_1^2 + a_{22}x_2^2 + ... + a_{nn}x_n^2 + \\ + a_{12}x_1x_2 + a_{12}x_1x_2 + ... + a_{1n}x_1x_n + \\ + a_{23}x_2x_3 + ... + a_{(n-1)n}x_{n-1}x_n, \end{aligned} \tag{48.16}$$

где $a_{ii} = f_{(0 \dots 2 \dots 0)}$ (двойка стоит на i-м месте); $a_{ij} = f_{(0 \dots 1 \dots 1 \dots 0)}$ ($i < j$; единицы стоят на местах с номерами i и j).

Замечание 48.11. Отметим следующие простые свойства однородных многочленов, связанные с алгебраическими действиями над ними.

1. Сумма нескольких форм одинаковой степени (если она не нулевая) сама является формой той же степени.

2. Произведение нескольких форм является формой (суммарной степени).

3. В частности, при возведении формы степени m в натуральную степень k получается форма степени mk.

48.5. Композиция многочленов (подстановка многочленов в многочлен). Рассмотрим многочлен $g(y_1, y_2, ..., y_r)$ вида

$$g(y) = \sum_{J \in \mathbb{Z}_+^r} g_J y^J,$$

степени m, от переменных $y = [y_1, y_2, ..., y_r]$, и еще r многочленов

$$h_k(x_1, x_2, ..., x_n) \ [k = 1, ...r]$$

от n переменных (48.9). Степени этих многочленов обозначим d_k.

Подставим вместо каждой из исходных переменных y_k в каждый из членов многочлена $g(y)$ выражения

$$y_k = h_k(x_1, x_2, ..., x_n). \tag{48.17}$$

Получим (для одночлена $g_J y^J$):

$$g_J y_1^{j_1} y_2^{j_2} ... y_r^{j_r} = g_J (h_1(x))^{j_1} (h_2(x))^{j_2} ... (h_r(x))^{j_r}, \tag{48.18}$$

являющееся многочленом от x степени $d_1 j_1 + ... + d_r j_r$.

Для многочлена $g(y)$ в целом получим после подстановки многочлен

$$f(x_1, ..., x_n) = g(h_1(x_1, ..., x_n), ..., h_r(x_1, ..., x_n)), \tag{48.19}$$

называемый *композицией* многочлена $g(y)$ и многочленов $h_k(x)$.

Степень многочлена $f(x)$ будет равняться наибольшему из чисел $d_1 j_1 + ... + d_r j_r$, при том что $J = (j_1 ... j_r)$ пробегает множество всех таких мультииндексов, которым отвечают ненулевые члены в $g(y)$.

§ 49. Симметрические многочлены

49.1. Определение симметрического многочлена

Определение 49.1. Многочлен

$$f(x_1, x_2, ..., x_n) \in P[x_1, x_2, ..., x_n]$$

называется симметрическим, если для любой перестановки

$$\sigma = \begin{pmatrix} 1 & 2 & ... & n \\ \sigma(1) & \sigma(2) & ... & \sigma(n) \end{pmatrix} \in S_n$$

справедливо равенство

$$f(x_{\sigma(1)}, x_{\sigma(2)}, ..., x_{\sigma(n)}) = f(x_1, x_2, ..., x_n). \tag{49.1}$$

Множество всех симметрических многочленов будем обозначать $P_{sym}[x]$ (где $x = [x_1, x_2, ..., x_n]$).

Замечание 49.1. Проанализируем данное выше определение. Равенство (49.1) имеет совершенно ясный смысл, если многочлен (от нескольких переменных) понимать как *функцию* (от нескольких переменных). (Каждая из переменных пробегает поле P, и значения функции принадлежат тому же полю.)

Но многочлен у нас — это не функция, а формальная сумма (см. определение 48.4). Левую часть (49.1) следует понимать как многочлен ${}^{\sigma}f(x)$, являющийся формальной суммой следующих (ненулевых) членов:

$$f_I \, x_{\sigma(1)}^{i_1} x_{\sigma(2)}^{i_2} ... x_{\sigma(n)}^{i_n}, \tag{49.2}$$

где I – мультииндекс такой, что $f_I \neq 0$. [В одночлене (49.2) переменные стоят не по порядку, но, как мы условились в замечании 48.6, переменные считаются коммутирующими символами и их можно как угодно переставлять.]

Тот факт, что $f(x)$ совпадает со всеми ${}^{\sigma}f(x)$, означает, что симметрический многочлен, вместе с каждым членом $f_I x^I$, содержит все члены вида (49.2).

Например, если симметрический многочлен содержит (ненулевой) член

$$a x_1^4 x_2^3 x_3 \ (a \in P),$$

то он обязан содержать еще пять членов (с одним и тем же коэффициентом):

$$a \cdot (x_1^4 x_3^3 x_2 + x_2^4 x_1^3 x_3 + x_2^4 x_3^3 x_1 + x_3^4 x_1^3 x_2 + x_3^4 x_2^3 x_1).$$

После перестановки сомножителей-переменных сумма всех шести членов будет выглядеть следующим образом:

$$a \cdot (x_1^4 x_2^3 x_3 + x_1^4 x_2 x_3^3 + x_1^3 x_2^4 x_3 + x_1 x_2^4 x_3^3 + x_1^3 x_2 x_3^4 + x_1 x_2^3 x_3^4).$$

Одночлены в скобке можно расположить лексикографически:

$$a \cdot (x_1^4 x_2^3 x_3 + x_1^4 x_2 x_3^3 + x_1^3 x_2^4 x_3 + x_1^3 x_2 x_3^4 + x_1 x_2^4 x_3^3 + x_1 x_2^3 x_3^4).$$

Обратите внимание на то обстоятельство, что мультистепень $I = (4\,3\,1)$ высшего члена является *монотонным* (неубывающим) мультииндексом. (Как мы убедимся ниже, в предложении 49.1, это совершенно не случайно.)

Замечание 49.2. К числу симметрических многочленов можно отнести все скалярные многочлены (в их числе и нулевой).

Если многочлены $f(x), g(x) \in P_{sym}[x]$, то их сумма и произведение также являются симметрическими многочленами. (Попробуйте осознать это в свете сказанного в предыдущем замечании о "строении" симметрических многочленов. Осознав, проведите доказательство последнего утверждения.)

Иначе это утверждение выражается следующим образом: множество симметрических многочленов $P_{sym}[x]$ является *подкольцом* в кольце $P[x]$ всех многочленов от n переменных.

Рассуждение можно продолжить и заметить сначала, что произвольная неотрицательная степень симметрического многочлена является симметрическим многочленом, а затем прийти к выводу о том, что, подставив в *произвольный* многочлен $g(y_1, ..., y_r)$ вместо переменных y_k *симметрические* многочлены $h_k(x_1, ..., x_n)$ (см. п. 48.6), мы получим *симметрический* многочлен-композицию (48.19).

Пример 49.1. Очевидным образом симметричны (и однородны) следующие *степенные суммы*:

$$\pi_k(x) = x_1^k + x_2^k + ... + x_n^k, \tag{49.3}$$

где $k = 1, 2, 3, ...$

[Многочлены (49.3) можно рассматривать над любым полем.]

Пример 49.2. Рассмотрим так называемые *элементарные симметрические (э.с.)* многочлены

$$\begin{cases} \sigma_1(x) = x_1 + x_2 + ... + x_n; \\ \sigma_2(x) = x_1x_2 + x_1x_3 + ... + x_1x_n + x_2x_3 + ... + x_{n-1}x_n; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \sigma_{n-1}(x) = x_1x_2...x_{n-1} + x_1x_2...x_{n-2}x_n + ... + x_2x_3...x_n; \\ \sigma_n(x) = x_1x_2...x_n. \end{cases} \tag{49.4}$$

Многочлен $\sigma_k(x)$ $[k = 1, ..., n]$ является суммой всех C_n^k возможных произведений по k (из общего числа n) переменных. Симметричность и однородность этих многочленов очевидны, как и то, что в формулах (49.4) они представлены в лексикографическом порядке.

[Многочлены (49.4) уже встречались нам в § 40, в *формулах Виета*; они также определены над произвольным полем.]

Пример 49.3. Еще один пример, в котором многочлен представлен не в стандартной записи, а разложенным на множители. (Именно в таком виде очевидна его симметричность.)

$$D_n(x) = (x_2 - x_1)^2(x_3 - x_1)^2(x_3 - x_2)^2 ... \cdot \\ \cdot (x_n - x_1)^2(x_n - x_2)^2 ... (x_n - x_{n-1})^2. \quad (49.5)$$

Этот многочлен есть не что иное, как квадрат определителя Вандермонда (см. п. 30а.5):

$$D_n(x_1, x_2, ..., x_n) = (\Delta_n(x_1, x_2, ..., x_n))^2. \quad (49.5а)$$

(Без возведения в квадрат, многочлен $\Delta_n(x)$ не является симметрическим. Он является *антисимметрическим.* Такие многочлены также представляют интерес и изучаются.)

Многочлен $D_n(x)$ определен над произвольным полем и является однородным, степени $2C_n^2$.

49.2. Лемма о высшем члене симметрического многочлена. Как и в названии п. 48.3, в заголовке данного пункта — традиционное наименование следующего утверждения.

Предложение 49.1. *Пусть* $f(x_1, x_2, ..., x_n)$ — *симметрический многочлен и*

$$f_I x^I = f_I x_1^{i_1} x_2^{i_2} ... x_n^{i_n} = \text{h.t.}(f(x)) \quad (49.6)$$

— *его высший член.*

Тогда мультииндекс I *является монотонным (неубывающим), т. е.*

$$i_1 \geqslant i_2 \geqslant ... \geqslant i_n. \quad (49.7)$$

Доказательство. Предположим противное. Тогда найдутся два соседних номера i_k, i_{k+1} $(k \leqslant n - 1)$ в мультииндексе I, такие, что $i_k < i_{k+1}$.

В силу замечания 49.1, наряду с (высшим) членом

$$f_I x_1^{i_1} x_2^{i_2} \dots x_k^{i_k} x_{k+1}^{i_{k+1}} \dots x_n^{i_n}, \tag{49.6'}$$

симметрический многочлен $f(x)$ обязан содержать член

$$f_I x_1^{i_1} x_2^{i_2} \dots x_{k+1}^{i_k} x_k^{i_{k+1}} \dots x_n^{i_n} = f_I x_1^{i_1} x_2^{i_2} \dots x_k^{i_{k+1}} x_{k+1}^{i_k} \dots x_n^{i_n}, \tag{49.8}$$

который выше (49.6′). Противоречие. □

49.3. Моногенные симметрические многочлены

Определение 49.2. *Моногенным* называется такой многочлен от n переменных, все члены которого получаются из одного из них ("порождающего") с помощью всевозможных перестановок переменных (причем, если различные перестановки приводят к одинаковым членам, то берется лишь один из них).

Моногенный многочлен, порожденный одночленом ax^I, обозначается $\mathcal{S}(ax^I)$.

Замечание 49.3. Ясно, что моногенный многочлен является симметрическим и однородным и что в качестве "порождающего" можно взять любой из его членов.

Операция $ax^I \mapsto \mathcal{S}(ax^I)$ может быть охарактеризована как "симметрическое размножение" одночлена, приводящее к моногенному многочлену.

Замечание 49.4. Согласно изложенному выше, в п. 48.4, всякий многочлен $f(x)$ степени m от n переменных разбивается в сумму однородных многочленов (форм), степень которых изменяется от 0 до m:

$$f(x) = f_{(0)} + f_{(1)}(x) + f_{(2)}(x) + \dots + f_{(m)}(x). \tag{49.9}$$

Очевидно, что если данный многочлен являлся симметрическим, то симметрическими будут и все получающиеся формы $f_{(k)}(x)$ (где $k = 1, \dots, m$). Некоторые из этих форм могут оказаться нулевыми.

Далее, в замечании 49.1, объяснено, что все (ненулевые) члены симметрического многочлена разбиваются на группы, составляющие *моногенные* многочлены.

Таким образом, каждая из (ненулевых) форм-слагаемых $f_{(k)}(x)$ разбивается на "еще более мелкие" (моногенные) симметрические формы, причем каждая из последних вполне определяется одним из своих членов. Естественно в качестве "порождающего" члена брать высший из них.

Пример 49.4. Многочлены, рассмотренные выше, в примерах 49.1, 49.2, являются моногенными:

$$\pi_k(x) = \mathcal{S}(x_1^k);\ \sigma_k(x) = \mathcal{S}(x_1 x_2 ... x_k). \tag{49.10}$$

Члены, указанные в формулах (49.10) в качестве порождающих, являются высшими в своих многочленах.

Заметим, что существует единственный (с точностью до пропорциональности) моногенный симметрический многочлен первой степени: $\pi_1(x) = \sigma_1(x)$.

Пример 49.5. Рассмотрим одночлен $x_1^2 x_2$. Если его считать многочленом от двух переменных, то он порождает моногенный многочлен $\mathcal{S}(x_1^2 x_2) = x_1^2 x_2 + x_1 x_2^2$.

Но этот одночлен можно рассматривать и как одночлен от трех переменных [соответствующий мультииндексу $(2\,1\,0)$], и тогда он породит моногенный многочлен

$$\mathcal{S}(x_1^2 x_2) = x_1^2 x_2 + x_1^2 x_3 + x_1 x_2^2 + x_1 x_3^2 + x_2^2 x_3 + x_2 x_3^2.$$

Еще один характерный случай, когда одночлен от n переменных порождает моногенный многочлен, содержащий менее чем $n!$ слагаемых (поскольку повторения исключаются):

$$\mathcal{S}(x_1^2 x_2^2 x_3) = x_1^2 x_2^2 x_3 + x_1^2 x_2 x_3^2 + x_1 x_2^2 x_3^2.$$

49.4. Основная теорема о симметрических многочленах. В силу замечания 49.2, если подставить в произвольный многочлен от n переменных $g(y_1, y_2, ..., y_n)$, вместо каждой из переменных y_k, э.с. многочлены $\sigma_k(x_1, x_2, ..., x_n)$ (от такого же количества переменных x), то получится симметрический многочлен

$$\begin{aligned} &f(x_1, x_2, ..., x_n) = \\ &= g(\sigma_1(x_1, x_2, ..., x_n), \sigma_2(x_1, x_2, ..., x_n), ..., \sigma_n(x_1, x_2, ..., x_n)). \end{aligned} \tag{49.11}$$

Настоящий пункт будет посвящен доказательству важной теоремы (заслужившей название *основной теоремы о симметрических многочленах*), которая будет утверждать, что и обратно: всякий симметрический многочлен $f(x) \in P_{sym}[x]$ можно представить в виде (49.11), т. е. "полиномиально выразить" через э.с. многочлены.

Теорема 49.1. *Всякий симметрический многочлен можно, причем единственным образом, представить в виде многочлена от э.с. многочленов, т. е. для любого симметрического многочлена*

$$f(x) \in P_{sym}[x_1, ..., x_n]$$

найдется (однозначно определенный) многочлен

$$g(y) \in P[y_1, ..., y_n],$$

такой, что выполняется (49.11).

Доказательство. 1. Докажем сначала *существование* представления (49.11).

Поскольку всякий многочлен представляется в виде суммы однородных, то достаточно доказать сформулированное утверждение для однородных многочленов положительной степени (для скалярных многочленов оно тривиально).

Пусть $f(x)$ — однородный симметрический многочлен степени m с высшим членом

$$\text{h.t.}(f) = a x_1^{i_1} x_2^{i_2} ... x_n^{i_n}. \tag{49.12}$$

Согласно предложению 49.1, мультииндекс $I = (i_1\, i_2\, ...\, i_n)$ является монотонным.

Первый шаг построения многочлена $g(y)$ будет состоять в подборе такого *одночлена*

$$g_{(1)}(y_1, y_2, ..., y_n) = A\, y_1^{j_1} y_2^{j_2} ... y_n^{j_n};\ A \in P, \tag{49.13}$$

после подстановки в который $y_k = \sigma_k(x)$ $(k = 1, ..., n)$ получится многочлен

$$p_{(1)}(x) = g_{(1)}(\sigma_1(x), \sigma_2(x), ..., \sigma_n(x)) = \\ = A\,(\,\sigma_1(x)\,)^{j_1}(\,\sigma_2(x)\,)^{j_2} ... (\,\sigma_n(x)\,)^{j_n}, \tag{49.14}$$

симметрический (в силу замечания 49.2) и такой, что его высший член совпадает с (49.12).

[Заметим дополнительно, что многочлен (49.14) является однородным степени $j_1 + 2j_2 + ... + nj_n$.]

Согласно предложению 48.5, высшим членом (49.14) будет

$$A\, x_1^{j_1} (x_1 x_2)^{j_2} ... (x_1 x_2 ... x_{n-1})^{j_{n-1}} (x_1 x_2 ... x_n)^{j_n} = \\ = A\, x_1^{j_1 + j_2 + ... + j_{n-1} + j_n} x_2^{j_2 + ... + j_{n-1} + j_n} ... x_{n-1}^{j_{n-1} + j_n} x_n^{j_n}, \tag{49.15}$$

и условие совпадения (49.15) с (49.12) приведет к совпадению коэффициентов $A = a$ и к системе линейных уравнений

$$\begin{cases} j_1 + j_2 + \ldots + j_{n-1} + j_n = i_1 \ ; \\ \quad\ j_2 + \ldots + j_{n-1} + j_n = i_2 \ ; \\ \qquad \ddots \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \qquad\qquad j_{n-1} + j_n = i_{n-1} \ ; \\ \qquad\qquad\quad\ j_n = i_n \ . \end{cases} \tag{49.16}$$

Система (49.16) легко решается:

$$j_1 = i_1 - i_2;\ j_2 = i_2 - i_3;\ \ldots\ ;\ j_{n-1} = i_{n-1} - i_n;\ j_n = i_n; \tag{49.17}$$

в силу монотонности I, все найденные целые числа j_k $(k = 1, ..., n)$ неотрицательны и, следовательно, образуют мультииндекс $J \in \mathbb{Z}_+^n$.

Таким образом, получено явное выражение для симметрического однородного многочлена (степени m):

$$p_{(1)}(x) = a\,(\sigma_1(x))^{i_1-i_2}(\sigma_2(x))^{i_2-i_3}\ldots\cdot \\ \cdot\,(\sigma_{n-1}(x))^{i_{n-1}-i_n}(\sigma_n(x))^{i_n}. \tag{49.18}$$

Если $f(x) = p_{(1)}(x)$, то цель достигнута на первом же шаге: данный симметрический многочлен $f(x)$ полиномиально выражен через э.с. многочлены.

Если же это не так, то рассмотрим разность

$$f_{(1)}(x) = f(x) - p_{(1)}(x). \tag{49.19}$$

Многочлен (49.19) снова является симметрическим и однородным (степени m), причем все его члены (как оставшиеся от $f(x)$, так и пришедшие от $p_{(1)}(x)$, либо полученные в результате вычитания и приведения подобных) будут ниже члена (49.12) (напомним, что высший член в вычитаемом был таким же).

Поэтому

$$\text{h.t.}(f) \succ \text{h.t.}(f_{(1)}). \tag{49.20}$$

Повторяем описанную выше процедуру устранения высшего члена применительно к многочлену $f_{(1)}(x)$ с высшим членом, более низким, чем (49.12).

Возникает последовательность многочленов

$$f(x);\ f_{(1)}(x) = f(x) - p_{(1)}(x);\ \dots;$$
$$f_{(k+1)}(x) = f_{(k)}(x) - p_{(k)}(x);\ \dots, \quad (49.21)$$

со строгим убыванием по высоте высших членов:

$$\text{h.t.}(f) \succ \text{h.t.}(f_{(1)}) \succ \dots \succ \text{h.t.}(f_{(k)}) \succ \dots \quad (49.22)$$

В силу предложения 48.4, не существует бесконечной последовательности строго убывающих по высоте одночленов. Поэтому последовательность многочленов (49.21) должна оборваться на некотором конечном шаге, скажем, с номером r.

Это может произойти, только если разность $f_{(r+1)}(x)$ окажется нулевой, и тогда для симметрического многочлена $f(x)$ будет получено представление

$$f(x) = p_{(1)}(x) + p_{(2)}(x) + \dots + p_{(r)}(x), \quad (49.23)$$

где каждое из слагаемых в правой части (а значит — и вся правая часть в целом) является многочленом от э.с. многочленов.

Существование представления (49.11) доказано.

Замечание 49.5. Прервем доказательство теоремы с тем, чтобы изложить важные соображения, касающиеся практической реализации описанного выше алгоритма.

Во-первых, заметим, что ссылка на предложение 48.4 была отнюдь не обязательной. Гарантировать обрыв на конечном шаге последовательности одночленов, строго убывающих по высоте, в данном случае можно было из более простых соображений: все эти одночлены имеют одинаковую *тотальную степень*, а существует лишь конечное число мультииндексов с фиксированной нормой.

Кроме того, по ходу работы алгоритма на каждом шаге фигурирует одночлен, который является высшим членом для соответствующего многочлена $f_{(k)}(x)$, и следовательно, его мультистепень является *монотонным* мультииндексом.

Таким образом, в разложении (49.23) могут фигурировать лишь одночлены (от э.с. многочленов), мультистепени J' которых получаются по формулам (49.17) из монотонных мультииндексов I', более ранних, чем мультистепень I высшего члена (49.12).

В простых задачах все такие (монотонные) одночлены

$$I = I_1 \succ I_2 \succ I_2 \succ ... \succ I_r \tag{49.24}$$

легко перечислить.

Соответствующие, уже не монотонные, мультииндексы [мультистепени одночленов $g_{(k)}(y)$]

$$J = J_1,\ J_2,\ ...,\ J_r \tag{49.25}$$

находятся по формулам типа (49.17).

Однако коэффициенты

$$A = A_1,\ A_2,\ ...\ ,\ A_r \tag{49.26}$$

одночленов $g_{(k)}(y)$ заранее не известны [кроме первого, $A = a$, который равен коэффициенту высшего члена (49.12)]; поэтому они должны рассматриваться как *неопределенные*.

Определяются неопределенные коэффициенты с помощью *пробных значений* для переменных x (подробности см. в примерах следующего пункта).

2. Возвращаемся к доказательству теоремы. Нам надо установить единственность полиномиального выражения для симметрического многочлена через э.с. многочлены.

Для этого достаточно доказать, что при подстановке в некоторый ненулевой многочлен $g(y_1, ..., y_n)$ вместо переменных y_k соответствующих э.с. многочленов $\sigma_k(x)$ получается ненулевой многочлен $f(x)$ [см. (49.11)].

(В самом деле, если тогда два многочлена от n переменных, $g(y)$ и $g^\circ(y)$, при подстановке в них э.с. многочленов будут давать одинаковые результаты, то разность этих многочленов даст нулевой результат и, следовательно, сама будет нулевой.)

Рассмотрим все ненулевые одночлены $g_J y^J$ $[J = (j_1 \dots j_n)]$ многочлена $g(y)$. Каждому из них, при подстановке $y_k = \sigma_k(x)$, соответствует многочлен от x с высшим членом (49.15). Легко видеть, что мультииндексу $J' = (j'_1 \dots j'_n)$, отличному от J, будет соответствовать высший член такого же вида (49.15), но отличный от того высшего члена, который соответствовал J.

Выберем из всех таких (попарно различных) высших членов самый высший. Он, очевидно, будет высшим членом во всем многочлене (49.11) и, следовательно, не погасится ни с каким из других членов в процессе приведения подобных.

Этим гарантируется тот факт, что многочлен $f(x)$ является ненулевым. Теперь теорема доказана полностью. □

Замечание 49.6. 1. Внимательный анализ проведенного выше доказательства позволяет получить важную дополнительную информацию: коэффициенты искомого многочлена $g(y)$ находятся как целочисленные линейные комбинации коэффициентов исходного симметрического многочлена $f(x)$.

2. В замечательном "антикварном" учебнике А. К. Сушкевича "Основы высшей алгебры" (М.; Л.: ОНТИ НКТП СССР, 1937) приведено целых три принципиально различных доказательства основной теоремы о симметрических многочленах. Доказательство, изложенное выше, известно как доказательство Гаусса.

49.5. Примеры выражения симметрических многочленов через элементарные симметрические. В этом пункте мы реализуем (на нескольких примерах) программу практического выражения симметрического многочлена через э.с. многочлены (49.4).

Пример 49.6. Рассмотрим однородный многочлен шестой степени от трех переменных, являющийся частным случаем многочлена (49.5) из примера 49.3:

$$f(x) = D_3(x) = (x_2 - x_1)^2(x_3 - x_1)^2(x_3 - x_2)^2.$$

По предложению 48.5, высшим членом этого многочлена является одночлен $x_1^4x_2^2$, имеющий мультистепень $I = I_1 = (4\,2\,0)$ и коэффициент $a = 1$.

Последовательность (49.24) монотонных мультииндексов выглядит, очевидно, следующим образом:

$$I_1 = (4\,2\,0) \succ I_2 = (4\,1\,1) \succ I_3 = (3\,3\,0) \succ I_4 = (3\,2\,1) \succ I_5 = (2\,2\,2).$$

Определяем по формулам вида (49.17) последовательность мультистепеней (49.25):

$$J_1 = (2\,2\,0);\ J_2 = (3\,0\,1);\ J_3 = (0\,3\,0);\ J_4 = (1\,1\,1);\ J_5 = (0\,0\,2)$$

и тем самым — вид многочлена $g(y)$:

$$g(y_1, y_2, y_3) = A_1y_1^2y_2^2 + A_2y_1^3y_3 + A_3y_2^3 + A_4y_1y_2y_3 + A_5y_3^2,$$

где $A_1 = a = 1$, а остальные коэффициенты A_2, A_3, A_4, A_5 являются неопределенными и подлежат определению из равенства многочленов

$$(x_2 - x_1)^2(x_3 - x_1)^2(x_3 - x_2)^2 = \\ = (\sigma_1(x))^2(\sigma_2(x))^2 + A_2(\sigma_1(x))^3(\sigma_3(x)) + \\ + A_3(\sigma_2(x))^3 + A_4\sigma_1(x)\sigma_2(x)\sigma_3(x) + A_5(\sigma_3(x))^2,$$

где

$$\sigma_1(x) = x_1 + x_2 + x_3;\ \sigma_2(x) = x_1x_2 + x_1x_3 + x_2x_3;\ \sigma_1(x) = x_1x_2x_3.$$

Равенство многочленов влечет равенство соответствующих полиномиальных функций. (Для случая многочленов от одной переменной это обсуждалось в п. 39.1. Для многочленов от нескольких переменных все рассуждения указанного пункта можно повторить практически дословно. Разумеется, полиномиальные функции также будут зависеть от нескольких переменных, пробегающих данное поле.)

В рассматриваемом примере три переменные x_1, x_2, x_3, независимо друг от друга, пробегают поле рациональных чисел $\mathbb{Q}$ (или любое более широкое поле).

Придавая переменным конкретные числовые значения, можно попытаться из получившихся равенств определить неопределенные коэффициенты.

Выбор "пробных значений"

$$x_1 = x_1^0;\ x_2 = x_2^0;\ x_3 = x_3^0$$

не всегда прост: надо стремиться к тому, чтобы несложно вычислялись значения $f(x_1^0, x_2^0, x_3^0)$ в левой части равенства и чтобы при этих значениях некоторые (но не все!) э.с. многочлены обращались в нуль (что упрощает правую часть равенства).

В рассматриваемом примере многочлен обращается в нуль тогда и только тогда, когда существует хотя бы одна пара переменных, принимающих равные значения (это верно для всех многочленов $D_n(x)$; см. пример 49.3).

Поэтому удобно выбирать пробные значения переменных так, чтобы два из них были одинаковы. Именно этому принципу мы следуем при заполнении табл. 49.1 (прил. 2). Заодно три из четырех проб произведены так, чтобы один из э.с. многочленов обратился в нуль.

Возможно, вас смущает необходимость "расчетливого" выбора пробных значений переменных. На самом деле это не необходимо, а лишь "полезно". Пробные значения можно, в принципе, выбирать почти произвольно. Надо только помнить о симметричности рассматриваемых многочленов. (Выбрав, скажем, значения переменных $x_1 = 1$, $x_2 = 1$, $x_3 = 0$, не следует ожидать ничего нового от значений $x_1 = 1$, $x_2 = 0$, $x_3 = 1$.)

И следует смириться с тем, что при "случайном" выборе пробных точек система для определения неопределенных коэффициентов может оказаться "сложной", что, конечно, не очень страшно. (Мало ли мы с вами с. л. у. перерешали? К тому же у нас всегда под рукой помощник — Maple.)

Решая образовавшуюся в среднем столбце табл. 49.1 систему уравнений, получим: $A_2 = -4$; $A_3 = -4$; $A_4 = 18$; $A_5 = -27$.

О т в е т:

$$D_3(x) = \sigma_1^2\sigma_2^2 - 4\sigma_1^3\sigma_3 - 4\sigma_2^3 + 18\sigma_1\sigma_2\sigma_3 - 27\sigma_3^2.$$

Пример 49.7. Рассмотрим степенную сумму $\pi_2(x)$ (см. пример 49.1). Чтобы выразить этот многочлен через э.с. многочлены, достаточно знать формулу для квадрата суммы (нескольких слагаемых):

$$\begin{aligned}&(x_1 + x_2 + ... + x_n)^2 = \\ &= (x_1^2 + x_2^2 + ... + x_n^2) + 2(x_1x_2 + ... + x_1x_n + x_2x_3 + ... + x_{n-1}x_n),\end{aligned}$$

из которой немедленно следует:

$$\pi_2 = \sigma_1^2 - 2\sigma_2. \tag{49.27}$$

Пример 49.8. Рассмотрим аналогичную задачу для суммы кубов $\pi_3(x)$.

Высшим членом этого многочлена является одночлен x_1^3 с мультистепенью $I = (3\,0\,...\,0)$.

Цепочка монотонных мультииндексов, более ранних, чем I, в случае $n \geqslant 3$ будет иметь вид

$$I_1 = I = (3\,0\,0\,0\,...\,0) \succ I_2 = (2\,1\,0\,0\,...\,0) \succ I_3 = (1\,1\,1\,0\,...\,0).$$

Соответствующие мультииндексы J_k $(k = 1, 2, 3)$:

$$J_1 = (3\,0\,0\,0\,...\,0);\ J_2 = (1\,1\,0\,0\,...\,0);\ J_3 = (0\,0\,1\,0\,...\,0).$$

Представление для $\pi_3(x)$ с неопределенными коэффициентами:

$$\pi_3(x) = (\sigma_1(x))^3 + A_2\sigma_1(x)\sigma_2(x) + A_3\sigma_3(x).$$

Пробная точка $x_1 = x_2 = 1$; $x_3 = ... = x_n = 0$ дает уравнение, из которого находится: $A_2 = -3$.

Точка $x_1 = x_2 = x_3 = 1$; $x_4 = ... = x_n = 0$ позволяет определить: $A_3 = 3$.

О т в е т:

$$\pi_3 = \sigma_1^3 - 3\sigma_1\sigma_2 + 3\sigma_3. \tag{49.28}$$

Заметим, что эту задачу, как и предыдущую, можно было решить элементарными средствами. (Сделайте это. Заодно разберитесь с тем, что получится в случае $n = 2$.)

Замечание 49.7. Можно вывести (см., например, задачники [4,10]) выражения для всех степенных сумм π_k через σ_k, а также — обратные формулы. Здесь мы приведем ("для красоты") одну из этих формул:

$$\sigma_k = \frac{1}{k!}\begin{vmatrix} \pi_1 & 1 & 0 & ... & 0 & 0 & 0 \\ \pi_2 & \pi_1 & 2 & ... & 0 & 0 & 0 \\ \pi_3 & \pi_2 & \pi_1 & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ \pi_{k-2} & \pi_{k-3} & \pi_{k-4} & ... & \pi_1 & k-2 & 0 \\ \pi_{k-1} & \pi_{k-2} & \pi_{k-3} & ... & \pi_2 & \pi_1 & k-1 \\ \pi_k & \pi_{k-1} & \pi_{k-2} & ... & \pi_3 & \pi_2 & \pi_1 \end{vmatrix}. \tag{49.29}$$

49.6. Значения симметрических многочленов от корней многочлена. "Виетовские мотивы", возможно, знакомы вам по школьным задачам типа: найти сумму кубов корней квадратного уравнения (не находя самих корней). В общем речь идет об отыскании каких-либо симметрических функций от корней квадратного трехчлена с помощью выражения этих функций через элементарные симметрические многочлены (здесь их всего два: сумма и произведение корней). Ниже исследуются аналогичные задачи для многочленов произвольной степени.

Рассмотрим многочлен

$$f(x) = a_0x_n + a_1x^{n-1} + a_2x^{n-2} + ... + a_{n-2}x^2 + a_{n-1}x + a_n, \tag{49.30}$$

над полем P, положительной степени n, разложимый над P на линейные множители, или, что равносильно, имеющий в поле P ровно n корней, которые мы представим в виде списка

$$c = [c_1,\, c_2,\, \ldots,\, c_n], \tag{49.31}$$

где каждый корень повторяется столько раз, какова его кратность.

Нормализация многочлена не отражается на списке его корней, поэтому можно заменить многочлен (49.30) на нормализованный многочлен:

$$f^\circ(x) = x_n + a_1^\circ x^{n-1} + a_2^\circ x^{n-2} + \ldots + a_{n-2}^\circ x^2 + a_{n-1}^\circ x + a_n^\circ, \tag{49.30°}$$

где $a_k^\circ = a_k/a_0;\ k = 1, \ldots, n$.

Согласно теореме Виета (см. теорему 40.1), коэффициенты многочлена (с точностью до знака) выражаются как э.с. многочлены от его корней:

$$a_k^\circ = (-1)^k \sigma_k(c_1,\, c_2,\, \ldots,\, c_n);\ k = 1, \ldots, n. \tag{49.32}$$

Эти соотношения можно использовать "в другую сторону": э.с. многочлены от корней (49.31), с точностью до знака, равны коэффициентам многочлена (49.30).

Предложение 49.2. *Всякий симметрический многочлен от корней (49.31) многочлена (49.30) можно представить в виде многочлена от его "нормализованных" коэффициентов (49.32).*

Доказательство немедленно следует из теоремы 49.1 и виетовских соотношений (49.32). □

Замечание 49.8. Подчеркнем, что в предложении 49.2 получается многочлен именно от n переменных — нормализованных коэффициентов a_k° $(k = 1, \ldots, n)$. По отношению к исходным $n+1$ переменным — коэффициентам a_k $(k = 0, 1, \ldots, n)$ — получается *рациональная дробь*, в знаменателе которой будет некоторая степень коэффициента a_0.

Важно бывает выяснить, какова эта степень (в случае, когда данный многочлен $f(c)$ является, скажем, однородным степени m).

Пусть $ac_1^{i_1} c_2^{i_2} \ldots c_n^{i_n}$ является высшим членом этого многочлена; $i_1 + i_1 + \ldots + i_n = m$.

Соответствующий член

$$a\sigma_1^{j_1}\sigma_2^{j_2}\ldots\sigma_n^{j_n} = a\left(-\frac{a_1}{a_0}\right)^{j_1}\left(\frac{a_2}{a_0}\right)^{j_2}\ldots\left((-1)^n\frac{a_n}{a_0}\right)^{j_n} \tag{49.33}$$

будет содержать a_0 в знаменателе в степени

$$j_1 + j_2 + ... + j_n \overset{(49.16)}{=\!=} i_1.$$

Для всех других членов $f(c)$ степень по c_1 будет не выше i_1, и следовательно, степень a_0 в знаменателе окончательной формулы будет также не выше i_1.

В ы в о д: если однородный симметрический многочлен от корней многочлена (49.30) выразить через коэффициенты этого многочлена и полученную формулу привести к общему знаменателю, то степень a_0 в этом знаменателе будет равна степени i_1 высшего члена многочлена $f(c)$ по "наивысшей из переменных" c_1.

Пример 49.9. Вычислим значение

$$\pi_3(c_1, c_2, c_3, c_4) = c_1^3 + c_2^3 + c_3^3 + c_4^3 \tag{49.34}$$

многочлена из примера 49.3 для корней многочлена четвертой степени

$$f(x) = a_0x^4 + a_1x^3 + a_2x^2 + a_3x + a_4.$$

(Другими словами: вычислим сумму кубов корней для многочлена четвертой степени.)

Согласно формулам (49.32),

$$\sigma_1(c) = -\frac{a_1}{a_0};\ \sigma_2(c) = \frac{a_2}{a_0};\ \sigma_3(c) = -\frac{a_3}{a_0};\ \sigma_4(c) = -\frac{a_4}{a_0}. \tag{49.35}$$

Подставляя значения (49.35) в (49.28), получим:

$$\begin{aligned}\pi_3(c) = (\sigma_1(\mathrm{c}))^3 - 3\sigma_1(\mathrm{c})\sigma_2(\mathrm{c}) + 3\sigma_3(\mathrm{c}) = \\ = \left(-\frac{a_1}{a_0}\right)^3 - 3\left(\frac{a_1}{a_0}\right)\left(\frac{a_2}{a_0}\right) + 3\left(-\frac{a_3}{a_0}\right) = \\ = -\frac{1}{a_0^3}(a_1^3 + 3a_0a_1a_2 + 3a_0^2a_3).\end{aligned} \tag{49.36}$$

49.5. Дискриминант многочлена (от одной переменной). Рассмотрим снова многочлен от одной переменной $f(x)$ вида (49.30), положительной степени n, имеющий (с учетом кратностей) n корней в поле P. (От последнего условия можно отказаться, если рассматривать данный многочлен над алгебраическим замыканием $\overline{P}$ поля P; см. замечание 40.4.)

Определение 49.3. *Дискриминантом* многочлена $f(x) \in P[x]$ вида (49.30) называется симметрический многочлен

$$D(f) = a_0^{2n-2} D_n(c_1, c_2, ..., c_n) = a_0^{2n-2} \prod_{1 \leqslant i < j \leqslant n} (c_j - c_i)^2 \qquad (49.37)$$

[см. (49.5)], если его представить в виде *многочлена*

$$(D(f))(a_0, a_1, ..., a_n)$$

от коэффициентов $a_0, a_1, ..., a_n$ многочлена $f(x)$.

Замечание 49.9. Данное выше определение является довольно "замысловатым": дискриминант, с одной стороны, может рассматриваться как симметрический многочлен (49.37) от n переменных $c_1, ..., c_n$, трактуемых как корни многочлена (49.30).

С другой стороны, согласно предложению 49.2, этот многочлен может быть представлен как (уже не обязательно симметрический) многочлен $(D(f))(a_0, a_1, ..., a_n)$ от $n+1$ переменных — коэффициентов многочлена $f(x)$.

Поскольку корни многочлена, как правило, заранее не известны (и вообще, некоторые из них могут принадлежать не данному полю P, а более широкому полю $\overline{P}$), "правильным" является второе представление дискриминанта.

Однако второе представление, по сути, равносильно первому, а при первом представлении очевидно *основное свойство дискриминанта*:

$$[\, D(f) = 0 \,] \Leftrightarrow [\text{многочлен } f(x) \text{ имеет кратные корни}]. \qquad (49.38)$$

Множитель a_0^{2n-2} в формуле (49.37) является не слишком существенным: он никак не сказывается на основном свойстве (49.38).

Вводится он с естественной целью, чтобы в случае, когда коэффициенты данного многочлена являются *целыми числами* (или *целыми элементами поля*; точного определения последнего понятия мы здесь не приводим), коэффициенты $D(f)$ также являлись целыми числами (целыми элементами поля).

Согласно замечанию 49.8, выражение для $D_n(c)$ через коэффициенты многочлена $f(x)$ будет содержать в знаменателе старший коэффициент a_0 в такой степени i_1, в которой c_1 входит в высший член многочлена $D_n(c)$.

Легко видеть, что этот показатель $i_1 = 2(n-1)$. Именно чтобы "погасить" знаменатель a_0^{2n-2}, и вводится такой же множитель перед $D_n(c)$ в определении дискриминанта.

Пример 49.10. Найдем дискриминант для многочлена третьей степени

$$f(x) = a_0x^3 + a_1x^2 + a_2x + a_3 = a_0(x - c_1)(x - c_2)(x - c_3).$$

Имеем, в силу примера 49.4 и формул (49.32):

$$\begin{aligned} D_3(c) &= (c_2 - c_1)^2(c_3 - c_1)^2(c_3 - c_2)^2 = \\ &= \sigma_1^2\sigma_2^2 - 4\sigma_1^3\sigma_3 - 4\sigma_2^3 + 18\sigma_1\sigma_2\sigma_3 - 27\sigma_3^2 = \\ &= \left(-\frac{a_1}{a_0}\right)^2\left(\frac{a_2}{a_0}\right)^2 - 4\left(-\frac{a_1}{a_0}\right)^3\left(-\frac{a_3}{a_0}\right) - 4\left(\frac{a_2}{a_0}\right)^3 + \\ &\quad + 18\left(-\frac{a_1}{a_0}\right)\left(\frac{a_2}{a_0}\right)\left(-\frac{a_3}{a_0}\right) - 27\left(-\frac{a_3}{a_0}\right)^2. \end{aligned}$$

Домножая на старший коэффициент в степени $2n - 2 = 4$, получим:

$$D(f) = a_1^2a_2^2 - 4a_1^3a_3 - 4a_0a_2^3 + 18a_0a_1a_2a_3 - 27a_0^2a_3^2. \tag{49.39}$$

В следующем параграфе нам понадобится частный случай этой формулы, относящийся к кубическому многочлену вида

$$f(x) = x^3 + px + q,$$

т. е. получающийся при $a_0 = 1$; $a_1 = 0$; $a_2 = p$; $a_3 = q$:

$$D(f) = -4p^3 - 27q^2. \tag{49.40}$$

Настоятельно рекомендуем читателям вычислить дискриминант квадратичного многочлена (чтобы вспомнить детство), а также дискриминант для общего многочлена степени 4 или хотя бы для следующего частного вида многочлена четвертой степени:

$$f(x) = x^4 + px^2 + qx + r.$$

Замечание 49.9. Можно доказать (см., например, учебники [5, 9]), что дискриминант многочлена $f(x)$, с точностью до постоянного множителя, совпадает с *результантом* (см. замечание 38.5) многочлена $f(x)$ и его *производной* $f'(x)$ (см. п. 47.2).

§ 50. Ферро, Тарталья, Кардано, Феррари и другие

50.1. Уравнения малых степеней (над полем $\mathbb{C}$). Основная теорема алгебры утверждает, что всякий многочлен степени $n \geqslant 1$ с комплексными коэффициентами имеет хотя бы один комплексный корень (а фактически — ровно n корней, с учетом кратностей). Другими словами, алгебраическое уравнение

$$a_0x^n+a_1x^{n-1}+...+a_{n-1}x+a_n = 0;\ a_k \in \mathbb{C}\ (k = 0, ..., n);\ a_0 \neq 0 \quad (50.1)$$

всегда имеет ровно n решений $c_k \in \mathbb{C}$ $(k = 1, ..., n)$.

Однако эта теорема не дает никакого способа отыскания всех корней уравнения (50.1), или хотя бы одного из них.

В то же время для уравнений первой и второй степени корни определяются простыми формулами.

Линейное уравнение

$$a_0x + a_1 = 0$$

имеет единственный корень

$$c_1 = -\frac{a_1}{a_0}.$$

Квадратное уравнение

$$a_0x^2 + a_1x + a_2 = 0$$

имеет два (возможно, совпадающих) корня

$$c_{1,2} = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_0a_2}}{2a_0}.$$

Очень долго и трудно математики искали аналогичную формулу для кубического уравнения

$$a_0x^3 + a_1x^2 + a_2x + a_3 = 0. \quad (50.2)$$

И в конце концов в XVI веке желанная формула была найдена. (См. следующий пункт.)

Вскоре после этого был найден способ решения для уравнения четвертой степени:

$$a_0x^4 + a_1x^3 + a_2x^2 + a_3x + a_4 = 0. \quad (50.3)$$

Формула, явно выражающая все корни уравнения через его коэффициенты (с использованием алгебраических операций сложения, умножения, возведения в натуральную степень и извлечения корня), в этом случае оказалась настолько сложной, что ее практически нигде не записывают. Вместо этого приводится "инструкция", разбивающая работу на ряд этапов. (Об этом — в п. 50.3.)

Принят такой термин: "уравнение решается в радикалах". Радикал — это символ $\sqrt[n]{\ }$, обозначающий операцию извлечения корня степени n из числа.

Лишь в XIX веке математической общественностью было понято и признано, что общих формул, выражающих "в радикалах" корни алгебраического уравнения степени 5 и выше через коэффициенты этого уравнения, не существует.

Таким образом, алгебраисты получили неожиданный результат, который имеет совсем особую природу: утверждается (и доказывается), что при $n > 4$ *в принципе не существует* никакой формулы, выражающей корни уравнения степени n через его коэффициенты (с использованием перечисленных выше алгебраических действий).

Устанавливается этот факт (и другие подобные "теоремы о несуществовании") весьма непросто. Построена особая математическая наука — *теория Галуа*, изучающая, в частности, условия разрешимости алгебраических уравнений в радикалах.

С началами этой науки вы можете познакомиться, например, по книге [5]. (Кстати, основой теории Галуа выступает *теория групп*, первичных понятий которой мы кратко касались в этом пособии.)

Данный пункт мы завершим описанием двух первых этапов упрощения общего уравнения степени n вида (50.1).

О первом этапе неоднократно уже говорилось: это — *нормализация*, состоящая в делении обеих частей уравнения на старший коэффициент a_0. При этом получается нормализованное уравнение

$$x^n + a_1^\circ x^{n-1} + ... + a_{n-1}^\circ x + a_n^\circ = 0;\ a_k^\circ = \frac{a_k}{a_0}\ (k = 1, ..., n). \quad (50.1^\circ)$$

Второй этап представляет собой замену неизвестной, устраняющую из уравнения второй (следующий за старшим) член.

Замена эта является сдвигом:

$$x = y - \frac{a_1^\circ}{n}. \quad (50.4)$$

Из формулы (50.4) следует, в силу формулы бинома Ньютона, что

$$x^n = y^n - ny^{n-1}\frac{a_1^\circ}{n} + (...) = y^n - a_1^\circ y^{n-1} + (...);$$
$$x^{n-1} = y^{n-1} + (...),$$

где символом (...) обозначены "младшие" члены (имеющие степень не выше $n-2$ по y).

При пересчете степеней x^k $(k \leqslant n-2)$ будут появляться только младшие члены.

В левой части уравнения (50.1°) получим:

$$\left[\, y^n - a_1^\circ y^{n-1} + (...) \,\right] + a_1^\circ \left[\, y^{n-1} + (...) \,\right] + (...) = y^n + (...).$$

Следовательно, при переходе к неизвестной y исчезнет член степени $n-1$ и уравнение приобретет "упрощенный" вид

$$y^n + p_2 y^{n-2} + ... + p_{n-1} y + p_n = 0. \tag{50.5}$$

В частности, для $n = 3$ и $n = 4$ уравнения (50.2) и (50.3) всегда могут быть приведены к упрощенным видам

$$y^3 + py + q = 0 \tag{50.6}$$

и

$$y^4 + py^2 + qy + r = 0 \tag{50.7}$$

соответственно.

Замечание 50.1. Замену переменной (50.4) можно выполнить по Горнеру, поскольку она есть не что иное, как переразложение многочлена в левой части уравнения по степеням

$$y = x - \left(-\frac{a_1^\circ}{n}\right).$$

50.2. Метод Ферро — Тартальи — Кардано решения уравнений третьей степени. Около 1515 года итальянский математик *дель Ферро* открыл метод решения кубических уравнений. Он долго хранил его в тайне, доверив лишь очень узкому кругу, и умер, так и не опубликовав свое открытие.

В 1535 году другой итальянский математик, *Тарталья*, готовясь к публичному диспуту (состязанию) с обладателем секрета дель Ферро, по-видимому, переоткрыл этот метод (и выиграл диспут).

Значительно более крупный деятель (врач, математик, механик), единственный человек из числа математиков, очень хорошо знакомый всем автомобилистам, *Кардано*, после долгих и скандальных переговоров завладел в 1539 году секретом Тартальи, дав священную клятву хранить тайну, но в 1545 году опубликовал (со ссылкой на дель Ферро и Тарталью) метод решения кубических уравнений.

Кардано и его ученик *Феррари* во многом развили и усовершенствовали метод Ферро — Тартальи, но они отнюдь не были его первооткрывателями.

Тем не менее, по традиции, открытая двумя итальянцами формула носит имя третьего итальянца, Кардано.

С подробностями яростной полемики между Тартальей и тандемом Кардано — Феррари, получившей в истории математики определение "великая контраверза", можно познакомиться, например, по книге Р. С. Гутера и Ю. Л. Полунова "Джироламо Кардано" (М.: Знание, 1980). Можно также обратиться к историческому обзору в начале еще одного "антикварного" учебника: Г. М. Шапиро. "Высшая алгебра" (М.: ГУПИ, 1936).

Приступаем к изложению метода Ферро — Тартальи — Кардано. (Разумеется — в современной трактовке и с использованием современных обозначений. В трудах итальянских алгебраистов все уравнения выражались словесно; например, фраза "Куб плюс вещь равны числу" описывала уравнение $x^3 + x = a$.)

Будем с самого начала считать, что кубическое уравнение имеет упрощенный вид

$$x^3 + px + q = 0, \tag{50.8}$$

где $p, q \neq 0$ (случай $p = 0$ сводится к задаче извлечения кубического корня из числа $-q$, которая уже разобрана в п. 32.3; случай $q = 0$ также не интересен, поскольку одно решение, $x = 0$, сразу находится и дело сводится к вычислению квадратного корня).

При $q \neq 0$ нуль не является корнем уравнения (50.8).

Далее используется знакомый оборот "будем искать", но уже в другой редакции: будем искать решение уравнения (50.8) в виде

$$x = u + v. \tag{50.9}$$

Таким образом, вместо одной неизвестной x, мы вводим *две* новые неизвестные, u и v. Чего же мы этим добиваемся? Неужели две

неизвестные искать легче, чем одну? Да, ибо мы *обретаем свободу*: к уже имеющемуся у нас уравнению (в котором теперь две неизвестные) мы можем добавить новое соотношение, связывающее эти неизвестные так, чтобы

1) из этого соотношения две неизвестные легко выражались одна через другую;

2) новое соотношение вместе с исходным уравнением давали легко решаемую систему уравнений.

После подстановки (50.9) в (50.8) получаем:

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

или

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

или

$$u^3 + v^3 + (3uv + p)(u+v) + q = 0. \tag{50.10}$$

Положим

$$3uv + p = 0. \tag{50.11}$$

Это и будет нашим вторым соотношением. В силу $p \neq 0$, из уравнения (50.11) вытекает, что $u, v \neq 0$, и следовательно, одну из этих неизвестных легко выразить через другую.

С учетом (50.11) уравнение (50.10) сводится к

$$u^3 + v^3 + q = 0. \tag{50.12}$$

Итак, для двух новых неизвестных получается система уравнений, которую мы запишем следующим образом:

$$\begin{cases} u^3 + v^3 & = -q; \\ uv & = -\frac{p}{3}. \end{cases} \tag{50.13}$$

Если u и v удовлетворяют системе (50.13), то их сумма будет корнем уравнения (50.8).

Второе из уравнений (50.13) влечет уравнение

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

и следовательно, система (50.13) влечет систему

$$\begin{cases} u^3 + v^3 & = -q; \\ u^3v^3 & = -\frac{p^3}{27}. \end{cases} \tag{50.14}$$

Систему (50.14) легко решить относительно u^3 и v^3. Эти величины можно найти как корни вспомогательного квадратного уравнения

$$z^2 + qz - \frac{p^3}{27} = 0. \tag{50.15}$$

(В "школьной" терминологии можно было бы выразиться так: мы применяем "обратную теорему Виета".)

Уравнение (50.15) решается по обычным формулам для корней квадратного уравнения, как мы это делали в § 31. Не забывайте о том, что мы находимся в поле $\mathbb{C}$. В частности, перечитайте замечание 31.6 о *главном значении* для квадратного корня из комплексного числа.

В силу симметрии, все равно какому из корней квадратного уравнения приравнивать u^3, а какому — v^3. Будем считать, что

$$\begin{cases} u^3 &= -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}; \\ v^3 &= -\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}. \end{cases} \tag{50.16}$$

[В формулах (50.16) радикал $\sqrt{\ }$ обозначает одно (любое; например, главное) значение квадратного корня. Второе значение также фигурирует в этих формулах как такой же радикал, но со знаком минус перед ним.]

Теперь можно найти неизвестные u и v, извлекая кубические корни из правых частей уравнений (50.16) (см. п. 32.3):

$$\begin{cases} u &= \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}; \\ v &= \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}. \end{cases} \tag{50.17}$$

Вот тут начинается самое тонкое место решения. Какой смысл придать знакам кубического корня $\sqrt[3]{\ }$ в формулах (50.17)? Кубический корень из ненулевого комплексного числа имеет три значения. Поэтому фактически у нас будет три значения для u:

$$u_0,\ u_1,\ u_2 \tag{50.18}$$

— и три значения для v:

$$v_0,\ v_1,\ v_2. \tag{50.19}$$

Девять возможных пар (u_j, v_k) $[j, k \in \{0, 1, 2\}]$ образуют множество всех решений системы (50.14). Но эта система является лишь следствием подлежавшей решению системы (50.13), и не все решения (50.14) будут удовлетворять (50.13). Выясним, какие будут.

Возьмем одно число из списка (50.18), все равно какое. Пусть именно оно было обозначено u_0. Тогда остальные числа в этом списке могут быть представлены (см. пункт 33.3, в частности, пример 33.1) в виде

$$u_1 = u_0\zeta;\ u_2 = u_0\zeta^2, \tag{50.20}$$

где (для краткости) обозначено:

$$\zeta = \zeta_3 = \frac{-\sqrt{3}+i}{2};\ \zeta^2 = \zeta_3^2 = \frac{-\sqrt{3}-i}{2}.$$

Аналогичным образом поступим со списком (50.19).

Рассмотрим теперь произведение u_0v_0. По построению

$$(u_0v_0)^3 = \left(-\frac{p}{3}\right)^3.$$

Если два числа имеют одинаковые кубы, то сами они могут отличаться лишь множителем, куб которого равен единице, т. е.

1) либо $u_0v_0 = -\frac{p}{3}$, и тогда пара (u_0, v_0) является решением системы (50.13);

2) либо $u_0v_0 = (-\frac{p}{3})\zeta$; тогда, домножая последнее равенство на ζ^2 и учитывая, что $\zeta^3 = 1$, мы получим: $u_0v_2 = -\frac{p}{3}$, и следовательно, пара (u_0, v_2) есть решение (50.13);

3) либо $u_0v_0 = (-\frac{p}{3})\zeta^2$; тогда, домножая равенство на ζ, получим: $u_0v_1 = -\frac{p}{3}$, и пара (u_0, v_1) будет решением (50.13).

Сменим нумерацию чисел (50.19) так, чтобы решением (50.13) являлась именно пара (u_0, v_0). Тогда можно будет выразить еще две пары, удовлетворяющие этой системе; это будут (u_1, v_2) и (u_2, v_1).

Вопрос практического подбора значения v_0, соответствующего произвольно выбранному значению u_0, решается [после того как существование такого значения в списке (50.19) доказано] очень просто. Надо взять

$$v_0 = -\frac{p}{3u_0}. \tag{50.21}$$

Теперь можно определить три корня уравнения (50.8):

$$x_0 = u_0 + v_0;$$
$$x_1 = u_1 + v_2;$$
$$x_2 = u_2 + v_1,$$

или

$$\begin{array}{rcccl} x_0 & = & u_0 & + & v_0; \\ x_1 & = & u_0\zeta & + & v_0\zeta^2; \\ x_2 & = & u_0\zeta^2 & + & v_0\zeta, \end{array} \tag{50.22}$$

где (напомним) u_0 — любое значение кубического корня

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}, \tag{50.23}$$

а v_0 — однозначно определяется по u_0 по формуле (50.21).

Окончательный ответ, содержащий все корни кубического уравнения, можно представить в виде следующей знаменитой формулы Ферро — Тартальи — Кардано (ФТК):

$$x = \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}. \tag{50.24}$$

Однако лишний раз подчеркнем, что формула (ФТК) — из тех, к которым, как говорится, "требуется еще голова". Здесь не срабатывает автоматическая подстановка своих данных в готовую формулу (подобно тому, как это было для случая квадратного уравнения), ибо необходим сознательный отбор трех правильных из девяти возможных значений правой части.

Замечание 50.2. Отметим, что в проведенных выше вычислениях под знаком квадратного корня фигурировала величина

$$\frac{q^2}{4} + \frac{p^3}{27} = -\frac{D}{108}, \tag{50.25}$$

где D — *дискриминант* (49.40) неполного кубического многочлена $x^3 + px + q$.

Если $D = 0$, то оба числа, u_0 и v_0, в формулах (50.22) являются кубическими корнями из $-\frac{q}{2}$. Поэтому

1) либо $v_0 = u_0$, и тогда $x_1 = x_2$,

2) либо $v_0 = u_0\zeta$, и тогда $x_0 = x_1$,

3) либо $v_0 = u_0\zeta^2$, и тогда $x_0 = x_2$.

Во всех трех случаях появляется корень кратности как минимум два. Возможен случай трехкратного корня. (Попробуйте выяснить, при каком условии это имеет место.)

Пример 50.1. Решим полное кубическое уравнение (с действительными коэффициентами):

$$x^3 - 9x^2 + 21x - 5 = 0.$$

Сделаем упрощающую замену неизвестной (50.4):

$$x = y + 3.$$

Относительно новой неизвестной y получим уравнение, не содержащее квадратичного члена:

$$y^3 - 6y + 4 = 0.$$

Выражение (50.25) под квадратным корнем в формуле (ФТК) будет иметь следующий вид:

$$\frac{q^2}{4} + \frac{p^3}{27} = 4 - 8 = -4.$$

Корень квадратный из этого числа будет иметь два значения: $\pm 2i$. Найдем одно (любое) значение u_0 кубического корня (50.23):

$$u = \sqrt[3]{-2 + 2i}.$$

Число

$$-2 + 2i = 2\sqrt{2}(\cos\,\frac{3\pi}{4} + i\sin\,\frac{3\pi}{4}) = (\sqrt{2})^3(\cos\,3\cdot\frac{\pi}{4} + i\sin\,3\cdot\frac{\pi}{4})$$

имеет в качестве главного значения кубического корня число

$$u_0 = \sqrt{2}(\cos\,\frac{\pi}{4} + i\sin\,\frac{\pi}{4}) = 1 + i.$$

По формуле (50.21) находим:

$$v_0 = -\frac{p}{3u_0} = -\frac{-6}{3(1+i)} = \frac{2}{1+i} = 1 - i.$$

Формулы (50.22) теперь дают:

$$y_0 = u_0 + v_0 = (1+i) + (1-i) = 2;$$

$$x_1 = u_0\zeta + v_0\zeta^2 = (1+i)\frac{-\sqrt{3}+i}{2} + (1-i)\frac{-\sqrt{3}-i}{2} = -1 - \sqrt{3};$$

$$x_2 = u_0\zeta^2 + v_0\zeta = (1+i)\frac{-\sqrt{3}-i}{2} + (1-i)\frac{-\sqrt{3}+i}{2} = -1 + \sqrt{3}.$$

Возвращаясь к старой неизвестной, получаем:

$$x_0 = y_0 + 3 = 5;\ x_1 = y_1 + 3 = 2 - \sqrt{3};\ x_2 = y_2 + 3 = 2 + \sqrt{3}.$$

О т в е т: $5;\ 2 - \sqrt{3};\ 2 + \sqrt{3}$.

Замечание 50.3. 1. Обратите внимание на то, что в разобранном выше примере кубический многочлен с действительными коэффициентами имел три действительных корня. Любопытно, однако, что при их отыскании пришлось извлекать квадратный корень из отрицательного действительного числа, а затем — кубический корень из комплексного числа.

Именно такая ситуация была камнем преткновения как для Тартальи, так и для Кардано. Последний, будучи более опытным и смелым в математике, фактически положил начало легализации *мнимых* чисел в алгебре. (Как ни странно, в ту эпоху не признавали также и *отрицательные действительные* числа, и это непризнание продержалось чуть ли не дольше, чем непризнание мнимых чисел.)

2. Можно во всех подробностях разобрать случай действительных коэффициентов в кубическом уравнении и выяснить, когда такое уравнение имеет:

— три действительных различных корня;

— три действительных корня, два из которых одинаковы;

— один действительный корень кратности три;

— один действительный корень и два сопряженных друг другу комплексных корня.

Решающую роль в анализе будет играть дискриминант.

(Мы настоятельно рекомендуем читателям ознакомиться с этим анализом и соответствующими примерами по какому-либо из учебников, указанных в списке литературы.)

Пример 50.2. Решим кубическое уравнение с комплексными коэффициентами:

$$x^3 + 3x - 2i = 0.$$

Это уравнение уже является упрощенным (не содержит квадратичного члена). Дискриминант многочлена в левой части:

$$D = -4p^3 - 27q^2 = -4 \cdot 27 - 27 \cdot (-4) = 0.$$

Следовательно, уравнение имеет кратные корни. Число u_0 находится как любое из значений кубического корня:

$$u = \sqrt[3]{-\frac{q}{2}} = \sqrt[3]{i}.$$

Кубический корень из i легко извлечь "без тригонометрии", если заметить, что $i^3 = -i$, и следовательно, $(-i)^3 = i$. Так что одним из значений искомого корня является

$$u_0 = -i.$$

Далее:

$$v_0 = -\frac{p}{3u_0} = -\frac{3}{3(-i)} = -i$$

и

$$x_0 = -2i;\ x_1 = x_2 = i.$$

О т в е т: $-2i;\ i;\ i$.

Пример 50.3. Maple знает метод (ФТК). Он может вам даже выдать общую формулу для решения полного кубического уравнения.

Наберите:

```
> solve( a[0]*x^ 3 + a[1]*x^ 2 + a[2]*x + a[3] = 0, x );
```

и полюбуйтесь предъявленным ответом.

Затем перерешайте примеры 50.1 и 50.2.

А теперь дадим Maple новое задание:

```
> solve( x^ 3 + (3-3*I)*x + (-2+I) = 0, x ); map( evalc, [ % ] );
```

$$\frac{1}{2} + I + \frac{1}{2}I\sqrt{3},\ \frac{1}{2} + I - \frac{1}{2}I\sqrt{3},\ -1 - 2I$$

$$\left[\frac{1}{2} + \left(1 + \frac{\sqrt{3}}{2}\right)I,\ \frac{1}{2} + \left(1 - \frac{\sqrt{3}}{2}\right)I,\ -1 - 2I\right]$$

(Вторая команда превращает ответ в список, все элементы которого суть комплексные числа в стандартной записи.)

Автор надеется, что вам захотелось теперь перерешать этот пример вручную.

50.3. Метод Феррари решения уравнений четвертой степени. Феррари, любимый (и очень способный) ученик Кардано, ознакомившись с формулой, доверенной его учителю Тартальей, довольно скоро понял, как решать уравнения четвертой степени.

Начинается решение, как и для кубических уравнений, с нормализации левой части и удаления из нее (с помощью описанной в п. 50.1 замены неизвестной) кубического члена.

Будем считать, что уравнение уже имеет упрощенный вид:

$$x^4 + px^2 + qx + r = 0. \tag{50.26}$$

Случай $r = 0$ сразу приводит к отысканию одного (нулевого) корня, после чего дело сводится к кубическому уравнению. Случай $q = 0$ также малоинтересен: это — биквадратное уравнение, легко сводимое к квадратному. Поэтому в дальнейшем можно считать, что $q, r \neq 0$.

Основным приемом в методе Феррари служит особый вариант *выделения полного квадрата*, не тот, что вы проходили в школе, а более сложный.

Сначала "разнесем" члены: два старших останутся слева, а два младших перенесем в правую часть. Затем к обеим частям добавим выражение $\frac{p^2}{4}$ и свернем в левой части полный квадрат. Получим:

$$x^4 + px^2 = -qx - r;\ x^4 + 2 \cdot x^2 \cdot \frac{p}{2} + \frac{p^2}{4} = -qx - r + \frac{p^2}{4};$$
$$\left(x^2 + \frac{p}{2}\right)^2 = -qx - r + \frac{p^2}{4}. \tag{50.27}$$

Это было первое выделение полного квадрата, вполне "школьное". Теперь наступает главный момент решения: к обеим частям (50.27) мы добавим выражение $2(x^2 + \frac{p}{2})\lambda + \lambda^2$, содержащее *параметр* λ (новую, совершенно "постороннюю" переменную, вводимую для "обретения свободы"). Затем в левой части снова свернем полный квадрат.

Получим:

$$\left(x^2 + \frac{p}{2}\right)^2 + 2(x^2 + \frac{p}{2})\lambda + \lambda^2 = 2(x^2 + \frac{p}{2})\lambda + \lambda^2 - qx - r + \frac{p^2}{4};$$
$$\left(x^2 + \frac{p}{2} + \lambda\right)^2 = 2\lambda x^2 - qx + \left(\lambda^2 + p\lambda - r + \frac{p^2}{4}\right). \tag{50.28}$$

Правую часть формулы (50.28) рассмотрим как квадратный трехчлен от переменной x, с коэффициентами

$$a = 2\lambda;\ b = -q;\ c = \lambda^2 + p\lambda - r + \frac{p^2}{4}. \tag{50.29}$$

Как известно из школьного курса алгебры, квадратный трехчлен ax^2+bx+c *является полным квадратом*, т. е. представляется в виде

$$ax^2+bx+c=a(x-x_0)^2,\ x_0=-\frac{b}{2a},$$

тогда и только тогда, когда его дискриминант $D=0$.

(Вообще-то мы здесь немного грешим против истины: в школе это утверждение доказывалось для многочленов над полем $\mathbb{R}$. Но вы должны были уже привыкнуть к тому, что все факты, доказательство которых опирается лишь на аксиомы поля, справедливы над любым полем. Сейчас мы работаем над $\mathbb{C}$.)

Подберем теперь значение параметра λ так, чтобы многочлен от x в правой части (50.28) являлся полным квадратом.

Вид этого полного квадрата должен быть следующим:

$$2\lambda\left(x+\frac{q}{4\lambda}\right)^2, \tag{50.30}$$

а условие $D=0$, из которого нам потребуется находить λ, выглядит так:

$$q^2-4\cdot 2\lambda\cdot\left(\lambda^2+p\lambda-r+\frac{p^2}{4}\right)=0$$

— или, если расположить члены по степеням λ и домножить на -1:

$$8\lambda^3+8p\lambda^2+2(p^2-4r)\lambda-q^2=0. \tag{50.31}$$

Уравнение (50.31) является кубическим относительно λ; оно называется *резольвентой* для исходного уравнения (50.26). Нам нужен один (любой) корень резольвенты.

Поскольку в предыдущем пункте мы обучились решению кубических уравнений, то будем считать, что некоторый корень (50.31) (обозначим его λ_0) мы уже нашли.

Теперь мы можем записать уравнение (50.28) при $\lambda=\lambda_0$, придав правой части вид (50.30):

$$\left(x^2+\frac{p}{2}+\lambda_0\right)^2=2\lambda_0\left(x+\frac{q}{4\lambda_0}\right)^2. \tag{50.32}$$

Подведем множитель $2\lambda_0$ "под квадрат"; получим:

$$\left(x^2+\frac{p}{2}+\lambda_0\right)^2=\left[\sqrt{2\lambda_0}\left(x+\frac{q}{4\lambda_0}\right)\right]^2, \tag{50.33}$$

где можно взять любое из двух значений корня $\sqrt{2\lambda_0}$.

Квадраты двух чисел совпадают тогда и только тогда, когда данные числа либо совпадают, либо отличаются знаком. Поэтому уравнение (50.33) равносильно следующей совокупности двух квадратных уравнений:

$$\left[\begin{array}{rcl} x^2+\frac{p}{2}+\lambda_0 &=& \sqrt{2\lambda_0}\left(x+\frac{q}{4\lambda_0}\right); \\ x^2+\frac{p}{2}+\lambda_0 &=& -\sqrt{2\lambda_0}\left(x+\frac{q}{4\lambda_0}\right). \end{array}\right. \tag{50.34}$$

Пусть $x_{1,2}$ — корни первого из уравнений совокупности (50.34), а $x_{3,4}$ — корни второго уравнения. Тогда четыре числа

$$x_1,\, x_2,\, x_3,\, x_4$$

будут корнями исходного уравнения (50.26).

Пример 50.4. Решим уравнение четвертой степени

$$x^4-2x^3+4x^2-2x+3.$$

Произведем замену

$$x=y+\frac{1}{2}$$

или, другим словами, переразложим (с помощью схемы Горнера; см. п. 41.2) левую часть по степеням $y=x-\frac{1}{2}$. Получим:

$$y^4+\frac{5}{2}y^2+y+\frac{45}{16}=0.$$

Разносим члены и дважды выделяем в левой части полный квадрат (вводя на втором этапе параметр λ):

$$y^4+\frac{5}{2}y^2=-y-\frac{45}{16};\quad y^4+\frac{5}{2}y^2+\frac{25}{16}=-y-\frac{45}{16}+\frac{25}{16};$$

$$\left(y^2+\frac{5}{4}\right)^2=-y-\frac{5}{4};$$

$$\left(y^2+\frac{5}{4}\right)^2+2\lambda\left(y^2+\frac{5}{4}\right)+\lambda^2=2\lambda\left(y^2+\frac{5}{4}\right)+\lambda^2-y-\frac{5}{4};$$

$$\left(y^2+\frac{5}{4}+\lambda\right)^2=2\lambda y^2-y+\left(\lambda^2+\frac{5}{2}\lambda-\frac{5}{4}\right).$$

Найдем дискриминант правой части последнего уравнения, рассматривая ее как квадратный трехчлен по y:

$$D = 1 - 8\lambda\left(\lambda^2 + \frac{5}{2}\lambda - \frac{5}{4}\right) = -8\lambda^3 - 20\lambda^2 + 10\lambda + 1.$$

Приравниваем этот дискриминант нулю, получая тем самым условие того, что правая часть представляет из себя полный квадрат.

После домножения на -1 указанное условие принимает вид:

$$8\lambda^3 + 20\lambda^2 - 10\lambda - 1 = 0.$$

Это уравнение и есть кубическая резольвента исходного уравнения четвертой степени.

Сейчас, начиная решение резольвенты, нам следовало бы, в соответствии с методом (ФТК), произвести замену $\lambda = u + v$; вывести систему уравнений для u и v; и т. д. Но данный (учебный) пример подобран таким образом, чтобы ранее изученный этап не был громоздким (не отвлекал читателей от вновь изучаемого метода).

Нам нужен один (любой) из корней резольвенты. В данном случае его можно найти простым подбором (по методу отыскания рациональных корней для многочлена с целыми коэффициентами; см. п. 42.3).

Резольвента в данном примере имеет рациональный корень

$$\lambda = \lambda_0 = \frac{1}{2}.$$

Уравнение (50.32) приобретает вид

$$\left(y^2 + \frac{7}{4}\right)^2 = \left(y - \frac{1}{2}\right)^2$$

и равносильно совокупности

$$\left[\begin{array}{rcl} y^2 + \frac{7}{4} & = & y - \frac{1}{2} \\ y^2 + \frac{7}{4} & = & -\left(y - \frac{1}{2}\right) \end{array}\right. ;\ .$$

Решая эти квадратные уравнения, находим четыре корня:

$$y_{1,2} = \frac{1}{2} \pm i\sqrt{2};\ y_{3,4} = -\frac{1}{2} \pm i.$$

Возвращаясь к неизвестной x, получаем

О т в е т:

$$x_{1,2} = 1 \pm i\sqrt{2};\ x_{3,4} = \pm i.$$

Пример 50.5. Решать уравнения четвертой степени Maple тоже умеет. Скажем, с уравнением из предыдущего примера он легко справляется:

```
> solve( x^4 - 2*x^3 + 4*x^2 - 2*x + 3 = 0, x );
```

$$I,\ -I,\ 1+\sqrt{2}\,I,\ 1-\sqrt{2}\,I$$

Однако, слегка изменив условие:

```
> solve( x^4 - 2*x^3 + 4*x^2 - 2*x + 5 = 0, x );
```

мы получим странный, на первый взгляд, ответ:

$$\begin{aligned}
&\mathrm{RootOf}(_Z^4 - 2\,_Z^3 + 4\,_Z^2 - 2\,_Z + 5, index = 1),\\
&\mathrm{RootOf}(_Z^4 - 2\,_Z^3 + 4\,_Z^2 - 2\,_Z + 5, index = 2),\\
&\mathrm{RootOf}(_Z^4 - 2\,_Z^3 + 4\,_Z^2 - 2\,_Z + 5, index = 3),\\
&\mathrm{RootOf}(_Z^4 - 2\,_Z^3 + 4\,_Z^2 - 2\,_Z + 5, index = 4)
\end{aligned}$$

Во-первых, Maple заменил нашу неизвестную x своей (системной) переменной $\_Z$. Во-вторых, он "отказался" показывать нам корни, которые имеют очень громоздкие выражения. Но он не отказался работать с ними! Эти громоздкие выражения занумерованы (*индексированы*) и хранятся в памяти; с ними в дальнейшем можно производить любые допустимые действия. (Другое дело, что результаты этих действий опять-таки будут содержать выражения типа RootOf.)

Есть, однако, возможность заставить Maple показать ответы. Для этого надо сменить (принятое по умолчанию) значение 'false' специальной *глобальной переменной* _EnvExplicit на значение 'true'.

Сделаем это:

```
> _EnvExplicit := true;
```

$$\_EnvExplicit := true$$

Теперь повторим команду "решить уравнение", и будет выдан "многоэтажный" ответ.

Приближенные значения корней можно, как обычно, получить с помощью команды evalf.

Замечание 50.4. Выдающиеся достижения итальянских математиков XVI века, связанные с открытием формул для корней уравнений третьей и четвертой степени, позволили глубже понять строение "математической вселенной".

Однако, как немедленно выяснилось, для практического вычисления значений корней эти формулы совершенно не пригодны.

Выразив корни через радикалы, мы вынуждены будем вычислять полученные радикалы. А это в большинстве случаев удается сделать лишь приближенно.

Но приближенные значения корней можно найти иначе (причем гораздо проще и с большей точностью) с помощью специально разработанных методов численного решения уравнений.

В настоящей книге мы практически не касались вопросов приближенных вычислений. Это особая, очень специфическая и очень важная сфера математической деятельности. С ней вы будете достаточно глубоко знакомиться в других математических дисциплинах.

Приложение 1

Рисунки к главе 5

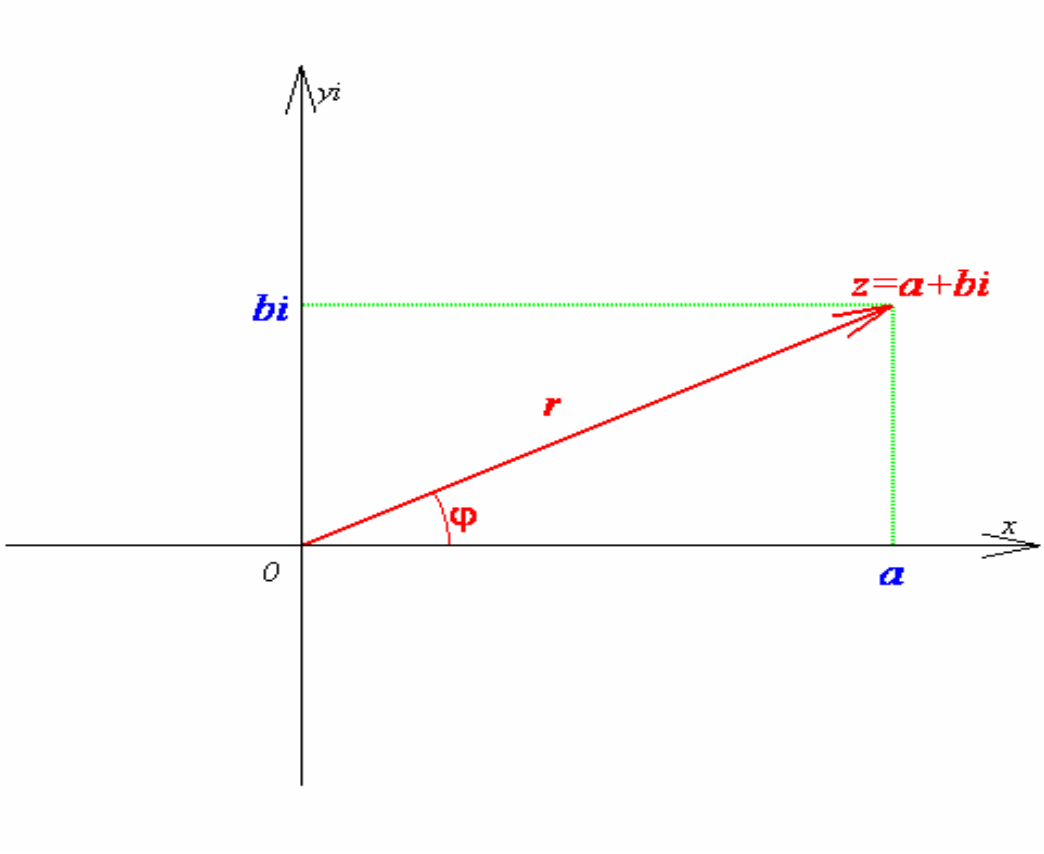


Рис. 32.1. Тригонометрическая форма комплексного числа

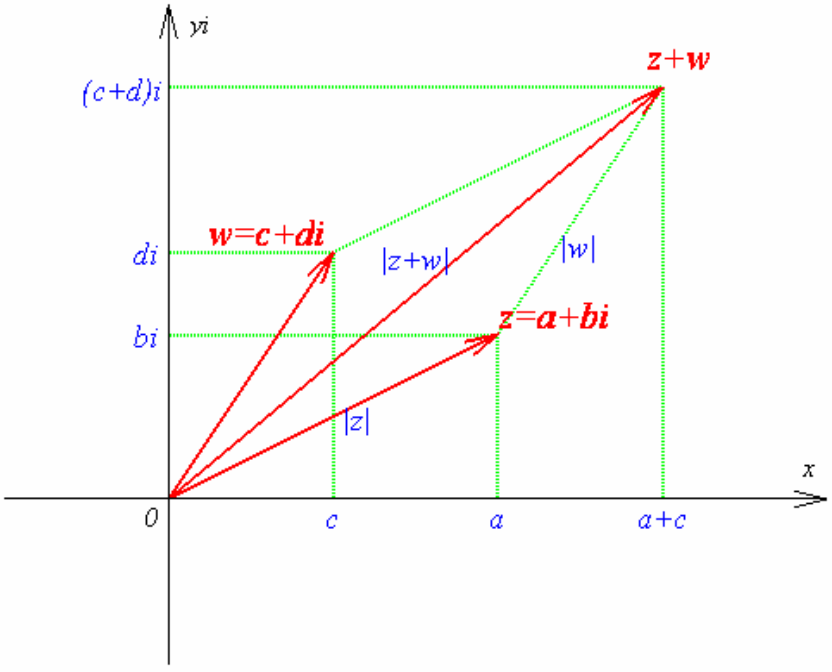


Рис. 32.2. Сложение комплексных чисел

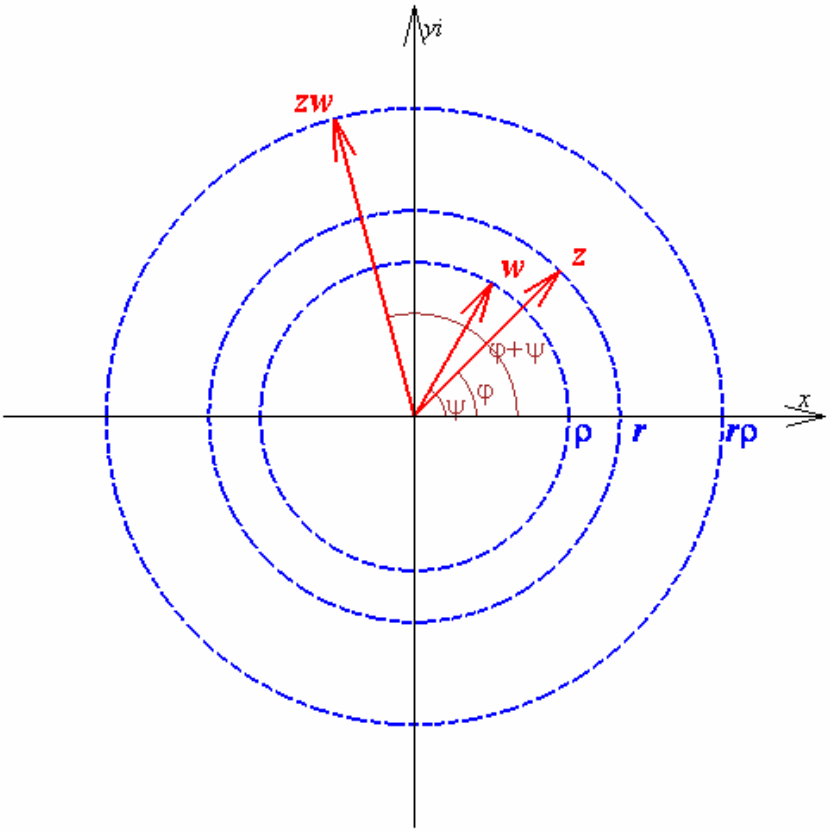


Рис. 32.3. Умножение комплексных чисел

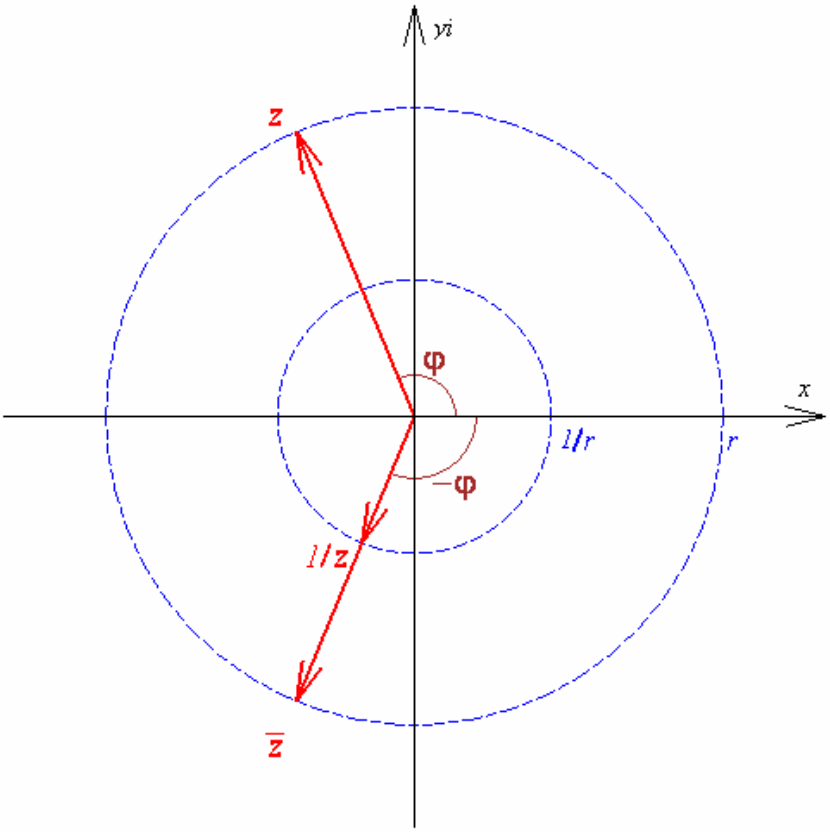


Рис. 32.4. Обращение комплексного числа

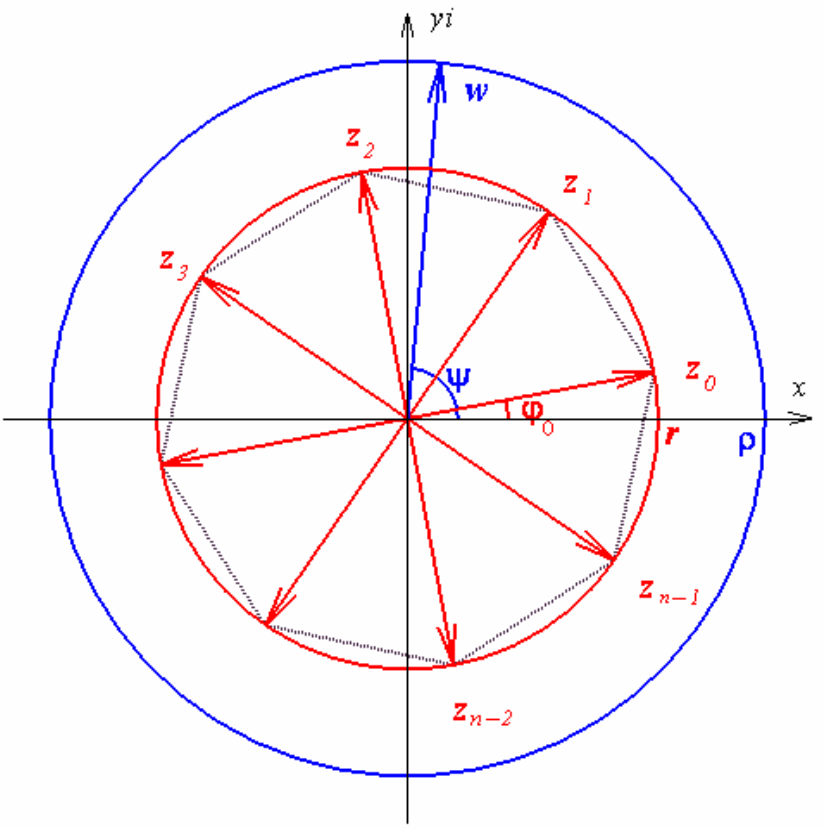


Рис. 32.5. Извлечение корня из комплексного числа

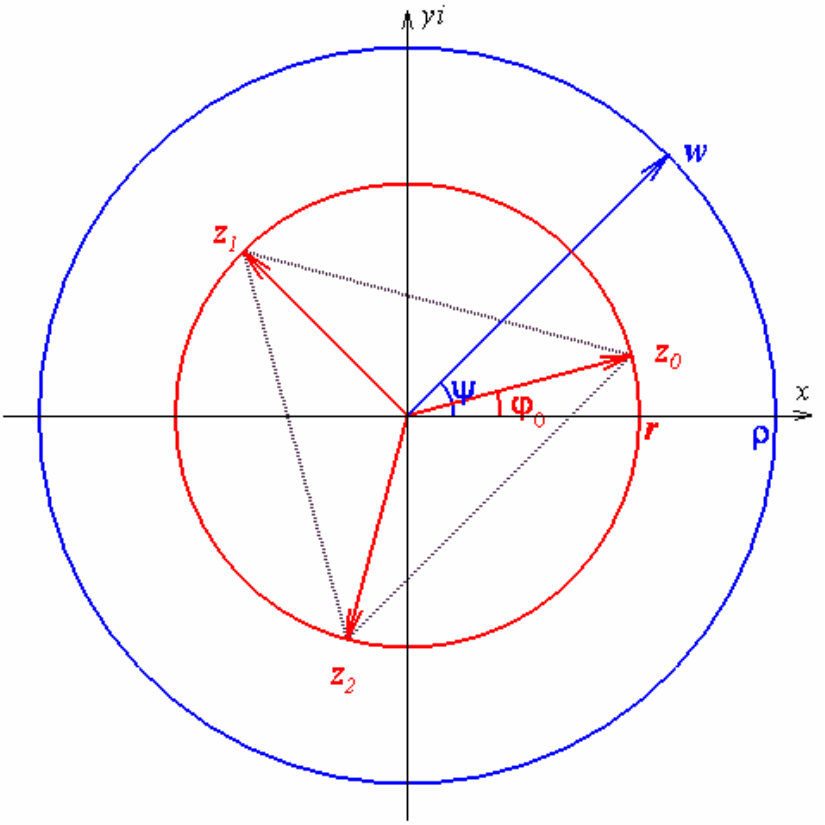


Рис. 32.6 (к примеру 32.5). Решения уравнения $z^3 = w$

$[w = \sqrt{2}(1+i);\ \rho = 2;\ \psi = \frac{\pi}{4};\ r = \sqrt[6]{2};\ \varphi_0 = \frac{\pi}{12}]$

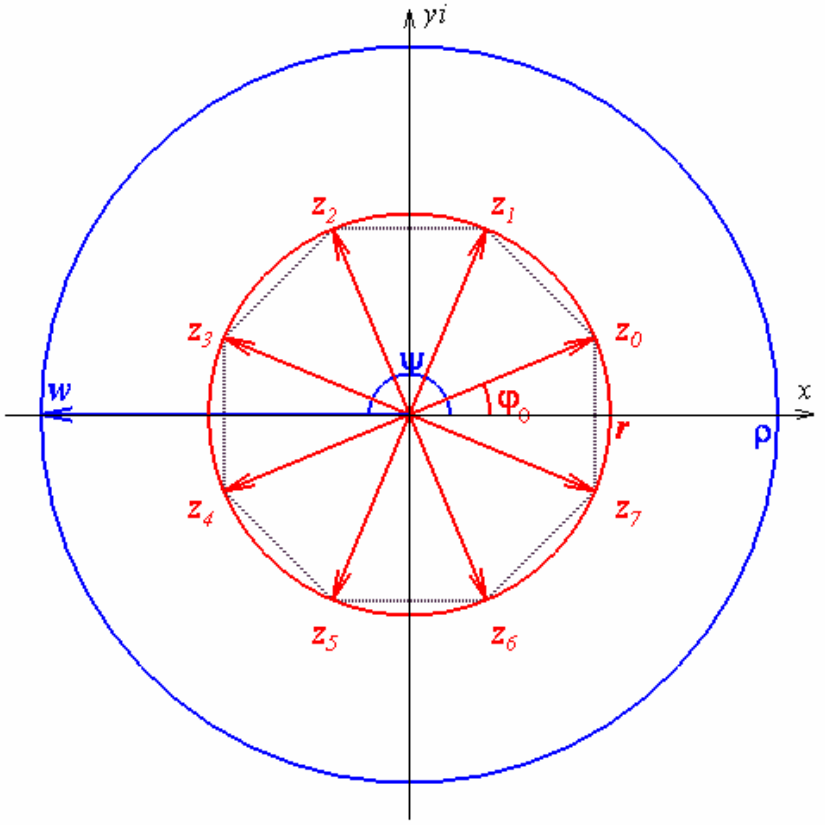


Рис. 32.7 (к примеру 32.6). Решения уравнения $z^8 = w$

$[\, w = -2\,;\ \rho = 2\,;\ \psi = \pi\,;\ r = \sqrt[8]{2}\,;\ \varphi_0 = \dfrac{\pi}{8}\,]$

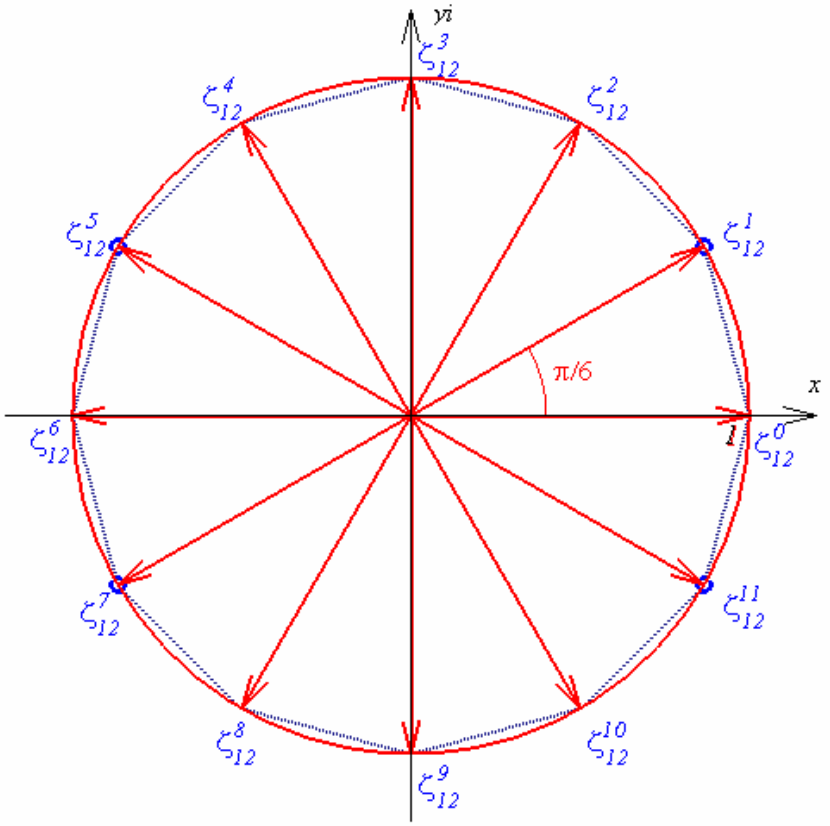


Рис. 33.1. Корни двенадцатой степени из единицы: выделены первообразные

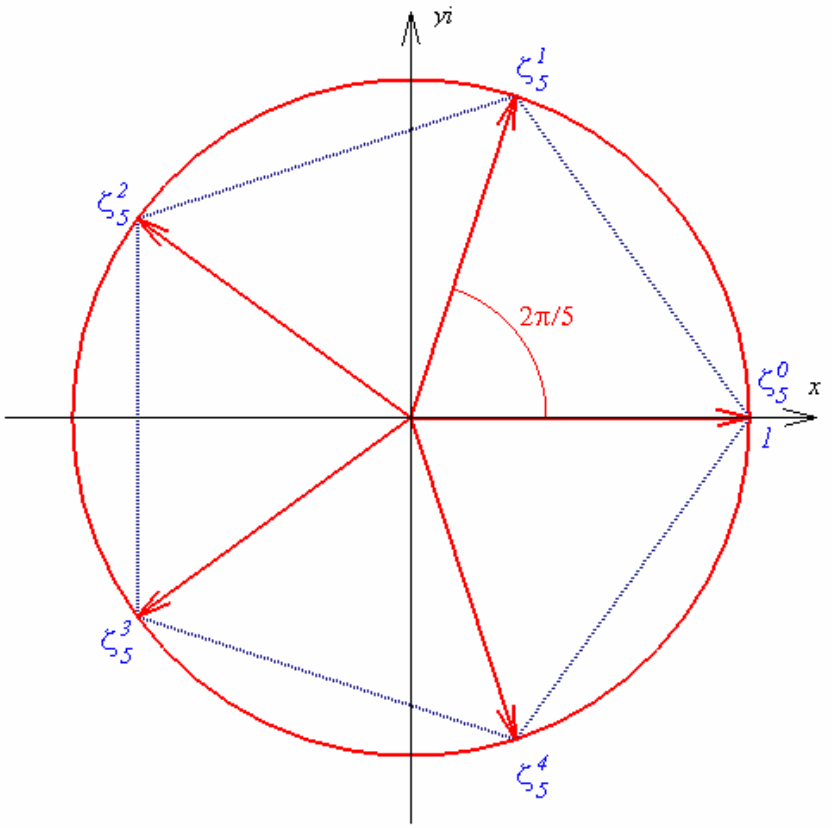


Рис. 33.2 (к примеру 33.1). Корни пятой степени из единицы

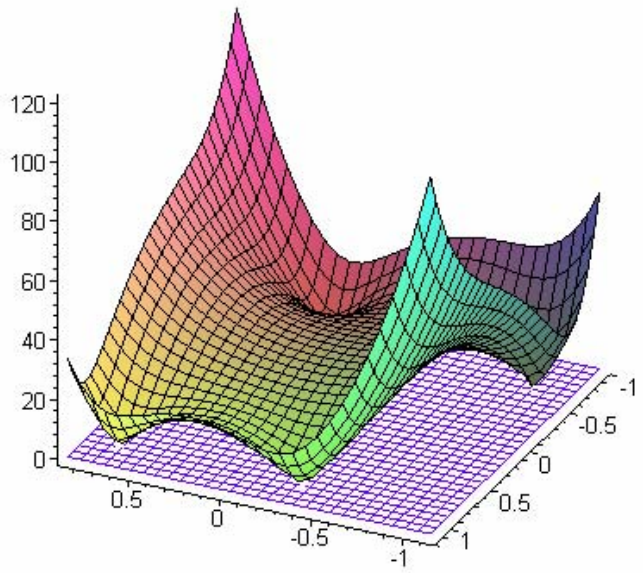


Рис 34.1. График модуля многочлена
$f(z)=(2+19i)z^4+(11+6i)z^3+(18+9i)z^2+(4+2i)z+(17+10i)$

Приложение 2

Таблицы к главе 6

Табл. 37.1а. **Деление многочленов с остатком: точное вычисление остатка и неполного частного**
(к примеру 37.1)

$6x^4$	$-2x^3$		$-3x$	$+2$	$2x^2$	$-x$	-3
$6x^4$	$-3x^3$	$-9x^2$			$3x^2$	$+\frac{1}{2}x$	$+\frac{19}{4}$
	x^3	$+9x^2$	$-3x$	$+2$			
	x^3	$-\frac{1}{2}x^2$	$-\frac{3}{2}x$				
		$\frac{19}{2}x^2$	$-\frac{3}{2}x$	$+2$			
		$\frac{19}{2}x^2$	$-\frac{19}{4}x$	$-\frac{57}{4}$			
			$\frac{13}{4}x$	$+\frac{65}{4}$			

Примечание. Многочлены заданы над кольцом целых чисел. Деление производится над полем рациональных чисел. Остаток: $r(x)=\frac{13}{4}x+\frac{65}{4}$; неполное частное: $q(x)=3x^2+\frac{1}{2}x+\frac{19}{4}$.

Табл. 37.1б. **Деление многочленов с остатком: определение остатка с точностью до пропорциональности**
(к примеру 37.1)

$6x^4$	$-2x^3$		$-3x$	$+2$	$2x^2$	$-x$	-3
$6x^4$	$-3x^3$	$-9x^2$			$3x^2$	$+x$	$+19$
	x^3	$+9x^2$	$-3x$	$+2$			
	$2x^3$	$+18x^2$	$-6x$	$+4$			
	$2x^3$	$-x^2$	$-3x$				
		$19x^2$	$-3x$	$+4$			
		$38x^2$	$-6x$	$+8$			
		$38x^2$	$-19x$	-57			
			$13x$	$+65$			

Примечание. Остаток: $r(x)=13x+65$; неполное частное искажено.

Табл. 38.1. **Вычисление НОД (60, 1008)**
(к примеру 38.1)

				1	0	0	8	6	0
					6	0		1	6
					4	0	8		
					3	6	0		
				6	0		4	8	
				4	8		1		
		4	8	1	2				
		4	8	4					
			0						

Примечание. $a=60;\ \ b=1008;\ \ q_1=16;\ r_1=48;\ \ q_2=1;\ r_2=12;\ \ q_3=4;\ r_3=0;$ $d=r_2=12$.

Табл. 38.2. **Вычисление НОД двух многочленов**
(к примеру 38.2)

				x^4	$+x^3$	$-3x^2$	$-4x$	-1	x^3	$+x^2$	$-x$	-1	
				x^4	$+x^3$	$-x^2$	$-x$		x				
						$-2x^2$	$-3x$	-1					
x^3	$+x^2$	$-x$	-1	$2x^2$	$+3x$	$+1$							
$2x^3$	$+2x^2$	$-2x$	-2	x	$+1$								
$2x^3$	$+3x^2$	$+x$											
	$-x^2$	$-3x$	-2										
	$2x^2$	$+6x$	$+4$										
	$2x^2$	$+3x$	$+1$										
		$3x$	$+3$										
$2x^2$	$+3x$	$+1$	x	$+1$									
$2x^2$	$+2x$		$2x$	$+1$									
	x	$+1$											
	x	$+1$											
		0											

Примечание. $f(x)=x^3+x^2-x-1;\ \ g(x)=x^4+x^3-3x^2-4x-1;$
$h_1(x)=2x^2+3x+1;\ \ h_2(x)=x+1;\ \ h_3(x)=0;\ \ d(x)=(f(x),g(x))=x+1$.

Табл. 41.1а. **Схема Горнера**
(общий вид)

	a_0	a_1	a_2	…	a_{n-2}	a_{n-1}	a_n
c	b_0	b_1	b_2	…	b_{n-2}	b_{n-1}	$f(c)$

Примечание. $b_0 = a_0;\ \ b_k = a_k + cb_{k-1};\ \ k = 1,...,n-1;\ \ f(c) = a_n + cb_{n-1}$.

Табл. 41.1б. **Схема Горнера**
(к примеру 41.1)

	2	*-1+2i*	*0*	*-2-3i*	*-4*
1+i	*2*	*1+4i*	*-3+5i*	*-10-i*	*-13-11i*

Примечание. $f(x) = 2x^4 - (1-2i)x^3 - (2+3i)x - 4;\ f(1+i) = -13-11i$.

Табл. 41.2а. **Разложение многочлена по степеням** $x-c$
(общая схема)

	a_0	a_1	a_2	…	a_{n-2}	a_{n-1}	a_n
c	$b_0 = a_0$	b_1	b_2	…	b_{n-2}	b_{n-1}	h_0
c	$c_0 = a_0$	c_1	c_2	…	c_{n-2}	h_1	
c	$d_0 = a_0$	d_1	d_2	…	h_2		
…	…	…	…	⋰			
c	$f_0 = a_0$	f_1	h_{n-2}				
c	$g_0 = a_0$	h_{n-1}					
c	$h_n = a_0$						

Примечание.

$$f(x) = a_0x^n + a_1x^{n-1} + ... + a_{n-1}x + a_n = h_0 + h_1x + ... + h_{n-1}x^{n-1} + h_nx^n\,.$$

Табл. 41.2б. **Разложение многочлена по степеням** $x-c$
(к примеру 41.2)

	8	−44	50	55	−55	−56	−12
2	8	−28	−6	43	31	6	0
2	8	−12	−30	−17	−3	0	
2	8	4	−22	−61	−125		
2	8	20	18	−25			
2	8	36	90				
2	8	52					
2	8						

Примечание. По выделенной диагонали идут коэффициенты Тейлора (в порядке от младших к старшим). Разложение:

$$f(x)=8x^6-44x^5+50x^4+55x^3-55x^2-56x-12=$$
$$=-125(x-2)^2-25(x-2)^3+90(x-2)^4+52(x-2)^5+8(x-2)^6.$$

Табл. 41.3. **Кратность корня**
(к примеру 41.3)

	A	B	0	0	1
1	A	A+B	A+B	A+B	A+B+1
1	A	2A+B	3A+2B	4A+3B	
1	A	3A+B	6A+4B		

Примечание. Многочлен $f(x)=Ax^4+Bx^3+1$ делится на $(x-2)^2$ тогда и только тогда, когда число 2 является его корнем кратности, не меньшей двух:

$$(x-2)^2 \,|\, f(x) \Leftrightarrow (A=3)\wedge(B=-4).$$

Табл. 42.1а. **Отыскание рациональных корней многочлена, подготовительный этап: проверка чисел ± 1**
(к примеру 42.1)

	12	−4	−17	−18	37	0	−14	4
1	12	8	−9	−27	10	10	−4	0
1	12	20	11	−16	−6	4	0	
1	12	32	43	27	21	25		
−1	12	8	3	−19	13	−9		

Примечание. Отыскание корней ± 1, если они есть:

$$f(x)=12x^7-4x^6-17x^5-18x^4+37x^3-14x+4;$$

$$f(x)=(x-1)^2 g(x);\ \ g(x)=12x^5+20x^4+11x^3-16x^2-6x+4.$$

Выделены "сработавшие" строки и значения $g(1)$ и $g(-1)$.

Табл. 42.1б. **Отыскание рациональных корней многочлена, определение всех рациональных корней**
(к примеру 42.1)

	12	20	11	−16	−6	4
2	12	44	99	182	358	720
−4	12	−28	123	−508	2026	−8100
1/2	12	26	24	−4	−8	0
1/2	12	32	40	16	0	
1/2	12	38	59	91/2		
−2/3	12	24	24	0		
−2/3	12	16	40/3			
−1/4	12	21	75/4			

Примечание. Определение всех рациональных корней многочлена $g(x)$ (с предварительным тестированием "подозрительных" дробей):

$$g(x)=12\left(x-\frac{1}{2}\right)^2\left(x+\frac{2}{3}\right)(x^2+2x+2);$$

$$f(x)=12(x-1)^2\left(x-\frac{1}{2}\right)^2\left(x+\frac{2}{3}\right)(x^2+2x+2);$$

выделены "сработавшие" строки.

Табл. 49.1. **Выражение многочлена $\Delta_3(x)$ через элементарные симметрические многочлены**
(к примеру 49.4)

x_1	x_2	x_3	$f(x)=\sigma_1^2\sigma_2^2+A_2\sigma_1^3\sigma_3+A_3\sigma_2^3+A_4\sigma_1\sigma_2\sigma_3+A_5\sigma_3^2$	σ_1	σ_2	σ_3
1	1	0	$0=4+A_3$	2	1	0
2	–1	–1	$0=-27A_3+4A_5$	0	–3	2
1	–2	–2	$0=-108A_2+16A_5$	–3	0	4
1	1	–1	$0=1-A_2-A_3+A_4+A_5$	1	–1	–1

Примечание. $f(x)=\Delta_3(x)=(x_2-x_1)^2(x_3-x_1)^2(x_3-x_2)^2$.

Список рекомендуемой литературы

Основной

1. *Кострикин А. И.* Введение в алгебру. М.: Наука, 1977.
2. *Кострикин А. И.* Введение в алгебру. Ч. I. Основы алгебры. М.: Физматлит, 2000.
3. *Курош А. Г.* Курс высшей алгебры. М.: Наука, 1973.
4. Сборник задач по алгебре / Под ред. А. И. Кострикина. М.: Факториал, 1995.

Дополнительный

5. *Винберг Э. Б.* Курс алгебры. М.: Факториал Пресс, 2001.
6. *Говорухин В., Цибулин В.* Компьютер в математическом исследовании: Maple, MATLAB, LaTeX: Учеб. курс. СПб.: Питер, 2001.
7. *Куликов Л. Я.* Алгебра и теория чисел. М.: Высш. шк., 1979.
8. *Куликов Л. Я., Москаленко А. И., Фомин А. А.* Сборник задач по алгебре и теории чисел. М.: Просвещение, 1993.
9. *Фаддеев Д. К.* Лекции по алгебре. М.: Наука, 1984.
10. *Фаддеев Д. К., Соминский И. С.* Сборник задач по высшей алгебре. М.: Наука, 1977.

ЯЦКИН Николай Иванович

А Л Г Е Б Р А

ТЕОРЕМЫ И АЛГОРИТМЫ

Учебное пособие

Директор издательства *Л. В. Михеева*
Редактор *О. В. Батова*
Технический редактор *И. С. Сибирева*

Подписано в печать 07.06.2006.
Формат 60 × 84$^{1}/_{16}$. Бумага офсетная. Печать офсетная.
Усл. печ. л. 29,53. Уч.-изд. л. 29,5. Тираж 150 экз.

Издательство «Ивановский государственный университет»
✉ 153025 Иваново, ул. Ермака, 39
☎ (4932) 35-63-81 E-mail: publisher@ivanovo.ac.ru

Типография ГОУ СПО «Ивановский энергетический колледж»
✉ 153025 Иваново, ул. Ермака, 41
☎ (4932) 37-52-44, 32-50-89