

Р. Грегори  
Е. Кришнамурти

# Безошибочные вычисления Методы и приложения



# Безошибочные вычисления

## Методы и приложения

Methods and Applications of  
Error-Free Computation

**R. T. Gregory**  
**E. V. Krishnamurthy**

Springer-Verlag  
New York Berlin Heidelberg Tokyo  
1984

Р. Грегори,  
Е. Кришнамурти

# Безошибочные вычисления

## Методы и приложения

*Перевод с английского*  
Х. Д. Икрамова и А. В. Князева  
*под редакцией*  
Х. Д. Икрамова



Москва «Мир»  
1988

ББК 22.193

Г79

УДК 518.5

Грегори Р., Кришнамурти Е.

Г79 Безошибочные вычисления. Методы и приложения:  
Пер. с англ. — М.: Мир, 1988. — 208 с., ил.

ISBN 5-03-001145-5

Книга известных специалистов по численным методам (США, Индия), посвященная методам и приложениям нового раздела вычислительной математики, так называемым «безошибочным вычислениям». Изложение отличается ясностью стиля и многообразием примеров. Книга может использоваться как учебное пособие.

Для специалистов по вычислительной алгебре и численным методам, для аспирантов и студентов университетов.

Г  $\frac{1702070000-174}{041(01)-88}$  32—88, ч. 1

ББК 22.193

*Редакция литературы по математическим наукам*

---

Научное издание

Роберт Тодд Грегори, Е. В. Кришнамурти

БЕЗОШИБОЧНЫЕ ВЫЧИСЛЕНИЯ.  
МЕТОДЫ И ПРИЛОЖЕНИЯ

Заведующий редакцией доктор физ.-мат. наук

профессор Б. В. Щабат

Зам. зав. редакцией А. С. Попов

Ст. науч. редакторы А. А. Бряндинская, Г. М. Ильичева

Художник А. А. Гейнце

Художественный редактор В. И. Шаповалов

Технический редактор О. Г. Лапко

Корректор С. А. Денисова

ИБ № 6413

Сдано в набор 13.05.87. Подписано к печати 29.03.88.

Формат 60 × 90<sup>1</sup>/<sub>16</sub>. Бумага книжно-журнальная. Печать высокая.

Гарнитура литературная.

Объем 6,50 бум. л. Усл. печ. л. 13,00. Усл. кр.-отг. 13,25. Уч.-изд. л. 10,05.

Изд. № 1/5368. Тираж 10000 экз. Зак. 689. Цена 1 р. 60 к.

ИЗДАТЕЛЬСТВО «МИР»

129820, ГСП, Москва, И-110, 1-й Рижский пер., 2

Ленинградская типография № 2 головное предприятие ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгения Соколовой Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 198052, г. Ленинград, Л-52, Измайловский проспект, 29.

---

ISBN 5-03-001145-5 (русск.) © Springer-Verlag New York Inc. 1984  
ISBN 0-387-90967-2 (англ.) All rights reserved. Authorized translation from English language edition published by Springer-Verlag Berlin Heidelberg New York Tokyo.

© перевод на русский язык, «Мир», 1988

## Предисловие редактора перевода

Эта книга посвящена сравнительно молодому разделу вычислительной математики, несколько двусмысленно названному авторами «безошибочными вычислениями». Впрочем, этот термин, как и его эквивалент — «точные вычисления» (exact computation), установился в иноязычной литературе. Теория безошибочных вычислений имеет дело с задачами, для которых входная информация представима набором целых чисел (или многочленов с целыми коэффициентами), а решение является рациональной функцией от этих чисел (или многочленов). К задачам такого типа относятся обращение (в обычном или обобщенном смысле) и построение характеристического многочлена целочисленной матрицы, а также решение линейной системы с целыми коэффициентами. Несколько систем безошибочных вычислений, описываемых в книге, нацелены на получение точных ответов в подобного рода задачах.

Роберт Грегори, недавно скончавшийся американский математик, известен алгебраистам прежде всего как один из авторов очень полезного «Собрания матриц для тестирования численных алгоритмов (1969 г., изд-во Wiley — Interscience; переиздание — 1978 г., изд-во Krieger). Вероятно, работа над «Собранием» и привела Грегори к безошибочным вычислениям; во всяком случае, первые его публикации в этой области появились как раз в конце 60-х г. Второй автор книги Е. Кришнамурти, сотрудник отделения прикладной математики научного института в г. Бангалоре (Индия), пришел к безошибочным вычислениям, занимаясь исследованием быстрых методов выполнения арифметических операций.

Краткий очерк модульной (говорят также «модулярной») арифметики дан в § 4.3.2 второго тома известной энциклопедии Д. Кнута «Искусство программирования для ЭВМ»<sup>1)</sup>. Первая глава данной книги содержит в основном тот же материал, излагаемый, однако, на существенно сниженном уровне, а именно уровне элементарного учебника. Кроме того, здесь показано, как применять модульную арифметику к ра-

---

<sup>1)</sup> Кнут Д. Искусство программирования для ЭВМ, Т. 2. Получисленные алгоритмы. — Пер. с англ. М.: Мир, 1977.

циональным дробям с ограниченными величинами числителей и знаменателей, так называемым дробям Фарея. Во второй главе изучается другая система безошибочных вычислений — конечно-разрядная  $p$ -адическая арифметика. В гл. III—VI обсуждаются приложения безошибочных вычислений к перечисленным выше задачам линейной алгебры, а также к целочисленному программированию и уравниванию химических реакций.

В последние годы интерес к методам безошибочных вычислений заметно возрос. В той же шпрингеровской серии «Text and monographs in computer science» в 1985 г. выпущена новая книга Е. Кришнамурти «Безошибочные вычисления с полиномиальными матрицами». В Алма-Ате издана книга В. М. Амербаева, И. Т. Пака «Параллельные вычисления в комплексной плоскости» (Алма-Ата: Наука, 1984). Часто появляются журнальные публикации. Все это связано с освоением сверхмощных компьютеров нового поколения и внутренним параллелизмом одной из систем безошибочных вычислений — многомодульной арифметики. Однако на этих вопросах авторы не останавливаются, ограничивая себя задачей написания вводного учебного пособия. Но и такое пособие полезно, позволяя читателю быстро войти в область, по-видимому, перспективную, но содержащую пока больше вопросов, чем ответов. Итогом ее разработки может стать существенное ускорение алгебраических алгоритмов и притом без вечной проблемы численной устойчивости.

Первые две главы книги переведены А. В. Князовым, остальные — мной.

*Х. Икрамов*



## Предисловие

Эта книга написана как введение в теорию вычислений без ошибок. Вдобавок в нее включено несколько глав, иллюстрирующих возможные практические приложения таких вычислений. Книга адресована студентам-старшекурсникам и начинающим аспирантам, специализирующимся в области машинных вычислений научного характера, а также тем, кто работает в этой области и желает получить начальные сведения о предмете.

Написать эту книгу нас побудило то обстоятельство, что существуют как обширные классы плохо обусловленных задач, так и численно неустойчивые алгоритмы. И в том и в другом случае при вычислении решений соответствующих задач нельзя допускать ошибок округления. Поэтому важно исследовать конечные системы машинных чисел, позволяющие проводить вычисления без ошибок округлений.

В гл. I мы обсудим числовые системы вычетов по одному модулю и по многим модулям, а также арифметику в этих системах; операндами могут быть целые или рациональные числа. В гл. II изучаются системы конечно-разрядных  $p$ -адических чисел и их связь с  $p$ -адическими числами Гензеля [Hensel, 1908]. Каждому рациональному числу из некоторого конечного множества приписывается единственный код Гензеля. Арифметические операции над кодами Гензеля математически эквивалентны одноименным операциям над соответствующими рациональными числами. Арифметика конечно-разрядных  $p$ -адических чисел, так же как и арифметика вычетов, свободна от ошибок округления.

В гл. III—VI показано, как вычисления без ошибок могут быть применены к получению точных решений некоторых классов задач вычислительной линейной алгебры.

Для понимания материала книги, как правило, достаточно знакомства с элементарной теорией чисел и простейшими численными алгоритмами; кроме того, нужно владеть (для понимания гл. III—VI. — *Ред.*) курсом линейной алгебры.

Римские цифры используются для нумерации глав, арабские — для нумерации параграфов каждой главы и нумерации внутри параграфов. В последнем случае уравнения, теоремы, следствия, таблицы, рисунки и т. д. нумеруются подряд как

элементы единого списка. Например, пятнадцатый элемент § 2 (любой главы) нумеруется как 2.15. Ссылаться на этот элемент можно по-разному. Если, например, речь идет об *уравнении*, пишем (2.15). Для элемента любого другого типа, скажем *леммы*, пишем «лемма 2.15». Если лемма содержится в гл. V, а ссылка на нее делается в другой главе, то пишем лемма 2.15 гл. V; в тексте же самой гл. V ссылаемся на лемму 2.15 без дополнительных пояснений.

В заключение мы хотели бы с признательностью отметить вклад всех, кто тем или иным образом повлиял на книгу. Это наши бывшие ученики Джо Энн Хауэлл, Т. М. Рао, К. Субраманиан, Рут Энн Льюис, Шу Хуахуан и Джон Смайр; Петер Корнеруп, который доказал теорему 6.39 гл. I, указал эффективный путь реализации обратного отображения и предоставил первому автору прекрасные возможности для научной работы в ходе двух его визитов в Орхус; Гермунд Дальквист, гостеприимный хозяин первого автора во время его визита в Стокгольм; Азриэль Розенберг, помогавший второму автору при многочисленных посещениях университета Мэриленда; Дэвид Матула и Карл Грегори, внесшие важный вклад в § 7 гл. I; Арне Франсен, который очень внимательно прочел всю рукопись и сделал много полезных предложений по ее улучшению. Мы признательны Университету Теннесси в Ноксвилле, Индийскому министерству образования и культуры и Индийскому институту науки в Бангалоре за финансовую помощь при подготовке рукописи. Мы благодарим Сюзи Уитинберджер за великолепную перепечатку рукописи, а также наших жен, которым посвящена эта книга.

Август 1983

Р. Т. Грегори,  
Е. В. Кришнамурти

## Список обозначений

$A^T$	транспонированная к $A$
$A^*$	сопряженная к $A$
$A^{-1}$	обратная для $A$
$A^-$	$g$ -обратная для $A$
$A_R^-$	рефлексивная $g$ -обратная для $A$
$A_L^-$	$g$ -обратная для $A$ со свойством наименьших квадратов
$A_M^-$	$g$ -обратная для $A$ со свойством минимальной нормы
$A^+$	$g$ -обратная Мура — Пенроуза для $A$
$A_I^-$	$A_R^-$ , если $AA_R^- \in \mathbb{I}^{mm}$ и $A_R^-A \in \mathbb{I}^{nn}$
$ A _m$	если $A = (a_{ij})$ , то $ A _m = ( a_{ij} _m)$
$ a _m$	неотрицательный вычет числа $a$ по модулю $m$
$ a/b _m$	симметричный вычет числа $a$ по модулю $m$
$ a/b _m$	$a \cdot b^{-1}(m) _m$
$ a _\beta$	$[ a _{m_1},  a _{m_2}, \dots,  a _{m_n}]$
$ a/b _\beta$	$[ a/b _{m_1},  a/b _{m_2}, \dots,  a/b _{m_n}]$
$\left \frac{a}{b}\right _\beta$	$\left[\left \frac{a}{b}\right _{m_1}^*, \left \frac{a}{b}\right _{m_2}^*, \dots, \left \frac{a}{b}\right _{m_n}^*\right]$
$a^{-1}(m)$ или $a^{-1}$	обратный (по модулю $m$ ) элемент для $a$
$a^{-1}(\beta)$	$[a^{-1}(m_1), a^{-1}(m_2), \dots, a^{-1}(m_n)]$
$\langle a \rangle_\rho$	$\langle d_0, d_1, \dots, d_{n-1} \rangle$ , цифры смешанного представления для $a$
$\beta$	$[m_1, m_2, \dots, m_n]$ , векторное основание
$\rho$	$[r_1, r_2, \dots, r_n]$ , основания смешанного представления
$\mathbb{F}_N$	множество дробей Фарея порядка $N$
$(a, b)$	наибольший общий делитель $a$ и $b$
$[a/b]$	целая часть частного $a/b$
$\ \alpha\ _\rho$	$p$ -адическая норма числа $\alpha$
$H(p, r, \alpha)$	обычный код Гензеля числа $\alpha$
$\hat{H}(p, r, \alpha)$	код Гензеля с плавающей точкой числа $\alpha$
$\mathbb{I}$	множество целых чисел
$\mathbb{I}_m$	$\{0, 1, 2, \dots, m-1\}$
$\tilde{\mathbb{I}}_m$	$\{ a/b _m: a/b \in \mathbb{F}_N\}$
$\mathbb{I}_\beta$	$\{ s _\beta: s \in \mathbb{I}\}$

---

$\Pi^m$	множество $m$ -мерных векторов над $\Pi$
$\Pi^{mn}$	множество матриц размера $m \times n$ над $\Pi$
$\Pi_r^{mn}$	множество матриц размера $m \times n$ и ранга $r$ над $\Pi$
$\mathbb{Q}$	множество рациональных чисел
$\mathbb{Q}_p$	множество $p$ -адических чисел
$\mathbb{R}$	множество вещественных чисел
$S_m$	$\{-(m-1)/2, \dots, -2, -1, 0, 1, 2, \dots, (m-1)/2\}$
$S_\beta$	$\{s/\beta: s \in \Pi\}$
$\text{tr } A$	след матрицы $A$
$\tilde{\mathbb{T}}_\beta$	$\{ a/b _\beta: a/b \in \mathbb{F}_N\}$

# Глава I

## Арифметика вычетов или модульная арифметика<sup>1)</sup>

### § 1. Введение

Автоматический цифровой компьютер является конечной машиной: он способен представлять, по сути дела, только конечное множество чисел. Таким образом, обречена на неудачу любая попытка использовать его для выполнения арифметических операций в поле вещественных чисел  $(\mathbb{R}, +, \cdot)$ , поскольку  $\mathbb{R}$  — бесконечное множество, большинство элементов которого непредставимо в вычислительной машине.

Сказанное не означает, однако, что нельзя пытаться *аппроксимировать* на компьютере арифметику в  $(\mathbb{R}, +, \cdot)$ . Часто для такой аппроксимации используется множество  $\mathbb{F}$  так называемых чисел с плавающей точкой (более подходящее название  $\mathbb{F}$  — множество *машинных чисел*). Множество  $\mathbb{F}$  является частью множества вещественных чисел со следующими свойствами.

(а)  $\mathbb{F}$  — конечное подмножество множества рациональных чисел  $\mathbb{Q}$ .

(б)  $\mathbb{F}$ , как правило, симметрично по отношению к числу нуль; в этом случае имеются два машинных представления для нуля.

(в) элементы  $\mathbb{F}$  распределены неравномерно на вещественной прямой. Интервал между двумя «соседними» машинными числами очень мал вблизи нуля, а при удалении от него постепенно увеличивается. Интервал между максимальным возможным машинным числом и соседним с ним очень велик (см., например, [Форсайт, Малькольм, Моулер, 1980, с. 24] и [Gregory, 1980, с. 9]).

(г) Почти все «привычные» рациональные числа исключены из  $\mathbb{F}$ . К примеру, на двоичной машине единственными кандидатами на принадлежность к  $\mathbb{F}$  являются рациональные числа вида  $p/q$ , где  $q$  — степень двойки. Таким образом, числа типа  $\frac{1}{10}$ ,  $\frac{1}{3}$ ,  $\frac{5}{6}$  и  $\frac{2}{7}$  в  $\mathbb{F}$  не входят.

(д) Система  $(\mathbb{F}, +, \cdot)$  не будет полем (главным образом из-за того, что нет замкнутости относительно обеих указанных бинарных операций).

---

<sup>1)</sup> В советской литературе наряду с термином «модульная арифметика» иногда используется термин «модулярная арифметика», — *Прим. перев.*

Итак, нет возможности сколько-нибудь детально отобразить континуум вещественных чисел. «Практичный» выход из этой трудности во многих случаях состоит в представлении вещественного числа  $x$  ближайшим к нему машинным числом  $\hat{x}$ ; тем самым вводится *ошибка округления*

$$(1.1) \quad \varepsilon = x - \hat{x}.$$

Из-за отсутствия замкнутости ошибки округления возникают также в результате арифметических операций над элементами  $\mathbb{F}$ . Например, если  $\hat{x}$  и  $\hat{y}$  — два «соседних» элемента  $\mathbb{F}$ , то число

$$(1.2) \quad z = \frac{\hat{x} + \hat{y}}{2}$$

уже не принадлежит  $\mathbb{F}$ . Его следует заменить на  $\hat{z}$  — элемент в  $\mathbb{F}$ , «ближайший» к  $z$ . В этом примере  $\hat{z}$  совпадает либо с  $\hat{x}$ , либо с  $\hat{y}$ .

Может показаться, что эффект неточного выполнения арифметических операций (и ошибок округления) не слишком серьезен. Однако хорошо известно, что вычисленное<sup>1)</sup> решение может быть интерпретировано, как точное решение слабо возмущенной задачи (такой взгляд характеризует *обратный анализ ошибок*; см., например, [Young, Gregory, 1972, с. 10]). Кроме того, существует класс задач, называемых *плохо обусловленными*, для которых решение (точное) предельно чувствительно к «малым» возмущениям данных. При решении таких задач эффект ошибок округления может быть катастрофическим.

К примеру, рассмотрим задачу вычисления определителя матрицы

$$(1.8) \quad A = \begin{bmatrix} -73 & 78 & 24 \\ & 92 & 66 & 25 \\ & & -80 & 37 & 10 \end{bmatrix}.$$

Плохая обусловленность этой задачи (см. [Gregory, Karney, 1978, с. 50]) видна из следующего обстоятельства. Пусть

$$(1.4) \quad E = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 10^{-2} \end{bmatrix}.$$

Тогда

$$(1.5) \quad \det(A) = 1,$$

в то время как

$$(1.6) \quad \det(A + E) = -118.94.$$

<sup>1)</sup> В этом обсуждении предполагается, что решение вычисляется посредством неточных арифметических операций над элементами  $\mathbb{F}$ .

Другими словами, если накопление ошибок округления соответствовало введению возмущающей матрицы  $E$ , то численное значение определителя будет равно  $-118.94$ , тогда как точное значение есть 1.

Попытки аппроксимировать при помощи конечного множества  $\mathbb{F}$  арифметику поля вещественных чисел  $(\mathbb{R}, +, \cdot)$  приводят к трудностям. Ошибки округления при решении плохо обусловленных задач вызывают неприятные эффекты. По этим причинам имеются серьезные поводы для использования числовых систем с арифметическими операциями, которые можно выполнить точно. Примером такой системы является система вычетов.

## § 2. Арифметика в одномодульной системе вычетов

Хорошо известно, что автоматические цифровые компьютеры могут выполнять некоторые арифметические операции точно, если операнды суть целые числа. Это побуждает рассмотреть целочисленную арифметику как средство избежать ошибок округления в надежде, что некоторые плохо обусловленные задачи станут решаться точно. Если мы оперируем с целыми числами и приводим результаты по модулю  $m$ , то такая арифметическая система называется *одномодульной арифметикой вычетов*<sup>1)</sup>. Целое число  $m > 1$  при этом называется *модулем* арифметической системы.

### Теоретические основы

Большую часть последующего материала составляет элементарная теория чисел. Будет предполагаться, что читатель знаком с простыми идеями этой теории (см., например, [Hardy, Wright, 1960]). Однако мы дадим краткий обзор нужных фактов, причем многие теоремы будут приводиться без доказательства. Большинство этих результатов, с акцентом на приложения к вычислительным машинам, можно найти в [Suabó, Tanaka, 1967; Young, Gregory, 1973, гл. 13]. Символ  $\mathbb{I}$  будет обозначать множество целых чисел на протяжении всей книги.

**2.1. Определение**<sup>2)</sup>. Пусть  $a, b$  и  $m \in \mathbb{I}$ ,  $m > 1$ . Если  $m$  является делителем  $b - a$ , то говорят, что  $b$  *сравнимо*<sup>3)</sup> с  $a$

<sup>1)</sup> В оригинале single-modulus residue arithmetic. — *Прим. перев.*

<sup>2)</sup> Использование в качестве модулей отрицательных целых чисел не имеет каких-либо преимуществ, поскольку  $b - a$  делится на  $-m$  тогда и только тогда, когда делится на  $m$ .

<sup>3)</sup> В оригинале is congruent. — *Прим. перев.*

по модулю  $m$ , и пишут

$$b \equiv a \pmod{m}.$$

В противном случае говорят, что число  $b$  не сравнимо с  $a$  по модулю  $m$ , и пишут

$$b \not\equiv a \pmod{m}.$$

Соотношения типа  $a \equiv b \pmod{m}$  называются *сравнениями*<sup>1)</sup>. Они имеют несколько свойств, перечисляемых в последующих теоремах, которые приводятся без доказательства.

**2.2. Теорема.** Пусть  $a, b, m \in \mathbb{I}$ ,  $m > 1$ . Тогда следующие три утверждения эквивалентны:

- (i)  $a \equiv b \pmod{m}$ ,
- (ii)  $b \equiv a \pmod{m}$ ,
- (iii)  $a - b \equiv 0 \pmod{m}$ ,

**2.3. Теорема.** Пусть  $a, b, c, m \in \mathbb{I}$ ,  $m > 1$ . Если

$$\begin{aligned} & a \equiv b \pmod{m} \\ \text{и} & \\ & b \equiv c \pmod{m}, \\ \text{то} & \\ & a \equiv c \pmod{m}. \end{aligned}$$

**2.4. Теорема.** Пусть  $a, b, c, d, x, y, m \in \mathbb{I}$ ,  $m > 1$ . Если

$$\begin{aligned} & a \equiv b \pmod{m} \\ \text{и} & \\ & c \equiv d \pmod{m}, \\ \text{то} & \\ & ax + cy \equiv bx + dy \pmod{m}. \end{aligned}$$

**2.5. Теорема.** Пусть  $a, b, c, d, m \in \mathbb{I}$ ,  $m > 1$ . Если

$$\begin{aligned} & a \equiv b \pmod{m} \\ \text{и} & \\ & c \equiv d \pmod{m}, \\ \text{то} & \\ & ac \equiv bd \pmod{m}. \end{aligned}$$

В *одномодульной* арифметике вычетов каждое целое  $b \in \mathbb{I}$  отображается на целое  $r$  в конечном множестве

$$(2.6) \quad \mathbb{I}_m = \{0, 1, 2, \dots, m-1\}.$$

<sup>1)</sup> В оригинале congruences. — Прим. перев.



Число  $r$  называется *наименьшим неотрицательным вычетом*<sup>1)</sup> числа  $b$  по модулю  $m$ . Отображение описывается следующим образом.

**2.7. Определение.** Отображение  $|\cdot|: \Pi \rightarrow \Pi_m$  определяется равенством

$$|b|_m = r.$$

Здесь  $0 \leq r < m$  и

$$b \equiv r \pmod{m}.$$

Легко видеть, что это отображение задает разбиение  $\Pi$  на  $m$  непересекающихся подмножеств  $\mathbb{R}_0, \mathbb{R}_1, \dots, \mathbb{R}_{m-1}$ , называемых *классами вычетов*, где

$$(2.8) \quad \mathbb{R}_k = \{b \in \Pi: |b|_m = k\}.$$

**2.9. Пример.** Если  $m = 7$ , то  $58 \in \mathbb{R}_2$ , поскольку

$$|58|_7 = 2.$$

Будем говорить, что число 2 есть *наименьший неотрицательный вычет* числа 58 по модулю 7 или что 58 *приведено к 2 по модулю 7*.

В следующих двух теоремах демонстрируются простые свойства введенного отображения.

**2.10. Теорема.** Пусть  $a, b, m \in \Pi$ ,  $m > 1$ . Тогда

- (i)  $|a|_m$  определяется единственным образом;
- (ii)  $|a|_m = |b|_m$  тогда и только тогда, когда  $a \equiv b \pmod{m}$ ;
- (iii)  $|km|_m = 0$  для любого  $k \in \Pi$ .

**2.11. Теорема.** Пусть  $a, b, m \in \Pi$ ,  $m > 1$ . Тогда

- (i)  $|a + b|_m = ||a|_m + |b|_m|_m = ||a|_m + |b|_m|_m = |a + |b|_m|_m$ ;
- (ii)  $|ab|_m = ||a|_m|b|_m|_m = ||a|_m|b|_m|_m = |a|b|_m|_m$ .

Отметим, что теорема 2.11 показывает, как выполнять сложение и умножение по модулю  $m$ . Ясно, что имеется полная свобода выбора места, где операнд приводится по модулю  $m$ . Например, есть четыре эквивалентных способа<sup>2)</sup> подсчета суммы 24 и 38 по модулю 3:

- (i)  $|24 + 38|_3 = |62|_3 = 2$ ;
- (ii)  $|24 + 38|_3 = |0 + 2|_3 = 2$ ;
- (iii)  $|24 + 38|_3 = |0 + 38|_3 = 2$ ;
- (iv)  $|24 + 38|_3 = |24 + 2|_3 = 2$ .

<sup>1)</sup> В оригинале least non-negative residue. — Прим. перев.

<sup>2)</sup> Они соответствуют четырем выражениям для суммы из утверждения теоремы 2.11. — Прим. перев.

Подобные формулы можно выписать и для вычисления произведения по модулю  $m$ .

Чтобы увидеть, как выполнять вычитание и деление по модулю  $m$ , потребуется теорема о множестве  $\Pi_m$  наименьших неотрицательных вычетов по модулю  $m$  (оно было определено равенством (2.6)).

**2.12. Теорема.** Система  $(\Pi_m, +, \cdot)$ , где  $+$  и  $\cdot$  означают сложение по модулю  $m$  и умножение<sup>1)</sup> по модулю  $m$  соответственно, представляет собой конечное коммутативное кольцо с единицей.

**Доказательство** заключается в простой проверке следующих свойств для всех  $a, b, c \in \Pi_m$ :

замкнутость  $|a + b|_m \in \Pi_m, |ab|_m \in \Pi_m$ ;  
 коммутативность  $|a + b|_m = |b + a|_m, |ab|_m = |ba|_m$ ;  
 ассоциативность  $|a + (b + c)|_m = |(a + b) + c|_m, |a(bc)|_m = |(ab)c|_m$ ;  
 единственность нуля и единицы  $|a + 0|_m = |a|_m, |a \cdot 1|_m = |a|_m$ ;  
 единственность противоположного элемента  $|a + \underline{a}|_m = 0$ ;  
 дистрибутивность  $|a(b + c)|_m = |ab + ac|_m$ .

Здесь противоположный по модулю  $m$  элемент определяется как

$$\underline{a} \equiv |-a|_m = m - a. \quad \square$$

В кольце  $(\Pi_m, +, \cdot)$  можно ввести вычитание, как сложение с противоположным (аддитивно обратным) по модулю  $m$  элементом.

### 2.13. Определение.

$$|a - b|_m \equiv |a + \underline{b}|_m.$$

Для некоторых элементов  $(\Pi_m, +, \cdot)$  возможно определение (мультипликативно) обратных по модулю  $m$ . Если такой элемент выступает в качестве делителя, то операцию деления можно ввести как умножение на обратный по модулю  $m$ . Однако важен вопрос: «Когда существует мультипликативно обратный по модулю  $m$  элемент?». Ответить на этот вопрос помогает следующая теорема.

**2.14. Теорема.** Конечное коммутативное кольцо  $(\Pi_m, +, \cdot)$  является конечным полем тогда и только тогда, когда число  $m$  простое.

**Доказательство.** См., например, [McCoу, 1948, с. 22]<sup>2)</sup>.

<sup>1)</sup> Часто символ  $\cdot$  опускают и пишут  $ab$  вместо  $a \cdot b$ .

<sup>2)</sup> См. также Понтрягин Л. С. Обобщения чисел. — Библиотечка «Квант», вып. 54. — М.: Наука, 1986, с. 71. — *Прим. перев.*

Следовательно, если  $m$  — простое, то  $(\mathbb{P}_m, +, \cdot)$  изоморфно полю Галуа  $\text{GF}(m)$  и каждый ненулевой элемент в  $\mathbb{P}_m$  имеет обратный по модулю  $m$ , который определяется следующим образом.

**2.15. Определение.** Если  $m$  — простое,  $b \neq 0$ ,  $b \in \mathbb{P}_m$ , то существует единственное целое  $c \in \mathbb{P}_m$ , удовлетворяющее уравнению

$$|cb|_m = |bc|_m = 1.$$

Число  $c$  называется обратным по модулю  $m$  для числа  $b$  и обозначается

$$c = b^{-1}(m),$$

или просто  $b^{-1}$ , когда модуль подразумевается.

Если  $m$  — составное число, то  $(\mathbb{P}_m, +, \cdot)$  уже не поле и ненулевой элемент может не иметь обратного. Чтобы узнать, в каких случаях он существует, потребуются приведенные ниже теорема и следствие. Выражение  $(a, b)$  обозначает *наибольший общий делитель*  $a$  и  $b$  (см. определение 6.1 в этой главе).

**2.16. Теорема.** Пусть  $b \in \mathbb{P}$ . Единственное целое  $c \in \mathbb{P}_m$ , удовлетворяющее равенствам

$$|bc|_m = |cb|_m = 1,$$

существует тогда и только тогда, когда  $|b|_m \neq 0$  и  $(b, m) = 1$ .

**2.17. Следствие.** Если  $b \in \mathbb{P}_m$  не равно нулю, то число  $b^{-1}(n) \in \mathbb{P}_m$  существует (и единственно) тогда и только тогда, когда  $b$  и  $m$  взаимно просты.

**2.18. Примеры.** Если  $m = 10$  (составное), то

$$\mathbb{P}_{10} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

и только 1, 3, 7 и 9 имеют обратные по модулю 10. Они соответственно равны 1, 7, 3, 9. Если  $m = 5$  (простое), то

$$\mathbb{P}_5 = \{0, 1, 2, 3, 4\}$$

и все ненулевые элементы (а именно, 1, 2, 3, и 4) имеют обратные по модулю 5, равные 1, 3, 2 и 4 соответственно. Очевидно, что  $(\mathbb{P}_5, +, \cdot)$  изоморфно полю Галуа  $\text{GF}(5)$ .

Резюмируем:  $(\mathbb{P}_m, +, \cdot)$  всегда является конечным коммутативным кольцом, а если  $m$  — простое, то и конечным полем, изоморфным полю Галуа  $\text{GF}(m)$ . Во всех случаях, когда  $b^{-1}$  существует, деление по модулю  $m$  определяется следующим образом.

## 2.19. Определение.

$$\left| \frac{a}{b} \right|_m = |ab^{-1}|_m.$$

Нужно отметить, что частное целых чисел в одномодульной арифметике вычетов, если оно существует, всегда целое даже в тех случаях, когда  $a$  не делится на  $b$  в  $(\mathbb{I}, +, \cdot)$ .

## 2.20. Пример.

$$|7/9|_{11} = |7 \cdot 9^{-1}|_{11} = |7 \cdot 5|_{11} = 2.$$

Может показаться, что одномодульная арифметика вычетов неприменима для вычисления дроби  $7/9$ . Однако следующий пример показывает, что результат деления в одномодульной арифметике из примера 2.20 нельзя считать бессмысленным<sup>1)</sup>.

## 2.21. Пример.

$$|7/9 \cdot 27|_{11} = ||7/9|_{11} \cdot |27|_{11}|_{11} = |2 \cdot 5|_{11} = 10.$$

Таким образом, целое число, полученное в примере 2.20, можно использовать в качестве промежуточного результата в выкладках примера 2.21. Это иллюстрирует тот факт, что одномодульная арифметика вычетов может применяться для выполнения последовательности арифметических операций над целыми в  $\mathbb{I}_m$ , даже если эта последовательность включает одну или несколько операций деления. Нужно лишь следить, чтобы  $m$  было взаимно простым с каждым целым, появляющимся в знаменателях дробей (чтобы существовали соответствующие обратные по модулю  $m$  элементы).

Трудность возникает только при интерпретации вычисленного результата. Если правильный (математический<sup>2)</sup>) результат принадлежит  $\mathbb{I}_m$ , то он просто совпадает с тем, что дают вычисления в одномодульной арифметике вычетов. В противном случае ответы не совпадают и для восстановления верного результата из того, который получен в модульной арифметике, требуется дополнительная информация. Отметим, что в предыдущем примере  $(7/9)(27) = 21$  и  $21 \equiv 10 \pmod{11}$ .

**2.22. Замечание.** Понятно, что желательно выбирать простой модуль  $m$ , поскольку это гарантирует, что  $(\mathbb{I}_m, +, \cdot)$  будет конечным полем, в котором все ненулевые элементы имеют обратные по модулю  $m$ .

<sup>1)</sup> См. § 5, где примеру 2.20 дается иная интерпретация.

<sup>2)</sup> То есть полученный при помощи обычных арифметических операций в  $(\mathbb{R}, +, \cdot)$ . — *Прим. перев.*

### Приложения теории

В определении 2.7  $|b|_m$  вводится как наименьший неотрицательный вычет  $b$  по модулю  $m$ . Вычисления в одномодульной арифметике вычетов просты в том смысле, что в них участвуют лишь неотрицательные целые. Однако нужно уметь обрабатывать отрицательные целые так же легко, как положительные, если требуется решать задачи типа вычисления определителя матрицы, некоторые элементы которой отрицательны, как в (1.3).

Одна из возможностей состоит во введении системы *симметричных вычетов* по модулю  $m$ . Определим множество

$$(2.23) \quad S_m = \left\{ -\frac{m-1}{2}, \dots, -2, -1, 0, 1, 2, \dots, \frac{m-1}{2} \right\},$$

где  $m$  должно быть *нечетным* целым. Каждое целое  $b \in \mathbb{I}$  может быть отображено на некоторое целое  $s \in S_m$  при помощи следующего правила.

**2.24. Определение.** Отображение  $/\cdot/_{/m}: \mathbb{I} \rightarrow S_m$  определяется формулой

$$/b/_{/m} = s$$

тогда и только тогда, когда

$$b \equiv s \pmod{m}$$

и

$$-\frac{m}{2} < s < \frac{m}{2}.$$

Число  $/b/_{/m}$  называется *симметричным вычетом*<sup>1)</sup> числа  $b$  по модулю  $m$ .

Легко проверить, что  $(S_m, +, \cdot)$  является конечным коммутативным кольцом и, в частности, конечным полем при простом  $m$ . Кроме того,  $(S_m, +, \cdot)$  изоморфно  $(\mathbb{I}_m, +, \cdot)$ . Следовательно, если в некоторой задаче исходные данные суть целые числа из  $S_m$ , то можно перевести их в  $\mathbb{I}_m$ , проделать вычисления<sup>2)</sup> и результат отобразить опять в  $S_m$ . Выпишем функции, осуществляющие изоморфизм  $S_m$  и  $\mathbb{I}_m$ :

$$(2.25) \quad |a|_m = \begin{cases} /a/_{/m}, & \text{если } 0 \leqslant /a/_{/m} < m/2, \\ /a/_{/m} + m & \text{в противном случае;} \end{cases}$$

$$(2.26) \quad /a/_{/m} = \begin{cases} |a|_m, & \text{если } 0 \leqslant |a|_m < m/2, \\ |a|_m - m & \text{в противном случае.} \end{cases}$$

<sup>1)</sup> В оригинале symmetric residue. — *Прим. перев.*

<sup>2)</sup> Все вычисления можно, конечно, выполнить и в  $(S_m, +, \cdot)$ , но при этом требуется следить за алгебраическими знаками промежуточных результатов.

Связь между  $S_m$  и  $\Pi_m$  для случая  $m = 11$  показана на рис. 2.27.

**2.28. Пример.** Пусть  $x = a/b + c$ , где  $a = 48$ ,  $b = 12$  и  $c = -24$ . Так как предполагается  $a, b, c \in S_m$ , то следует

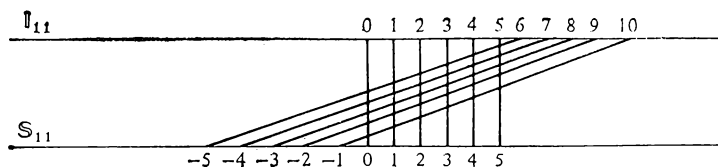


Рис. 2.27. Соответствие между  $S_{11}$  и  $\Pi_{11}$ .

отобразить все данные в  $\Pi_m$  при помощи (2.25) в самом начале вычислений. Допустим, что выбирается  $m = 103$ . Тогда

$$\begin{aligned} |x|_{103} &= |48/12 + (-24)|_{103} = ||48 \cdot 12^{-1}|_{103} + 79|_{103} = \\ &= ||48 \cdot 43|_{103} + 79|_{103} = |4 + 79|_{103} = 83. \end{aligned}$$

В заключение переводим этот результат обратно в  $S_{103}$ , используя (2.26), и получаем правильный ответ:

$$|x|_{103} = -20.$$

**2.29. Замечание.** Важно выбрать  $m$  настолько большим, чтобы все данные и решение задачи содержались в  $S_m$ . В противном случае некоторые решения могут быть неверными, но они будут сравнимы по модулю  $m$  с правильными ответами. Такая ситуация называется *псевдопереполнением*<sup>1)</sup>.

**2.30. Замечание.** В определении 2.15 определен обратный элемент числа  $b$  по модулю  $m$ . В дальнейшем были даны необходимые и достаточные условия его существования и единственности, приводились примеры его использования. Однако алгоритм вычисления обратных элементов по модулю  $m$  не был указан. Простой и практичный алгоритм будет описан ниже (в § 5), где его появление представляется более естественным.

<sup>1)</sup> Этот термин был предложен Т. М. Рао. Мы используем название «псевдопереполнение», а не «переполнение» потому, что отличие от истинных значений для промежуточных результатов вообще не играет роли, а при неверном конечном результате правильный ответ обычно может быть получен на основе дополнительной информации.

## Упражнения 1.2

## 1. Вычислить

- (a)  $|279|_5$ ; (e)  $|279/5|$ ;  
 (b)  $|-279|_5$ ; (f)  $|-279/5|$ ;  
 (c)  $|10^6|_{37}$ ; (g)  $|10^6/37|$ ;  
 (d)  $|48|_{11}$ ; (h)  $|48/11|$ .

2. Вычислить  $|27 \cdot 17|_{13}$  четырьмя способами.

3. Определить, какие числа в  $\Pi_{16}$  имеют обратные по модулю 16, и вычислить соответствующие обратные числа.

§ 3. Многомодульная арифметика вычетов<sup>1</sup>

Лучший способ устранить проблему псевдопереполнения, описанную в замечании 2.29, заключается в использовании арифметики вычетов по нескольким модулям одновременно. Дело в том, что можно показать эквивалентность такой арифметики и арифметики вычетов по одному модулю, которым является наименьшее общее кратное этих нескольких модулей.

Например, рассмотрим упорядоченное множество

$$(3.1) \quad \beta = [m_1, m_2, \dots, m_n],$$

составленное из  $n$  (различных) модулей  $m_1, m_2, \dots, m_n$ . Предположим, что модули попарно взаимно просты, т. е. что

$$(3.2) \quad (m_i, m_j) = 1, \quad i \neq j.$$

В случаях когда  $\beta$  вида (3.1) удовлетворяет условиям (3.2), множество  $\beta$  называется *векторным основанием*<sup>2)</sup> для числовой системы вычетов, к описанию которой мы переходим.

**3.3. Определение.** Для каждого целого  $s$  (единственный) упорядоченный набор из  $n$  вычетов

$$|s|_\beta = [|s|_{m_1}, |s|_{m_2}, \dots, |s|_{m_n}]$$

называется *стандартным*<sup>3)</sup> *представлением*<sup>4)</sup> числа  $s$  по отношению к векторному основанию  $\beta$ . Каждый из вычетов  $|s|_{m_i}$  называется *цифрой стандартного представления*<sup>5)</sup>  $s$  по отношению к  $\beta$ .

<sup>1)</sup> В оригинале multiple-modulus residue arithmetic. — Прим. перев.

<sup>2)</sup> В оригинале base vector. — Прим. перев.

<sup>3)</sup> Термин «стандартное» используется, чтобы можно было отличить это представление от «симметричного», которое вводится в замечании 3.14.

<sup>4)</sup> В оригинале standard residue representation. — Прим. перев.

<sup>5)</sup> В оригинале standard residue digit. — Прим. перев.

**3.4. Пример.** Пусть  $\beta = [5, 7, 9]$  и  $s = 34$ . Тогда

$$|34|_{\beta} = [4, 6, 7].$$

Определим число  $M$  как произведение всех модулей, входящих в векторное основание  $\beta$ :

$$(3.5) \quad M = \prod_{i=1}^n m_i.$$

Теперь мы готовы сформулировать важную теорему и ее следствие. И теорема и следствие будут также справедливы, когда модули  $m_i$  не являются взаимно простыми, но в этом случае  $M$  следует определить как наименьшее общее кратное всех модулей.

**3.6. Теорема.** *Два целых числа  $s$  и  $t$  имеют одинаковое стандартное представление по отношению к  $\beta$  (т. е.  $|t|_{\beta} = |s|_{\beta}$ ) тогда и только тогда, когда*

$$s \equiv t \pmod{M}.$$

**3.7. Следствие.** *Если  $t = |s|_M$ , то  $t$  и  $s$  имеют одинаковое стандартное представление по отношению к  $\beta$ :*

$$|t|_{\beta} = |s|_{\beta}.$$

**3.8. Пример.** Пусть  $\beta = [3, 5, 7]$ , так что  $M = 105$ . Возьмем  $s = 403$ . Тогда

$$|403|_{105} = 88.$$

Легко проверить (прямыми вычислениями), что

$$|403|_{\beta} = |88|_{\beta} = [1, 3, 4].$$

Для каждого  $a \in \mathbb{I}$  рассмотрим единственное целое число  $r$ , удовлетворяющее условиям

$$(3.9) \quad a \equiv r \pmod{M},$$

$$(3.10) \quad 0 \leq r < M.$$

Тогда это число

$$(3.11) \quad r = |a|_M$$

принадлежит множеству неотрицательных вычетов по модулю  $M$ :

$$(3.12) \quad \mathbb{I}_M = \{0, 1, 2, \dots, M-1\}.$$

**3.13. Определение.** *Стандартной системой вычетов для векторного основания  $\beta$  назовем множество (из  $M$  элементов) стандартных представлений*

$$\mathbb{I}_{\beta} = \{|s|_{\beta} : s \in \mathbb{I}\}.$$



Из теоремы 3.6 и следствия 3.7, очевидно, вытекает, что элементы конечной числовой системы  $\Pi_\beta$  находятся во взаимно однозначном соответствии с элементами  $\Pi_M$ . Согласно теореме 2.12, система  $(\Pi_M, +, \cdot)$ , где  $+$  и  $\cdot$  обозначают операции сложения и умножения по модулю  $M$ , является конечным коммутативным кольцом. Несложно показать, что над элементами  $\Pi_\beta$  можно так определить бинарные операции  $\oplus$  и  $\odot$ , что система  $(\Pi_\beta, \oplus, \odot)$  также образует конечное коммутативное кольцо (например, см. теоремы 3.15 и 3.16).

Можно доказать, что кольца  $(\Pi_M, +, \cdot)$  и  $(\Pi_\beta, \oplus, \odot)$  изоморфны и следовательно, многомодульная арифметика вычетов с векторным основанием  $\beta$  эквивалентна одномодульной арифметике вычетов с числом  $M$  в качестве модуля.

**3.14. Замечание.** Если набор  $\beta = [m_1, m_2, \dots, m_n]$  включает в себя только нечетные модули, то на основе (2.23) и определения 2.24 можно ввести соответствующую *симметричную систему вычетов* для векторного основания  $\beta$ . В этом случае любое целое  $s$  имеет *симметричное представление*

$$/s/\beta = [/s/m_1, /s/m_2, \dots, /s/m_n],$$

где каждый вычет  $/s/m_i$  в отдельности называется *цифрой симметричного представления*. Симметричная система вычетов определяется как множество всех (а именно,  $M$ ) различных симметричных представлений:

$$S_\beta = \{/s/\beta: s \in \mathbb{Z}\}.$$

В  $S_\beta$  каждое целое  $s$  представляется единственным упорядоченным набором из  $n$  чисел  $/s/\beta$ , и это представление взаимно однозначно для целых чисел из множества

$$S_M = \left\{ -\frac{M-1}{2}, \dots, -2, -1, 0, 1, 2, \dots, \frac{M-1}{2} \right\}.$$

Для колец  $(\Pi_M, +, \cdot)$  и  $(\Pi_\beta, \oplus, \odot)$  стандартного представления имеются симметричные аналоги — конечные коммутативные кольца  $(S_M, +, \cdot)$  и  $(S_\beta, \oplus, \odot)$ . Можно показать, что все эти кольца изоморфны между собой (таким образом, имеется взаимно однозначное соответствие между стандартными и симметричными системами вычетов; см. табл. 3.26).

Заметим, что для каждого целого  $a$  связь между  $|a|_\beta$  и  $/a/\beta$  задается отображениями (2.25) и (2.26), применяемыми к соответствующим цифрам стандартного и симметричного представлений  $|a|_{m_i}$  и  $/a/m_i$  для  $i = 1, \dots, n$ .

### Арифметика в $\Pi_\beta$

В следующих двух теоремах будут описаны операции сложения, вычитания и умножения в  $\Pi_\beta$ .

**3.15. Теорема.** Пусть  $a$  и  $b$  — целые числа. Стандартным представлением числа  $a \pm b$  по отношению к векторному основанию  $\beta$  является упорядоченный набор из  $n$  чисел

$$|a \pm b|_\beta = [z_1, z_2, \dots, z_n],$$

где

$$z_i = ||a|_{m_i} \pm |b|_{m_i}|_{m_i}, \quad i = 1, 2, \dots, n.$$

**3.16. Теорема.** Пусть  $a$  и  $b$  — целые числа. Стандартным представлением числа  $ab$  по отношению к векторному основанию  $\beta$  является упорядоченный набор из  $n$  чисел

$$|ab|_\beta = [w_1, w_2, \dots, w_n],$$

где

$$w_i = ||a|_{m_i}|b|_{m_i}|_{m_i}, \quad i = 1, 2, \dots, n.$$

Таким образом, чтобы получить цифры стандартного представления суммы, разности или произведения двух целых чисел, мы просто *покомпонентно* складываем, вычитаем или умножаем цифры стандартного представления данных операндов и приводим каждый результат по нужному модулю.

**3.17. Пример.** Пусть  $\beta = [3, 5, 7]$ ; тогда  $M = 105$ . Также пусть  $a = 24$  и  $b = 20$ . Тогда

$$|24|_\beta = [0, 4, 3],$$

$$|20|_\beta = [2, 0, 6].$$

Таким образом,

$$|24 + 20|_\beta = [|0 + 2|_3, |4 + 0|_5, |3 + 6|_7] = [2, 4, 2].$$

Аналогично

$$|24 - 20|_\beta = [|0 - 2|_3, |4 - 0|_5, |3 - 6|_7] = [1, 4, 4]$$

и

$$|(24)(20)|_\beta = [| (0)(2) |_3, | (4)(0) |_5, | (3)(6) |_7] = [0, 0, 4].$$

В качестве проверки отметим, что

$$|44|_\beta = [2, 4, 2],$$

$$|4|_\beta = [1, 4, 4],$$

$$|480|_\beta = [0, 0, 4].$$

**3.18. Замечание.** В описанных выше вычислениях сумма, разность и произведение чисел 24 и 20 равны 44, 4 и 480 со-

ответственно. Первые два результата принадлежат множеству  $\Pi_{105}$ , а последний — нет. Поэтому при отображении <sup>1)</sup> в  $\Pi_{105}$  представлений  $[2, 4, 2]$  и  $[1, 4, 4]$  получаются правильные ответы 44 и 4 соответственно, однако  $[0, 0, 4]$  перейдет в 60 (а не в 480), поскольку

$$0 \leq 60 < 105, \quad 480 \equiv 60 \pmod{105}.$$

Заметим, что величина  $M = 105$  недостаточно велика, чтобы избежать псевдопереполнения при вычислении произведения. Единственный выход состоит, конечно, в увеличении значения  $M$  за счет выбора либо *большого количества* модулей, либо *их большей величины*.

Чтобы научиться выполнять деление в системе стандартных представлений вычетами, требуется несколько обобщить процедуру, которая применялась в одномодульной арифметике.

**3.19. Определение.** Пусть  $a$  — целое, и пусть существуют обратные элементы  $a^{-1}(m_1), a^{-1}(m_2), \dots, a^{-1}(m_n)$ . Тогда упорядоченный набор из  $n$  чисел

$$a^{-1}(\beta) = [a^{-1}(m_1), a^{-1}(m_2), \dots, a^{-1}(m_n)]$$

назовем *стандартным представлением обратного к  $a$  числа по отношению к векторному основанию  $\beta$* .

Чтобы набор  $a^{-1}(\beta)$  существовал, требуется просто существование всех чисел  $a^{-1}(m_i)$ ,  $i = 1, 2, \dots, n$ . Очевидна *единственность* стандартного представления  $a^{-1}(\beta)$ .

**3.20. Теорема.** Пусть  $a$  и  $b$  — целые и  $a^{-1}(\beta)$  существует. Тогда

$$\left| \frac{b}{a} \right|_{\beta} = [c_1, c_2, \dots, c_n],$$

где

$$c_i = \left| b \right|_{m_i} a^{-1}(m_i) \Big|_{m_i}, \quad i = 1, 2, \dots, n.$$

Отметим, что

$$(3.21) \quad \left| \frac{a}{a} \right|_{\beta} = [1, 1, \dots, 1],$$

как и следовало ожидать.

**3.22. Пример.** Пусть  $\beta = [3, 5, 7]$ ,  $M = 105$  и  $a = 23$ ,  $b = 46$ . В этом примере  $b/a = 2$ . Поскольку

$$\begin{aligned} |46|_{\beta} &= [1, 1, 4], \\ 23^{-1}(\beta) &= [2, 2, 4], \end{aligned}$$

---

<sup>1)</sup> Метод, реализующий отображение стандартных представлений вычетами в целых числах, будет описан в следующем параграфе.

то

$$|46/23|_{\beta} = [|1 \cdot 2|_3, |1 \cdot 2|_5, |4 \cdot 4|_7] = [2, 2, 2],$$

что является стандартным представлением числа 2.

**3.23. Замечание.** Если  $b$  не делится на  $a$ , то положение аналогично тому, которое сложилось в одномодульной арифметике вычетов. Ответ может быть трудно интерпретируемым (без дополнительной информации), но он вполне пригоден как промежуточный результат в дальнейших вычислениях. См. примеры 2.20 и 2.21.

**3.24. Замечание.** Следует подчеркнуть важное различие, существующее между многомодульной и одномодульной (по модулю  $M$ ) арифметиками, хотя они в определенном смысле эквивалентны. Выбор  $m$  простым всегда гарантирует, что  $(\Pi_m, +, \cdot)$  — конечное поле. С другой стороны,  $M$  не может быть простым (по определению), и поэтому  $(\Pi_M, +, \cdot)$  является только конечным коммутативным кольцом. Следует ожидать, что при делении в многомодульной арифметике могут возникнуть проблемы, связанные с тем, что для некоторых векторных оснований  $\beta$  и некоторых целых чисел  $a$  набор  $a^{-1}(\beta)$  не будет существовать. См. теоремы 7.16 и 7.17 в § 7.

**3.25. Замечание.** Описание многомодульной арифметики вычетов с приложениями к решению систем линейных алгебраических уравнений можно найти в работах [Howell, Gregory, 1969, 1970] и [Joung, Gregory, 1973, гл. 13].

3.26. Таблица. Представления для целых чисел при  $\beta = [3, 5]$  и  $M = 15$

$\mathbb{I}_M$	$\mathbb{S}_M$	$\mathbb{I}_{\beta}$	$\mathbb{S}_{\beta}$
0	0	[0, 0]	[0, 0]
1	1	[1, 1]	[1, 1]
2	2	[2, 2]	[-1, 2]
3	3	[0, 3]	[0, -2]
4	4	[1, 4]	[1, -1]
5	5	[2, 0]	[-1, 0]
6	6	[0, 1]	[0, 1]
7	7	[1, 2]	[1, 2]
8	-7	[2, 3]	[-1, -2]
9	-6	[0, 4]	[0, -1]
10	-5	[1, 0]	[1, 0]
11	-4	[2, 1]	[-1, 1]
12	-3	[0, 2]	[0, 2]
13	-2	[1, 3]	[1, -2]
14	-1	[2, 4]	[-1, -1]

## Упражнения I.3

1. Пусть  $\beta = [3, 5, 7]$ . Вычислить стандартные представления
 

(a) $ 137 _{\beta}$ ;	(d) $ 137/\beta $ ;
(b) $ -137 _{\beta}$ ;	(e) $ -137/\beta $ ;
(c) $ 537 _{\beta}$ ;	(f) $ 537/\beta $ .
2. Пусть  $\beta = [5, 7, 11]$ ,  $a = 34$ ,  $b = 408$ . Найти стандартные представления
 

(a) $ a + b _{\beta}$ ;	(e) $ a + b/\beta $ ;
(b) $ a - b _{\beta}$ ;	(f) $ a - b/\beta $ ;
(c) $ ab _{\beta}$ ;	(g) $ ab/\beta $ ;
(d) $\left  \frac{b}{a} \right _{\beta}$ ;	(h) $\left  \frac{b}{a} \right _{\beta}$ .
3. В каких примерах из указанных выше возникает псевдопереполнение?
4. Доказать теорему 3.6.
5. Доказать следствие 3.7.
6. Доказать теорему 3.15.
7. Доказать теорему 3.16.
8. Доказать теорему 3.20.

### § 4. Отображение стандартных представлений вычетами в целые числа

В замечании 3.18 было указано, что при  $\beta = [3, 5, 7]$  стандартные представления  $[2, 4, 2]$ ,  $[1, 4, 4]$  и  $[0, 0, 4]$  отображаются в целые числа 44, 4 и 60 соответственно, но не был описан алгоритм отображения. Цель настоящего параграфа — ответить на вопрос: «Как осуществить отображение стандартного представления вычетами (по отношению к  $\beta$ ) в *единственное* целое число в  $\Pi_m$ ?».

Один из старейших известных алгоритмов (но не самый быстрый), решающих задачу отображения, основан на классической теореме теории чисел, называемой *китайской теоремой об остатках* (например, см. [Young, Gregory, 1973, с. 874]<sup>1)</sup>). Мы не будем, однако, приводить здесь этот алгоритм<sup>2)</sup> и вместо него опишем способ, в котором используется представление целых чисел в *системе счисления со смешанным основанием*<sup>3)</sup>.

<sup>1)</sup> См. также Кнут Д. Искусство программирования для ЭВМ. Т. 2. Получисленные алгоритмы. — М.: Мир, 1977, с. 305. — *Прим. ред.*

<sup>2)</sup> Подобный алгоритм приведен в книге Ахо А., Хопкрофт Дж., Ульман Дж. Построение и анализ вычислительных алгоритмов. — М.: Мир, 1979, с. 329. — *Прим. перев.*

<sup>3)</sup> В оригинале mixed-radix number representation. — *Прим. перев.*

Рассмотрим упорядоченный набор из  $n$  целых чисел

$$(4.1) \quad \rho = [r_1, r_2, \dots, r_n],$$

компоненты которого  $r_1, r_2, \dots, r_n$  назовем *основаниями*. Пусть  $R$  есть произведение оснований, т. е.

$$(4.2) \quad R = \prod_{i=1}^n r_i.$$

Хорошо известно (например, см. [Szabó, Такака, 1967, с. 41]), что каждое целое число  $s$ , такое, что

$$(4.3) \quad 0 \leq s < R,$$

можно единственным образом представить в виде

$$(4.4) \quad s = d_0 + d_1(r_1) + d_2(r_1 r_2) + \dots + d_{n-1}(r_1 r_2 \dots r_{n-1}),$$

где  $d_0, d_1, \dots, d_{n-1}$  являются *цифрами стандартного представления для смешанного основания* и удовлетворяют неравенствам

$$(4.5) \quad 0 \leq d_i < r_{i+1}, \quad i = 0, 1, \dots, n-1.$$

Заметим, что основная роль  $r_n$  служить границей для  $d_{n-1}$ .

Упорядоченный набор цифр  $d_0, d_1, \dots, d_{n-1}$  для данного  $s$  записывается в виде

$$(4.6) \quad \langle s \rangle_\rho = \langle d_0, d_1, \dots, d_{n-1} \rangle.$$

Например, если  $\rho = [2, 3, 5]$ , то  $R = 30$ : Следовательно, из

$$(4.7) \quad 29 = 1 + 2(2) + 4(2 \cdot 3)$$

имеем  $d_0 = 1, d_1 = 2, d_2 = 4$ . Отсюда

$$(4.8) \quad \langle 29 \rangle_\rho = \langle 1, 2, 4 \rangle.$$

**4.9. Определение.** *Стандартной системой счисления со смешанным основанием для  $\rho$  назовем множество всевозможных наборов цифр типа  $\langle s \rangle_\rho$  для целых чисел  $s \in [0, R)$ .*

В частном случае  $r_1 = r_2 = \dots = r_n$  приходим к известному представлению числа в позиционной системе с фиксированным основанием. Если каждое основание равно десяти, например, то это просто десятичное представление.

Более интересен (для наших целей) специальный случай  $r_i = m_i, i = 1, 2, \dots, n$ , где  $m_i$  — элементы векторного основания  $\beta$  многомодульной системы вычетов. Иная формулировка:

$$(4.10) \quad \rho = \beta.$$

В этом случае  $R = M$  и следовательно, многомодульная система вычетов и система со смешанным основанием здесь представляют один и тот же диапазон целых чисел. Это обстоятельство очень важно, поскольку мы хотим заменить первую систему второй.

Рассмотрим стандартную систему со смешанным основанием и соответствующую ей стандартную систему вычетов с векторным основанием  $\beta = [m_1, m_2, \dots, m_n]$ . Пусть целое число  $s$  имеет представление

$$(4.11) \quad s = d_0 + d_1(m_1) + d_2(m_1 m_2) + \dots + d_{n-1}(m_1 m_2 \dots m_{n-1})$$

с набором (единственным) цифр

$$(4.12) \quad \langle s \rangle_\beta = \langle d_0, d_1, \dots, d_{n-1} \rangle,$$

с одной стороны, и представление в вычетах

$$(4.13) \quad |s|_\beta = [|s|_{m_1}, |s|_{m_2}, \dots, |s|_{m_n}]$$

с другой. Соответствующие цифры  $d_{i-1}$  и  $|s|_{m_i}$  в этих двух представлениях принадлежат одному и тому же замкнутому интервалу  $[0, m_i - 1]$ ,  $i = 1, 2, \dots, n$ , что видно из (4.5) и (2.6).

Предположим, что задано представление  $|s|_\beta$  вида (4.13) и требуется найти  $\langle s \rangle_\beta$  вида (4.12). Другими словами, предположим известными цифры  $|s|_{m_i}$  из (4.13) и поставим задачу определения цифр  $d_{i-1}$ ,  $i = 1, 2, \dots, n$ , представления со смешанным основанием (4.12). Чтобы получить  $d_0$ , положим  $t_1 = s$  и запишем, используя (4.11),

$$(4.14) \quad t_1 = s = d_0 + m_1[d_1 + d_2(m_2) + \dots + d_{n-1}(m_2 \dots m_{n-1})] = \\ = d_0 + m_1 t_2.$$

Тогда из теоремы 2.10 следует равенство

$$(4.15) \quad |t_1|_{m_1} = |d_0 + m_1 t_2|_{m_1} = d_0.$$

Заметим, что  $d_0 = |t_1|_{m_1} = |s|_{m_1}$ , т. е. первые цифры в обоих представлениях совпадают и *никакие вычисления не требуются*.

Для вычисления  $d_1$  привлечем (4.14):

$$(4.16) \quad t_2 = d_1 + m_2[d_2 + d_3(m_3) + \dots + d_{n-1}(m_3 \dots m_{n-1})] = \\ = d_1 + m_2 t_3.$$

Следовательно, при помощи теоремы 2.10 заключаем

$$(4.17) \quad |t_2|_{m_2} = |d_1 + m_2 t_3|_{m_2} = d_1.$$

Таким образом, можно предложить рекурсивную процедуру для определения цифр представления со смешанным основанием. Выберём начальные значения  $t_1 = s$  и  $d_0 = |t_1|_{m_1}$  и проведем вычисления по формулам

$$(4.18) \quad \begin{aligned} t_{i+1} &= \frac{t_i - d_{i-1}}{m_i}, & i &= 1, 2, \dots, n-1. \\ d_i &= |t_{i+1}|_{m_{i+1}}, \end{aligned}$$

Важно отметить, что в алгоритме (4.18) используется величина  $s$  (напомним, что  $t_1 = s$ ), которая неизвестна и подлежит определению. Следовательно, алгоритм (4.18) *непосредственно неприменим*. Поскольку вместо  $s$  задано представление в вычетах  $|s|_\beta$ , следует заменить в (4.18) обычные арифметические операции на операции в арифметике вычетов.

*Вычисление цифр  $d_i$  с использованием арифметики вычетов*

Пусть  $|s|_\beta$  определено формулой (4.13). Если положить  $s = t_1$  и привлечь (4.15), то можно записать

$$(4.19) \quad |t_1|_\beta = [d_0, |t_1|_{m_2}, \dots, |t_1|_{m_n}].$$

По определению

$$(4.20) \quad |d_0|_\beta = [|d_0|_{m_1}, |d_0|_{m_2}, \dots, |d_0|_{m_n}],$$

что позволяет вычислить

$$(4.21) \quad |t_1 - d_0|_\beta = [0, |z_2|_{m_2}, |z_3|_{m_3}, \dots, |z_n|_{m_n}],$$

где

$$(4.22) \quad z_i = |t_1|_{m_i} - |d_0|_{m_i}, \quad i = 2, 3, \dots, n.$$

Если ввести сокращенное векторное основание

$$(4.23) \quad \beta_1 = [m_2, m_3, \dots, m_n],$$

то можно представить  $t_1 - d_0$  в вычетах по отношению к  $\beta_1$ . В этом случае

$$(4.24) \quad |t_1 - d_0|_{\beta_1} = [|z_2|_{m_2}, |z_3|_{m_3}, \dots, |z_n|_{m_n}].$$

Чтобы вычислить  $t_2$ , требуется определить  $m_1^{-1}(\beta_1)$ . Этот обратный элемент существует, потому что число  $m_1$  взаимно просто с любым числом из набора  $\beta_1$ . Таким образом,

$$(4.25) \quad m_1^{-1}(\beta_1) = [m_1^{-1}(m_2), m_1^{-1}(m_3), \dots, m_1^{-1}(m_n)]$$



и поэтому

$$(4.26) \quad |t_2|_{\beta_1} = |(t_1 - d_0)/m_1|_{\beta_1} = [|w_2|_{m_2}, |w_3|_{m_3}, \dots, |w_n|_{m_n}],$$

где

$$(4.27) \quad w_i = |z_i|_{m_i} m_i^{-1}(m_i), \quad i = 2, 3, \dots, n.$$

Из (4.17) и (4.26) находим вторую цифру представления со смешанным основанием:

$$(4.28) \quad |w_2|_{m_2} = |t_2|_{m_2} = d_1.$$

Если подставить этот результат в (4.26) и учесть, что  $|w_i|_{m_i} = |t_2|_{m_i}$ , то можно записать

$$(4.29) \quad |t_2|_{\beta_1} = [d_1, |t_2|_{m_3}, \dots, |t_2|_{m_n}].$$

По аналогии с (4.20) имеем

$$(4.30) \quad |d_1|_{\beta_1} = [|d_1|_{m_2}, |d_1|_{m_3}, \dots, |d_1|_{m_n}],$$

что дает

$$(4.31) \quad |t_2 - d_1|_{\beta_1} = [0, |v_3|_{m_3}, \dots, |v_n|_{m_n}],$$

где

$$(4.32) \quad v_i = |t_2|_{m_i} - |d_1|_{m_i}, \quad i = 3, 4, \dots, n.$$

Введем сокращенное векторное основание

$$(4.33) \quad \beta_2 = [m_3, m_4, \dots, m_n]$$

подобно тому, как было определено  $\beta_1$ . Представим  $t_2 - d_1$  в вычетах по отношению к  $\beta_2$ :

$$(4.34) \quad |t_2 - d_1|_{\beta_2} = [|v_3|_{m_3}, |v_4|_{m_4}, \dots, |v_n|_{m_n}].$$

Для вычисления  $t_3$  требуется определить  $m_2^{-1}(\beta_2)$ . Этот обратный элемент существует, поскольку число  $m_2$  взаимно просто с любым числом из набора  $\beta_2$ . Таким образом,

$$(4.35) \quad m_2^{-1}(\beta_2) = [m_2^{-1}(m_3), m_2^{-1}(m_4), \dots, m_2^{-1}(m_n)]$$

и поэтому

$$(4.36) \quad |t_3|_{\beta_2} = |(t_2 - d_1)/m_2|_{\beta_2} = [|u_3|_{m_3}, |u_4|_{m_4}, \dots, |u_n|_{m_n}],$$

где

$$(4.37) \quad u_i = |v_i|_{m_i} m_2^{-1}(m_i), \quad i = 3, 4, \dots, n.$$

Из (4.18) следует, что мы нашли третью цифру представления со смешанным основанием:

$$(4.38) \quad |u_3|_{m_3} = |t_3|_{m_3} = \bar{d}_2.$$

Продолжая этот алгоритм, можно в конце концов вычислить каждую цифру искомого представления (4.12).

**4.39. Задача.** Пусть  $\beta = [13, 11, 7]$  и  $|s|_\beta = [4, 2, 4]$ . Найти  $\langle s \rangle_\beta$  и вычислить  $s$ .

**Решение.** Поскольку  $M_\beta = 1001$ , допустим, что нас интересует единственное целое число  $s \in [0, 1001)$ . Вычисления, описанные формулами (4.19)–(4.38), можно свести в таблицу.

$\beta$	$m_1 = 13$	$m_2 = 11$	$m_3 = 7$	
$ t_1 _\beta$	$\begin{smallmatrix} \cdot & \cdot & \cdot \\ \cdot & 4 & \cdot \\ \cdot & \cdot & \cdot \end{smallmatrix}$	2	4	
$ d_0 _\beta$	4	4	4	вычесть
$ t_1 - d_0 _\beta$	0	9	0	
$m_1^{-1}(\beta_1)$		6	6	умножить
$ t_2 _{\beta_1} =  (t_1 - d_0)/m_1 _{\beta_1}$		$\begin{smallmatrix} \cdot & \cdot & \cdot \\ \cdot & 10 & \cdot \\ \cdot & \cdot & \cdot \end{smallmatrix}$	0	
$ d_1 _{\beta_1}$		10	3	вычесть
$ t_2 - d_1 _{\beta_1}$		0	4	
$m_2^{-1}(\beta_2)$			2	умножить
$ t_3 _{\beta_2} =  (t_2 - d_1)/m_2 _{\beta_2}$			$\begin{smallmatrix} \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot \end{smallmatrix}$	

Элементы таблицы, обведенные пунктиром, суть  $d_0$ ,  $d_1$  и  $d_2$ ; таким образом,  $\langle s \rangle_\beta = \langle 4, 10, 1 \rangle$ . Следовательно, из (4.11) получаем

$$s = 4 + 10(13) + 1(13)(11) = 277,$$

что и дает правильный ответ.

**4.40. Замечание.** Теперь нам известен алгоритм вычисления цифр представления  $\langle s \rangle_\beta$  по заданному  $|s|_\beta$ . Чтобы найти  $s$  при помощи  $\langle s \rangle_\beta$ , достаточно подсчитать значение правой части (4.11). В предыдущем примере расчеты настолько просты, что выполняются непосредственно. В общем случае нужно использовать следующую рекурсию:

$$\begin{aligned} s_1 &= d_{n-1}, \\ s_2 &= d_{n-2} + s_1 m_{n-1}, \\ s_3 &= d_{n-3} + s_2 m_{n-2}, \\ &\vdots \\ s_n &= d_0 + s_{n-1} m_1, \end{aligned}$$

в которой  $s = s_n$ .

**4.41. Замечание.** В начале § 3 было указано, что многомодульная арифметика с векторным основанием  $\beta$  эквивалентна одномодульной арифметике с числом  $M$  в качестве модуля<sup>1)</sup>. Другими словами, арифметики в  $(\Pi_\beta, \oplus, \odot)$  и  $(\Pi_M, +, \cdot)$  эквивалентны.

Предположим, что требуется выполнять арифметические операции одновременно с положительными и отрицательными целыми числами. Для этой ситуации в § 2 была предложена процедура, в соответствии с которой целые в  $S_m$  отображаются в  $\Pi_m$  при помощи (2.25), так что арифметические операции можно выполнять в  $(\Pi_m, +, \cdot)$ . Полученные результаты затем отображаются обратно в  $S_m$  при помощи (2.26).

Подобным образом можно поступить и сейчас. Целые из  $S_M$  переводятся в  $\Pi_M$  отображением, аналогичным (2.25). Затем они представляются в  $\Pi_\beta$ ; арифметические операции выполняются в  $(\Pi_\beta, +, \cdot)$ , а результат отображается в  $\Pi_M$ . На последнем этапе результат переводится из  $\Pi_M$  обратно в  $S_M$  отображением, аналогичным (2.26).

**4.42. Замечание.** Мы обсудили четыре основные арифметические операции многомодульной арифметики, переход от целых чисел к их представлениям по отношению к векторному основанию и обратный переход. Обратим внимание, что арифметические операции выполняются покомпонентно. Это наводит на мысль, что вычислительные машины, ориентированные на использование многомодульной арифметики, могут иметь преимущества перед обычными. Если имеется  $n$  процессоров, каждый из которых предназначен для выполнения операций с вычетами целых чисел, то  $n$  компонент модульного представления могут обрабатываться одновременно. Создатели новейшего поколения компьютеров с векторными операциями и параллельными процессорами сделали шаг в нужном направлении. Однако ни в одном из этих компьютеров приведение результата по модулю  $m$  не выполняется непосредственно аппаратурой. Это приведение можно осуществить лишь программными средствами.

**4.43. Замечание.** Мы показали, что в многомодульной арифметике сложение, вычитание и умножение являются чрезвычайно простыми операциями, а деление — несколько

---

<sup>1)</sup> Фактически модуль в эквивалентной одномодульной системе является наименьшим общим кратным модулей из векторного основания. Однако поскольку последние предполагаются попарно взаимно простыми, то число  $M$ , определенное как произведение модулей из векторного основания, одновременно оказывается и их наименьшим общим кратным.

более сложным (см. замечание 3.24). Имеются также три другие ситуации, приводящие к некоторым осложнениям:

(а) сравнение величин двух целых чисел  $s$  и  $t$ ,

(б) определение знака числа  $s$ ,

(в) восстановление  $s$  по  $|s|_\beta$ .

Обсуждение ситуаций (а) и (б) читатель найдет в гл. 4 книги [Szabó, Тапака, 1967]. Ситуация (в) обсуждалась в настоящем параграфе, а (б) рассматривалась в замечании 4.41.

### Упражнения I. 4

1. Пусть  $\beta = [3, 5, 2]$ .

(а) Найти  $\langle 29 \rangle_\beta$ .

(б) Вычислить  $s$ , если задано  $\langle s \rangle_\beta = \langle 1, 2, 3 \rangle$ .

2. Пусть  $\beta = [5, 7, 3]$  и  $|s|_\beta = [0, 3, 1]$ . Найти  $\langle s \rangle_\beta$  и  $s$ .

3. Пусть  $\beta = [3, 7, 5]$  и  $|s|_\beta = [1, 2, 0]$ . Найти  $\langle s \rangle_\beta$  и  $s$ .

## § 5. Одномодульная арифметика вычетов для рациональных чисел

Оказывается, что одномодульная арифметика вычетов может использоваться и для выполнения арифметических операций над некоторыми рациональными числами. Основная идея состоит в том, чтобы отобразить рациональные операнды в множество целых чисел  $\Pi_m$ , определенное в (2.6), произвести арифметические операции в  $(\Pi_m, +, \cdot)$  и затем отобразить целочисленные результаты в соответствующие рациональные числа.

При таком использовании арифметики вычетов будет удобным выбрать модуль вида  $m = p^r$ , где  $p$  — простое и  $r$  — положительное целое число. Рациональные числа  $a/b$ , для которых  $(b, p) = 1$ , отображаются на  $\Pi_m$ , причем используется тот факт, что  $b^{-1}(m)$  существует тогда и только тогда, когда  $(b, p) = 1$ . Очевидно, если  $r = 1$ , то  $m$  совпадает с простым  $p$ .

**5.1. Определение.** Если  $x = a/b$  и  $(b, p) = 1$ , так что  $b^{-1}(m)$  существует, то

$$|x|_m = \left| \frac{a}{b} \right|_m = |ab^{-1}|_m.$$

Вспоминая определение 2.19, теперь можно интерпретировать пример 2.20, как пример отображения рационального числа  $7/9$  на целое число 2 из  $\Pi_{11}$ .

Пусть  $\hat{\mathbb{Q}}$  — множество тех рациональных чисел, которые допускают отображение в  $\Pi_m$ , т. е.

$$(5.2) \quad \hat{\mathbb{Q}} = \left\{ \frac{a}{b} : (b, p) = 1 \right\}.$$

Определение 5.1 описывает отображение  $|\cdot|_m: \hat{\mathbb{Q}} \rightarrow \Pi_m$ .

Каждое целое число  $k \in \Pi_m$  оказывается образом бесконечного множества элементов из  $\hat{\mathbb{Q}}$ , которое будем обозначать  $\mathbb{Q}_k$ . Следовательно, для  $k = 0, 1, \dots, m-1$  по определению

$$(5.3) \quad \mathbb{Q}_k = \left\{ \frac{a}{b} \in \hat{\mathbb{Q}} : \left| \frac{a}{b} \right|_m = k \right\},$$

откуда вытекает, что

$$(5.4) \quad \hat{\mathbb{Q}} = \bigcup_{k=0}^{m-1} \mathbb{Q}_k.$$

Например, множество  $\mathbb{Q}_0$  (рациональные числа из  $\mathbb{Q}$ , которые переводятся в нуль), состоит из тех элементов  $a/b$  при  $(b, p) = 1$ , для которых  $|a|_m = 0$ . Непересекающиеся подмножества  $\mathbb{Q}_0, \mathbb{Q}_1, \dots, \mathbb{Q}_{m-1}$  назовем *обобщенными классами вычетов по модулю  $m$* , поскольку они включают в себя обычные классы вычетов, определенные в (2.8), как собственные подмножества:

$$(5.5) \quad \mathbb{R}_k \subset \mathbb{Q}_k, \quad k = 0, 1, 2, \dots, m-1.$$

В приведенных ниже теореме и следствии характеризуются элементы обобщенного класса вычетов.

**5.6. Теорема.** Пусть  $x = a/b$  и  $y = c/d$ , где  $b^{-1}(m)$  и  $d^{-1}(m)$  существуют. Равенство

$$|x|_m = |y|_m$$

имеет место тогда и только тогда, когда

$$ad \equiv bc \pmod{m}.$$

**Доказательство.** Предположим, что  $ad \equiv bc \pmod{m}$ . Умножая обе части этого сравнения на  $b^{-1}d^{-1}$ , получаем

$$ab^{-1} \equiv cd^{-1} \pmod{m},$$

что влечет за собой равенство

$$|ab^{-1}|_m = |cd^{-1}|_m.$$

Таким образом, в силу определения 5.1 имеем

$$|x|_m = |y|_m.$$

Доказательство обратного утверждения предлагается читателю в качестве упражнения.  $\square$

**5.7. Следствие.** Пусть  $x = a/b$  и  $y = c/d$  принадлежат  $\hat{\mathbb{Q}}$ . Тогда  $x$  и  $y$  являются элементами одного обобщенного класса вычетов  $\mathbb{Q}_k$  при условии

$$ad \equiv bc \pmod{m},$$

которое будет необходимым и достаточным.

**Доказательство.** В силу определения два рациональных числа  $x$  и  $y$  принадлежат одному обобщенному классу вычетов тогда и только тогда, когда

$$|x|_m = |y|_m,$$

что эквивалентно условию

$$ad \equiv bc \pmod{m},$$

как утверждается в теореме 5.6. Доказательство завершено.

Отметим, что не каждое рациональное число  $a/b$  удовлетворяет условию  $(b, p) = 1$  и поэтому не каждое рациональное число принадлежит  $\hat{\mathbb{Q}}$ , т. е. какому-либо из непересекающихся подмножеств  $\mathbb{Q}_0, \mathbb{Q}_1, \dots, \mathbb{Q}_{m-1}$ , отображаемых на  $\Pi_m$ .

**5.8. Замечание.** Обращая внимание на тот факт, что  $(ka)/(kb) = a/b$  для любого целого  $k \neq 0$ , мы предполагаем всюду, что каждое рациональное число уже приведено. Таким образом, если  $x = (2p)/(3p)$  и  $p > 3$ , то  $|x|_m$  существует, поскольку  $x = 2/3$ . С учетом указанного соглашения все подмножество рациональных чисел  $x = a/b$ , для которых  $|x|_m$  не существует, полностью описывается условием, что  $b$  делится на  $p$ .

**5.9. Лемма.** Пусть  $x = a/b$  и  $y = c/d$  принадлежат  $\hat{\mathbb{Q}}$ ; тогда  $x + y$  и  $xy$  принадлежат  $\hat{\mathbb{Q}}$ .

**Доказательство.** Если  $b \neq 0$  и  $d \neq 0$  не делятся на  $p$ , то  $bd \neq 0$  также не делится на  $p$ . Поэтому  $xy = (ac)/(bd)$  и  $x + y = (ad + bc)/(bd)$  являются элементами  $\hat{\mathbb{Q}}$ .  $\square$

**5.10. Теорема.** Система  $(\hat{\mathbb{Q}}, +, \cdot)$  образует коммутативное кольцо с единицей.

**Доказательство.** Из леммы 5.9 вытекает, что множество  $\hat{\mathbb{Q}}$  замкнуто относительно операций сложения и умножения. Кроме того, коммутативный и ассоциативный законы для сложения и умножения и дистрибутивные законы справедливы в  $\hat{\mathbb{Q}}$  как в подмножестве  $\mathbb{Q}$ . Наконец, легко проверить, что 0 и 1 принадлежат  $\hat{\mathbb{Q}}$  и для любого элемента  $x \in \hat{\mathbb{Q}}$  элемент  $-x$  также входит в  $\hat{\mathbb{Q}}$ .  $\square$

Теперь докажем лемму, в которой устанавливается связь колец  $(\hat{\mathbb{Q}}, +, \cdot)$  и  $(\Pi_m, +, \cdot)$ .

**5.11. Лемма.** Пусть  $x = a/b$  и  $y = c/d$  содержатся в  $\hat{\mathbb{Q}}$ . Тогда

$$(i) \quad ||x|_m \cdot |y|_m|_m = |xy|_m$$

и

$$(ii) \quad ||x|_m + |y|_m|_m = |x + y|_m.$$

$$\begin{aligned} \text{Доказательство.} \quad ||x|_m \cdot |y|_m|_m &= ||ab^{-1}|_m \cdot |cd^{-1}|_m|_m = \\ &= |ab^{-1}cd^{-1}|_m = \\ &= |ac(bd)^{-1}|_m = |xy|_m. \end{aligned}$$

Подобным образом

$$\begin{aligned} ||x|_m + |y|_m|_m &= ||ab^{-1}|_m + |cd^{-1}|_m|_m = |ab^{-1} + cd^{-1}|_m = \\ &= |ab^{-1}dd^{-1} + cd^{-1}bb^{-1}|_m = \\ &= |(ad + bc)(bd)^{-1}|_m = |x + y|_m. \quad \square \end{aligned}$$

Эта лемма позволяет установить следующий фундаментальный результат.

**5.12. Теорема.** Отображение  $|\cdot|_m: \hat{\mathbb{Q}} \rightarrow \Pi_m$  задает гомоморфизм по отношению к операциям сложения и умножения.

Другими словами,  $\Pi_m$  есть гомоморфный образ  $\hat{\mathbb{Q}}$ , и арифметическим операциям в кольце  $(\hat{\mathbb{Q}}, +, \cdot)$  соответствуют те же арифметические операции в кольце  $(\Pi_m, +, \cdot)$ . Напомним, что при  $r = 1$  число  $m = p^r$  является простым  $p$  и в этом случае  $(\Pi_p, +, \cdot)$  есть конечное поле, изоморфное полю Гауа  $\text{GF}(p)$ .

Для наших целей было бы идеально, если бы соответствие между  $(\hat{\mathbb{Q}}, +, \cdot)$  и  $(\Pi_m, +, \cdot)$  являлось изоморфизмом. Однако это требует невозможного, а именно — существования взаимно однозначного соответствия между  $\hat{\mathbb{Q}}$  и  $\Pi_m$ .

*Выбор подходящего подмножества в  $\hat{\mathbb{Q}}$*

Отображение  $|\cdot|_m: \hat{\mathbb{Q}} \rightarrow \Pi_m$  не является взаимно однозначным, поскольку каждое целое  $k \in \Pi_m$  является образом бесконечного подмножества  $\mathbb{Q}_k$  рациональных чисел. Следовательно, это отображение не имеет обратного. Возникает вопрос: можно ли некоторым естественным образом выбрать по одному элементу из каждого обобщенного класса вычетов  $\mathbb{Q}_k$ , чтобы обеспечить взаимно однозначное отображение ме-

жду этими элементами и всеми целыми числами из  $\Pi_m$ ? Если подобный выбор возможен, то такое отображение будет иметь обратное.

К несчастью, предлагаемый нами способ выбора охватывает *только некоторые* из обобщенных классов вычетов, но не все эти классы. Поэтому приходится ограничиться взаимно однозначным отображением выбранных элементов из некоторых обобщенных классов вычетов на множество их образов из  $\Pi_m$ .

**5.13. Определение.** Конечное подмножество  $\mathbb{F}_N$  множества  $\hat{\mathbb{Q}}$ , задаваемое следующим образом:

$$\mathbb{F}_N = \{a/b \in \hat{\mathbb{Q}}: (a, b) = 1, \quad 0 \leq |a| \leq N, \quad 0 < |b| \leq N\},$$

где  $N > 0$  — целое число, назовем множеством *дроби Фарей порядка  $N^1$* ).

Оказывается, что при подходящем выборе величины  $N$  каждый обобщенный класс вычетов  $\mathbb{Q}_k$  содержит не больше одной дроби Фарей.

**5.14. Теорема.** Пусть  $N$  — максимальное целое число, для которого выполнено неравенство

$$2N^2 + 1 \leq m,$$

и пусть обобщенный класс вычетов  $\mathbb{Q}_k$  содержит некоторую дробь Фарей  $x = a/b$  порядка  $N$ . Тогда  $x$  — единственная дробь Фарей порядка  $N$  в множестве  $\mathbb{Q}_k$ .

**Доказательство.** Предположим наличие в  $\mathbb{Q}_k$  двух дробей Фарей  $x = a/b$  и  $y = c/d$  порядка  $N$ . Тогда

$$|x|_m = |y|_m = k.$$

По теореме 5.6 это влечет за собой сравнение

$$ad \equiv bc \pmod{m},$$

или

$$|ad - bc|_m = 0.$$

Далее, поскольку

$$0 \leq |ad - bc| \leq |a| \cdot |d| + |b| \cdot |c| \leq 2N^2 \leq m - 1,$$

то  $ad - bc = 0$ . Таким образом  $a/b = c/d$ , или  $x = y$ ; следовательно,  $x$  — единственная дробь Фарей порядка  $N$  в  $\mathbb{Q}_k$ .  $\square$

**5.15. Замечание.** Если величина  $N$  выбрана в соответствии с условием теоремы 5.14, то число элементов в множестве

<sup>1)</sup> В оригинале order- $N$  Farey fractions. — Прим. перев.



$\mathbb{F}_N$  меньше  $m$ . Следовательно, не каждый из обобщенных классов вычетов  $Q_0, Q_1, \dots, Q_{m-1}$  может содержать элемент из  $\mathbb{F}_N$ . Однако если некоторый класс  $Q_k$  все же содержит дробь Фарея порядка  $N$ , то в этом классе такая дробь только одна.

Теперь мы готовы указать взаимно однозначное отображение выбранных элементов из  $\hat{Q}$  на множество их образов из  $\Pi_m$ . Обозначим через

$$(5.16) \quad \hat{\Pi}_m = \{ |a/b|_m : a/b \in \mathbb{F}_N \}$$

образ множества дробей Фарея порядка  $N$  при гомоморфизме  $|\cdot|_m: \hat{Q} \rightarrow \Pi_m$ . (Далее всюду, за исключением специально оговариваемых случаев, будем считать выполненным ограничение на величину  $N$ , указанное в теореме 5.14.) Справедлив следующий результат.

**5.17. Теорема.** *Отображение  $|\cdot|_m: \mathbb{F}_N \rightarrow \hat{\Pi}_m$  является взаимно однозначным и переводит  $\mathbb{F}_N$  во все множество  $\hat{\Pi}_m$ . Таким образом, оно имеет обратное<sup>1)</sup>.*

Доказательство предоставляется читателю в качестве упражнения.  $\square$

**5.18. Пример.** Пусть  $m = 19$ . Тогда  $N = 3$ . В следующей таблице в левых частях каждой колонки записаны все дроби Фарея третьего порядка, в правой — их образы при отображении  $|\cdot|_{19}: \mathbb{F}_3 \rightarrow \hat{\Pi}_{19}$ .

0	0		
1	1	-1	18
2	2	-2	17
3	3	-3	16
.	.	.	.
$-\frac{1}{3}$	6	$\frac{1}{3}$	13
$\frac{2}{3}$	7	$-\frac{2}{3}$	12
$-\frac{2}{3}$	8	$\frac{2}{3}$	11
$-\frac{1}{3}$	9	$\frac{1}{3}$	10

Многоточие в таблице на месте чисел 4, 5, 14 и 15 призвано напомнить, что  $\hat{\Pi}_{19} \subset \Pi_{19}$ . Именно эти числа из  $\Pi_{19}$  не входят в  $\hat{\Pi}_{19}$ , т. е. не являются образами каких-либо дробей Фарея третьего порядка.

<sup>1)</sup> Это отображение можно было бы назвать изоморфизмом, если бы числовые системы  $(\mathbb{F}_N, +, \cdot)$  и  $(\hat{\Pi}_m, +, \cdot)$  являлись подкольцами колец  $(\hat{Q}, +, \cdot)$  и  $(\Pi_m, +, \cdot)$ . К сожалению, это не так, поскольку указанные числовые системы незамкнуты. — *Прим. перев.*

**5.19. Замечание.** Мы построили взаимно однозначное отображение конечного подмножества рациональных чисел в  $\hat{\mathbb{Q}}$  (а именно, так называемых дробей Фарея порядка  $N$ ) на (конечное) подмножество целых чисел в  $\hat{\mathbb{I}}_m$ . Следовательно, поскольку  $\mathbb{F}_N \subset \hat{\mathbb{Q}}$  и  $\hat{\mathbb{I}}_m \subset \hat{\mathbb{I}}_m$ , то вступает в действие отмеченная ранее эквивалентность арифметических операций в кольцах  $(\hat{\mathbb{Q}}, +, \cdot)$  и  $(\hat{\mathbb{I}}_m, +, \cdot)$ . Если выбрано очень большое целое  $m$ , то множество  $\mathbb{F}_N$  дробей Фарея порядка  $N$  достаточно представительно. Если для некоторой задачи все ее данные и ответы содержатся в  $\mathbb{F}_N$ , то мы имеем право:

- (i) перевести операнды из  $\mathbb{F}_N$  в  $\hat{\mathbb{I}}_m$ ,
- (ii) выполнить арифметические операции в кольце  $(\hat{\mathbb{I}}_m, +, \cdot)$  и
- (iii) перевести целые результаты обратно в  $\mathbb{F}_N$ .

Если некоторые из ответов задачи не принадлежат  $\mathbb{F}_N$ , то возникает ситуация, подобная той, что в замечании 2.29 названа псевдопереполнением.

В следующих задачах иллюстрируются как нормальный ход вычислений, так и ситуация псевдопереполнения.

**5.20. Задача.** Найти  $x$ , если

$$x = \frac{1}{3} - \frac{2}{3} = \frac{1}{3} + \left(-\frac{2}{3}\right).$$

**Решение.** Если выбрать  $m = 19$ , то  $N = 3$  и можно использовать отображение, представленное в таблице примера 5.18. Тогда

$$|x|_{19} = \left|\frac{1}{3} + \left(-\frac{2}{3}\right)\right|_{19} = |13 + 12|_{19} = 6.$$

Поскольку  $6 \in \hat{\mathbb{I}}_{19}$ , то, применяя обратное отображение, получаем  $x = -1/3$ , что совпадает с правильным ответом.

**5.21. Задача.** Найти  $x$ , если

$$x = \frac{1}{2} - \frac{2}{3} = \frac{1}{2} + \left(-\frac{2}{3}\right).$$

**Решение.** Как и ранее, выбираем  $m = 19$ ,  $N = 3$ . Однако теперь решение задачи не является дробью Фарея третьего порядка, поэтому возникает псевдопереполнение. Чтобы убедиться к  $x = 3$ , что неверно (правильный ответ  $x = -1/6$ ).  $\square$

$$|x|_{19} = \left|\frac{1}{2} + \left(-\frac{2}{3}\right)\right|_{19} = |10 + 12|_{19} = 3.$$

Поскольку  $3 \in \hat{\mathbb{I}}_{19}$ , то, применяя обратное отображение, приходим к  $x = 3$ , что неверно (правильный ответ  $x = -1/6$ ).

Интересно отметить, однако, что в последней задаче

$$(5.22) \quad \left|-\frac{1}{6}\right|_{19} = |(-1)6^{-1}|_{19} = |(-1)(16)|_{19} = 3.$$

Это показывает, что вычисленный ответ ( $x=3$ ) и правильный ответ ( $x=-1/6$ ) принадлежат одному обобщенному классу вычетов  $\mathbb{Q}_3$ .

Псевдопереполнение также может проявляться в том, что вычисления в  $(\Pi_m, +, \cdot)$  приводят к элементам  $\Pi_m$ , не входящим в  $\Pi_m$ . Так, в  $\Pi_{19}$  результат, равный одному из чисел 4, 5, 14 и 15, не может быть отображен в какой-либо элемент  $\mathbb{F}_3$ .

**5.23. Замечание.** Если рациональные числа, играющие роль операндов или возникающие на *промежуточных этапах*, не принадлежат  $\mathbb{F}_N$ , но *конечный ответ* все же принадлежит  $\mathbb{F}_N$ , то результат наших вычислений оказывается верным. Это демонстрируется в следующем примере:

$$x = \frac{1}{2} - \frac{2}{3} - \frac{1}{6} = \frac{1}{2} + \left(-\frac{2}{3}\right) + \left(-\frac{1}{6}\right).$$

В  $(\Pi_{19}, +, \cdot)$  получаем

$$|x|_{19} = \left| \frac{1}{2} + \left(-\frac{2}{3}\right) + \left(-\frac{1}{6}\right) \right|_{19} = |10 + 12 + 3|_{19} = 6.$$

Применяя обратное отображение, представленное в примере 5.18, получаем правильный ответ  $x=-1/3$ . Отметим, что  $-1/6$  не входит в  $\mathbb{F}_3$ , как и результат сложения в исходном выражении для  $x$  первых двух операндов (см. пример псевдопереполнения в задаче 5.21). Однако в качестве промежуточного результата эта «неверная» сумма не мешает вычислению правильного конечного результата. (Воскресим в памяти пример 2.21.)

**5.24. Замечание.** Очевидно, что в практических задачах<sup>1)</sup> для выполнения прямого отображения  $\mathbb{F}_N \rightarrow \Pi_m$  и обратного отображения  $\Pi_m \rightarrow \mathbb{F}_N$  невыгодно составлять, хранить и использовать таблицу типа той, что приведена в примере 5.18. В следующем параграфе будут предложены алгоритмы, осуществляющие эти отображения. Как частный случай алгоритма для прямого отображения по ходу дела рассматривается способ вычисления обратного элемента  $b^{-1}(m)$ .

### Упражнения I. 5

1. Завершить доказательство теоремы 5.6.
2. Доказать теорему 5.17.
3. Построить таблицу, подобную таблице из примера 5.18 и соответствующую отображению  $\mathbb{F}_5$  на  $\Pi_{53}$ .

<sup>1)</sup> То есть для больших значений  $m$  и  $N$ . — Прим. перев.

4. Используя вычисления в конечном поле  $(\mathbb{P}_{53}, +, \cdot)$ , найти  $x$  для случаев

(a)  $x = 1/3 - 2/3$ ;

(b)  $x = 1/2 + 3/4$ ;

(c)  $x = (2/3) \cdot (5/2)$ ;

(d)  $x = 1/2 - 2/3 - 1/6$  (см. замечание 5.23).

### § 6. Прямое и обратное отображения

В этом параграфе описан алгоритм прямого отображения  $\mathbb{F}_N \rightarrow \hat{\mathbb{P}}_m$ , т. е. для данного  $a/b \in \mathbb{F}_N$  определяется элемент  $|a/b|_m \in \hat{\mathbb{P}}_m$ . Этот алгоритм основан на алгоритме Евклида и в частном случае  $a = 1$  позволяет вычислять обратный элемент  $b^{-1}(m)$ .

Мы описываем также алгоритм для обратного отображения  $\hat{\mathbb{P}}_m \rightarrow \mathbb{F}_N$ . Как и предыдущий, он основан на алгоритме Евклида и был независимо разработан Корнерупом и Кришнамурти. Для заданного целого числа  $k \in \hat{\mathbb{P}}_m$  алгоритм порождает сразу несколько рациональных чисел из  $\mathbb{Q}_k$ , в том числе и единственную в  $\mathbb{Q}_k$  дробь Фарей порядка  $N$ . Эту дробь Фарей легко выделить, привлекая определение 5.13.

#### Алгоритм Евклида

Для начала рассмотрим алгоритм Евклида применительно к нахождению наибольшего общего делителя двух целых чисел  $a$  и  $b$ .

**6.1. Определение.** Наибольшим общим делителем  $(a, b)$  двух целых чисел  $a$  и  $b$  (не равных одновременно нулю) назовем наибольшее положительное целое число из тех, которые делят одновременно  $|a|$  и  $|b|$ .

Наибольший общий делитель удовлетворяет следующим условиям:

$$(6.2) \quad \begin{aligned} (0, 0) & \text{ не определен;} \\ (0, b) & = |b| \text{ при } b \neq 0; \\ (a, b) & = (b, a); \\ (a, b) & = (|a|, |b|). \end{aligned}$$

Следовательно, без потери общности можно считать, что  $a$  и  $b$  — неотрицательные целые, не обращающиеся одновременно в нуль<sup>1)</sup>.

<sup>1)</sup> Поскольку  $(a, a) = |a|$ , то нетривиальная ситуация определяется условиями  $a > b > 0$ . Это соглашение фактически принимается далее. — *Прим. перев.*

Стандартный прием описания алгоритма Евклида для двух неравных друг другу положительных целых чисел  $a > b > 0$  основан на использовании следующего *свойства деления*: существуют такие целые числа  $q$  и  $r$ , что  $q > 0$ ,  $0 \leq r < b$  и

$$(6.3) \quad a = bq + r.$$

Здесь  $q$  — частное и  $r$  — остаток при делении  $a$  на  $b$ . Применяя последовательно это свойство, приходим к системе уравнений и неравенств:

$$(6.4) \quad \begin{aligned} a &= bq_1 + r_1, & 0 < r_1 < b, \\ b &= r_1q_2 + r_2, & 0 < r_2 < r_1, \\ r_1 &= r_2q_3 + r_3, & 0 < r_3 < r_2, \\ &\vdots \\ r_{n-2} &= r_{n-1}q_n + r_n, & 0 < r_n < r_{n-1}, \\ r_{n-1} &= r_nq_{n+1}, \end{aligned}$$

где  $r_n \neq 0$  (но  $r_{n+1} = 0$ ). Эта система, трактуемая как вычислительный процесс, называется алгоритмом Евклида. Последний ненулевой остаток  $r_n$  в (6.4), как известно, удовлетворяет следующему равенству.

### 6.5. Теорема.

$$r_n = (a, b).$$

**Доказательство.**

$$(a, b) = (b, r_1) = (r_1, r_2) =$$

$$\vdots$$

$$= (r_{n-1}, r_n) = (r_n, 0) = r_n. \quad \square$$

Наибольший общий делитель можно характеризовать следующей важной теоремой.

**6.6. Теорема.** Для двух целых  $a > b > 0$  величина  $(a, b)$  совпадает с наименьшим положительным числом  $d$  вида

$$d = ax + by,$$

где  $x, y \in \mathbb{Z}$ .

**Доказательство.** См., например, [Pettofrezzo, Byrkit, 1970, с. 34].  $\square$

Заметим, что целые числа  $x$  и  $y$  в представлении для  $d$  из этой теоремы определяются неоднозначно, потому что для

любого  $t \in \Pi$  выполнено

$$(6.7) \quad d = a(x + bt) + b(y - at).$$

Ясно также, что не каждая линейная комбинация  $a$  и  $b$  приводит к  $d$ , поскольку при

$$(6.8) \quad d = ax + by$$

имеем для любого  $k \in \Pi$  равенство

$$(6.9) \quad (kd) = a(kx) + b(ky).$$

Чтобы найти пару целых чисел  $x$  и  $y$  для (6.8), используем (6.4) для определения остатков  $r_1, r_2, \dots, r_{n+1}$  как функций от  $a$  и  $b$ . Получаем

$$(6.10) \quad \begin{aligned} r_1 &= a + b(-q_1), \\ r_2 &= b + r_1(-q_2) = \\ &= a(-q_2) + b(1 + q_1q_2), \\ r_3 &= r_1 + r_2(-q_3) = \\ &= a(1 + q_2q_3) + b(-q_1 - q_3 - q_1q_2q_3), \\ &\vdots \\ r_n &= r_{n-2} + r_{n-1}(-q_n) = \\ &= ax + by, \\ 0 &= r_{n-1} + r_n(-q_{n+1}) = \\ &= au + bv. \end{aligned}$$

Эти вычисления можно записать в виде таблицы.

6.11. Таблица. Вычисление  $r_n, x$  и  $y$

	$a$	$1$	$0$
	$b$	$0$	$1$
$q_1$	$r_1$	$1$	$-q_1$
$q_2$	$r_2$	$-q_2$	$1 + q_1q_2$
$q_3$	$r_3$	$1 + q_2q_3$	$-q_1 - q_3 - q_1q_2q_3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$q_n$	$r_n$	$x$	$y$
$q_{n+1}$	$0$	$u$	$v$

Столбец табл. 6.11, который начинается с чисел  $a$  и  $b$ , содержит набор остатков, порожденных алгоритмом Евклида (6.4). Если одновременно с остатками вычисляются еще какие-либо величины (например, в нашем случае это последо-

вательности, приводящие к  $x$  и  $y$ ), то подобный алгоритм называется [Кнут, 1976, с. 42] *обобщенным* или *расширенным алгоритмом Евклида*; см. алгоритм 6.26, например.

В табл. 6.11 мы фактически генерируем последовательность  $n+1$  целочисленных троек, рассматривая их как векторы-строки и исходя из *начальной матрицы*

$$\begin{bmatrix} a & 1 & 0 \\ b & 0 & 1 \end{bmatrix}.$$

Так,

$$\begin{aligned} [r_1, 1, -q_1] &= [1, -q_1] \begin{bmatrix} a & 1 & 0 \\ b & 0 & 1 \end{bmatrix}, \\ (6.12) \quad [r_2, -q_2, 1+q_1q_2] &= [1, -q_2] \begin{bmatrix} b & 0 & 1 \\ r_1 & 1 & -q_1 \end{bmatrix}, \\ [r_3, 1+q_2q_3, -q_1-q_3-q_1q_2q_3] &= [1, -q_3] \begin{bmatrix} r_1 & 1 & -q_1 \\ r_2 & -q_2 & 1+q_1q_2 \end{bmatrix} \end{aligned}$$

и т. д. В этих равенствах величины  $q_1, q_2, \dots, q_{n+1}$  определяются формулами<sup>1)</sup>

$$\begin{aligned} (6.13) \quad q_1 &= [a/b], \\ q_2 &= [b/r_1], \\ q_3 &= [r_1/r_2], \\ &\vdots \\ q_{n+1} &= [r_{n-1}/r_n]. \end{aligned}$$

**6.14. Пример.** Чтобы найти (9,11) вместе с  $x$  и  $y$  (см. табл. 6.11), используем начальную матрицу

$$\begin{bmatrix} 19 & 1 & 0 \\ 11 & 0 & 1 \end{bmatrix}.$$

Результаты вычислений сведены в следующую таблицу:

	19	1	0
	11	0	1
1	8	1	-1
1	3	-1	2
2	2	3	-5
1	1	-4	7
2	0	11	-19

<sup>1)</sup> Символ  $[a/b]$  обозначает целую часть дроби  $a/b$ .

Из этой таблицы видим, что

$$\begin{aligned}(19, 11) &= 1, \\ x &= -4, \\ y &= 7.\end{aligned}$$

Для контроля проверяем

$$1 = (19)(-4) + (11)(7).$$

Кроме того, поскольку  $u = 11$  и  $v = -19$ , то справедливо равенство <sup>1)</sup>

$$0 = (19)(11) + (11)(-19).$$

*Обратный по модулю  $m$  элемент*

Вычислительную схему, отраженную формулами (6.10) и табл. 6.11, можно применить для отыскания обратного элемента к целому числу  $b$  в конечном коммутативном кольце  $(\Pi_m, +, \cdot)$  или конечном поле  $(\Pi_p, +, \cdot)$ , когда  $m$  совпадает с простым  $p$ . Из следствия 2.17 известно, что  $b^{-1}(m)$  существует для  $b \neq 0$  тогда и только тогда, когда  $(m, b) = 1$ . Поэтому существенна следующая теорема.

**6.15. Теорема.** Если  $(m, b) = 1$  и

$$1 = mx + by,$$

то

$$b^{-1}(m) = |y|_m.$$

**Доказательство.** Запишем равенства

$$1 = |mx + by|_m = |by|_m = |b|_m |y|_m,$$

из которых следует по определению

$$|y|_m = b^{-1}(m). \quad \square$$

Следовательно, в примере 6.14 из равенства  $y = 7$  вытекает, что  $11^{-1}(19) = |7|_{19} = 7$ . Для контроля проверяем  $|11 \cdot 7|_{19} = 1$ .

**6.16. Пример.** Предположим  $m = 5^4 = 625$  и вычислим обратный элемент к 342 по модулю 625. Этот обратный элемент существует, поскольку  $(342, 5) = 1$ . В данном случае роль начальной матрицы играет матрица

$$\begin{bmatrix} 625 & 0 \\ 342 & 1 \end{bmatrix}$$

<sup>1)</sup> Обобщение этого факта приводится в задаче 2 из упражнений I, 6.



и мы получаем следующую таблицу:

	625	0
	342	1
1	283	-1
1	59	2
4	47	-9
1	12	11
3	11	-42
1	1	53
11	0	-625

Следовательно, имеем  $(625, 342) = 1$  и  $y = 53$ , что влечет за собой равенства

$$342^{-1} \equiv 53 \pmod{625}.$$

Для контроля проверяем

$$|53 \cdot 342|_{625} = 1.$$

### Прямое отображение

Прямое отображение  $\mathbb{F}_N \rightarrow \hat{\mathbb{I}}_m$  можно выполнить в два этапа. А именно, на первом этапе вычисление  $|d/c|_m$  определяем  $c^{-1}(m)$ , используя начальную матрицу

$$\begin{bmatrix} m & 0 \\ c & 1 \end{bmatrix},$$

как в примере 6.16. Затем по определению 5.1 вычисляем  $|d/c|_m = |d \cdot c^{-1}|_m$ .

С другой стороны, эти этапы совмещаются, если в расширенном алгоритме Евклида использовать начальную матрицу

$$\begin{bmatrix} m & 0 \\ c & d \end{bmatrix}.$$

Это обстоятельство подсказал авторам Корнеруп (см. [Корнеруп, Gregory, 1983]).

В частном случае  $d = 1$  будет вычисляться обратный элемент

$$(6.17) \quad \left| \frac{1}{c} \right|_m = c^{-1}(m).$$

В следующем примере иллюстрируются как двух-, так и одноэтапный методы.

6.18. **Пример.** Требуется вычислить  $|-3/2|_{19}$ . Для нахождения  $2^{-1}(19)$  используем начальную матрицу

$$\begin{bmatrix} 19 & 0 \\ 2 & 1 \end{bmatrix}$$

и получаем таблицу

$$\begin{array}{c|cc} & 19 & 0 \\ & 2 & 1 \\ \hline 9 & 1 & -9 \\ \hline 2 & 0 & 19 \end{array}$$

Таким образом,

$$2^{-1} = |-9|_{19} = 10.$$

Следовательно,

$$|-\frac{3}{2}|_{19} = |(-3)(2^{-1})|_{19} = |-30|_{19} = 8.$$

Для одноэтапного метода возьмем начальную матрицу

$$\begin{bmatrix} 19 & 0 \\ 2 & -3 \end{bmatrix}.$$

Получаем таблицу

$$\begin{array}{c|cc} & 19 & 0 \\ & 2 & -3 \\ \hline 9 & 1 & 27 \\ \hline 2 & 0 & -57 \end{array}$$

Каждый элемент третьего столбца таблицы умножен на  $-3^1$ ); следовательно,

$$|-\frac{3}{2}|_{19} = |27|_{19} = 8.$$

### *Некоторые свойства расширенного алгоритма Евклида*

Выберем четыре целых числа  $a$ ,  $b$ ,  $c$  и  $d$  и произвольную последовательность целых чисел  $\{q_1, q_2, \dots\}$ . Построим последовательность пар целых чисел при помощи рекурсии

$$(6.19a) \quad \begin{aligned} a_i &= a_{i-2} - q_i a_{i-1}, \\ b_i &= b_{i-2} - q_i b_{i-1}, \end{aligned} \quad i = 1, 2, \dots$$

<sup>1)</sup> Если сравнивать с таблицей двухэтапного метода, — *Прим. перев.*

На языке матриц можно записать

$$(6.196) \quad [a_i, b_i] = [1, -q_i] \begin{bmatrix} a_{i-2} & b_{i-2} \\ a_{i-1} & b_{i-1} \end{bmatrix}$$

где начальная матрица задана условием

$$(6.20) \quad \begin{bmatrix} a_{-1} & b_{-1} \\ a_0 & b_0 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Эти вычисления можно также представить в виде следующей таблицы:

	$a$	$b$
	$c$	$d$
$q_1$	$a_1$	$b_1$
$q_2$	$a_2$	$b_2$
$q_3$	$a_3$	$b_3$
$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$

6.21. **Пример.** Положим  $a = -3$ ,  $b = -2$ ,  $c = -2$ ,  $d = 5$ ,  $q_1 = 2$ ,  $q_2 = -1$ ,  $q_3 = 4$ , ... . Проведем вычисления, получаем таблицу

	$-3$	$-2$
	$-2$	$5$
$2$	$1$	$-12$
$-1$	$-1$	$-7$
$4$	$5$	$16$
$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$

Если наложить на начальную матрицу условие типа

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \equiv 0 \pmod{w}$$

для некоторого  $w$ , то при произвольной последовательности  $\{q_1, q_2, \dots\}$  будет справедлив следующий результат.

6.23. **Лемма.** Если  $ad - bc \equiv 0 \pmod{w}$ , то

$$a_i b_{i-1} - a_{i-1} b_i \equiv 0 \pmod{w}, \quad i = 1, 2, \dots$$

**Доказательство.** Имеем

$$\begin{aligned}
 a_i b_{i-1} - a_{i-1} b_i &= (a_{i-2} - q_i a_{i-1}) b_{i-1} - a_{i-1} (b_{i-2} - q_i b_{i-1}) = \\
 &= (a_{i-2} b_{i-1} - a_{i-1} b_{i-2}) + q_i \cdot 0 = \\
 &\quad \vdots \\
 &= (-1)^{i-1} (a_{-1} b_0 - a_0 b_{-1}) \\
 &= (-1)^{i-1} (ab - cd) \equiv 0 \pmod{w}. \quad \square
 \end{aligned}$$

Утверждение леммы можно иллюстрировать на примере 6.21:

$$\begin{aligned}
 (6.24) \quad \begin{vmatrix} -3 & -2 \\ -2 & 5 \end{vmatrix} &\equiv \begin{vmatrix} -2 & 5 \\ 1 & -12 \end{vmatrix} \equiv \begin{vmatrix} 1 & -12 \\ -1 & -7 \end{vmatrix} \equiv \begin{vmatrix} -1 & -7 \\ 5 & 16 \end{vmatrix} \equiv \\
 &\equiv 0 \pmod{19}.
 \end{aligned}$$

Другими словами, если из пар целых чисел примера 6.21 составить последовательность соответствующих рациональных чисел

$$\left\{ \frac{2}{3}, -\frac{5}{2}, -12, 7, \frac{16}{5}, \dots \right\},$$

то все члены этой последовательности принадлежат одному обобщенному классу вычетов (в арифметике с модулем, равным 19), а именно классу  $\mathbb{Q}_7$  (см. следствие 5.7).

Теперь определим последовательность  $\{q_1, q_2, \dots\}$  специальным образом:

$$(6.25) \quad q_i = \left[ \frac{a_{i-2}}{a_{i-1}} \right],$$

$a_{i-1} \neq 0$ , т. е. как в (6.13). Здесь  $a_{-1} = a$  и  $a_0 = c$  предполагаются положительными целыми числами. В этом случае легко проверить, что (конечная) последовательность  $\{a_1, a_2, \dots, a_n, a_{n+1}\}$  совпадает с последовательностью частичных остатков, порожденных алгоритмом Евклида для задачи вычисления величины  $(a, c)$  (см. табл. 6.11). Следовательно,  $a_n = (a, c)$  и  $a_{n+1} = 0$ .

6.26. Алгоритм (расширенный алгоритм Евклида).

(1) Выбираем начальную матрицу

$$\begin{bmatrix} a_{-1} & b_{-1} \\ a_0 & b_0 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

где  $a$  и  $c$  — положительные целые,  $a, b$  и  $d$  — произвольные целые числа.

(2) Для  $i = 1, 2, \dots, n+1$ , пока <sup>1)</sup>  $a_{i-1} \neq 0$ , определяем  $q_i$  как частное и  $a_i$  как неотрицательный остаток при делении  $a_{i-2}$  на  $a_{i-1}$ . Тогда

$$a_i = a_{i-2} - q_i a_{i-1}.$$

(3) Аналогично определяем  $b_i$ :

$$b_i = b_{i-2} - q_i b_{i-1}.$$

(4) Завершаем вычисления при  $a_{n+1} = 0$ . Имеем  $a_n = (a, c)$ .

Как было продемонстрировано в примере 6.18, этот расширенный алгоритм можно использовать для выполнения прямого отображения, т. е. для вычисления величины  $|r/s|_m = |rs^{-1}|_m$  — наименьшего неотрицательного вычета дроби  $r/s$  по модулю  $m$ . Чтобы доказать в общем случае применимость расширенного алгоритма для реализации прямого отображения  $\mathbb{F}_N \rightarrow \mathbb{F}_m$ , потребуется следующий вспомогательный результат.

**6.27. Лемма.** В алгоритме 6.26 положим  $a = m$ ,  $b = 0$  и  $0 < c < m$ , причем потребуем, чтобы  $(c, m) = 1$ . Тогда справедливы равенства

$$\left| \frac{b_i}{a_i} \right|_m = \left| \frac{d}{c} \right|_m, \quad i = 1, 2, \dots, n.$$

Если  $0 < |d| < m$ , причем  $(d, m) = 1$ , то верны и равенства

$$\left| \frac{a_i}{b_i} \right|_m = \left| \frac{c}{d} \right|_m.$$

**Доказательство.** Поскольку  $a = m$  и  $b = 0$ , то

$$ad \equiv bc \pmod{m}.$$

Следовательно, по лемме 6.23

$$a_1 d \equiv b_1 c \pmod{m}.$$

Отсюда уже вытекает утверждение леммы, поскольку необходимым и достаточным условием равенства

$$\left| \frac{u}{v} \right|_m = \left| \frac{r}{s} \right|_m$$

является сравнение

$$us \equiv vr \pmod{m}.$$

□

<sup>1)</sup> Именно это условие определяет длину цикла вычислений и, тем самым, число  $n$ , которое заранее неизвестно. — *Прим. перев.*

6.28. **Пример.** Пусть  $m = 19$  и  $d/c = 8/10$ . Тогда вычисления с начальной матрицей

$$\begin{bmatrix} 19 & 0 \\ 10 & 8 \end{bmatrix}$$

приводят к результатам, собранным в следующей таблице:

	19	0
	10	8
1	9	-8
1	1	16
9	0	-152

Таким образом,

$$\left| \frac{8}{10} \right|_{19} = \left| \frac{-8}{9} \right|_{19} = |16|_{19} = 16.$$

6.29. **Замечание.** Следует еще раз подчеркнуть то важное обстоятельство, связанное с рекурсией (6.19), что при условии

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \equiv 0 \pmod{w},$$

все пары целых чисел в алгоритме представляют рациональные числа из одного обобщенного класса вычетов  $\mathbb{Q}_k$  для некоторого  $k \in [0, w)$ .

Теперь переходим к главному результату.

6.30. **Теорема.** Пусть заданы некоторое рациональное число  $r/s$  ( $s > 0$ ) и целое  $m$ , такие, что  $(s, m) = 1$ . Тогда алгоритм 6.26 с начальной матрицей

$$\begin{bmatrix} m & 0 \\ s & r \end{bmatrix}$$

закончится на некотором шаге  $n + 1$  (характеризуемом условием  $a_{n+1} = 0$ ) и будет вычислен результат

$$\left| \frac{r}{s} \right|_m = |b_n|_m.$$

**Доказательство.** Поскольку  $a_n = (s, m) = 1$ , из леммы 6.27 следует, что

$$\left| \frac{b_n}{a_n} \right|_m = |b_n|_m = \left| \frac{r}{s} \right|_m. \quad \square$$

6.31. **Пример.** Для вычисления  $|10/13|_{625}$  используем в алгоритме 6.26 начальную матрицу

$$\begin{bmatrix} 625 & 0 \\ 13 & 10 \end{bmatrix}.$$

Результаты вычислений представлены в следующей таблице:

	625	0
	13	10
48	1	-480
13	0	6250

Отсюда заключаем, что

$$\left| \frac{10}{13} \right|_{625} = |-480|_{625} = 145.$$

Укажем, что в этом примере  $m = 5^4 = 625$  не является простым числом, но  $(625, 13) = 1$  и теорема 6.30 применима.

В примере 6.31 исходное рациональное число  $10/13$  является дробью Фарея порядка 17. Однако в общем случае дробь  $r/s$  в теореме 6.30 не обязана быть дробью Фарея порядка  $N$ . Выберем, например, в примере 6.31 вместо  $10/13$  число  $-5/56$ . Применяя алгоритм 6.26, опять получим целое число  $145 \in \mathbb{P}_{625}$ . Таким образом, дроби  $10/13$  и  $-5/56$  принадлежат одному обобщенному классу вычетов, а именно классу  $\mathbb{Q}_{145}$ . См. пример 6.33.

### Обратное отображение (Корнеруп, Кришнамурти)

В алгоритме (6.19) генерируется последовательность пар целых чисел  $\{(a_1; b_1), (a_2; b_2), \dots\}$  и с каждой парой можно связать рациональное число. Последовательность таких рациональных чисел

$$\left\{ \frac{b_1}{a_1}, \frac{b_2}{a_2}, \dots \right\}$$

имеет свойство

$$(6.32) \quad \left| \frac{b_i}{a_i} \right|_m = \left| \frac{r}{s} \right|_m, \quad i = 1, 2, \dots,$$

если в (6.19) за начальную матрицу выбрана

$$\begin{bmatrix} m & 0 \\ s & r \end{bmatrix},$$

причем последовательность  $\{q_1, q_2, \dots\}$  может быть произвольной. В этом можно убедиться, привлекая одновременно леммы 6.23 и 6.27 (как в теореме 6.30).

Таким образом, полагая  $(s; r) = (k; 1)$ , где  $k \in [0, m)$ , в начальной матрице алгоритма (6.19), можно вычислить разнообразные дроби из одного обобщенного класса вычетов  $Q_k$ . Если возможен выбор (единственной) дроби Фарея порядка  $N$  среди этих элементов  $Q_k$ , то мы имеем тем самым метод для выполнения обратного отображения  $\hat{\Pi}_m \rightarrow \mathbb{F}_N$ .

**6.33. Пример.** «Обратим» отображение, найденное в предыдущем примере, т. е. переведем целое  $145 \in \hat{\Pi}_{625}$  в дробь  $10/13 \in \mathbb{F}_{17}$ .

Поскольку  $k = 145$ , начальная матрица имеет вид

$$\begin{bmatrix} 625 & 0 \\ 145 & 1 \end{bmatrix}.$$

Результаты вычислений отражены в следующей таблице (значения  $q_i$  выбирались при помощи (6.25), поэтому таблица получилась конечной длины).

	625	0
	145	1
4	45	-4
3	10	13
4	5	-56
2	0	125

Пара чисел, отвечающих дроби  $10/13$ , выделена. Отметим, что числа 145 и 625 не являются взаимно простыми и

$$a_n = (625, 145) = 5,$$

как описано в алгоритме 6.26.

В предыдущей таблице из множества

$$\left\{ -\frac{45}{4}, \frac{10}{13}, -\frac{5}{56} \right\} \subset Q_{145}$$

легко выделить число  $10/13$ , являющееся дробью Фарея порядка  $17^1$ ).

Напомним, что теоремой 5.14 гарантируется единственность дроби Фарея порядка  $N$  в классе  $Q_k$ . Теперь возникает очевидный вопрос: если предположить существование дроби

<sup>1)</sup> Например, последовательно проверяя для каждого элемента, выполнены ли условия определения (5.13) дробей Фарея. — *Прим. перев.*



Фарея порядка  $N$  в  $\mathbb{Q}_k$ , то обязана ли эта дробь Фарея присутствовать среди рациональных чисел, порожденных алгоритмом 6.26 (в частности, среди чисел, представленных в предыдущей таблице)? Если последовательность целых чисел  $\{q_1, q_2, \dots, q_{n+1}\}$  из алгоритма 6.26 определяется в соответствии с правилом (6.25), то ответ оказывается утвердительным. Для доказательства этого потребуются некоторые дополнительные утверждения, связанные с понятием цепной дроби (см. [Koenig, Gregory, 1983]).

Расширим начальную матрицу

$$(6.34) \quad \begin{bmatrix} a_{-1} & b_{-1} & c_{-1} \\ a_0 & b_0 & c_0 \end{bmatrix} = \begin{bmatrix} m & 0 & -1 \\ k & 1 & 0 \end{bmatrix}$$

и определим последовательность троек целых чисел  $\{(a_i, b_i, c_i)\}$  при помощи следующей рекурсии: пока  $a_{i-1} \neq 0$ , полагаем

$$(6.35) \quad q_i = \left\lfloor \frac{a_{i-2}}{a_{i-1}} \right\rfloor \quad \text{и} \quad \begin{aligned} a_i &= a_{i-2} - q_i a_{i-1}, \\ b_i &= b_{i-2} - q_i b_{i-1}, \\ c_i &= c_{i-2} - q_i c_{i-1} \end{aligned}$$

для  $i = 1, 2, \dots, n+1$ . Введем последовательность дробей

$$(6.36) \quad \left\{ \frac{|b_1|}{|c_1|}, \frac{|b_2|}{|c_2|}, \dots, \frac{|b_{n+1}|}{|c_{n+1}|} \right\}.$$

В терминологии теории цепных дробей это полная совокупность подходящих дробей всех порядков для рационального числа  $m/k$  (см. [Hardy, Wright, 1960<sup>1)</sup>]). В дальнейшем мы используем соотношение<sup>2)</sup>

$$(6.37) \quad a_i = kb_i - mc_i, \quad i = 1, 2, \dots, n+1,$$

и следующую теорему, которая описывает свойство «наилучшей рациональной аппроксимации», характеризующее подходящие дроби.

**6.38. Теорема.** Если дробь  $r/s$  удовлетворяет неравенству

$$\left| \alpha - \frac{r}{s} \right| < \frac{1}{2s^2},$$

то она является подходящей дробью числа  $\alpha$ .

**Доказательство.** См. [Hardy, Wright, 1960, с. 153]<sup>3)</sup>.  $\square$

<sup>1)</sup> См. также Хинчин А. Я. Цепные дроби. — 4-е изд. — М.: Наука, 1978, с. 9—11. — *Прим. перев.*

<sup>2)</sup> См. задачу 2 из упражнений 1.6 (а также цитированную в предыдущем примечании книгу А. Я. Хинчина, с. 35—39. — *Прим. перев.*).

<sup>3)</sup> См. также цитированную выше книгу А. Я. Хинчина, с. 42. — *Прим. перев.*

Теперь мы можем доказать следующую теорему, в которой устанавливается, что при наличии в обобщенном классе вычетов  $\mathbb{Q}_k$  дроби Фарея порядка  $N$  эта дробь обязательно появится в последовательности, порожденной алгоритмом 6.26.

**6.39. Теорема (Корнеруп).** *Если некоторая дробь Фарея  $r/s$  порядка  $N$  удовлетворяет условию*

$$\left| \frac{r}{s} \right|_m = k,$$

где  $k \in \hat{\Pi}_m$  и

$$0 < r \leq N, \quad 0 < |s| \leq N,$$

то существует номер  $i$ , при котором

$$(r; s) = (a_i; b_i),$$

где  $\{(a_j; b_j)\}$ ,  $j = 1, 2, \dots, n+1$  — набор пар целых чисел, генерированных алгоритмом 6.26<sup>1)</sup> с начальной матрицей

$$\begin{bmatrix} m & 0 \\ k & 1 \end{bmatrix}.$$

**Доказательство.** Расширим начальную матрицу по аналогии с (6.34) и определим последовательность  $\{c_i\}$ ,  $i = 1, 2, \dots, n+1$ , по формулам (6.35); тогда в (6.36) будет представлен полный набор подходящих дробей рационального числа  $m/k$  при  $k \neq 0$ . По предположению

$$\left| \frac{r}{s} \right|_m = |k|_m = k \in \hat{\Pi}_m.$$

Следовательно,

$$r \equiv ks \pmod{m},$$

откуда можно сделать вывод о существовании единственного целого числа  $t$ , такого, что

$$r = ks - mt.$$

Это позволяет провести следующую цепочку выкладок:

$$\begin{aligned} \left| \frac{k}{m} - \frac{t}{s} \right| &= \left| \frac{ks - mt}{ms} \right| = \left| \frac{r}{ms} \right| \leq \frac{1}{s^2} \cdot \frac{|s| \cdot N}{2N^2 + 1} \leq \\ &\leq \frac{1}{s^2} \cdot \frac{N^2}{2N^2 + 1} < \frac{1}{2s^2}. \end{aligned}$$

Таким образом, используя теорему 6.38, заключаем, что или  $t/s$ , или  $(-t)/(-s)$  является подходящей дробью числа  $k/m$ .

<sup>1)</sup> Со значениями  $q_i$  из (6.25). — Прим. перев.

Поскольку последовательность (6.36) составлена из подходящих дробей числа  $m/k$ , то последовательность

$$\left\{ \frac{0}{1}, \frac{|c_1|}{|b_1|}, \dots, \frac{|c_{n+1}|}{|b_{n+1}|} \right\}$$

составлена из подходящих дробей числа  $k/m$ . Следовательно, существует такой номер  $i \in [1, n+1]$ , что  $|b_i| = |s|$  и

$$\frac{t}{s} = \frac{c_i}{b_i}.$$

Наконец, из (6.37) имеем

$$\frac{a_i}{b_i} = k - m \frac{c_i}{b_i}.$$

Заметим также, что равенство  $r = ks - mt$  эквивалентно равенству

$$\frac{r}{s} = k - m \frac{t}{s}.$$

Таким образом,

$$\frac{a_i}{b_i} = \frac{r}{s},$$

и тогда

$$(r; s) = (a_i; b_i).$$

поскольку  $r$  и  $a_i$  положительны. □

#### Метод «общего знаменателя»

Если дробь Фарея  $a/b$  порядка  $N$  является образом целого числа  $k \in \hat{\Pi}_m$  и если можно найти число, кратное  $b$ , то существует более простой (по сравнению с алгоритмом 6.26) метод для вычисления  $a/b$  по заданному  $k$ .

**6.40 Теорема.** Предположим, что  $x = a/b$  — дробь Фарея порядка  $N$  и

$$k = |ab^{-1}|_m.$$

Целое число  $k$  можно отобразить на  $x$  следующим образом. Предположим, что известно произведение  $t \cdot b$ , где  $t$  — целое число из полуинтервала  $(0, N]$ . Тогда

$$ta = |(tb)k|_m,$$

где  $| \cdot |$  обозначает отображение, введенное определением 2.24. Ясно, что

$$x = \frac{ta}{tb}.$$

**Доказательство.** По предположению  $ta = (tb)x$ , откуда следует

$$/ta/m = /(tb)x/m = /(tb)|x|_m/m = /(tb)k/m.$$

Однако поскольку  $|a| \leq N$  и  $t \in (0, N]$ , то

$$|ta| \leq N^2 \leq (m-1)/2.$$

Поэтому  $/ta/m = ta$ , что влечет за собой равенства

$$ta = /(tb)k/m$$

и

$$x = \frac{ta}{tb}. \quad \square$$

6.41. **Пример.** Вернемся к примеру из замечания 5.23:

$$x = \frac{1}{2} + \left(-\frac{2}{3}\right) + \left(-\frac{1}{6}\right).$$

В арифметике  $(\mathbb{I}_{19}, +, \cdot)$  получаем

$$|x|_{19} = \left| \frac{1}{2} + \left(-\frac{2}{3}\right) + \left(-\frac{1}{6}\right) \right|_{19} = |10 + 12 + 3|_{19} = 6,$$

т. е.  $k = 6$ . Заметим, что общий знаменатель трех дробей в исходном выражении для  $x$  равен 6. Выбираем  $tb = 6$  и получаем

$$ta = /(tb)k/_{19} = /6 \cdot 6/_{19} = -2.$$

Следовательно,

$$x = \frac{ta}{tb} = -\frac{1}{3}.$$

В этих вычислениях  $m = 19$ ,  $N = 3$  и  $t = 2$ . Условие  $t \in (0, N]$ , очевидно, выполнено, поэтому получается правильный ответ. Общий метод априорного определения числа  $t \in (0, N]$  неизвестен. Однако можно пытаться наугад выбирать какое-либо кратное  $b$  и затем проверять, является результат дробью Фарея порядка  $N$  или нет.

6.42. **Пример.** Положим  $m = 19$ ,  $N = 3$ ,  $k = |x|_{19} = 8$ , тогда дробь Фарея, которую мы ищем, равна

$$x = -\frac{3}{2}.$$

Можно использовать тот факт, что  $b = 2$  (конечно, в общем случае такой информации нет). Покажем, что получится, если выбрать  $tb = 6$  или  $tb = 8$ .

Для случая  $tb = 6$  имеем

$$ta = /6 \cdot 8/_{19} = -9,$$

что влечет за собой равенство

$$x = -\frac{3}{2}.$$

Поскольку  $x \in \mathbb{F}_3$ , то ответ верен (здесь  $t = 3$ ).

Для случая  $tb = 8$  имеем

$$ta = /8 \cdot 8/_{19} = 7;$$

следовательно,

$$x = \frac{7}{8}.$$

Поскольку и числитель, и знаменатель превосходят  $N = 3$ , то налицо псевдопереполнение. Это неудивительно, так как здесь  $t = 4 > N = 3$ .

Отметим, однако, что оба ответа  $-3/2$  и  $7/8$  принадлежат одному обобщенному классу вычетов потому, что

$$\left| -\frac{3}{2} \right|_{19} = \left| \frac{7}{8} \right|_{19} = 8.$$

Очевидно, что  $-3/2$  является единственной дробью Фарея порядка 3 в  $\mathbb{Q}_8$ , а  $7/8$  — просто один из элементов бесконечного множества  $\mathbb{Q}_8$ .

$$\left| \frac{a}{b} \right|_{19} = 8.$$

**6.43. Замечание.** В теореме 6.40 условие

$$t \in (0, N]$$

является *достаточным*, но отнюдь не *необходимым*. Чтобы увидеть это, вернемся к примеру 6.41, в котором

$$|x|_{19} = 6.$$

Выбирая  $tb = 27$ , получаем

$$ta = /27 \cdot 6/_{19} = -9,$$

откуда вытекает

$$x = -\frac{1}{3}.$$

Это правильный ответ; именно он был получен ранее. В рассматриваемом случае  $t = 9$ , что превосходит число  $N = 3$ .

### Упражнения 1.6

1. Пусть  $a = 49$  и  $b = 63$ . Найти целые числа  $x$  и  $y$ , для которых  $(a, b) = ax + by$ .

2. Перепишем табл. 6.11 в других обозначениях:

	$a$	1	0
	$b$	0	1
$q_1$	$r_1$	$s_1$	$t_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$q_n$	$r_n$	$s_n$	$t_n$
$q_{n+1}$	0	$s_{n+1}$	$t_{n+1}$

Доказать, что  $r_i = as_i + bt_i$ ,  $i = 1, 2, \dots, n+1$ .

3. Вычислить обратный элемент числа 321 по модулю 625.

4. Вычислить  $|-15/14|_{625}$  и  $|-4/5|_{81}$ .

5. Пусть  $m = 625$ ,  $N = 17$ . Какие дроби Фарея порядка  $N$  соответствуют целым числам 259, 367 и 530? (Использовать обратное отображение, описанное в данном параграфе.)

6. Пусть  $m = 81$ ,  $N = 6$ . Какие дроби Фарея порядка  $N$  соответствуют целым числам 16, 39 и 66? (См. указание к задаче 5.)

7. Вычислить значение

$$x = \frac{\frac{3}{7} - \frac{5}{11}}{\frac{7}{11} - \frac{3}{11}},$$

применяя одномодульную арифметику вычетов с модулем  $m = 625$ .

8. Доказать, что 369 не входит в  $\hat{\Pi}_{625}$ , хотя принадлежит  $\Pi_{625}$ .

## § 7. Многомодульная арифметика вычетов для рациональных чисел

Как и в § 3, введем векторное основание  $\beta$  с компонентами равными модулям  $m_1, m_2, \dots, m_n$ , и положим  $M = m_1 m_2 \dots m_n$ . В этом параграфе будем предполагать, что  $N$  является наибольшим целым числом, удовлетворяющим неравенству

$$2N^2 + 1 \leq M.$$

Также предположим простоту каждого модуля  $m_i$ , что обеспечивает выполнение условия

$$(m_i, m_j) = 1, \quad i, j = 1, \dots, n, \quad i \neq j.$$

В § 5 было показано, что система  $(\hat{\mathbb{Q}}, +, \cdot)$  является коммутативным кольцом с единицей (теорема 5.10) и существует гомоморфизм, отображающий это кольцо на конечное коммутативное кольцо  $(\Pi_m, +, \cdot)$ ; здесь  $m$  — число вида  $p^r$ , где  $p$  — простое. С другой стороны, многомодульная арифметика вычетов с векторным основанием  $\beta$  эквивалентна одномодульной арифметике в  $(\Pi_M, +, \cdot)$ . Чтобы развить многомодульную арифметику вычетов для рациональных чисел, необходимо

для некоторого коммутативного кольца с единицей  $(\tilde{\mathbb{Q}}, +, \cdot)$  (его точное определение приводится ниже) предложить гомоморфизм на конечное коммутативное кольцо  $(\Pi_M, +, \cdot)$  по аналогии с определением гомоморфизма из § 5.

Изложение будет полностью параллельно тому, что дано в § 5, с небольшими вариациями, связанными с тем, что в § 5 имеем модуль  $m = p^r$  — степень простого числа, тогда как в настоящем параграфе модуль  $M = m_1 m_2 \dots m_n$  — произведение  $n$  различных простых чисел.

Для перевода рационального числа  $x = a/b$  в целое из  $\Pi_M$  требуется, чтобы существовал элемент  $b^{-1}(M)$ , ибо, как в определении 5.1,

$$(7.3) \quad \left| \frac{a}{b} \right|_M = |ab^{-1}|_M.$$

В нашем случае  $b^{-1}(M)$  существует тогда и только тогда, когда  $(b, M) = 1$ , а это эквивалентно условиям

$$(7.4) \quad (b, m_i) = 1, \quad i = 1, 2, \dots, n.$$

Следовательно, в  $\Pi_M$  нельзя отобразить те и только те рациональные числа, для которых (7.4) нарушено хотя бы при одном  $i$ .

Пусть  $\tilde{\mathbb{Q}}$  — множество всех рациональных чисел, которые можно перевести в  $\Pi_M$ , т. е. пусть

$$(7.5) \quad \tilde{\mathbb{Q}} = \left\{ \frac{a}{b} : (b, m_i) = 1, i = 1, 2, \dots, n \right\}.$$

Тогда равенство (7.3) определяет отображение  $|\cdot|_M: \tilde{\mathbb{Q}} \rightarrow \Pi_M$ .

Оказывается, что каждое целое число  $j \in \Pi_M$  есть образ некоторого бесконечного подмножества  $\mathbb{Q}_j \subset \tilde{\mathbb{Q}}$ :

$$(7.6) \quad \mathbb{Q}_j = \left\{ \frac{a}{b} \in \tilde{\mathbb{Q}} : \left| \frac{a}{b} \right|_M = j \right\},$$

а тогда

$$(7.7) \quad \tilde{\mathbb{Q}} = \bigcup_{j=0}^{M-1} \mathbb{Q}_j.$$

Например, множество  $\mathbb{Q}_0$  (рациональных чисел из  $\tilde{\mathbb{Q}}$ , которые переходят в нуль) состоит в точности из тех чисел вида  $a/b$  с  $(b, m_i) = 1, i = 1, 2, \dots, n$ , для которых  $|a|_M = 0$ . Как и в § 5, непересекающиеся подмножества  $\mathbb{Q}_0, \mathbb{Q}_1, \dots, \mathbb{Q}_{M-1}$  назовем *обобщенными классами вычетов (по модулю  $M$ )*. По аналогии с теоремой 5.6 и следствием 5.7 имеем следующие результаты.

7.8. **Теорема.** Пусть  $x = a/b$  и  $y = c/d$ , причем  $b^{-1}(M)$  и  $d^{-1}(M)$  существуют. Равенство

$$|x|_M = |y|_M$$

имеет место тогда и только тогда, когда

$$ad \equiv bc \pmod{M}.$$

**Доказательство** такое же, как в теореме 5.6.

7.9. **Следствие.** Пусть  $x = a/b$  и  $y = c/d$  принадлежат  $\tilde{Q}$ . В таком случае  $x$  и  $y$  входят в один обобщенный класс вычетов  $Q_i$  тогда и только тогда, когда

$$ad \equiv bc \pmod{M}.$$

**Доказательство** такое же, как в следствии 5.7.

7.10. **Замечание.** Для числа  $x = a/b$  представление  $|x|_M$  не существует тогда и только тогда, когда  $(b, m_i) \neq 1$  хотя бы для одного  $i$ . Поскольку предполагается  $(a, b) = 1$ , то можно заключить, что представление  $|x|_M$  не существует тогда и только тогда, когда  $b$  является целым кратным хотя бы одного из числа  $m_i$ .

Таким образом, устанавливая результаты, аналогичные лемме 5.9, теореме 5.10 и лемме 5.11, приходим к следующему утверждению, которое соответствует теореме 5.12.

7.11. **Теорема.** Отображение  $|\cdot|_M: \tilde{Q} \rightarrow \Pi_M$  есть гомоморфизм по отношению к операциям сложения и умножения.

**Доказательство.** См. текст, предшествующий теореме 5.12.

Теорема 7.11 гласит, что арифметическим операциям в кольце  $(\tilde{Q}, +, \cdot)$  соответствуют те же арифметические операции в кольце  $(\Pi_M, +, \cdot)$ . Следовательно, как и в § 5, представляет интерес замена действий с рациональными числами из  $\tilde{Q}$  на действия с целыми числами из  $\Pi_M$ .

Нетрудно доказать теорему, аналогичную теореме 5.14, о единственности дроби Фарея порядка  $N$ , если  $N$  удовлетворяет (7.1), в обобщенном классе вычетов. Кроме того, поскольку число различных дробей Фарея порядка  $N$  меньше, чем  $M$  — число различных обобщенных классов вычетов, то не каждый класс содержит дробь Фарея из  $\mathbb{F}_N$ .

Введем обозначение для множества образов всех дробей Фарея порядка  $N$ :

$$\hat{\Pi}_M = \left\{ \left| \frac{a}{b} \right|_M : \frac{a}{b} \in \mathbb{F}_N \right\}.$$



Отображение  $|\cdot|_M: \mathbb{F}_N \rightarrow \tilde{\mathbb{P}}_M$  взаимно однозначно и его область значений совпадает с  $\tilde{\mathbb{P}}_M$ ; поэтому оно имеет обратное (напомним теорему 5.17).

**7.13. Пример.** Пусть  $m_1 = 5$ ,  $m_2 = 7$ , так что  $M = 35$ . Тогда  $N = 4$ . Действие отображения  $|\cdot|_{35}: \mathbb{F}_4 \rightarrow \tilde{\mathbb{P}}_{35}$  иллюстрируется следующей таблицей:

0	0		
1	1	-1	34
2	2	-2	33
3	3	-3	32
4	4	-4	31
.	.	.	.
$-\frac{3}{4}$	8	$\frac{3}{4}$	27
$-\frac{1}{4}$	9	$-\frac{1}{4}$	26
.	.	.	.
$-\frac{2}{3}$	11	$\frac{2}{3}$	24
$-\frac{1}{3}$	12	$-\frac{1}{3}$	23
$-\frac{4}{3}$	13	$-\frac{4}{3}$	22
.	.	.	.
$-\frac{3}{2}$	16	$\frac{3}{2}$	19
$-\frac{1}{2}$	17	$\frac{1}{2}$	18

Очевидно<sup>1)</sup>, что  $\tilde{\mathbb{P}}_{35} \subset \mathbb{P}_{35}$ . Заметим, что дроби Фарея порядка 4 в таблице расположены группами. В каждой группе не менее двух элементов. Внутри группы соседние элементы отличаются друг от друга на единицу. При сравнении двух столбцов таблицы можно также заметить «косую симметрию». Эти свойства также проявляются в примере 5.18 и в таблицах из работы [Rao, Gregory, 1981].

### Система с двумя модулями

Все, что было установлено до сих пор, аналогично материалу § 5 для одномодульных систем с той только разницей, что модуль  $m = p'$  заменен на модуль  $M = m_1 m_2 \dots m_n$ . Теперь рассмотрим отображение дробей Фарея порядка  $N$  не на множество  $\tilde{\mathbb{P}}_M$ , а на меньшие множества  $\mathbb{P}_{m_1}$ ,  $\mathbb{P}_{m_2}$ , ...,  $\mathbb{P}_{m_n}$ .

**7.14. Пример.** Продолжим изучение примера 7.13 с  $m_1 = 5$ ,  $m_2 = 7$ . Рассматривая отображения в  $\mathbb{P}_m$  для  $m_1 = 5$  или

<sup>1)</sup> В этой таблице, как и в таблице из примера 5.18, многоточие поставлено на месте чисел из  $\tilde{\mathbb{P}}_{35}$ , не входящих в  $\mathbb{P}_{35}$ , и призвано напомнить, что  $\tilde{\mathbb{P}}_{35} \subset \mathbb{P}_{35}$ . — Прим. перев.

$m_2 = 7$ , находим, что многие дроби Фарея порядка 4 имеют одинаковые образы. При  $m_1 = 5$  имеем

0	0				
1	$-\frac{1}{4}$	-4	1	$-\frac{3}{4}$	$-\frac{3}{2}$
2	$-\frac{4}{3}$	$\frac{3}{4}$	-3	2	$\frac{1}{3}$ $-\frac{1}{2}$
3	$\frac{1}{2}$	$-\frac{1}{3}$	-2	3	$-\frac{3}{4}$ $\frac{4}{3}$
4	$\frac{3}{2}$	$\frac{2}{3}$	-1	4	$\frac{1}{4}$

а при  $m_2 = 7$

0	0				
1	$-\frac{4}{3}$	1	$-\frac{3}{4}$		
2	$-\frac{1}{3}$	2	$\frac{1}{4}$	$-\frac{3}{2}$	
3	$\frac{2}{3}$	-4	3	$-\frac{1}{2}$	
4	$\frac{1}{2}$	-3	4	$-\frac{2}{3}$	
5	$\frac{3}{2}$	$-\frac{1}{4}$	-2	$\frac{1}{3}$	
6	$\frac{3}{4}$	-1	$\frac{4}{3}$		

В этом примере и  $m_1 = 5$ , и  $m_2 = 7$  больше, чем  $N = 4$ . Следовательно, у ненулевых дробей Фарея порядка 4 знаменатели не нарушают условия (7.4). Таким образом, существуют обратные элементы  $b^{-1}(m_i)$  для  $i = 1$  и для  $i = 2$ , и тем самым существуют представления  $|a/b|_{m_i}$   $i = 1, 2$ , для всех дробей Фарея. Напрашивается вопрос: всегда ли можно выбрать модули  $m_1, m_2, \dots, m_n$  так, чтобы

$$(7.15) \quad m_i > N, \quad i = 1, 2, \dots, n?$$

Ответ оказывается положительным для  $n = 2$  и отрицательным при  $n > 2$ .

**7.16. Теорема.** Пусть  $M = m_1 m_2$ , где  $m_1$  и  $m_2$  — соседние простые числа, и пусть  $N > 0$  является наибольшим целым числом, для которого верно неравенство

$$2N^2 + 1 \leq M.$$

Тогда и  $m_1$ , и  $m_2$  больше, чем  $N$ .

**Доказательство** (Матула). Пусть  $m_2$  — простое число, следующее за  $m_1$  в ряду натуральных чисел. Тогда из посту-

лата Бертрана (см. [Hardy, Wright, 1960, с. 343]<sup>1)</sup>) следует, что

$$m_1 < m_2 < 2m_1.$$

Следовательно,

$$2m_1^2 > m_1 m_2 \geq 2N^2 + 1$$

в силу предположения. Но тогда  $m_1 > N$ . Таким образом,  $m_1$  и  $m_2$  больше, чем  $N$ .

**7.17. Теорема.** Пусть  $m_1 < m_2 < \dots < m_n$  — произвольные простые числа,  $M$  — их произведение, и пусть  $N > 0$  является наименьшим целым числом, для которого верно неравенство

$$2N^2 + 1 \leq M.$$

Если  $n > 2$ , то  $m_1 \leq N$ .

**Доказательство.** В силу предположения

$$N \leq \left[ \frac{(m_1 m_2 \dots m_n) - 1}{2} \right]^{1/2} < N + 1,$$

откуда вытекает

$$\frac{(m_1 m_2 \dots m_n) - 1}{2} < (N + 1)^2.$$

Следовательно,

$$m_1 m_2 \dots m_n < 2N^2 + 4N + 3.$$

Если  $m_2 \leq N$ , то утверждение теоремы, очевидно, верно. Предположим  $m_2 > N$ , тогда

$$m_1 < \frac{2N^2 + 4N + 3}{N^{n-1}} = \frac{2}{N^{n-3}} + \frac{4}{N^{n-2}} + \frac{3}{N^{n-1}}.$$

Поскольку тремя наименьшими простыми числами являются числа 2, 3 и 5, то  $N \geq 3$ . Таким образом, при  $n > 2$  получаем  $m_1 < 2 + 4/3 + 1/3 = 11/3$ , что влечет за собой неравенства  $m_1 \leq 3 \leq N$ .  $\square$

**7.18. Замечание.** Всегда найдутся два соседних простых числа  $m_1, m_2$ , превосходящие заданное  $N$ , для которых верно неравенство <sup>2)</sup>

$$m_1 m_2 \geq 2N^2 + 1,$$

<sup>1)</sup> Постулат Бертрана: при натуральном  $n > 3$  существует простое число, большее  $n$  и меньшее  $2n - 2$ . Более слабая формулировка: при любом  $x > 1$  в интервале  $(x, 2x)$  имеется простое число. См. Математическая энциклопедия, т. 1. — М.: Советская энциклопедия, 1977, с. 433. — Прим. перев.

<sup>2)</sup> Где  $N$  нельзя увеличить, не нарушая это неравенство. — Прим. перев.

что следует из теоремы 7.16. Однако подходящая пара простых чисел не обязательно составлена из соседей, что иллюстрируется в следующей таблице:

$N$	$2N^2 + 1$	$M = m_1 m_2$	Другой выбор
3	19		
4	33	$35 = 5 \cdot 7$	
5	51		
6	73	$77 = 7 \cdot 11$	$91 = 7 \cdot 13$
7	99		
8	129	$143 = 11 \cdot 13$	
9	163	$187 = 11 \cdot 17$	
10	201	$209 = 11 \cdot 19$	$221 = 13 \cdot 17$
11	243	$247 = 13 \cdot 19$	
12	289	$299 = 13 \cdot 23$	$323 = 17 \cdot 19$

7.19. Задача. Заново рассмотрим задачу 5.20 и вычислим сумму

$$x = \frac{1}{3} + \left(-\frac{2}{3}\right)$$

при помощи системы вычетов с векторным основанием

$$\beta = [5, 7].$$

В  $(\Pi_5, +, \cdot)$  вычисляем

$$|x|_5 = \left| \left| \frac{1}{3} \right|_5 + \left| -\frac{2}{3} \right|_5 \right|_5 = |2 + 1|_5 = 3,$$

и в  $(\Pi_7, +, \cdot)$  вычисляем

$$|x|_7 = \left| \left| \frac{1}{3} \right|_7 + \left| -\frac{2}{3} \right|_7 \right|_7 = |5 + 4|_7 = 2.$$

Эти выкладки эквивалентны записи операндов в виде

$$\left| \frac{1}{3} \right|_\beta = [2, 5],$$

$$\left| -\frac{2}{3} \right|_\beta = [1, 4]$$

и формированию покомпонентной суммы

$$|x|_\beta = [3, 2].$$

Теперь используем метод, показанный в задаче 4.39, чтобы найти целое число  $|x|_{35} \in \tilde{\Pi}_{35}$ , которое соответствует решению  $x \in \mathbb{F}_4$ . Ход вычислений представлен в следующей таблице:

$\beta$	$m_1 = 5$	$m_2 = 7$	
$ t_1 _\beta$	$\begin{smallmatrix} \vdots & 3 & \vdots \end{smallmatrix}$	2	
$ d_0 _\beta$	3	3	вычесть
$ t_1 - d_0 _\beta$	0	6	
$m_1^{-1}(\beta_1)$		3	умножить
$ t_2 _{\beta_1} =  (t_1 - d_0)/m_1 _{\beta_1}$		$\begin{smallmatrix} \vdots & 4 & \vdots \end{smallmatrix}$	

Следовательно,

$$|x|_{35} = 3 + 4 \cdot 5 = 23,$$

и это целое число в  $\tilde{\mathbb{P}}_{35}$  переводится в дробь  $-1/3$  из  $\mathbb{F}_4$  при помощи обратного отображения, представленного в примере 6.33. Действительно,

	35	0
	23	1
1	12	-1
1	11	2
1	$\begin{smallmatrix} \vdots & 1 & \vdots \end{smallmatrix}$	$\begin{smallmatrix} \vdots & -3 & \vdots \end{smallmatrix}$
11	0	35

Заметим, что  $x = -1/3$  — единственная дробь Фарея порядка 4 среди дробей  $-12/1$ ,  $11/2$  и  $-1/3$ , принадлежащих классу  $\mathbb{Q}_{23}$ .

### Многомодульная система вычетов с $n > 2$

Весь дальнейший материал до конца параграфа написан после того, как рукопись книги была сдана в издательство. Таким образом, нигде больше в книге (в частности, в «прикладных разделах») этот материал не используется. Однако, считая приведенные результаты важными, мы постарались включить их в текст.

Исходя из теорем 7.16 и 7.17, можно предположить, что в случае рациональных операндов нецелесообразно использовать многомодульные системы вычетов с более чем двумя модулями. Однако это не так, и мы опишем практичный метод вычислений для систем с тремя и более модулями. Этот метод разработал Дэвид Матула и модифицировал Карл Грегори. Мы опишем здесь метод без точного обоснования, поскольку оно будет представлено в статье названных авторов.

Чтобы описать алгоритм Матулы — Грегори, введем векторное основание

$$(7.20) \quad \beta = [m_1, m_2, \dots, m_n],$$

где  $m_1, m_2, \dots, m_n$  — различные простые числа и  $M$  — их произведение. Рассмотрим произвольное ненулевое рациональное число  $a/b$ . Для каждого модуля  $m_i$  можно записать

$$(7.21) \quad \frac{a}{b} = \frac{a_i}{b_i} \cdot (m_i)^{r_i},$$

где

$$(7.22) \quad (a_i, b_i) = (a_i, m_i) = (b_i, m_i) = 1.$$

Очевидно, что целое число  $r_i$  может быть положительным, отрицательным или нулем.

Теперь определим  $N$  как наибольшее целое число, для которого

$$(7.23) \quad 2N^2 + 1 \leq M.$$

Наша цель — установить биективное отображение  $\mathbb{F}_N$  на множество наборов чисел с  $n$  компонентами. Поступим следующим образом.

**7.24. Определение.** Для произвольного ненулевого рационального числа  $a/b$  положим

$$\left| \frac{a}{b} \right|_{\beta} = \left[ \left| \frac{a}{b} \right|_{m_1}^*, \left| \frac{a}{b} \right|_{m_2}^*, \dots, \left| \frac{a}{b} \right|_{m_n}^* \right],$$

где каждая компонента является упорядоченной парой чисел:

$$\left| \frac{a}{b} \right|_{m_i}^* = \left( \left| \frac{a_i}{b_i} \right|_{m_i}; r_i \right), \quad i = 1, 2, \dots, n.$$

Для представления нуля можно использовать любое целое  $z$  (обычно выбирается  $z = 0$ ) в записи  $|0|_{m_i}^* = (0; z_i)$ ,  $i = 1, 2, \dots, n$ .

**7.25. Пример.** Пусть  $\beta = [2, 3, 5, 7]$  и  $a/b = 3/7$ . Поскольку

$$\begin{aligned} \left| \frac{3}{7} \right|_2^* &= (1; 0), \\ \left| \frac{3}{7} \right|_3^* &= (1; 1), \\ \left| \frac{3}{7} \right|_5^* &= (4; 0), \\ \left| \frac{3}{7} \right|_7^* &= (3; -1), \end{aligned}$$

то имеем

$$\left| \frac{3}{7} \right|_{\beta} = [(1; 0), (1; 1), (4; 0), (3; -1)].$$

Отметим, что при  $r_i > 0$  модуль  $m_i$  является сомножителем числителя, а при  $r_i \leq 0$  — сомножителем знаменателя.

Показатель  $r_i$  равен нулю, когда ни числитель, ни знаменатель не делятся на  $m_i$ .

Из (7.21) и определения 7.24 видно, что

$$(7.26) \quad \left| \frac{b}{a} \right|_{m_i}^* = \left( \left| \frac{b_i}{a_i} \right|_{m_i}; -r_i \right), \quad i = 1, 2, \dots, n.$$

Это позволяет по заданному представлению ненулевого рационального числа получить представление обратного к нему.

**7.27. Пример.** В примере 7.25 выписано представление  $|3/7|_\beta$ . Чтобы получить  $|7/3|_\beta$ , вычислим обратный элемент к первой компоненте и сменим знак на противоположный у второй компоненты в каждой паре. Таким образом, поскольку  $\beta = [2, 3, 5, 7]$  и

$$\begin{aligned} (1; 0) &\rightarrow (1; 0), \\ (1; 1) &\rightarrow (1; -1), \\ (4; 0) &\rightarrow (4; 0), \\ (3; -1) &\rightarrow (5; 1), \end{aligned}$$

то можно записать

$$\left| \frac{7}{3} \right|_\beta = [(1; 0), (1; -1), (4; 0), (5; 1)].$$

По заданному представлению любого рационального числа легко также получить представление противоположного элемента при помощи формулы

$$(7.28) \quad \left| -\frac{a}{b} \right|_{m_i}^* = \left( \left| -\frac{a_i}{b_i} \right|_{m_i}; r_i \right), \quad i = 1, 2, \dots, n.$$

**7.29. Пример.** В примере 7.25 выписано представление  $|3/7|_\beta$ . Чтобы получить  $|-3/7|_\beta$ , вычислим противоположный элемент для первой компоненты и сохраним без изменений вторую компоненту в каждой паре. Таким образом, поскольку  $\beta = [2, 3, 5, 7]$  и

$$\begin{aligned} (1; 0) &\rightarrow (1; 0), \\ (1; 1) &\rightarrow (2; 1), \\ (4; 0) &\rightarrow (1; 0), \\ (3; -1) &\rightarrow (4; -1), \end{aligned}$$

то можно записать

$$\left| -\frac{3}{7} \right|_\beta = [(1; 0), (2; 1), (1; 0), (4; -1)].$$

Рассмотрим множество образов всех рациональных чисел

$$(7.30) \quad \mathbb{T}_\beta = \left\{ \left| \frac{a}{b} \right|_\beta : \frac{a}{b} \in \mathbb{Q} \right\}$$

и его конечное подмножество, состоящее из образов дробей Фарея порядка  $N$

$$(7.31) \quad \tilde{\mathbb{T}}_\beta = \left\{ \left| \frac{a}{b} \right|_\beta : \frac{a}{b} \in \mathbb{F}_N \right\}.$$

Согласно (7.31), отображение  $|\cdot|_\beta: \mathbb{F}_N \rightarrow \tilde{\mathbb{T}}_\beta$  есть отображение «на» (сюръекция). Покажем, что оно взаимно однозначно (инъекция), описав конструкцию обратного отображения  $\tilde{\mathbb{T}}_\beta \rightarrow \mathbb{F}_N$ .

**7.32. Определение.** Для рассматриваемого представления  $|a/b|_\beta$  введем величины  $M_r$ ,  $M_0$  и  $M_{-r}$  следующим образом:

$$(i) \quad M_r = \prod_i m_i,$$

где сомножителями являются те и только те модули  $m_i$ , для которых  $r_i > 0$  в определении 7.24. Если таких модулей нет, то полагаем  $M_r = 1$ . Далее

$$(ii) \quad M_0 = \prod_j m_j,$$

где сомножителями являются те и только те модули  $m_j$ , для которых  $r_j = 0$  в определении 7.24. Если таких нет, то необходимо выбрать другое векторное основание. Наконец,

$$(iii) \quad M_{-r} = \prod_k m_k,$$

где модули  $m_k$  отвечают  $r_k < 0$  в определении 7.24. Если таких нет, то  $M_{-r} = 1$ .

Очевидно, что

$$(7.33) \quad M = M_r M_0 M_{-r}.$$

**7.34. Определение.** Пусть задано представление  $|a/b|_\beta$ . Целое число  $q$  из множества  $\{0, 1, \dots, M_0 - 1\}$  определяется условиями

$$|q|_{m_j}^* = \left| \frac{a}{b} \right|_{m_j}^*,$$

которые должны выполняться для всех модулей  $m_j$  из  $M_0$ .

**7.35. Определение.** Положим

$$q' = [q M_{-r} M_r^{-1} (M_0)]_{M_0}.$$



В следующей теореме дается алгоритм для выполнения отображения  $\mathbb{T}_\beta$  на  $\mathbb{F}_N$ .

**7.36. Теорема.** По заданному  $|a/b|_\beta$  из  $\mathbb{T}_\beta$  можно найти единственную дробь Фарея  $a/b$  порядка  $N$ , если применить обратное отображение, описанное в § 6, с начальной матрицей

$$\begin{bmatrix} M_r M_0 & 0 \\ M_r q' & M_{-r} \end{bmatrix}.$$

**Доказательство** будет представлено в упоминавшейся выше работе Матулы и Грегори.

**7.37. Пример.** Пусть  $\beta = [2, 3, 5, 7]$ ; тогда  $M = 210$  и  $N = 10$ . Покажем, что представление

$$[(1; 0), (1; 1), (4; 0), (3; -1)]$$

соответствует дроби  $3/7$  в  $\mathbb{F}_{10}$ .

(i) Вычисляем  $M_r = 3$ ,  $M_0 = 10$ ,  $M_{-r} = 7$ .

(ii) Находим целое  $q \in [0, 9]$ , для которого

$$|q|_2^* = (1; 0),$$

$$|q|_5^* = (4; 0).$$

Иными словами, определяем  $q$  из условий

$$|q|_2 = 1,$$

$$|q|_5 = 4.$$

При помощи описанной в задаче 7.19 процедуры с использованием смешанного основания получаем  $q = 9$ .

(iii) Вычисляем  $M_r^{-1}(M_0) = 3^{-1}(10) = 7$ .

(iv) Определяем значение  $q' = |9 \cdot 7 \cdot 7|_{10} = 1$ .

(v) Из этих величин составляем начальную матрицу

$$\begin{bmatrix} M_r M_0 & 0 \\ M_r q' & M_{-r} \end{bmatrix} = \begin{bmatrix} 30 & 0 \\ 3 & 7 \end{bmatrix}$$

При такой начальной матрице имеем

$$\begin{array}{c|cc} & 30 & 0 \\ \hline & \vdots & \vdots \\ & 3 & 7 \\ \hline 10 & 0 & -70 \end{array}$$

Тем самым мы показали, что

$$[(1; 0), (1; 1), (4; 0), (3; -1)] \rightarrow \frac{8}{7}.$$

Арифметика в  $\tilde{T}_\beta$ .

Множество  $\tilde{T}_\beta$  состоит из  $n$ -элементных наборов, причем каждый элемент представляет собой упорядоченную пару чисел (см. определение 7.24). Арифметические операции с этими наборами выполняются поэлементно в соответствии с правилами, представленными ниже. Перед тем как сформулировать эти правила, мы должны ввести некоторые обозначения.

Для каждого элемента с номером  $i = 1, 2, \dots, n$  запишем, используя (7.21) и (7.22),

$$(7.38) \quad \frac{a}{b} = \frac{a_i}{b_i} \cdot (m_i)^{r_i},$$

$$(7.39) \quad \frac{c}{d} = \frac{c_i}{d_i} \cdot (m_i)^{s_i}.$$

Пусть также

$$(7.40) \quad f_i = \left| \frac{a_i}{b_i} \right|_{m_i}, \quad i = 1, 2, \dots, n,$$

$$(7.41) \quad g_i = \left| \frac{c_i}{d_i} \right|_{m_i}, \quad i = 1, 2, \dots, n.$$

Умножение (поэлементное) производится в соответствии с правилом

$$(7.42) \quad (f_i; r_i)(g_i; s_i) = (|f_i g_i|_{m_i}; r_i + s_i), \quad i = 1, 2, \dots, n.$$

**7.43. Пример.** Покажем, что  $3/7 \cdot 7/3 = 1$ , используя результаты примеров 7.25 и 7.27, где  $\beta = [2, 3, 5, 7]$ :

$$\frac{[(1; 0), (1; 1), (4; 0), (3; -1)] \odot [(1; 0), (1; -1), (4; 0), (5; 1)]}{[(1; 0), (1; 0), (1; 0), (1; 0)]}.$$

Ответ является представлением единицы.

Правило сложения (поэлементного) несколько сложнее правила умножения. Чтобы выразить сумму  $(f_i, r_i) + (g_i, s_i)$ , введем обозначение

$$(7.44) \quad h_i = |f_i + g_i|_{m_i}$$

для  $i = 1, 2, \dots, n$ . Правило сложения лучше всего описывается таблицей 7.45. Отметим, что всего имеются только четыре типа элементов (упорядоченных пар): первый тип — элементы вида  $(0, z)$ , остальные три типа — элементы с ненулевой первой компонентой и с положительной, нулевой или от-

## 7.45. Таблица. Правило сложения упорядоченных пар

+	$(0, z)$	$(f_i, r_i)$	$(f_i, 0)$	$(f_i, -r_i)$
$(0, z)$	$(0, z)$	$(f_i, r_i)$	$(f_i, 0)$	$(f_i, -r_i)$
$(g_i, s_i)$	$(g_i, s_i)$	$\begin{cases} (h_i, r_i), r_i = s_i \\ (g_i, s_i), r_i > s_i \\ (f_i, r_i), r_i < s_i \end{cases}$	$(f_i, 0)$	$(f_i, -r_i)$
$(g_i, 0)$	$(g_i, 0)$	$(g_i, 0)$	$(h_i, 0)$	$(f_i, -r_i)$
$(g_i, -s_i)$	$(g_i, -s_i)$	$(g_i, -s_i)$	$(g_i, -s_i)$	$\begin{cases} (h_i, -r_i), r_i = s_i \\ (f_i, -r_i), r_i > s_i \\ (g_i, -s_i), r_i < s_i \end{cases}$

рицательной (соответственно типу) второй компонентой. Предполагается, что  $r_i$  и  $s_i$  положительны и  $f_i g_i \neq 0^1$ .

7.46. **Пример.** Покажем, что  $3/7 - 3/7 = 0$ , используя результаты из примеров 7.25 и 7.29, где  $\beta = [2, 3, 5, 7]$ :

$$\begin{array}{r} [(0; 1), (1; 1), (4; 0), (3; -1)] \\ \oplus [(0; 1), (2; 1), (1; 0), (4; -1)] \\ \hline [(0; 0), (0; 1), (0; 0), (0; -1)] \end{array}.$$

Поскольку у ответа первая компонента в каждой упорядоченной паре равна нулю, то из определения 7.24 заключаем, что это представление числа нуль.

7.47. **Пример.** Покажем, что  $3/7 + 1 = 10/7$ , используя векторное основание  $\beta = [2, 3, 5, 7]$ :

$$\begin{array}{r} [(1; 0), (1; 1), (4; 0), (3; -1)] \\ \oplus [(1; 0), (1; 0), (1; 0), (1; 0)] \\ \hline [(0; 0), (1; 0), (0; 0), (3; -1)] \end{array}.$$

(i) Сначала вычисляем  $M_r = 1$ ,  $M_0 = 30$ ,  $M_{-r} = 7$ .

(ii) Затем находим  $q = 10$ .

(iii) Поскольку  $M_r = 1$ , то  $M_r^{-1} = 1$ .

(iv) Тогда  $q' = |10 \cdot 7 \cdot 1|_{30} = 10$ .

(v) Наконец

$$\begin{bmatrix} M_r M_0 & 0 \\ M_r q' & M_{-r} \end{bmatrix} = \begin{bmatrix} 30 & 0 \\ 10 & 7 \end{bmatrix}.$$

Следовательно, с такой начальной матрицей получаем

$$\begin{array}{c|cc} & 30 & 0 \\ \hline & 10 & 7 \\ \hline 3 & 0 & -21 \end{array}$$

<sup>1)</sup> Если отбросить предположения о знаках чисел  $r_i$  и  $s_i$ , то в нетривиальном случае  $f_i g_i \neq 0$  можно дать компактную формулировку правила сложения: сумма  $(f_i, r_i) + (g_i, s_i)$  совпадает с первым слагаемым при  $r_i > s_i$  и равна  $(h_i, t_i)$  при  $r_i = s_i = t_i$ . — Прим. перев.

Тем самым мы показали, что

$$[(0; 0), (1; 0), (0; 0), (3; -1)] \rightarrow \frac{10}{7}.$$

**7.48. Замечание.** Если использовать определение 7.24, то имеем

$$\frac{10}{7} \rightarrow [(1; 1), (1; 0), (1; 1), (3; -1)].$$

Этот результат можно назвать *каноническим представлением* для  $10/7$ , а результат, полученный в примере 7.47 (при сложении представлений для  $3/7$  и  $1$ ), можно назвать *неканоническим представлением* для  $10/7$ . Отметим, что оба представления переводятся в правильный ответ  $10/7$ .

Связь между различными представлениями рационального числа детально обсуждается в упоминавшейся выше работе Матулы и Грегори.

**7.49. Замечание.** Очевидно, что при практической реализации многомодульной арифметики вычетов следует использовать большое число простых модулей максимальной величины (насколько позволяет длина машинного слова в компьютере с фиксированным числом разрядов в слове). Схема вычислений идеально подходит для параллельных ЭВМ, поскольку для выполнения поэлементной операции можно одновременно привлечь несколько процессоров так, чтобы вычисления по каждому модулю проводились на отдельном процессоре.

### Упражнения I.7

1. Определить значение

$$x = \frac{1}{2} - \frac{2}{3} - \frac{1}{6},$$

используя многомодульную арифметику вычетов с основанием  $\beta = [11, 13]$ .

2. Найти значение  $x$  в задаче 7 из упражнений I.6, используя многомодульную арифметику вычетов с основанием  $\beta = [23, 29]$ .

3. Доказать теорему 7.8.

4. Доказать следствие 7.9.

5. Доказать теорему 7.11.

6. Повторить задачу 1 с  $\beta = [2, 3, 5, 7]$ .

7. Повторить задачу 2 с  $\beta = [3, 5, 7, 11]$ .

# Глава II

## Конечноразрядная $p$ -адическая арифметика

### § 1. Введение

В этой главе мы хотим представить альтернативную числовую систему, а именно конечноразрядную систему  $p$ -адических чисел, введенную в работах [Krishnamurthy, Rao, Subramanian, 1975a, 1975b; Alparslan, 1975]. Эта числовая система конечна, ее связь с (бесконечной) системой  $p$ -адических чисел [Hensel, 1908] <sup>1)</sup> обсуждается в последующих параграфах.

С математической точки зрения конечноразрядная  $p$ -адическая арифметика эквивалентна одномодульной арифметике вычетов, если модуль является целым числом вида  $m = p^r$ , где  $p$  — простое, а  $r$  — натуральное. Однако  $p$ -адическое представление имеет преимущества, когда мы хотим распространить обсуждение с рациональных чисел на полиномы.

В этой числовой системе каждое рациональное число из некоторого конечного множества (дробей Фарея порядка  $N$ , введенных в гл. I) отображается в единственное закодированное представление, которое называется кодом Гензеля <sup>2)</sup>. Арифметические операции над парами рациональных чисел из этого множества можно свести к соответствующим арифметическим операциям над их кодами Гензеля. Следует отметить, что в конечноразрядной  $p$ -адической арифметике нет ошибок округления, как и в арифметике вычетов.

### § 2. Поле $p$ -адических чисел

Пусть  $\mathbb{K}$  — произвольное поле и  $\mathbb{R}$  — поле вещественных чисел. Определим *норму* на  $\mathbb{K}$  как следующее отображение.

**2.1. Определение.** Нормой на поле  $\mathbb{K}$  (которое рассматривается как векторное пространство над самим собой) называется отображение  $\|\cdot\|: \mathbb{K} \rightarrow \mathbb{R}$ , такое, что для всех  $\alpha, \beta$  в  $\mathbb{K}$  выполнены соотношения

- (i)  $\|\alpha\| \geq 0$ , причем  $\|\alpha\| = 0$  тогда и только тогда, когда  $\alpha = 0$ ,
- (ii)  $\|\alpha\beta\| = \|\alpha\| \cdot \|\beta\|$ ,
- (iii)  $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$ .

---

<sup>1)</sup> См. также Понтрягин Л. С. Обобщения чисел. — Библиотечка «Квант», вып. 54. — М.: Наука, 1986, где также изложены основы теории  $p$ -адических чисел. — *Прим. перев.*

<sup>2)</sup> По имени немецкого математика Гензеля (K. Hensel, 1861—1941).

Например, в поле рациональных чисел  $\mathbb{Q}$  абсолютная величина задает отображение, являющееся нормой на  $\mathbb{Q}$ . Для нас сейчас больший интерес представляет другая норма на  $\mathbb{Q}$ , построение которой можно связать со следующим наблюдением. Если  $\alpha = a/b \in \mathbb{Q}$ ,  $\alpha \neq 0$  и  $(a, b) = 1$ , то число  $\alpha$  можно представить единственным образом в виде

$$(2.2) \quad \alpha = \frac{c}{d} \cdot p^e,$$

где  $p$  — заданное простое число,  $c, d, e$  — целые,  $(c, d) = 1$  и  $p$  не делит ни  $c$ , ни  $d$ . Основываясь на этом представлении числа  $\alpha$ , имеем следующий результат.

**2.3. Теорема.** *Отображение  $\|\cdot\|_p: \mathbb{Q} \rightarrow \mathbb{R}$ , определяемое равенством*

$$\|\alpha\|_p = \begin{cases} p^{-e}, & \text{если } \alpha \neq 0, \\ 0, & \text{если } \alpha = 0, \end{cases}$$

*является нормой на  $\mathbb{Q}$ .*

**Доказательство** см. в книге [Коблиц, 1982, с. 11].  $\square$

**2.4. Определение.** Норма, введенная в теореме 2.3, называется  $p$ -адической нормой на  $\mathbb{Q}$ .

Следует отметить, что некоторые свойства  $p$ -адической нормы вступают в противоречие с интуитивными представлениями о норме чисел, поскольку большому положительному значению  $e$  в (2.2) соответствует малая величина  $p$ -адической нормы числа  $\alpha$ .

### Метрическое пространство

Введем понятия *метрики* и *метрического пространства*.

**2.5. Определение.** Метрическое пространство есть пара  $(H, d)$  из непустого множества  $H$  и метрики (или функции расстояния)  $d: H \times H \rightarrow \mathbb{R}$ , такой, что при всех  $x, y, z \in H$  выполнены следующие аксиомы

- (i)  $d(x, y) = 0$  тогда и только тогда, когда  $x = y$ ,
- (ii)  $d(x, y) = d(y, x)$ ,
- (iii)  $d(x, z) \leq d(x, y) + d(y, z)$ .

Свойства (i)–(iii) иногда называют постулатами Хаусдорфа<sup>1)</sup>. Из них нетрудно вывести четвертое свойство: для

<sup>1)</sup> По имени немецкого математика Хаусдорфа (F. Hausdorff, 1868—1942).

всех  $x, y$  в  $\mathbb{H}$  имеет место неравенство

$$(iv) \quad d(x, y) \geq 0.$$

**2.6. Определение.** Последовательность  $\{x_n\} = \{x_1, x_2, \dots\}$ , где  $x_n \in \mathbb{H}$  для всех  $n$ , называется последовательностью Коши<sup>1)</sup> в метрическом пространстве  $(\mathbb{H}, d)$  тогда и только тогда, когда

$$d(x_n, x_m) \rightarrow 0 \quad (n, m \rightarrow \infty),$$

т. е. для любого  $\varepsilon > 0$  существует номер  $N = N(\varepsilon)$ , такой, что для всех  $n, m > N$  выполнено неравенство

$$d(x_n, x_m) < \varepsilon.$$

**2.7. Определение.** Последовательность  $\{x_n\} = \{x_1, x_2, \dots\}$  в метрическом пространстве  $(\mathbb{H}, d)$  называется *сходящейся* (к  $x$ ) тогда и только тогда, когда существует элемент  $x \in \mathbb{H}$ , такой, что

$$d(x_n, x) \rightarrow 0 \quad (n \rightarrow \infty).$$

Этот элемент  $x$  называется *пределом* данной последовательности; сам факт сходимости записывают в виде  $x_n \rightarrow x$ .

Заметим, что из определения последовательности Коши не следует, что она обязана сходиться<sup>2)</sup>. Действительно, хорошо известно, что не каждая последовательность Коши в произвольном метрическом пространстве сходится. Однако если в некотором метрическом пространстве сходится каждая последовательность Коши, то такому метрическому пространству дается специальное название.

**2.8. Определение.** Метрическое пространство  $(\mathbb{H}, d)$  называется *полным* тогда и только тогда, когда каждая последовательность Коши сходится (к элементу из  $\mathbb{H}$ ). Более точно, требуется, чтобы при условии

$$d(x_n, x_m) \rightarrow 0 \quad (n, m \rightarrow \infty)$$

существовал элемент  $x \in \mathbb{H}$ , такой, что

$$d(x_n, x) \rightarrow 0 \quad (n \rightarrow \infty).$$

### *Пример конкретного метрического пространства*

Можно получить метрическое пространство, если выбрать  $\mathbb{H} = \mathbb{Q}$  и определить метрику  $d: \mathbb{Q} \times \mathbb{Q} \rightarrow \mathbb{R}$  в терминах  $p$ -ади-

<sup>1)</sup> По имени французского математика Коши (A. L. Cauchy, 1789—1857).

<sup>2)</sup> Под сходимостью понимается сходимость к некоторому элементу данного пространства; см. определение 2.7.

ческой нормы на  $\mathbb{Q}$ . Таким образом, если положить

$$(2.9) \quad d(\alpha, \beta) = \|\alpha - \beta\|_p$$

для всех  $\alpha, \beta \in \mathbb{Q}$ , то пара  $(\mathbb{Q}, d)$  образует метрическое пространство.

**2.10. Определение.** Метрика (2.9), индуцированная  $p$ -адической нормой  $\|\cdot\|_p$ , называется  $p$ -адической метрикой.

Особый интерес в метрическом пространстве  $(\mathbb{Q}, d)$  представляет последовательность степеней простого числа  $p$ :

$$(2.11) \quad \{p^n\} = \{p, p^2, p^3, \dots\}.$$

Любопытно, что эта последовательность сходится к нулю, так как в смысле  $p$ -адической нормы имеют место соотношения

$$(2.12) \quad d(p^n, 0) = \|p^n\|_p = p^{-n}$$

и  $p^{-n} \rightarrow 0$  при  $n \rightarrow \infty$ <sup>1)</sup>. (Как уже отмечалось,  $p$ -адическая норма вступает в противоречие с интуицией.)

### Полношение метрического пространства

Из теории метрических пространств хорошо известно, что по неполному метрическому пространству (в котором не каждая последовательность Коши сходится) можно построить полное метрическое пространство (в котором каждая последовательность Коши сходится). Последнее метрическое пространство называется пополнением первого; см., например, [Коблиц, 1982].

**2.13. Пример.** Рассмотрим метрическое пространство  $(\mathbb{Q}, \hat{d})$ , где  $\hat{d}$  — абсолютное значение:

$$\hat{d}(\alpha, \beta) = |\alpha - \beta|.$$

Пусть  $\hat{S}$  — множество последовательностей Коши в этом метрическом пространстве. Две последовательности Коши  $s_1 = \{a_1, a_2, \dots\}$  и  $s_2 = \{b_1, b_2, \dots\}$  считаются по определению эквивалентными (что записывается в виде  $s_1 \sim s_2$ ), тогда и только тогда, когда  $|a_i - b_i| \rightarrow 0$  при  $i \rightarrow \infty$ .

Отношение эквивалентности  $\sim$  имеет следующие свойства:

- (i)  $s_1 \sim s_1$ ,
- (ii)  $s_1 \sim s_2 \Rightarrow s_2 \sim s_1$ ,
- (iii)  $s_1 \sim s_2$  и  $s_2 \sim s_3 \Rightarrow s_1 \sim s_3$ .

<sup>1)</sup> В данном случае сходимость  $p^{-n} \rightarrow 0$  понимается в обычном смысле, т. е. в смысле абсолютного значения. — *Прим. перев.*



Говорят, что две последовательности  $s_1$  и  $s_2$  принадлежат одному классу эквивалентности, если  $s_1 \sim s_2$ . Теперь определим  $\mathbb{R}$  как множество классов эквивалентности последовательностей Коши в  $(\mathbb{Q}, \check{d})$ . На этих классах эквивалентности можно определить операции сложения и умножения и ввести понятия противоположного и обратного классов таким образом, что система  $(\mathbb{R}, +, \cdot)$  будет полем (например, см. [Коблиц, 1982]). Это в точности поле вещественных чисел, и метрическое пространство  $(\mathbb{R}, \check{d})$  является пополнением метрического пространства  $(\mathbb{Q}, \check{d})$ .

Мы получим поле  $p$ -адических чисел, если определим пополнение множества рациональных чисел по отношению к  $p$ -адической метрике вместо метрики абсолютного значения, которая использована выше. В этом случае строится пополнение метрического пространства  $(\mathbb{Q}, d)$ , где метрика  $d$  определена в (2.9). Пусть  $\mathbb{Q}_p$  обозначает множество классов эквивалентности последовательностей Коши в  $(\mathbb{Q}, d)$ , где последовательности  $s_1$  и  $s_2$  эквивалентны тогда и только тогда, когда  $\|a_i - b_i\|_p \rightarrow 0$  при  $i \rightarrow \infty$ . Если определить подходящим образом сложение, умножение и обратные элементы относительно этих операций (см. [Коблиц, 1982, с. 22—23]), то числовая система  $(\mathbb{Q}_p, +, \cdot)$  образует поле. Это поле  $p$ -адических чисел, и метрическое пространство  $(\mathbb{Q}_p, d)$  является пополнением метрического пространства  $(\mathbb{Q}, d)$ .

**2.14. Определение.** Каждый элемент множества  $\mathbb{Q}_p$  называется  $p$ -адическим числом.

Для определения множества  $p$ -адических чисел  $\mathbb{Q}_p$  мы использовали довольно абстрактный подход. Возможно, следующая теорема о разложении  $p$ -адических чисел охарактеризует их более конкретно. Это разложение до некоторой степени аналогично десятичному разложению вещественных чисел.

**2.15. Теорема.** Любому  $p$ -адическому числу  $\alpha \in \mathbb{Q}_p$  соответствует представление

$$\alpha = \sum_{i=n}^{\infty} a_i p^i,$$

где каждое число  $a_i \in \mathbb{Z}$  и номер  $n$  выбран из условия  $\|\alpha\|_p = p^{-n}$ . Кроме того, если каждое  $a_i \in \{0, 1, 2, \dots, p-1\}$ , то это разложение единственно (в этом случае оно является каноническим представлением для  $\alpha$ ).

**Доказательство** см. в книге [Bachman, 1964, с. 34—35]<sup>1)</sup>.  $\square$

<sup>1)</sup> Это доказательство приводится в книге [Коблиц, 1982, с. 24—28]. — Прим. перев.

Мы интересуемся главным образом  $p$ -адическими разложениями рациональных чисел. Поскольку поле рациональных чисел вложено в поле  $p$ -адических чисел<sup>1)</sup>, имеет место соответствующее следствие.

**2.16. Следствие.** Любое рациональное число  $\alpha \in \mathbb{Q}$  имеет единственное  $p$ -адическое разложение

$$\alpha = \sum_{j=n}^{\infty} a_j p^j,$$

где каждое число  $a_j \in \{0, 1, 2, \dots, p-1\}$  и номер  $n$  выбран из условия  $\|\alpha\|_p = p^{-n}$  (ряд сходится к  $\alpha$  в  $p$ -адической метрике).

**Доказательство.** Следствие непосредственно вытекает из теоремы 2.15.  $\square$

**2.17. Пример.** Рассмотрим следующее разложение в степенной ряд:

$$\begin{aligned} \alpha &= 2 + 3p + p^2 + 3p^3 + p^4 + 3p^5 + \dots = \\ &= 2 + 3p(1 + p^2 + p^4 + \dots) + p^2(1 + p^2 + p^4 + \dots) = \\ &= 2 + (3p + p^2)(1 + p^2 + p^4 + \dots). \end{aligned}$$

Поскольку в  $p$ -адической метрике ряд  $1 + p^2 + p^4 + \dots$  сходится к  $(1 - p^2)^{-1}$  (см. задачу 5 из упражнений к данному параграфу), приходим к равенству

$$\alpha = 2 + \frac{3p + p^2}{1 - p^2}.$$

При  $p = 5$  получаем  $\alpha = 1/3$  и разложение  $\alpha$  обычно записывают в сокращенной форме:

$$1/3 = .23131313\dots \quad (p=5).$$

Точка в левой части записи называется  $p$ -адической точкой.

**2.18. Замечание.** Укажем на взаимно однозначное соответствие между разложением в степенной ряд

$$\alpha = a_n p^n + a_{n+1} p^{n+1} + a_{n+2} p^{n+2} + \dots$$

и сокращенным представлением

$$\alpha = a_n a_{n+1} a_{n+2} \dots,$$

---

<sup>1)</sup> Вспоминая определение  $p$ -адического числа как класса эквивалентности последовательностей Коши, укажем вид стандартного вложения: рациональному числу  $\alpha$  отвечает класс последовательностей, эквивалентных стационарной последовательности  $\{\alpha, \alpha, \dots\}$ . Характеризация рациональных чисел в  $\mathbb{Q}_p$  при помощи  $p$ -адического разложения будет приведена далее. — *Прим. перев.*

где сохранены только коэффициенты при степенях  $p$ . Это соответствие позволяет заменять разложения в степенные ряды на сокращенные представления, и наоборот. Фактически мы будем называть каждое из них  $p$ -адическим разложением для  $\alpha$ . Сокращенное представление полностью аналогично представлению вещественного числа в виде десятичной дроби. Усилим эту аналогию, введя  $p$ -адическую точку так, чтобы ее местоположение указывало знак числа  $n$ . Таким образом, записываем

$$\alpha = \begin{cases} a_n a_{n+1} \dots a_{-2} a_{-1} \cdot a_0 a_1 a_2 \dots & \text{для } n > 0, \\ \dots a_0 a_1 a_2 \dots & \text{для } n = 0, \\ 0 \dots 0 a_n a_{n+1} \dots & \text{для } n < 0. \end{cases}$$

Как известно, вещественное число является рациональным тогда и только тогда, когда его десятичное разложение периодически. Подобным образом  $p$ -адическое число рационально тогда и только тогда, когда его  $p$ -адическое разложение периодически (см. [Bachman, 1964, с. 40]<sup>1)</sup>). Следовательно, поскольку мы интересуемся в основном  $p$ -адическими разложениями рациональных чисел, нам будут встречаться только периодические  $p$ -адические разложения.

### *Дополнительное представление отрицательного рационального числа*

Связь между  $p$ -адическим разложением для чисел  $\alpha$  и  $-\alpha$  видна из следующего результата.

#### **2.19. Теорема. Если**

$$\alpha = a_n p^n + a_{n+1} p^{n+1} + a_{n+2} p^{n+2} + \dots,$$

то

$$-\alpha = b_n p^n + b_{n+1} p^{n+1} + b_{n+2} p^{n+2} + \dots,$$

где  $b_n = p - a_n$  и  $b_j = (p - 1) - a_j$ ,  $j > n$ .

**Доказательство.** Дадим набросок доказательства, а именно проверим, что сумма представлений для  $\alpha$  и  $-\alpha$  равна 0. Запишем

$$\begin{aligned} -\alpha &= (p - a_n) p^n + (p - 1 - a_{n+1}) p^{n+1} + \\ &\quad + (p - 1 - a_{n+2}) p^{n+2} + \dots \end{aligned}$$

<sup>1)</sup> См. также [Коблиц, 1982, с. 35]. — Прим. перев.

Формируя сумму  $\alpha + (-\alpha)$ , получаем  $p$ -адическое разложение с нулевыми коэффициентами:

$$\begin{aligned} 0 &= p \cdot p^n + (p-1)p^{n+1} + (p-1)p^{n+2} + \dots = \\ &= 0 + \quad rp^{n+1} \quad + (p-1)p^{n+2} + \dots = \\ &= 0 + \quad 0 \quad + \quad rp^{n+2} \quad + \dots = \\ &= 0 + \quad 0 \quad + \quad 0 \quad + \dots \end{aligned}$$

(сколько бы мы ни продолжали это разложение, везде будут стоять нули).  $\square$

**2.20. Пример.** Напомним 5-адическое разложение для  $1/3$  в примере 2.17:

$$\frac{1}{3} = 2 + 3 \cdot 5 + 1 \cdot 5^2 + 3 \cdot 5^3 + 1 \cdot 5^4 + \dots$$

Отсюда выводим

$$-\frac{1}{3} = 3 + 1 \cdot 5 + 3 \cdot 5^2 + 1 \cdot 5^3 + 3 \cdot 5^4 + \dots$$

и, выполняя сложение, получаем

$$\begin{aligned} 0 &= 5 + 4 \cdot 5 + 4 \cdot 5^2 + 4 \cdot 5^3 + \dots \\ &= 0 + 5 \cdot 5 + 4 \cdot 5^2 + 4 \cdot 5^3 + \dots \\ &= 0 + 0 + 5 \cdot 5^2 + 4 \cdot 5^3 + \dots \\ &= 0 + 0 + 0 + 5 \cdot 5^3 + \dots \\ &= 0 + 0 + 0 + 0 + \dots \end{aligned}$$

*$p$ -адическое представление основной дроби*

Рациональное число  $\alpha = a/b$  с  $(a, b) = 1$  называется *основной дробью*<sup>1)</sup>, если знаменатель  $b$  есть степень числа  $p$ . Представляет интерес случай положительной основной дроби.

**2.21. Теорема.**  $p$ -адическое число  $\alpha$  можно представить в виде конечного<sup>2)</sup>  $p$ -адического разложения тогда и только тогда, когда  $\alpha$  является положительной основной дробью.

**Доказательство** см. в работе [MacDuffee, 1938]<sup>3)</sup>.  $\square$

**2.22. Следствие.** Положительное целое число представляется конечным  $p$ -адическим разложением.

<sup>1)</sup> В оригинале radix fraction. — Прим. перев.

<sup>2)</sup> Конечным называется  $p$ -адическое разложение с конечным числом  $p$ -адических цифр.

<sup>3)</sup> Теорему 2.21 в книге [Коблиц, 1982, с. 35] предлагается доказать в качестве упражнения. — Прим. перев.

**Доказательство.** Положительное целое число, очевидно, будет и положительной основной дробью.  $\square$

Заметим, что дополнительное представление (см. теорему 2.19) конечного  $p$ -адического разложения оказывается бесконечным  $p$ -адическим разложением. Таким образом, предыдущие теорема и следствие имеют отношение только к положительным основным дробям (и целым числам). Например,

$$(2.23) \quad \frac{199}{125} = 442.100000 \dots \quad (p=5)$$

и

$$(2.24) \quad -\frac{199}{125} = 102.344444 \dots \quad (p=5).$$

Легко проверить справедливость равенств

$$(2.25) \quad 199 = .442100000 \dots \quad (p=5)$$

и

$$(2.26) \quad -199 = .10234444 \dots \quad (p=5),$$

которые иллюстрируют следующий факт. Пусть  $\alpha = a/b$ , где  $(a, b) = 1$  и  $b = p^k$ ; тогда  $p$ -адическое разложение числа  $\alpha$  можно получить из  $p$ -адического разложения числителя  $a$ , просто сдвигая  $p$ -адическую точку вправо на  $k$  позиций. Здесь имеется *полная аналогия*<sup>1)</sup> со случаем десятичных дробей, в которых знаменатели — степени десяти.

**2.27. Замечание.** Поскольку целое положительное число  $h$  единственным образом представляется в виде суммы степеней числа  $p$ ,

$$h = d_0 + d_1p + d_2p^2 + \dots + d_kp^k,$$

с целыми коэффициентами  $d_i \in \mathbb{P}_p$ , то в сущности, нет различий между его представлениями в  $p$ -ичной системе счисления и  $p$ -адическим. Действительно, единственный нюанс заключается в том, что в сокращенной записи  $p$ -адического разложения *цифры записываются в обратном порядке*. Например,

$$14 = 2 + 3 + 3^2,$$

откуда вытекает, что троичное представление имеет вид

$$14_{\text{десять}} = 112_{\text{три}}.$$

Однако 3-адическое разложение есть

$$14_{\text{десять}} = .2110000 \dots \quad (p=3).$$

---

<sup>1)</sup> Только сдвиг производится в обратном направлении. — *Прим. перев.*

Поскольку ненулевых цифр конечное число и нули справа отбрасываются, записываем

$$14_{\text{десять}} = .211 \quad (p=3).$$

Представления 3-ичное и 3-адическое являются зеркальными образами друг друга.

*Вычисление цифр  $p$ -адического разложения*

Пусть  $\alpha$  имеет  $p$ -адическое разложение

$$(2.28) \quad \alpha = a_n p^n + a_{n+1} p^{n+1} + a_{n+2} p^{n+2} + \dots = \\ = p^n (a^n + a_{n+1} p + a_{n+2} p^2 + \dots) = p^n \left( \frac{c_1}{d_1} \right),$$

где  $(c_1, d_1) = 1$  и  $p$  не делит ни  $c_1$ , ни  $d_1$ . Дробь  $c_1/d_1$  имеет  $p$ -адическое разложение

$$(2.29) \quad \frac{c_1}{d_1} = a_n + a_{n+1} p + a_{n+2} p^2 + \dots,$$

и тогда

$$(2.30) \quad \left| \frac{c_1}{d_1} \right|_p = |a_n + a_{n+1} p + a_{n+2} p^2 + \dots|_p = a_n.$$

Другими словами, мы получаем  $a_n$ , вычисляя

$$(2.31) \quad a_n = \left| \frac{c_1}{d_1} \right|_p.$$

Затем, используем равенство (2.29) и образуем выражение

$$(2.32) \quad \frac{c_1}{d_1} - a_n = p(a_{n+1} + a_{n+2} p + a_{n+3} p^2 + \dots) = p \left( \frac{c_2}{d_2} \right),$$

в котором  $(c_2, d_2) = 1$ . Дробь  $c_2/d_2$  имеет  $p$ -адическое разложение

$$(2.33) \quad \frac{c_2}{d_2} = a_{n+1} + a_{n+2} p + a_{n+3} p^2 + \dots,$$

и тогда

$$(2.34) \quad \left| \frac{c_2}{d_2} \right|_p = |a_{n+1} + a_{n+2} p + a_{n+3} p^2 + \dots|_p = a_{n+1}.$$

Другими словами, мы получаем  $a_{n+1}$ , вычисляя

$$(2.35) \quad a_{n+1} = \left| \frac{c_2}{d_2} \right|_p.$$

Вообще говоря, эта процедура продолжается дальше и приводит к ответу за конечное число шагов только тогда, когда  $a$  — положительная основная дробь. В нашем случае,

когда  $\alpha$  — рациональное число, его  $p$ -адическое разложение будет периодическим и процедуру можно закончить, когда период уже выявлен.

**2.36. Пример.** Пусть  $\alpha = 2/3$  и  $p = 5$ . В этом случае  $c_1 = 2$ ,  $d_1 = 3$  и  $n = 0$ . Таким образом, используя прямое отображение из гл. I, получаем

$$a_0 = |2/3|_5 = 4.$$

Затем образуем выражение

$$\frac{c_1}{d_1} - a_0 = \frac{2}{3} - 4 = 5 \left( \frac{-2}{3} \right).$$

В этом случае  $c_2 = -2$  и  $d_2 = 3$ . Таким образом,

$$a_1 = |-2/3|_5 = 1.$$

Теперь образуем выражение

$$\frac{c_2}{d_2} - a_1 = \frac{-2}{3} - 1 = 5 \left( \frac{-1}{3} \right).$$

В этом случае  $c_3 = -1$  и  $d_3 = 3$ . Таким образом,

$$a_2 = |-1/3|_5 = 3.$$

Продолжая эту процедуру, получим  $a_3 = 1$  и  $a_4 = 3$ . Период  $p$ -адического разложения уже обнаружен, и здесь выполнение процедуры можно прервать. Следовательно, имеем разложение

$$2/3 = .4131313 \dots \quad (p = 5).$$

**2.37. Замечание.** Мы уже указывали некоторые общие черты  $p$ -адических и десятичных чисел. Одно из различий между ними состоит в том, что  $p$ -адическое разложение определяется  $p$ -адическим числом *однозначно*, тогда как десятичную дробь с конечным числом знаков можно записать также в виде бесконечной с цифрой девять в периоде, например  $1 = 0.9999 \dots$ . Эти два десятичных разложения определяют одно и то же число.

## Упражнения II.2

1. Рассмотрим поле рациональных чисел  $(\mathbb{Q}, +, \cdot)$ . Доказать, что отображение абсолютного значения  $|\cdot|: \mathbb{Q} \rightarrow \mathbb{R}$  является нормой на  $\mathbb{Q}$ .
2. (a) Найти 5-ичное представление числа 14.  
 (b) Найти 5-адическое представление числа 14.  
 (c) Найти 5-адическое представление числа  $-14$ .

3. Вычислить 5-адические разложения следующих чисел:

(a)  $\alpha = 4/3$ ; (d)  $\alpha = 1/6$ ;

(b)  $\alpha = 5/2$ ; (e)  $\alpha = -9$ ;

(c)  $\alpha = -5/2$ ; (f)  $\alpha = -10$ .

4. Доказать неравенство  $d(x, y) \geq 0$ , используя определение 2.5.

5. Показать, что ряд  $1 + p^2 + p^4 + \dots$  сходится к  $(1 - p^2)^{-1}$  в  $p$ -адической метрике.

### § 3. Арифметика в $\mathbb{Q}_p$

Операции сложения, вычитания, умножения и деления  $p$ -адических чисел очень схожи с одноименными операциями над десятичными дробями. Основное различие заключается в том, что в  $\mathbb{Q}_p$  операции поразрядно выполняются «слева направо», а не «справа налево», как с десятичными разложениями.

#### Сложение

В примере 2.20 было показано, как вычислить сумму двух чисел,  $1/3$  и  $-1/3$ , используя разложения в степенные ряды. Переходя к общему случаю, предположим известными разложения двух произвольных  $p$ -адических чисел:

$$(3.1) \quad \alpha = a_n p^n + a_{n+1} p^{n+1} + a_{n+2} p^{n+2} + \dots$$

и

$$(3.2) \quad \beta = b_n p^n + b_{n+1} p^{n+1} + b_{n+2} p^{n+2} + \dots,$$

где  $p$ -адические цифры  $a_i$  и  $b_i$  принадлежат множеству  $\Pi_p$ . Укажем, что одна из цифр  $a_n$  или  $b_n$  может равняться нулю, но не обе сразу. Последующие цифры могут быть произвольными.

Образуюм сумму

$$(3.3) \quad \alpha + \beta = (a_n + b_n) p^n + (a_{n+1} + b_{n+1}) p^{n+1} + \dots = \\ = c_n p^n + c_{n+1} p^{n+1} + \dots,$$

где

$$(3.4) \quad c_i = a_i + b_i, \quad i = n, n+1, n+2, \dots$$

Предположим, что все значения  $c_n, c_{n+1}, \dots, c_{k-1}$  меньше  $p$ , но  $c_k \geq p$ . Тогда

$$(3.5) \quad c_k = p + d_k,$$



где  $0 \leq d_k < p$ . В этом случае имеем

$$\begin{aligned}
 (3.6) \quad \alpha + \beta &= c_n p^n + \dots + c_{k-1} p^{k-1} + (p + d_k) p^k + \\
 &\quad + c_{k+1} p^{k+1} + c_{k+2} p^{k+2} + \dots = \\
 &= c_n p^n + \dots + c_{k-1} p^{k-1} + d_k p^k + \\
 &\quad + (c_{k+1} + 1) p^{k+1} + c_{k+2} p^{k+2} + \dots,
 \end{aligned}$$

т. е.  $p^k$  соответствует цифра  $d_k$ . Обратим внимание на возникший «перенос» в соседний разряд, вследствие которого к значению  $c_{k+1}$  прибавилась единица. Теперь следует проверить справедливость неравенства  $c_{k+1} + 1 < p$ . Если оно не выполнено, то «перенос» распространяется на следующий разряд и т. д.

Эта ситуация подобна возникающей при сложении десятичных дробей. Однако в  $p$ -адическом случае применительно к сокращенному представлению цифры суммируются в  $p$ -ичной арифметике и, начиная с первых разрядов, слева направо (а не справа налево, как в случае десятичных дробей).

**3.7. Пример.** Прибавим  $5/6$  к  $2/3$  в  $\mathbb{Q}_5$ . Легко проверить, что  $p$ -адические разложения этих двух операндов задаются равенствами

$$\begin{aligned}
 2/3 &= .4131313 \dots \quad (p=5), \\
 5/6 &= .0140404 \dots \quad (p=5).
 \end{aligned}$$

Теперь вычислим сумму этих двух  $p$ -адических разложений поразрядно в  $p$ -ичной арифметике (продвигаясь слева направо). Получаем

$$\begin{array}{r}
 .4131313 \dots \\
 .0140404 \dots \\
 \hline
 .4222222 \dots
 \end{array}$$

В качестве проверки убедимся, что  $2/3 + 5/6 = 3/2$  и

$$3/2 = .4222222 \dots \quad (p=5).$$

### Вычитание

Чтобы выполнить вычитание, найдем дополнительное представление к вычитаемому (при помощи теоремы 2.19) и вычислим сумму этого представления с уменьшаемым, т. е. используем преобразование  $\alpha - \beta = \alpha + (-\beta)$ .

**3.8. Пример.** Вычтем  $5/6$  из  $2/3$  в  $\mathbb{Q}_5$ . Дополнение к представлению для дроби  $5/6$  из предыдущего примера имеет вид

$$-5/6 = .0404040 \dots \quad (p=5).$$

Таким образом, если сложить  $p$ -адические разложения чисел  $2/3$  и  $-5/6$  поразрядно, используя 5-ичную арифметику и продвигаясь слева направо, то получим

$$\begin{array}{r} .4131313 \dots \\ .0404040 \dots \\ \hline .4040404 \dots \end{array}$$

В качестве контроля убеждаемся, что  $2/3 - 5/6 = -1/6$  и  $-1/6 = .4040404 \dots$  ( $p = 5$ ).

### Умножение

Прежде чем приступить к обсуждению операции умножения, заметим, что произвольное  $p$ -адическое число  $\gamma$  всегда можно записать в виде

$$(3.9) \quad \begin{aligned} \gamma &= g_n p^n + g_{n+1} p^{n+1} + g_{n+2} p^{n+2} + \dots = \\ &= p^n (g_n + g_{n+1} p + g_{n+2} p^2 + \dots) = p^n \alpha. \end{aligned}$$

Последнее равенство служит определением числа  $\alpha$ , а величина  $n$  может оказаться положительной, отрицательной или равной нулю. Если сменить обозначения так, чтобы индексы  $p$ -адических цифр соответствовали степеням числа  $p$  у выражения в скобках из (3.9), то можно записать

$$(3.10) \quad \alpha = a_0 + a_1 p + a_2 p^2 + \dots$$

**3.11. Определение.** Любое  $p$ -адическое число, не содержащее членов с отрицательными степенями  $p$  в  $p$ -адическом разложении, называется  $p$ -адическим *целым*. Любое  $p$ -адическое целое с ненулевой первой цифрой называется  $p$ -адической *единицей*.

Так, в (3.10)  $\alpha$  есть  $p$ -адическая единица. Из (3.9) вытекает, что любое  $p$ -адическое число представимо в виде произведения  $p$ -адической единицы и некоторой степени  $p$ .

Для произведения числа  $\gamma$  из (3.9) на число  $\delta$  типа

$$(3.12) \quad \delta = p^m \beta,$$

где  $\alpha$  и  $\beta$  суть  $p$ -адические единицы, справедлива формула

$$(3.13) \quad \gamma \delta = p^{n+m} \alpha \beta.$$

Таким образом, без потери общности можно считать сомножители  $p$ -адическими единицами.

Пусть  $\alpha$  и  $\beta$  суть  $p$ -адические единицы с  $p$ -адическими разложениями

$$(3.14) \quad \begin{aligned} \alpha &= a_0 + a_1p + a_2p^2 + \dots, \\ \beta &= b_0 + b_1p + b_2p^2 + \dots, \quad a_0b_0 \neq 0. \end{aligned}$$

Тогда

$$(3.15) \quad \begin{aligned} \alpha\beta &= (a_0 + a_1p + a_2p^2 + \dots)(b_0 + b_1p + b_2p^2 + \dots) = \\ &= c_0 + c_1p + c_2p^2 + \dots, \end{aligned}$$

где

$$(3.16) \quad \begin{aligned} c_0 &= a_0b_0, \\ c_1 &= a_0b_1 + a_1b_0, \\ c_2 &= a_0b_2 + a_1b_1 + a_2b_0, \\ &\vdots \\ c_k &= a_0b_k + a_1b_{k-1} + \dots + a_kb_0, \\ &\vdots \end{aligned}$$

Эти равенства можно записать в матричной форме (см. (3.23)).

Хотя  $p$ -адические цифры  $a_i$  и  $b_i$  принадлежат множеству  $\Pi_p$ , нельзя гарантировать этого же относительно целых чисел  $c_i$  (и, вообще говоря, это не так). Поэтому представим  $c_0$  в виде

$$(3.17) \quad c_0 = a_0b_0 = d_0 + t_1p,$$

где  $0 \leq d_0 < p$  и  $t_1 \geq 0$ . Тогда целое число  $d_0$  является первой цифрой  $p$ -адического разложения произведения  $\alpha\beta$ , а  $t_1$  «переносится» в следующий разряд, т. е. прибавляется к  $c_1$ . На следующем шаге имеем

$$(3.18) \quad c_1 + t_1 = (a_0b_1 + a_1b_0) + t_1 = d_1 + t_2p,$$

где  $0 \leq d_1 < p$  и  $t_2 \geq 0$ . Тогда  $d_1$  является второй цифрой  $p$ -адического разложения для  $\alpha\beta$  и  $t_2$  прибавляется к  $c_2$ . Продолжая эту процедуру, получим (единственное)  $p$ -адическое разложение

$$(3.19) \quad \alpha\beta = d_0 + d_1p + d_2p^2 + \dots,$$

в котором  $0 \leq d_i < p$  для всех  $i \geq 0$ .

Вновь укажем на аналогию со случаем умножения десятичных дробей. Однако в  $p$ -адическом случае арифметические операции в (3.16) и (3.18) выполняются в  $p$ -ичной арифметике и применительно к сокращенному представлению по-

разрядное выполнение операции происходит слева направо (а не справа налево, как в случае десятичных дробей).

**3.20. Пример.** Умножить  $2/3$  на  $1/6$  в  $\mathbb{Q}_5$ . В примере 3.7 найдены 5-адические разложения для дробей  $2/3$  и  $5/6$ . Легко проверить, что 5-адическое разложение для  $1/6$  получается из 5-адического разложения для  $5/6$  простым сдвигом 5-адической точки на один разряд вправо. Следовательно,

$$1/6 = .14040404 \dots \quad (p=5).$$

Умножая разложения для  $2/3$  и  $1/6$ , получаем

$$\begin{array}{r} .4131313131313\dots \\ .1404040404040\dots \\ \hline 4131313131313\dots \\ 1231313131313\dots \\ 1231313131313\dots \\ 1231313131313\dots \\ 1231313131313\dots \\ 1231313131313\dots \\ 1231313131313\dots \\ \hline .4201243201243\dots \end{array}$$

Для контроля убеждаемся, что  $(2/3) \cdot (1/6) = 1/9$  и

$$1/9 = .4201243201243 \dots \quad (p=5).$$

### Деление

Без ограничения общности можно рассматривать алгоритм деления только для  $p$ -адических единиц, как и в случае умножения. Соответствующая аргументация аналогична приведенной выше для умножения. Итак, рассмотрим  $p$ -адические единицы

$$(3.21) \quad \begin{aligned} \delta &= d_0 + d_1p + d_2p^2 + \dots, \\ \beta &= b_0 + b_1p + b_2p^2 + \dots \end{aligned}$$

с  $d_0b_0 \neq 0$ . Отношение  $\alpha = \delta/\beta$  можно записать в виде

$$(3.22) \quad \begin{aligned} \alpha &= \frac{d_0 + d_1p + d_2p^2 + \dots}{b_0 + b_1p + b_2p^2 + \dots} = \\ &= a_0 + a_1p + a_2p^2 + \dots, \end{aligned}$$

где предполагается, что все коэффициенты разложения  $a_0, a_1, a_2, \dots$  являются  $p$ -адическими цифрами.

Заметим, что  $\delta = \alpha\beta$  и имеют место соотношения (3.15). Равенства (3.16) можно записать в эквивалентной мат-

ричной форме

$$(3.23) \quad \begin{bmatrix} b_0 & & & & \\ b_1 & b_0 & & & \\ b_2 & b_1 & b_0 & & \\ \dots & \dots & \dots & \dots & \\ b_k & b_{k-1} & b_{k-2} & \dots & b_0 \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \\ \vdots \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_k \\ \vdots \end{bmatrix}.$$

Чтобы получить из коэффициентов  $c_0, c_1, \dots$  цифры  $d_0, d_1, \dots$   $p$ -адического разложения (3.19), следует использовать равенства типа (3.17), (3.18). Основываясь на этих рассуждениях, выведем формулу для цифры  $a_0$ .

Из (3.16) или (3.23) видим, что  $a_0 b_0 = c_0$ . Однако  $c_0 = d_0 + t_1 p$ , как указано в (3.17). Следовательно,

$$(3.24) \quad a_0 b_0 = d_0 + t_1 p,$$

откуда выводим равенство  $|a_0 b_0|_p = d_0$ . Поэтому

$$(3.25) \quad a_0 = |d_0 b_0^{-1}|_p.$$

Другими словами, чтобы получить первую цифру частного, мы вычисляем  $b_0^{-1}(p)$ , умножаем эту величину на  $d_0$  и приводим результат по модулю  $p$ .

Это правило оказывается справедливым при вычислении каждой цифры разложения  $\alpha$  в (3.22). На каждом шаге обычной процедуры «длинного деления» мы умножаем  $b_0^{-1}(p)$  на первую цифру частичного остатка и приводим результат по модулю  $p$  (на первом шаге в качестве частичного остатка выступает делимое).

**3.26. Пример.** Разделить  $2/3$  на  $1/12$  в  $\mathbb{Q}_5$ . Имеем разложения

$$2/3 = .4131313 \dots, \quad 1/12 = .3424242 \dots \quad (p=5).$$

Первая цифра делителя равна  $b_0 = 3$ , обратный элемент для нее есть

$$b_0^{-1}(p) = 3^{-1}(5) = 2.$$

Первая цифра частичного остатка (на первом шаге это делимое) равна  $d_0 = 4$ ; тогда получаем первую цифру частного:

$$a_0 = |4 \cdot 2|_5 = 3.$$

При вычислении частичного остатка в процедуре «длинного деления» каждое вычитание заменяем сложением с числом, дополнительным к вычитаемому. Таким образом, пер-

вый шаг деления в данном случае выглядит следующим образом <sup>1)</sup>):

$$\begin{array}{r} .3 \\ .3424242\dots \overline{)4131313\dots} \\ \underline{1111111\dots} \\ 342424\dots \end{array}$$

и новый частичный остаток равен 342424 ...

Чтобы получить вторую цифру частного, умножим  $b_0^{-1}(p)$  на первую цифру частичного остатка и приведем результат умножения по модулю  $p$ :

$$a_1 = |3 \cdot 2|_5 = 1.$$

Таким образом, второй шаг процедуры деления приводит к таблице

$$\begin{array}{r} .31 \\ .3424242\dots \overline{)4131313\dots} \\ \underline{1111111\dots} \\ 342424\dots \\ \underline{202020\dots} \\ 00000\dots \end{array}$$

В этом конкретном примере на втором шаге получился частичный остаток, равный нулю. Следовательно, на этом процесс обрывается, все последующие цифры частного нулевые. В общем случае обрыва не происходит и вычисления продолжают, пока в разложении частного не проявится период.

В качестве контроля находим, что

$$2/3 \div 1/12 = 8 = .3100000 \dots \quad (p=5).$$

**3.27. Замечание.** Заострим внимание на том, что процедура деления  $p$ -адических чисел носит детерминированный характер и не требует подбора и проверок в отличие от деления десятичных дробей. Это связано с тем, что для вычисления каждой цифры частного, когда уже найдено  $b_0^{-1}(p)$  для первой цифры  $b_0$  делителя, применяется очень специфиче-

<sup>1)</sup> Приведем более привычную (еще со школьной скамьи) запись этого шага:

$$\begin{array}{r|l} .4131313\dots & .3424242\dots \\ -4333333\dots & \underline{.3} \\ \hline 342424\dots & \end{array}$$

Здесь число  $.4333333\dots$  является противоположным (дополнительным) к числу  $.1111111\dots$ . — *Прим. перев.*

ский<sup>1)</sup> алгоритм: первая цифра частичного остатка умножается на  $b_0^{-1}(p)$  и результат умножения приводится по модулю  $p$ .

### Упражнения II.3

В каждой задаче использовать арифметику в  $\mathbb{Q}_5$ .

1. Прибавить  $4/3$  к  $1/6$ .
2. Прибавить  $1/3$  к  $5/2$ .
3. Вычесть  $4/3$  из  $1/6$ .
4. Вычесть  $1/3$  из  $5/2$ .
5. Умножить  $4/3$  на  $1/6$ .
6. Умножить  $1/3$  на  $5/2$ .
7. Разделить  $4/3$  на  $1/6$ .
8. Разделить  $1/3$  на  $5/2$ .

## § 4. Конечноразрядная система $p$ -адических чисел

Конечная числовая система, основанная на системе  $p$ -адических чисел (которая описана в двух предыдущих параграфах), предложена в сравнительно недавних работах [Krishnamurthy, Rao, Subramanian, 1975a, 1975b; Alparslan., 1975]. В этой числовой системе множество  $\mathbb{F}_N$  (дробей Фарея порядка  $N$ , см. определение 5.13 гл. I) отображается на множество кодированных представлений, которые называются кодами Гензеля. Арифметические операции над кодами Гензеля эквивалентны соответствующим арифметическим операциям над дробями Фарея порядка  $N$ .

Целое число  $N$ , определяющее мощность множества  $\mathbb{F}_N$ , выбирается как наибольшее положительное число, удовлетворяющее неравенству

$$(4.1) \quad 2N^2 + 1 \leq m,$$

где

$$(4.2) \quad m = p',$$

$p$  — простое, а  $r$  — натуральное.

### Коды Гензеля

Код Гензеля рационального числа есть просто конечный отрезок его (бесконечного)  $p$ -адического разложения; количество цифр в коде определяется числом  $r$ . Через  $H(p, r, \alpha)$

<sup>1)</sup> В котором ход вычислений заранее предопределен, т. е. условные операторы отсутствуют. — *Прим. перев.*

обозначается код Гензеля с  $r$  цифрами для  $p$ -адического числа  $\alpha$ .

Например, (бесконечное)  $p$ -адическое разложение для  $\alpha = 1/3$  дается равенством (см. пример 2.17 настоящей главы)

$$(4.3) \quad 1/3 = .2313131313 \dots \quad (p=5).$$

Выбирая  $p=5$ ,  $r=4$ , получаем код Гензеля

$$(4.4) \quad H(5, 4, 1/3) = .2313.$$

В табл. 4.22 (см. ниже) приводятся коды Гензеля  $H(5, 4, \alpha)$  для дробей Фарея  $\alpha$  порядка 17.

#### *Представление в вычетах, эквивалентное коду Гензеля*

Опишем алгоритм отображения рационального числа  $\alpha$  на его код Гензеля  $H(p, r, \alpha)$ , не требующий знания (бесконечного)  $p$ -адического разложения для  $\alpha$ . Он основан на следующем результате.

#### 4.5. Теорема. Пусть

$$\alpha = \frac{a}{b} = \frac{c}{d} \cdot p^n,$$

где  $(c, d) = (c, p) = (d, p) = 1$ . Обозначим код Гензеля для  $c/d$  через

$$H(p, r, c/d) = .a_0 a_1 \dots a_{r-1}.$$

Тогда  $a_{r-1} \dots a_1 a_0$  есть  $p$ -ичное представление целого числа  $|cd^{-1}|_m$ ,  $m = p^r$ . Другими словами,

$$|cd^{-1}|_m = a_0 + a_1 p + a_2 p^2 + \dots + a_{r-1} p^{r-1}.$$

**Доказательство.** Пусть дробь  $c/d$  имеет  $p$ -адическое разложение

$$\begin{aligned} \frac{c}{d} &= \sum_{j=0}^{\infty} a_j p^j = \\ &= (a_0 + a_1 p + \dots + a_{r-1} p^{r-1}) + p^r R_r. \end{aligned}$$

Тогда

$$c = d(a_0 + a_1 p + \dots + a_{r-1} p^{r-1}) + p^r (d \cdot R_r).$$

Следовательно, при  $m = p^r$  имеем

$$|c|_m = |d(a_0 + a_1 p + \dots + a_{r-1} p^{r-1})|_m,$$

что приводит к требуемому равенству

$$|cd^{-1}|_m = a_0 + a_1 p + \dots + a_{r-1} p^{r-1}.$$

□



4.6. **Пример.** Пусть  $p = 5$ ,  $r = 4$ ,  $m = 625$ . Вычислим код Гензеля для дроби  $\alpha = 1/3$  следующим образом:

$$|1/3|_{625} = |1 \cdot 3^{-1}|_{625} = |417|_{625} = 417 = 2 + 3 \cdot 5 + 1 \cdot 5^2 + 3 \cdot 5^3.$$

Таким образом,  $H(5, 4, 1/3) = .2313$ . Другими словами,

$$|1/3|_{625} = 417_{\text{десять}} = 3132_{\text{пять}}.$$

Записывая цифры этого 5-ичного представления в обратном порядке, приходим к коду Гензеля.

4.7. **Замечание.** Для вычисления величины  $|1/3|_{625}$  можно использовать алгоритм 6.26 из гл. I. В этом алгоритме строится таблица

	625	0
	3	1
208	1	-208
3	0	625

из которой заключаем, что

$$3^{-1}(625) = |-208|_{625} = 417.$$

4.8. **Замечание.** Конечноразрядное  $p$ -адическое представление (код Гензеля) для дроби Фарея порядка  $N$  эквивалентно одномодульному представлению вычетов с модулем  $m = p^r$ . Таким образом, для кодов Гензеля остаются в силе все рассуждения из § 5 гл. I. Основное различие заключается в том, что теперь *всегда* выбирается  $r > 1$  и после отображения дроби Фарея порядка  $N$  на целое число в  $\mathbb{P}_m$  это целое представляется соответствующим кодом Гензеля. Единственность кода Гензеля для каждой дроби Фарея порядка  $N$  гарантируется теоремой 5.14 гл. I.

В замечании 2.18 определено, что роль  $p$ -адической точки в разложении для  $\alpha$  — указывать знак порядка  $n$ , когда рациональное число  $\alpha = a/b$  записывается в форме.

$$(4.9) \quad \frac{a}{b} = \frac{c}{d} \cdot p^n,$$

где  $(c, d) = (c, p) = (d, p) = 1$ . При определении кода Гензеля, когда используются только первые  $r$  цифр  $p$ -адического разложения,  $p$ -адическая точка должна оставаться на своем месте. Чтобы прояснить детали, рассмотрим следующие три случая.

*Случай I.  $n = 0$ .*

Имеем в (4.9)  $a = c$  и  $b = d$ . На первом шаге построения кода Гензеля  $H(p, r, \alpha)$  вычисляем целое число  $|cd^{-1}|_m$ . На втором шаге это целое число (десятичное) переводим в  $p$ -ичное целое. На третьем шаге записываем цифры полученного  $p$ -ичного представления в обратном порядке<sup>1)</sup>.

Например, пусть  $\alpha = 2/3$ ,  $p = 5$ ,  $r = 4$ . Тогда  $c = 2$ ,  $d = 3$ ,  $n = 0$  и  $p^r = 625$ . Следовательно, при помощи алгоритма 6.26 из гл. I получаем

$$(4.10) \quad |2/3|_{625} = 209.$$

Отметим, что число 209 является представлением числа  $2/3$  вычетом по модулю 625. Затем записываем равенство

$$(4.11) \quad 209_{\text{десять}} = 1314_{\text{пять}}.$$

Меняя в 5-ичном представлении порядок цифр на обратный, получаем код Гензеля

$$(4.12) \quad H(5, 4, 2/3) = .4131.$$

Цифры этого кода естественно совпадают с первой четверкой  $p$ -адических цифр представления для  $2/3$  в примере 2.36.

*Случай II.  $n < 0$ .*

Имеем представление  $\alpha = (c/d)p^{-k}$ , где  $k = -n$  является положительным целым числом. Используя указанные выше в случае I три шага, можно вычислить код Гензеля  $H(p, r, c/d)$ . Теперь сдвигаем  $p$ -адическую точку вправо на  $k$  разрядов и получаем тем самым искомый код Гензеля  $H(p, r, \alpha)$ <sup>2)</sup>.

Например, пусть  $\alpha = 2/15$ ,  $p = 5$ ,  $r = 4$ , тогда  $p^r = 625$ . Запишем  $\alpha$  в виде  $\alpha = (2/3)5^{-1}$ , что дает значения  $c = 2$ ,  $d = 3$ ,  $k = 1$ . Поскольку код Гензеля  $H(5, 4, 2/3)$  дается формулой (4.12), сдвигаем 5-адическую точку вправо на один разряд и получаем

$$(4.13) \quad H(5, 4, 2/15) = 4.131.$$

*Случай III.  $n > 0$ .*

Имеем  $\alpha = (c/d)p^k$ , где  $k = n > 0$ . Находим за три шага код Гензеля  $H(p, r, c/d)$ , как в случае I, и сдвигаем  $p$ -адическую точку на  $k$  разрядов влево.

<sup>1)</sup> И наконец, ставим перед ними  $p$ -адическую точку. — *Прим. перев.*

<sup>2)</sup> Если не ограничиваться рассмотрением дробей Фарея, то в принципе не исключена возможность  $k > r$ , когда при сдвиге точки вправо возникает ситуация, которую можно назвать «переполнением». Избежать возникновения подобных ситуаций позволяет код Гензеля с плавающей точкой (см. ниже). — *Прим. перев.*

Например, пусть  $\alpha = 10/3$ ,  $p = 5$ ,  $r = 4$ , тогда  $p^r = 625$ . Запишем  $\alpha$  в виде  $\alpha = (2/3)5$ , что дает значение  $c = 2$ ,  $d = 3$ ,  $k = 1$ . Код Гензеля  $H(5, 4, 2/3)$  уже известен из (4.12); сдвигаем 5-адическую точку влево на один разряд и получаем

$$(4.14) \quad H(5, 4, 10/3) = .0413.$$

Отметим, что при  $r = 4$  в коде Гензеля сохраняются только четыре цифры; в данном случае одна нулевая, крайняя правая цифра кода Гензеля  $H(5, 4, 2/3)$  при сдвиге точки утрачивается<sup>1)</sup>.

#### *Коды Гензеля для отрицательных рациональных чисел*

В теореме 2.19 установлена связь (бесконечных)  $p$ -адических представлений для чисел  $\alpha$  и  $-\alpha$ . В примере 2.20 в качестве иллюстрации указаны представления

$$(4.15) \quad \begin{aligned} 1/3 &= .2313131 \dots \quad (p=5), \\ -1/3 &= .3131313 \dots \quad (p=5). \end{aligned}$$

Следовательно, соответствующие коды Гензеля (с  $p = 5$ ,  $r = 4$ ) имеют вид

$$(4.16) \quad \begin{aligned} H(5, 4, 1/3) &= .2313, \\ H(5, 4, -1/3) &= .3131. \end{aligned}$$

Заметим, что крайняя левая (ненулевая) цифра кода Гензеля для положительного рационального числа дополняется по отношению к целому  $p$ , а каждая из последующих цифр — по отношению к  $p - 1$ . Таким образом, из (4.12) — (4.14), например, получаем

$$(4.17) \quad \begin{aligned} H(5, 4, -2/3) &= .1313, \\ H(5, 4, -2/15) &= 1.313, \\ H(5, 4, -10/3) &= .0131. \end{aligned}$$

**4.18. Замечание.** Следует указать, что алгоритм, который мы использовали для получения представлений (4.12) — (4.14), можно применять и в случае отрицательных дробей, а также при  $(a, b) \neq 1$  и  $(a, p) \neq 1$ . Приведем два примера.

Пусть  $\alpha = -2/4$ ,  $p = 5$ ,  $r = 4$ ,  $m = 625$ . Тогда имеет место представление

$$|-2/4|_{625} = 312,$$

<sup>1)</sup> При  $k > r$  происходит потеря всех значащих цифр. Эта ситуация, как и переполнение, исключается в кодах Гензеля с плавающей точкой (см. ниже). — *Прим. перев.*

которое можно получить при помощи алгоритма 6.26 из гл. I с начальной матрицей

$$\begin{bmatrix} 625 & 0 \\ 4 & -2 \end{bmatrix}.$$

Определим 5-ичную запись этого целого числа:

$$312_{\text{десять}} = 2222_{\text{пять}},$$

и выписываем код Гензеля

$$H(5, 4, -2/4) = .2222.$$

Легко проверить, что все дроби  $-1/2$ ,  $-3/6$ ,  $-4/8$  и т. п. имеют один код Гензеля. Наконец, заметим, что

$$H(5, 4, 1/2) = .3222$$

и что представления  $.2222$  и  $.3222$  взаимно дополнительные.

Подобным образом, если  $\alpha = 10/3$  (как в примере (4.14) для случая III выше), то

$$|10/3|_{625} = 420,$$

в чем можно убедиться, применяя алгоритм 6.26 из главы I с начальной матрицей

$$\begin{bmatrix} 625 & 0 \\ 3 & 10 \end{bmatrix}.$$

Поскольку 5-ичная запись этого целого числа есть

$$420_{\text{десять}} = 3140_{\text{пять}},$$

получаем код Гензеля

$$H(5, 4, 10/3) = .0413,$$

что согласуется с результатом (4.14).

### *Коды Гензеля с плавающей точкой*

При обсуждении представления (4.14) в случае III выше было сказано, что код Гензеля  $H(5, 4, 10/3)$  можно получить из кода  $H(5, 4, 2/3)$  простым сдвигом 5-адической точки, причем одна «значащая цифра» кода Гензеля для  $2/3$  утрачивается<sup>1)</sup>. Подобных неприятностей можно избежать, если ввести понятие нормализованных кодов Гензеля с плавающей точкой.

<sup>1)</sup> Это необратимая утрата в том смысле, что из кода Гензеля  $H(5, 4, 10/3)$  нельзя получить код  $H(5, 4, 2/3)$  сдвигом 5-адической точки в обратном направлении. — *Прим. перев.*

4.19. **Определение.** Пусть  $\alpha = a/b = (c/d)p^n$  и

$$(c, d) = (c, p) = (d, p) = 1.$$

*Нормализованным кодом Гензеля  $\hat{H}(p, r, \alpha)$  с плавающей точкой для числа  $\alpha$  называется пара*

$$\hat{H}(p, r, \alpha) = (m_\alpha, e_\alpha),$$

*состоящая из мантиссы*

$$m_\alpha = H(p, r, c/d)$$

*и показателя*

$$e_\alpha = n.$$

4.20. **Пример.** Вместо представлений (4.12)—(4.14) получаем

$$\hat{H}(5, 4, 2/3) = (.4131, 0),$$

$$\hat{H}(5, 4, 2/15) = (.4131, -1),$$

$$\hat{H}(5, 4, 10/3) = (.4131, 1),$$

а вместо (4.17)

$$\hat{H}(5, 4, -2/3) = (.1313, 0),$$

$$\hat{H}(5, 4, -2/15) = (.1313, -1),$$

$$\hat{H}(5, 4, -10/3) = (.1313, 1).$$

Заметим, что цифра кода  $H(5, 4, 2/3)$ , которая была утеряна в (4.14) при формировании кода  $H(5, 4, 10/3)$ , сохраняется, если мы используем для представления дроби  $10/3$  код Гензеля с плавающей точкой. Это окажется важным, когда речь пойдет об арифметических операциях над кодами Гензеля.

4.21. **Замечание.** Мантисса  $m_\alpha$  в нормализованном коде Гензеля с плавающей точкой есть обычный код для дроби  $c/d$  (в определении 4.19). Слева от  $r$  цифр мантиссы  $m_\alpha$  стоит  $p$ -адическая точка, и крайняя левая цифра (ближайшая к  $p$ -адической точке) отлична от нуля. Другими словами, дробь  $c/d$  является  $p$ -адической единицей (см. определение 3.11), и мантиссу можно трактовать как конечный<sup>1)</sup> отрезок беско-

<sup>1)</sup> Из  $r$  цифр. — *Прим. перев.*

нечного  $p$ -адического разложения этой дроби. Таким образом, мантиссы играют роль  $p$ -адических единиц при переходе от бесконечной системы  $p$ -адических чисел к системе нормализованных кодов Гензеля с плавающей точкой.

4.22. Таблица. Обычные коды Гензеля  $H(5,4, a/b)^{1)}$ 

$a \backslash b$	1	2	3	4	5	6	7	8
1	.1000	.2000	.3000	.4000	.0100	.1100	.2100	.3100
2	.3222	.1000	.4222	.2000	.0322	.3000	.1322	.4000
3	.2313	.4131	.1000	.3313	.0231	.2000	.4313	.1231
4	.4333	.3222	.2111	.1000	.0433	.4222	.3111	.2000
5	1.000	2.000	3.000	4.000	.1000	1.100	2.100	3.100
6	.1404	.2313	.3222	.4131	.0140	.1000	.2404	.3313
7	.3302	.1214	.4021	.2423	.0330	.3142	.1000	.4302
8	.2414	.4333	.1303	.3222	.0241	.2111	.4030	.1000
9	.4201	.3012	.2313	.1124	.0420	.4131	.3432	.2243
10	3.222	1.000	4.222	2.000	.3222	3.000	1.322	4.000
11	.1332	.2120	.3403	.4240	.0133	.1411	.2204	.3041
12	.3424	.1404	.4333	.2313	.0342	.3222	.1202	.4131
13	.2034	.4014	.1143	.3123	.0203	.2232	.4212	.1341
14	.4101	.3302	.2013	.1214	.0410	.4021	.3222	.2423
15	2.313	4.131	1.000	3.313	.2313	2.000	4.313	1.231
16	.1234	.2414	.3104	.4333	.0123	.1303	.2042	.3222
17	.3043	.1132	.4121	.2210	.0304	.3342	.1431	.4420

$a \backslash b$	9	10	11	12	13	14	15	16	17
1	.4100	.0200	.1200	.2200	.3200	.4200	.0300	.1300	.2300
2	.2322	.0100	.3322	.1100	.4322	.2100	.0422	.3100	.1422
3	.3000	.0413	.2231	.4000	.1413	.3231	.0100	.2413	.4231
4	.1433	.0322	.4111	.3000	.2433	.1322	.0211	.4000	.3433
5	4.100	.2000	1.200	2.200	3.200	4.200	.3000	1.300	2.300
6	.4222	.0231	.1140	.2000	.3404	.4313	.0322	.1231	.2140
7	.2214	.0121	.3423	.1330	.4142	.2000	.0402	.3214	.1121
8	.3414	.0433	.2303	.4222	.1241	.3111	.0130	.2000	.4414
9	.1000	.0301	.4012	.3313	.2124	.1420	.0231	.4432	.3243
10	2.322	.1000	3.322	1.100	4.322	2.100	.4222	3.100	1.422
11	.4324	.0212	.1000	.2332	.3120	.4403	.0340	.1133	.2411
12	.2111	.0140	.3020	.1000	.4424	.2404	.0433	.3313	.1342
13	.3321	.0401	.2430	.4410	.1000	.3034	.0114	.2143	.4123
14	.1134	.0330	.4431	.3142	.2343	.1000	.0201	.4302	.3013
15	3.000	.4131	2.231	4.000	1.413	3.231	.1000	2.413	4.231
16	.4402	.0241	.1421	.2111	.3340	.4030	.0310	.1000	.2234
17	.2024	.0113	.3102	.1240	.4234	.2323	.0412	.3401	.1000

<sup>1)</sup> См. [Krishnamurthy, Rao, Rubramanian, 1975 а, с. 70].

*Два рациональных числа с одинаковым кодом Гензеля*

Рассмотрим два разных рациональных числа  $\alpha$  и  $\beta$  и их канонические  $p$ -адические разложения

$$(4.23) \quad \begin{aligned} \alpha &= a_n p^n + a_{n+1} p^{n+1} + \dots, \\ \beta &= b_n p^n + b_{n+1} p^{n+1} + \dots \end{aligned}$$

Хотя  $\alpha \neq \beta$ , вполне возможно, что в этих разложениях первые  $r$  коэффициентов совпадают; тогда совпадают и коды Гензеля  $H(p, r, \alpha)$  и  $H(p, r, \beta)$ . Эта ситуация охарактеризована в следующей теореме.

**4.24. Теорема.** *Коды Гензеля двух рациональных чисел  $\alpha$  и  $\beta$  совпадают,  $H(p, r, \alpha) = H(p, r, \beta)$ , тогда и только тогда, когда число  $p^r$  делит разность  $\alpha - \beta$ <sup>1)</sup>.*

**Доказательство.** Пусть коды Гензеля совпадают; тогда

$$\begin{aligned} \alpha &= a_n p^n + \dots + a_{n+r-1} p^{n+r-1} + a_{n+r} p^{n+r} + \dots, \\ \beta &= b_n p^n + \dots + b_{n+r-1} p^{n+r-1} + b_{n+r} p^{n+r} + \dots, \end{aligned}$$

где число  $n$  может быть положительным, отрицательным или нулем. Для разности  $\alpha - \beta$  получаем разложение

$$\begin{aligned} \alpha - \beta &= (a_{n+r} - b_{n+r}) p^{n+r} + (a_{n+r+1} - b_{n+r+1}) p^{n+r+1} + \dots = \\ &= p^r [(a_{n+r} - b_{n+r}) p^n + (a_{n+r+1} - b_{n+r+1}) p^{n+1} + \dots], \end{aligned}$$

значит,  $p^r$  делит  $\alpha - \beta$ . Доказательство обратного утверждения предоставляется читателю.

Обратим внимание на тот факт, что дроби  $\alpha = a/b$  и  $\beta = g/h$  с  $(b, p) = (h, p) = 1$  представляются целыми числами  $|ab^{-1}|_m$  и  $|gh^{-1}|_m$  соответственно, где  $m = p^r$ . В замечании 4.18 установлено, что по этим целым числам, записанным в  $p$ -ичной системе, можно построить коды Гензеля для  $\alpha$  и  $\beta$  (меняя порядок цифр на обратный). На основе этих рассуждений легко доказать следующий результат.

**4.25. Теорема.** *Пусть  $\alpha = a/b$ ,  $\beta = g/h$ ,  $(b, p) = (h, p) = 1$  и  $m = p^r$ . Коды совпадают,  $H(p, r, \alpha) = H(p, r, \beta)$ , тогда и только тогда, когда*

$$|a \cdot b^{-1}|_m = |g \cdot h^{-1}|_m.$$

<sup>1)</sup> Приведенная формулировка неудачна, например, потому, что не указано, в каком смысле понимается выражение « $p^r$  делит (рациональное число!)  $\alpha - \beta$ ». В действительности имеется в виду, что  $\|\alpha - \beta\|_p = p^{-k} \|\alpha\|_p = p^{-k} \|\beta\|_p$ , где  $k \geq r$ . — Прим. ред.

Последнее равенство можно также записать в виде сравнения:

$$a \cdot b^{-1} \equiv g \cdot h^{-1} \pmod{p^r}.$$

4.26. **Следствие.** Пусть  $\alpha = a/b$ ,  $\beta = g/h$ ,  $(b, p) = (h, p) = 1$  и  $t = p^r$ . Коды совпадают,  $H(p, r, \alpha) = H(p, r, \beta)$ , тогда и только тогда, когда

$$|ah|_m = |bg|_m.$$

Последнее равенство можно также записать в виде сравнения:

$$ah \equiv bg \pmod{p^r}.$$

**Доказательство.** В теореме 4.25 обе части равенства и сравнения умножаем на  $bh$  и упрощаем.

Читателю рекомендуется сравнить эти теорему и следствие с теоремой 5.6 и следствием 5.7 гл. I.

4.27. **Пример** (предложен в работе [Lewis, 1979]). Пусть  $p = 5$ ,  $r = 4$ , так что  $p^r = 625$ . Рассмотрим два рациональных числа  $\alpha = 10/13$  и  $\beta = -35/17$ . Очевидно, число 625 делит разность  $\alpha - \beta = 625/221$ . Заметим также, что

$$\begin{aligned} 10 \cdot 13^{-1}(625) &= 10 \cdot 577 = 5770, \\ -35 \cdot 17^{-1}(625) &= -35 \cdot 478 = -16730 \end{aligned}$$

и что

$$5770 \equiv -16730 \pmod{625}.$$

Наконец, заметим, что

$$10 \cdot 17 \equiv 13(-35) \pmod{625}.$$

Следовательно, из теоремы 4.24, теоремы 4.25 или следствия 4.26 вытекает, что коды совпадают:

$$H(5, 4, 10/13) = H(5, 4, -35/17).$$

Отметим, что (см. пример 6.31 из гл. I)

$$|10 \cdot 13^{-1}|_{625} = |-35 \cdot 17^{-1}|_{625} = 145.$$

Запишем это десятичное число в 5-ичной форме

$$145_{\text{десять}} = 1040_{\text{пять}}.$$

Получаем

$$H(5, 4, 10/13) = H(5, 4, -35/17) = .0401,$$

что согласуется со значением кода  $H(5, 4, 10/13)$  в табл. 4.22.

Очевидно, каждый элемент обобщенного класса вычетов  $\mathbb{Q}_{145}$  (см. (5.3), гл. I), каждое рациональное число  $a/b$ , для



которого

$$(4.28) \quad |a \cdot b^{-1}|_{625} = 145,$$

имеет тот же код Гензеля, что и дробь  $10/13$ . Однако, как установлено в § 5 главы I, если класс  $\mathbb{Q}_k$  вообще содержит дробь Фарея порядка  $N$ , то только одну. В нашем примере именно число  $\alpha = 10/13$  оказывается единственной дробью Фарея порядка 17 в классе  $\mathbb{Q}_{145}$ , а дробь  $-35/17$  просто является другим представителем этого (бесконечного) класса.

**4.29. Замечание.** Следует обратить внимание на то обстоятельство, что в теоремах 4.24 и 4.25 фигурирует обычный код Гензеля, а не нормализованный код с плавающей точкой. Например,

$$\hat{H}(5, 4, 10/13) = (.4014, 1),$$

в то время как

$$\hat{H}(5, 4, -35/17) = (.4013, 1),$$

и мантиссы различаются на единицу в последнем знаке. Ясно, что совпадение нормализованных кодов Гензеля с плавающей точкой вызывает совпадение соответствующих обычных кодов, хотя обратное не всегда верно.

### Упражнения II.4

1. Выписать все дроби Фарея порядка 7, т. е. множество  $\mathbb{F}_7$ . Указание: использовать двумерный массив.

2. Если  $p = 7$  и  $r = 4$ , то чему равно  $N$ ?

3. Вычислить  $\hat{H}(5, 4, \alpha)$  для следующих случаев:

- |                       |                         |
|-----------------------|-------------------------|
| (a) $\alpha = 1/7$ ;  | (e) $\alpha = 13/15$ ;  |
| (b) $\alpha = 2/14$ ; | (f) $\alpha = -17/3$ ;  |
| (c) $\alpha = 17/3$ ; | (g) $\alpha = -15/7$ ;  |
| (d) $\alpha = 15/7$ ; | (h) $\alpha = -13/15$ . |

Указание: использовать алгоритм 6.26 гл. I. Сверить результаты с элементами табл. 4.22.

4. Вычислить  $\hat{H}(5, 4, \alpha)$  для следующих случаев:

- |                         |                        |
|-------------------------|------------------------|
| (a) $\alpha = 2/14$ ;   | (d) $\alpha = -15/7$ ; |
| (b) $\alpha = 15/7$ ;   | (e) $\alpha = 26/30$ ; |
| (c) $\alpha = -13/15$ ; | (f) $\alpha = -2/5$ .  |

5. В примере 4.27 показано, что  $H(5, 4, 10/13) = H(5, 4, -35/17)$ . Найти еще одну дробь  $\alpha = a/b$ , для которой

$$H(5, 4, \alpha) = H(5, 4, 10/13).$$

6. Завершить доказательство теоремы 4.24.

7. Доказать теорему 4.25.

### § 5. Арифметические операции над кодами Гензеля

В § 3 обсуждалось выполнение арифметических операций в поле  $\mathbb{Q}_p$   $p$ -адических чисел. Рассматривались операции сложения, вычитания, умножения и деления с (бесконечными)  $p$ -адическими разложениями в качестве операндов. Поскольку коды Гензеля являются просто конечными отрезками  $p$ -адических разложений, не вызывает удивления, что они могут быть операндами в арифметических действиях. Кроме того, не удивляет близкое сходство правил арифметики в  $\mathbb{Q}_p$  и арифметики кодов Гензеля.

В настоящем параграфе операнды предполагаются нормализованными кодами Гензеля при  $p=5$  и  $r=4$ , если не оговаривается противное. Имеем  $m=625$ ; тогда  $N=17$ .

#### Сложение

Рассмотрим следующий численный пример:

$$(5.1) \quad \alpha = 2/3 + 1/5.$$

Сложение этих дробей можно выполнить, используя коды Гензеля

$$(5.2) \quad \begin{aligned} \hat{H}(5, 4, 2/3) &= (.4131, 0), \\ \hat{H}(5, 4, 1/5) &= (.1000, -1). \end{aligned}$$

Мы просто выравниваем по вертикали  $p$ -адические точки и выполняем сложение цифр в 5-ичной арифметике, продвигаясь слева направо. Поскольку  $r=4$ , получаем

$$\begin{array}{r} .4131 \\ 1.000 \\ \hline 1.413 \end{array}$$

Таким образом, искомая сумма имеет мантиссу .1413 и показатель  $-1$ , т. е.

$$(5.3) \quad \hat{H}(5, 4, \alpha) = (.1413, -1).$$

Поскольку

$$(5.4) \quad \hat{H}(5, 4, 13/15) = (.1413, -1)$$

и число  $\alpha = 13/15$  является единственной дробью Фарея порядка 17, для которой код Гензеля имеет вид (5.3), мы отображаем<sup>1)</sup> код  $(.1413, -1)$  в

$$(5.5) \quad \alpha = 13/15.$$

Это правильный ответ; псевдопереполнение не возникает.

<sup>1)</sup> Алгоритм отображения кодов Гензеля в рациональные числа обсуждается в § 7.

*Вычитание*

Как и в поле  $\mathbb{Q}_p$ , вычитание выполняется как операция, «дополнительная к сложению» в том смысле, что вычитаемое заменяется дополнительным представлением, которое прибавляется к уменьшаемому. Например, чтобы вычислить разность

$$(5.6) \quad \alpha = 2/3 - 1/5$$

при помощи кодов Гензеля, требуется дополнительное представление

$$(5.7) \quad \hat{H}(5, 4, -1/5) = (.4444, -1).$$

Теперь, как и раньше при сложении, выравниваем  $p$ -адические точки и слева направо выполняем поразрядное сложение в 5-ичной арифметике. Поскольку  $r = 4$ , получаем

$$\begin{array}{r} .4131 \\ 4.444 \\ \hline 4.313 \end{array}$$

Таким образом, искомый ответ имеет мантиссу 4313 и показатель  $-1$ , т. е.

$$(5.8) \quad \hat{H}(5, 4, \alpha) = (.4313, -1).$$

Поскольку выполнено равенство

$$(5.9) \quad \hat{H}(5, 4, 7/15) = (.4313, -1)$$

и число  $\alpha = 7/15$  является единственной дробью Фарея порядка 17, для которой код Гензеля имеет вид (5.8), мы отображаем код  $(.4313, -1)$  в

$$(5.10) \quad \alpha = 7/15.$$

Это правильный ответ; псевдопереполнение не возникает.

*Умножение*

Рассмотрим следующий численный пример:

$$(5.11) \quad \alpha = (1/3) \cdot (6/5).$$

Умножение этих чисел выполним, используя коды Гензеля с плавающей точкой

$$(5.12) \quad \begin{cases} \hat{H}(5, 4, 1/3) = (.2313, 0), \\ \hat{H}(5, 4, 6/5) = (.1100, -1). \end{cases}$$

Алгоритм заключается в *умножении* мантисс и *сложении* показателей операндов. Таким образом, производя вычисления

в 5-ичной арифметике поразрядно (слева направо), получаем

$$\begin{array}{r} .2313 \\ \underline{.1100} \\ 2313 \\ \underline{231} \\ .2000 \end{array}$$

как мантиссу произведения и число  $0 + (-1) = -1$  как его показатель, т. е.

$$(5.13) \quad \hat{H}(5, 4, \alpha) = (.2000, -1).$$

Поскольку имеет место равенство

$$(5.14) \quad \hat{H}(5, 4, 2/5) = (.2000, -1)$$

и число  $\alpha = 2/5$  является единственной дробью Фарея порядка 17 с кодом Гензеля (5.13), мы отображаем код  $(.2000, -1)$  в

$$(5.15) \quad \alpha = 2/5.$$

Это правильный результат; псевдопереполнение не возникает.

### Деление

Рассмотрим следующий численный пример:

$$(5.16) \quad \alpha = 2/15 \div 4/15.$$

Деление этих чисел выполним, используя коды Гензеля с плавающей точкой

$$(5.17) \quad \begin{aligned} \hat{H}(5, 4, 2/15) &= (.4131, -1), \\ \hat{H}(5, 4, 4/15) &= (.3313, -1). \end{aligned}$$

Алгоритм заключается в *делении* мантисс и *вычитании* показателей *операндов*. Деление мантисс производится по правилам деления  $p$ -адических единиц (см. § 3, в частности, пример 3.26). Требуется определить число, обратное по модулю  $p$  к крайней левой цифре делителя. В данном примере мы делим .4131 на .3313, так что это число равно

$$(5.18) \quad 3^{-1}(5) = 2.$$

Выкладки проводятся по аналогии с примером 3.26 с той только разницей, что после вычисления  $r$  цифр частного про-

цесс заканчивается. Таким образом, имеем

$$\begin{array}{r}
 .3222 \\
 .3313 \overline{) 4131} \\
 \underline{1444} \\
 131 \\
 421 \\
 13 \\
 \underline{42} \\
 1 \\
 4 \\
 0
 \end{array}$$

т. е. мантисса частного равна .3222. При вычитании показателей операндов получаем нуль. Следовательно,

$$(5.19) \quad \hat{H}(5, 4, \alpha) = (.3222, 0).$$

Поскольку выполнено равенство

$$(5.20) \quad \hat{H}(5, 4, 1/2) = (.3222, 0)$$

и число  $\alpha = 1/2$  является единственной дробью Фарея порядка 17 с кодом (5.19), мы отображаем этот код (.3222, 0) в

$$(5.21) \quad \alpha = 1/2.$$

Это правильный ответ; псевдопереполнение не возникает.

**5.22. Пример.** Как частный случай деления рассмотрим нахождение обратного элемента

$$\beta = 1 \div 4/15 = 15/4$$

к числу  $\alpha = 4/15$ , т. е.  $\beta = 1/\alpha$ . Если использовать коды Гензеля

$$\hat{H}(5, 4, 1) = (.1000, 0),$$

$$\hat{H}(5, 4, 4/15) = (.3313, -1)$$

и произвести деление мантисс и вычитание показателей этих кодов, то ответом будет код

$$\hat{H}(5, 4, \beta) = (.2111, 1).$$

Поскольку число 15/4 является единственной дробью Фарея порядка 17, для которой

$$\hat{H}(5, 4, 15/4) = (.2111, 1),$$

мы отображаем код (.2111, 1) в

$$\beta = 15/4 = 1/\alpha.$$

Это правильный ответ; псевдопереполнение не возникает.

*Вычисление обратного к коду Гензеля  
при помощи метода Ньютона*

Мантиссу .2111 обратного числа в примере 5.22 можно найти уже известным нам способом «длинного деления»:

$$\begin{array}{r} .2111 \\ .3313 \overline{) 1000} \\ \underline{4213} \phantom{00} \\ 313 \phantom{00} \\ \underline{213} \phantom{00} \\ 31 \phantom{00} \\ \underline{21} \phantom{00} \\ 3 \phantom{00} \\ \underline{2} \phantom{00} \\ 0 \end{array}$$

в котором частное вычисляется цифра за цифрой. Однако этот последовательный процесс можно ускорить, если применить обобщение метода Ньютона нахождения обратного к вещественному числу на случай кодов Гензеля, см. [Krishnamurthy, 1970, 1971; Krishnamurthy, Murthy, 1983].

Напомним, что метод Ньютона для решения уравнения  $f(x) = 0$  состоит в построении последовательности приближений  $x_1, x_2, \dots$  при помощи итерационной формулы

$$(5.23) \quad x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}, \quad i = 0, 1, 2, \dots,$$

по заданному начальному приближению  $x_0$ , «выбранному соответственным образом». Выбирая функцию

$$(5.24) \quad f(x) = \frac{1}{x} - a,$$

приходим от (5.23) к итерационной процедуре

$$(5.25) \quad x_{i+1} = x_i(2 - ax_i), \quad i = 0, 1, 2, \dots$$

Последовательность  $\{x_i\}$  приближений метода (5.25) сходится к числу  $1/a$ . Метод Ньютона имеет квадратичную сходимость (в случае простых нулей функции  $f(x)$ ); поэтому если в начальном приближении  $x_0$  хотя бы одна цифра верна, то на каждой итерации количество верных цифр удваивается.

Чтобы обобщить метод Ньютона на случай  $p$ -адических чисел (и кодов Гензеля), нужно гарантировать, что:

(i) метод Ньютона применим к кодам Гензеля и обоснован в  $p$ -адической метрике и

(ii) количество верных цифр в мантиссе кода Гензеля удваивается на каждой итерации.

Поскольку в примерах типа 5.22 используется только мантисса нормализованного кода Гензеля с плавающей точкой и в требовании (ii) речь идет только о цифрах мантиссы, без ограничения общности далее можно рассматривать задачу обращения чисел с нулевым показателем в коде Гензеля. В этом случае мантисса нормализованного кода Гензеля совпадает с обычным кодом Гензеля для данного числа. Поэтому пункт (ii) можно переписать в виде<sup>1)</sup>

$$H(p, s, 1/\alpha) \rightarrow H(p, 2s, 1/\alpha).$$

Обоснованиями утверждений (i) и (ii) служат соответственно лемма Гензеля (см., например, [Коблиц, 1982, с. 31]) и теорема 5.36, приведенная ниже.

Для удобства выпишем код

$$(5.26) \quad H(p, r, \alpha) = .a_0 a_1 \dots a_{r-1},$$

который соответствует целому числу

$$(5.27) \quad a = a_0 + a_1 p + a_2 p^2 + \dots + a_{r-1} p^{r-1}.$$

( $a_0 \neq 0$ ), и код

$$(5.28) \quad H\left(p, r, \frac{1}{\alpha}\right) = .c_0 c_1 \dots c_{r-1},$$

который соответствует целому числу

$$(5.29) \quad c = c_0 + c_1 p + c_2 p^2 + \dots + c_{r-1} p^{r-1}$$

( $c_0 \neq 0$ ). Отметим равенство

$$(5.30) \quad |a \cdot c|_p = 1,$$

которое влечет за собой равенство  $c = a^{-1}(p^r)$ , т. е. число  $c$  является обратным к  $a$  по модулю  $p^r$ . Отметим также тот факт, что произведение кодов Гензеля (5.26) и (5.28) есть код .100 ... 0, откуда следует справедливость соотношения

$$(5.31) \quad c_0 = a_0^{-1}(p).$$

---

<sup>1)</sup> То есть на каждой итерации должно удваиваться количество знаков в кодах Гензеля числа  $1/\alpha$ . — *Прим. перев.*

Введем обозначения

$$\begin{aligned}
 b_1 &= c_0, & H\left(p, 1, \frac{1}{\alpha}\right) &= .c_0, \\
 b_2 &= c_0 + c_1p, & H\left(p, 2, \frac{1}{\alpha}\right) &= .c_0c_1, \\
 b_4 &= c_0 + c_1p + c_2p^2 + c_3p^3, & H\left(p, 4, \frac{1}{\alpha}\right) &= .c_0c_1c_2c_3, \\
 &\vdots & & \\
 b_r &= c_0 + c_1p + \dots + c_{r-1}p^{r-1}, & H\left(p, r, \frac{1}{\alpha}\right) &= .c_0c_1\dots c_{r-1},
 \end{aligned}
 \tag{5.32}$$

где  $r = 2^{i-1}$ . Имеем  $H(p, r, 2) = .200 \dots 0$ . Вспоминая соглашения (5.26) и (5.27), по аналогии с методом Ньютона (5.25) записываем итерационную схему

$$(5.33) \quad b_{2^k} = |b_{2^{k-1}}(2 - ab_{2^{k-1}})|_{p^{2^k}}, \quad k = 1, 2, \dots, i,$$

или, в эквивалентной форме,

$$\begin{aligned}
 (5.34) \quad H\left(p, 2^k, \frac{1}{\alpha}\right) &= H\left(p, 2^{k-1}, \frac{1}{\alpha}\right) * \\
 &\quad * [H(p, 2^k, 2) - H(p, 2^k, \alpha)] * \\
 &\quad * H\left(p, 2^{k-1}, \frac{1}{\alpha}\right)],
 \end{aligned}$$

где символ  $*$  обозначает умножение кодов Гензеля и «недостающие» разряды в коде  $H(p, 2^{k-1}, 1/\alpha)$  заполняются нулями<sup>2)</sup>.

Формальное обоснование итерационной схемы (5.33) будет дано в теореме 5.36. Прежде чем перейти к формулировке и доказательству этой теоремы, рассмотрим работу алгоритма (5.34) на численном примере.

**5.35. Пример.** Дан код  $H(5, 4, 3/11) = .3403$ ; требуется определить код  $H(5, 4, 11/3)$ . Вычисления проводятся сле-

<sup>1)</sup> Величина  $i$  здесь фиксирована и никак не связана с номерами  $i$  в (5.23) и (5.24). Роль номера итерации далее будет играть индекс  $k = 1, 2, \dots, i$ . — *Прим. перев.*

<sup>2)</sup> Поясним последнее замечание. Арифметические операции над кодами Гензеля с разным числом разрядов не определялись, поэтому в более аккуратной формулировке записи (5.34) следует заменить код  $H(p, 2^{k-1}, 1/\alpha)$  на код Гензеля с  $2^k$  разрядами, у которого первые  $2^{k-1}$  цифр совпадают с цифрами кода  $H(p, 2^{k-1}, 1/\alpha)$ , а остальные цифры нулевые. — *Прим. перев.*



дующим образом. Сначала, поскольку  $a_0 = 3$  и  $c_0 = a_0^{-1}(5)$ , находим значение  $c_0 = 2$ . Тем самым имеем код

$$H(5, 1, 11/3) = .2,$$

который аппроксимирует ответ с одной верной цифрой. Затем вычисляем код

$$\begin{aligned} H(5, 2, 11/3) &= (.20) * [(.20) - (.34) * (.20)] = \\ &= (.20) * [(.20) - (.14)] = \\ &= (.20) * (.11) = .22 \end{aligned}$$

уже с двумя верными цифрами. Наконец, получаем результат

$$\begin{aligned} H(5, 4, 11/3) &= (.2200) * [(.2000) - (.3403) * (.2200)] = \\ &= (.2200) * [(.2000) - (.1013)] = \\ &= (.2200) * (.1041) = .2231. \end{aligned}$$

**5.36. Теорема.** *Существует последовательность целых чисел*

$$\{b_1, b_2, b_4, \dots, b_r\}$$

с  $r = 2^i$ , такая, что справедливы равенства

$$|a \cdot b_{2^k}|_{p^{2^k}} = 1, \quad k = 0, 1, 2, \dots, i,$$

в которых число  $a$  из (5.27) отвечает коду Гензеля  $H(p, r, \alpha)$  из (5.26) и число  $b_{2^k}$  отвечает коду Гензеля  $H(p, 2^k, 1/\alpha)$  из (5.32).

**Доказательство** (по индукции). По определению  $b_1 = a_0^{-1}(p)$ ; тогда <sup>1)</sup>

$$\begin{aligned} |ab_1|_p &= |(a_0 + a_1p + \dots + a_{r-1}p^{r-1})a_0^{-1}|_p = \\ &= |a_0 \cdot a_0^{-1}|_p = 1. \end{aligned}$$

Сформулируем предположение индукции:

$$|a \cdot b_{2^{k-1}}|_{p^{2^{k-1}}} = 1.$$

Имеем

$$\begin{aligned} |a \cdot b_{2^k}|_{p^{2^k}} &= |a| (2 - a \cdot b_{2^{k-1}}) b_{2^{k-1}}|_{p^{2^k}} = \\ &= |a \cdot b_{2^{k-1}} (2 - a \cdot b_{2^{k-1}})|_{p^{2^k}}. \end{aligned}$$

---

<sup>1)</sup> Вместо  $a_0^{-1}(p)$  будем писать просто  $a_0^{-1}$ , когда значение  $p$  ясно.

В последнем выражении можно осуществить замену, используя тот вытекающий из предположения индукции факт, что для некоторого целого  $m$  верно равенство

$$a \cdot b_{2^{k-1}} = 1 + m \cdot p^{2^{k-1}}.$$

Проводя замену, получаем

$$\begin{aligned} |a \cdot b_{2^k}|_{p^{2^k}} &= |(1 + m \cdot p^{2^{k-1}})(2 - [1 + m \cdot p^{2^{k-1}}])|_{p^{2^k}} = \\ &= |(1 + m \cdot p^{2^{k-1}})(1 - m \cdot p^{2^{k-1}})|_{p^{2^k}} = 1. \end{aligned}$$

Таким образом, мы проверили утверждение при  $k = 0$  и доказали, что если оно верно для  $k - 1$ , то верно и для  $k$ . Следовательно, утверждение имеет место для всех номеров  $k \geq 0$ .

### *Повышение порядка сходимости*

Описанная выше итерационная процедура имеет квадратичную сходимость, т. е. количество верных  $p$ -адических цифр в приближениях удваивается на каждой итерации. В работе [Krishnamurthy, 1971] описана более общая процедура, позволяющая получить третий или даже более высокий порядок сходимости.

Чтобы описать это обобщение, повышающее порядок сходимости, приведем соотношение (5.33) к форме

$$\begin{aligned} (5.37) \quad b_{2^k} &= |b_{2^{k-1}} [1 + (1 - a \cdot b_{2^{k-1}})]|_{p^{2^k}} = \\ &= |b_{2^{k-1}} [1 + d_{k-1}]|_{p^{2^k}}, \quad k = 1, 2, \dots, i, \end{aligned}$$

где

$$(5.38a) \quad d_{k-1} = 1 - a \cdot b_{2^{k-1}}.$$

Сходимость порядка  $q > 2$  получаем при замене основания 2 в (5.38a) на  $q$ :

$$(5.38b) \quad d_{k-1} = 1 - a \cdot b_{q^{k-1}}.$$

Получаем итерационную схему

$$(5.39) \quad b_{q^k} = |b_{q^{k-1}} [1 + d_{k-1} (1 + d_{k-1} (1 + \dots) \dots)]|_{p^{q^k}},$$

в которой выражение в квадратных скобках имеет уровень вложенности  $q - 1$ . Например, кубическая сходимость обеспечивается, если использовать схему

$$\begin{aligned} (5.40) \quad b_{3^k} &= |b_{3^{k-1}} [1 + d_{k-1} (1 + d_{k-1})]|_{p^{3^k}} = \\ &= |b_{3^{k-1}} [1 + (1 - a \cdot b_{3^{k-1}})(2 - a \cdot b_{3^{k-1}})]|_{p^{3^k}}. \end{aligned}$$

5.41. **Пример.** Дан код  $H(5, 9, 7) = .210000000$ ; требуется найти код  $H(5, 9, 1/7)$ . Вычисления проводятся следующим образом. Сначала, поскольку  $a_0 = 2$ , определяем значение  $c_0 = 3$ ; тем самым код

$$H(5, 1, 1/7) = .3$$

аппроксимирует ответ с одной верной цифрой. Затем вычисляем код

$$\begin{aligned} H(5, 3, \tfrac{1}{7}) &= (.300) * [(.100) + \{(.100) - (.210) * (.300)\} * \\ &\quad * \{(.200) - (.210) * (.300)\}] = \\ &= (.300) * [(.100) + \{(.100) - (.140)\} * \\ &\quad * \{(.200) - (.140)\}] = \\ &= (.300) * [(.100) + (.014) * (.114)] = \\ &= (.300) * [(.100) + (.010)] = \\ &= (.300) * (.110) = \\ &= .330 \end{aligned}$$

уже с тремя верными цифрами. Наконец, получаем результат

$$\begin{aligned} H(5, 9, \tfrac{1}{7}) &= (.330000000) * [(.100000000) + \\ &\quad + \{(.100000000) - (.210000000) * (.330000000)\} * \\ &\quad * \{(.200000000) - (.210000000) * (.330000000)\}] = \\ &= (.330000000) * [(.100000000) + \\ &\quad + \{(.100000000) - (.100100000)\} * \\ &\quad * \{(.200000000) - (.100100000)\}] = \\ &= (.330000000) * [(.100000000) + \\ &\quad + \{.000444444\} * \{.100444444\}] = \\ &= (.330000000) * [(.100000000) + (.000444000)] = \\ &= (.330000000) * (.100444000) = \\ &= .330214230 \end{aligned}$$

## Упражнения II.5

1. При помощи конечноразрядной  $p$ -адической арифметики с  $p = 5$  и  $r = 4$ :

- |                              |                                |
|------------------------------|--------------------------------|
| (a) сложить $4/3$ и $1/6$ ;  | (e) умножить $4/3$ на $1/6$ ;  |
| (b) сложить $1/3$ и $5/2$ ;  | (f) умножить $1/3$ на $5/2$ ;  |
| (c) вычесть $4/3$ из $1/6$ ; | (g) разделить $4/3$ на $1/6$ ; |
| (d) вычесть $1/3$ из $5/2$ ; | (h) разделить $1/3$ на $5/2$ . |

2. Найти код Гензеля для  $\alpha$  и затем вычислить:

(а)  $H(7, 8, 1/\alpha)$ , где  $\alpha = 3/4$ , при помощи метода Ньютона с квадратичной сходимостью;

(б)  $H(7, 9, 1/\alpha)$ , где  $\alpha = 3/4$ , при помощи аналога метода Ньютона с кубической сходимостью;

(с)  $H(101, 8, 1/\alpha)$ , где  $\alpha = 2/7$ , при помощи метода Ньютона с квадратичной сходимостью;

(д)  $H(601, 9, 1/\alpha)$ , где  $\alpha = 7/11$ , при помощи аналога метода Ньютона с кубической сходимостью.

### § 6. Удаление первого нуля в коде Гензеля

В определении 4.19 предполагалось, что имеет место представление

$$(6.1) \quad \alpha = \left(\frac{c}{d}\right) p^n,$$

причем  $(c, d) = (c, p) = (d, p) = 1$ . Это позволяло записать нормализованный код Гензеля в виде

$$(6.2) \quad \hat{H}(p, r, \alpha) = (m_\alpha, e_\alpha),$$

где

$$(6.3) \quad m_\alpha = H\left(p, r, \frac{c}{d}\right).$$

Другими словами, мантисса в коде  $\hat{H}(p, r, \alpha)$  есть обычный код Гензеля для дроби  $c/d$ . В этом случае, конечно, крайняя левая (первая) цифра мантиссы  $m_\alpha$  отлична от нуля, что позволяет использовать  $m_\alpha$  в качестве делителя в операции деления.

Если число  $\alpha$  явилось результатом сложения или вычитания, то может оказаться, что мантисса не нормализована, т. е. первой цифрой мантиссы будет нуль. Если в последующих вычислениях эта мантисса является делителем, то возникают известные трудности, поскольку первая цифра кода Гензеля для делителя должна иметь обратный элемент по модулю  $p$ .

**6.4. Пример.** Требуется вычислить величину

$$x = \frac{a}{b + c}$$

при  $b = 1/2$  и  $c = 1/8$ . Если использовать коды Гензеля с плавающей точкой

$$\hat{H}(5, 4, 1/2) = (.3222, 0),$$

$$\hat{H}(5, 4, 1/8) = (.2414, 0),$$

то мантисса кода с плавающей точкой для суммы  $y = b + c$  является суммой мантисс операторов:

$$\begin{array}{r} .3222 \\ .2414 \\ \hline .0241 \end{array}$$

Таким образом, для числа  $y$  получается *ненормализованный* код Гензеля с плавающей точкой  $(.0241, 0)$ .

Отметим, что этот ненормализованный код нельзя использовать в качестве делителя, так как разложение мантиссы начинается с нуля (см. § 5). Следовательно, код необходимо нормализовать.

Очевидная процедура нормализации кода Гензеля включает два этапа. Сначала ненормализованный код переводится в дробь Фарея порядка  $N$ , затем эта дробь отображается в нормализованный код с плавающей точкой. Например, метод, который будет описан в следующем параграфе, позволяет перевести код  $(.0241, 0)$  в дробь

$$y = 5/8 = (1/8) 5.$$

Код Гензеля для дроби  $1/8$  уже выписывался ранее, и используя мантиссу этого кода, отображаем дробь  $5/8$  в нормализованный код с плавающей точкой:

$$\hat{H}(5, 4, 5/8) = (.2414, 1).$$

**6.5. Замечание.** Ненормализованный код Гензеля  $(.0241, 0)$  с первой цифрой мантиссы, равной нулю, нельзя нормализовать простым сдвигом мантиссы влево и согласованием порядка потому, что в новой записи

$$(.241x, 1)$$

четвертая цифра неизвестна. Следовательно, необходимо использовать метод из примера 6.4, чтобы убедиться в *эквивалентности* ненормализованного кода  $(.0241, 0)$  и нормализованного кода  $(.2414, 1)$  в том смысле, что они отвечают одной и той же дроби Фарея порядка  $N$ ; см. ниже замечание 7.2.

## § 7. Отображение кода Гензеля в единственную дробь Фарея порядка $N$

В предыдущем параграфе было показано, что мы сможем нормализовать ненормализованные коды Гензеля, коль скоро научимся отображать ненормализованные мантиссы в соответствующие единственные дроби Фарея порядка  $N$ . В настоящем параграфе рассматривается алгоритм такого отобра-

жения. (В этом алгоритме оказывается несущественным, нормализована мантисса или нет.)

Алгоритм включает следующие операции:

(i) отображение мантиссы в  $p$ -ичное целое число за счет обращения порядка цифр;

(ii) перевод  $p$ -ичного целого в  $\beta$ -ичное целое число, где  $\beta$  — основание системы счисления, в которой проводится следующий шаг (в наших примерах будем считать  $\beta = 10$ );

(iii) восстановление единственной дроби Фарея порядка  $N$  по этому целому числу из класса  $\hat{\Pi}_m$  (где  $m = p^r$ ) при помощи методов, описанных в § 6 гл. I.

**7.1. Пример.** В примере 6.4 упоминалось, что ненормализованному коду Гензеля с плавающей точкой  $(.0241, 0)$  отвечает дробь  $y = 5/8$  (напомним, что  $p = 5$ ,  $r = 4$ ,  $m = 625$ ,  $N = 17$ ). Величина  $y$  вычисляется следующим образом:

(i) ненормализованная мантисса отображается в пятеричное целое

$$.0241 \rightarrow 1420_{\text{пять}};$$

(ii) это пятеричное число переводится в десятичное

$$1420_{\text{пять}} = 235_{\text{десять}};$$

(iii) используется алгоритм 6.26 гл. I или метод «общего знаменателя», чтобы по целому числу 235 восстановить величину  $5/8$  — дробь Фарея порядка 17.

Алгоритм 6.26 приводит к таблице

	625	0
	235	1
2	155	-2
1	80	3
1	75	-5
1	5	8
15	0	-125

из которой видим, что  $y = 5/8$ .

В методе «общего знаменателя» требуется угадать общий знаменатель суммы<sup>1)</sup>

$$y = 1/2 + 1/8.$$

Очевидно, что эта сумма является рациональным числом  $u/v$  и в качестве общего знаменателя можно взять хотя бы число

<sup>1)</sup> Именно приведенная ниже сумма определяет значение  $y$  в примере 6.4. — *Прим. перев.*

$2 \cdot 8 = 16$ . Положим  $tv = 16$ . На шаге (ii) было получено значение

$$k = |u \cdot v^{-1}|_{625} = 235.$$

Отсюда заключаем, что

$$tu = /(tv) k /_{625} = /16 \cdot 235 /_{625} = 10,$$

и тогда  $y = 5/8$ .

**7.2. Замечание.** Когда мы уже отобразили мантиссу кода Гензеля с плавающей точкой (нормализованную или нет) в дробь Фарея порядка  $N$ , остается задача умножения этой дроби Фарея на  $p^n$ , где  $n$  — показатель в заданном коде Гензеля. Например, в замечании 6.5 утверждалось, что коды  $(.0241, 0)$  и  $(.2414, 1)$  эквивалентны. Теперь мы можем проверить это.

Как уже было установлено, ненормализованная мантисса  $.0241$  отображается в дробь Фарея порядка 17  $y = 5/8$ . Поскольку в соответствующем коде Гензеля нулевой показатель, имеем

$$(.0241, 0) \rightarrow 5/8.$$

Легко видеть, что нормализованная мантисса  $.2414$  отображается в дробь Фарея порядка 17  $\tilde{y} = 1/8$ , так как

$$.2414 \rightarrow 4142_{\text{пять}} = 547_{\text{десять}},$$

и применение алгоритма 6.26 гл. I приводит к таблице

	625	0
	547	1
1	78	-1
7	$\vdots 1 \vdots \vdots 8 \vdots$	
78	0	-625

Соответствующий код Гензеля имеет показатель  $n = 1$ ; тогда  $y = 5\tilde{y}$  и, следовательно,

$$(.2414, 1) \rightarrow 5/8.$$

Это показывает, что оба заданных кода Гензеля, нормализованный и ненормализованный, отвечают одному и тому же рациональному числу.

**7.3. Пример.** Пусть числу

$$\alpha = \frac{u}{v} = \left(\frac{c}{d}\right) p^n,$$

где  $(c, d) = (c, p) = (d, p) = 1$ , отвечают обычный код Гензеля

$$H(5, 4, \alpha) = 3.222$$

и нормализованный код Гензеля с плавающей точкой

$$\hat{H}(5, 4, \alpha) = (.3222, -1).$$

Мантисса .3222 отображается на целое число

$$2223_{\text{пять}} = 313_{\text{десять}},$$

следовательно

$$|c \cdot d^{-1}|_{625} = 313.$$

Далее рассмотрим два варианта вычисления.

(а). Используя алгоритм 6.26 гл. I, получаем таблицу

	625	0
	313	1
1	312	-1
1	1	2
312	0	-625

из которой выписываем значение

$$c/d = 1/2.$$

Наконец, учитывая величину показателя  $n = -1$ , умножаем  $c/d$  на  $1/5$ ; тогда

$$\alpha = 1/10,$$

что согласуется с табл. 4.22.

(б). Пусть нам известно, что  $gv = 20$ . Поскольку целое кратное числа  $v$  является целым кратным числа  $d$ , то  $td = 20$  для некоторого целого  $t$ . Следовательно, можно вычислить величину  $tc$ , применяя метод «общего знаменателя»:

$$tc = /20 \cdot 313/_{625} = 10,$$

откуда вытекает, что

$$c/d = 1/2.$$

Наконец, как и в случае (а), заключаем, что

$$\alpha = 1/10.$$

Заметим, что метод «общего знаменателя» из варианта (б) основан на предположении, что каким-то образом известно целое кратное числа  $v$ , скажем  $gv$ , где  $0 < g \leq N$ .



К сожалению, у нас нет способа априорно определять, лежит ли величина  $g$  в указанных границах. Единственное, что можно сделать — использовать алгоритм (не взирая на возможность  $g > N$ ) и затем проверить, является ли дробью Фарея порядка  $N$  полученная дробь  $u/v$ . Если нет, то вместо искомой дроби Фарея мы нашли другую дробь с тем же самым кодом Гензеля.

Предположим, что известна вся цепочка арифметических операций, приводящая к дроби  $u/v$ . Один из подходов к определению величины  $gv$  заключается в вычислении «текущего значения знаменателя» по ходу выполнения каждой арифметической операции. Когда в цепочке операций, определяющая  $u/v$ , пройдена, можно положить величину  $gv$  равной полученному знаменателю.

Более точно, пусть для начальных данных (рациональных чисел) известны не только их коды Гензеля, но и коды Гензеля их знаменателей (для операндов-делителей нужны коды Гензеля их числителей). Тогда при выполнении арифметической операции с очередной парой кодов Гензеля можно также вычислять и запоминать код Гензеля текущего значения общего знаменателя. Например, при выполнении сложения или вычитания текущее значение знаменателя есть наименьшее общее кратное знаменателей операндов. Детальное обсуждение этой процедуры см. в работах [Rao, 1975, с. 149—151, Krishnamurthy et al., 1975 b, s. 170—171].

Для некоторых конкретных алгоритмов можно предположить, используя их специфику, лучшие и более простые способы вычисления общего знаменателя. Например, пусть задана невырожденная матрица  $A = (a_{ij})$  с целыми элементами  $a_{ij}$ ; тогда в представлении для обратной матрицы  $A^{-1} = (\det A)^{-1} A^{\text{adj}}$  присоединенная<sup>1)</sup> матрица  $A^{\text{adj}}$  имеет только целые элементы. Следовательно, целое число  $\det A$  является общим знаменателем для рациональных элементов матрицы  $A^{-1}$ .

Аналогично при решении системы линейных алгебраических уравнений  $Ax = b$  имеем

$$(7.4) \quad x = A^{-1}b = (\det A)^{-1} A^{\text{adj}}b.$$

Таким образом, если целое число  $d$  оказывается общим знаменателем рациональных чисел, составляющих правую часть — вектор  $b$ , то рациональные компоненты решения  $x$  имеют общий знаменатель  $d \cdot (\det A)$ . При использовании метода Гаусса (т. е. метода последовательного исключения не-

<sup>1)</sup> Транспонированная к матрице алгебраических дополнений для  $A$ . — *Прим. перев.*

известных) величину  $\det A$  можно определить по ходу вычислений как произведение главных элементов исключения.

**7.5. Пример.** Требуется решить систему уравнений  $Ax = b$ , где

$$A = \begin{bmatrix} 2 & 2 & -1 \\ -3 & 0 & 2 \\ 4 & -5 & -1 \end{bmatrix} \quad \text{и} \quad b = \begin{bmatrix} 3 \\ -\frac{7}{2} \\ \frac{1}{2} \end{bmatrix}.$$

Применяем метод Гаусса (с делением ведущей строки на главный элемент). Расширенная матрица системы в прямом ходе процесса изменяется следующим образом:

$$\begin{aligned} & \begin{bmatrix} 2 & 2 & -1 & 3 \\ -3 & 0 & 2 & -\frac{7}{2} \\ 4 & -5 & -1 & \frac{1}{2} \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & -\frac{1}{2} & \frac{3}{2} \\ -3 & 0 & 2 & -\frac{7}{2} \\ 4 & -5 & -1 & \frac{1}{2} \end{bmatrix} \\ & \rightarrow \begin{bmatrix} 1 & 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 3 & \frac{1}{2} & 1 \\ 0 & -9 & 1 & -\frac{11}{2} \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & \frac{1}{6} & \frac{1}{3} \\ 0 & -9 & 1 & -\frac{11}{2} \end{bmatrix} \\ & \rightarrow \begin{bmatrix} 1 & 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & \frac{1}{6} & \frac{1}{3} \\ 0 & 0 & \frac{5}{2} & -\frac{5}{2} \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & \frac{1}{6} & \frac{1}{3} \\ 0 & 0 & 1 & -1 \end{bmatrix}. \end{aligned}$$

Последняя расширенная матрица отвечает системе уравнений

$$\begin{bmatrix} 1 & 1 & -1/2 \\ 0 & 1 & 1/6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 1/3 \\ -1 \end{bmatrix}.$$

Отметим, что операция деления производилась только при масштабировании строк и знаменателями были главные элементы 2, 3 и  $5/2$ . Произведение этих элементов дает величину определителя

$$\det A = 15.$$

Теперь, решая путем обратной подстановки приведенную выше систему уравнений с треугольной матрицей, получим

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \\ -1 \end{bmatrix}.$$

Предположим, что эти вычисления проводятся в конечно-разрядной  $p$ -адической арифметике. Можно показать (на-

пример, см. [Young, Gregory, 1973, с. 883]), что достаточно выбрать  $p$  и  $r$  из условия

$$p^r \geq 2 \left[ \sum_{j=1}^n |b_j| \right] \prod_{i=1}^n \left[ \sum_{k=1}^n a_{ik}^2 \right]^{1/2}.$$

Однако это условие не является необходимым и часто приводит к завышенным значениям  $p$  и  $r$ . В нашем случае оказывается возможным положить  $p = 5$  и  $r = 4$ , и это не приведет к псевдопереполнению, хотя  $p^r = 625$ , а правая часть в неравенстве примерно равна 981.

Начнем с записи расширенной матрицы  $[A, b]$  в  $p$ -адической форме, используя нормализованные коды Гензеля с плавающей точкой (код Гензеля  $(.0000, 0)$  соответствует числу нуль):

$$[A, b] = \begin{bmatrix} (.2000, 0) & (.2000, 0) & (.4444, 0) & (.3000, 0) \\ (.2444, 0) & (.0000, 0) & (.2000, 0) & (.4122, 0) \\ (.4000, 0) & (.4444, 1) & (.4444, 0) & (.3222, 0) \end{bmatrix}.$$

Применяя метод Гаусса (с делением ведущей строки на главный элемент), получаем матрицы

$$\begin{bmatrix} (.1000, 0) & (.1000, 0) & (.2222, 0) & (.4222, 0) \\ (.0000, 0) & (.3000, 0) & (.3222, 0) & (.1000, 0) \\ (.0000, 0) & (.1344, 0) & (.1000, 0) & (.2122, 0) \end{bmatrix},$$

$$\begin{bmatrix} (.1000, 0) & (.1000, 0) & (.2222, 0) & (.4222, 0) \\ (.0000, 0) & (.1000, 0) & (.1404, 0) & (.2313, 0) \\ (.0000, 0) & (.0000, 0) & (.3222, 1) & (.2222, 1) \end{bmatrix},$$

и

$$\begin{bmatrix} (.1000, 0) & (.1000, 0) & (.2222, 0) & (.4222, 0) \\ (.0000, 0) & (.1000, 0) & (.1404, 0) & (.2313, 0) \\ (.0000, 0) & (.0000, 0) & (.1000, 0) & (.4444, 0) \end{bmatrix}.$$

Произведение главных элементов равно

$$\hat{H}(5, 4, \det A) = (.2000, 0) \cdot (.3000, 0) \cdot (.3222, 1) = (.3000, 1);$$

таким образом,  $\det A = 15$ .

Чтобы решить систему уравнений с треугольной матрицей, проведем обратную подстановку в кодах Гензеля:

$$\hat{H}(5, 4, x_3) = (.4444, 0),$$

$$\begin{aligned} \hat{H}(5, 4, x_2) &= (.2313, 0) - (.4444, 0) \cdot (.1404, 0) = \\ &= (.3222, 0), \end{aligned}$$

$$\begin{aligned} \hat{H}(5, 4, x_1) &= (.4222, 0) - (.4444, 0) \cdot (.2222, 0) - \\ &- (.3222, 0) \cdot (.1000, 0) = (.3222, 0). \end{aligned}$$

Получаем

$$|x_3|_{625} = 624, \quad |x_2|_{625} = 313, \quad |x_1|_{625} = 313.$$

Поскольку  $b = [3, -7/2, 1/2]^T$ , общий знаменатель компонент вектора  $b$  равен  $d = 2$ . Таким образом, общий знаменатель компонент вектора  $x$  равен

$$d \cdot (\det A) = 30.$$

Следовательно, компоненты имеют значения

$$x_1 = \frac{1}{30} / 30 \cdot 624 / 625,$$

$$x_2 = \frac{1}{30} / 30 \cdot 313 / 625,$$

$$x_3 = \frac{1}{30} / 30 \cdot 313 / 625.$$

Тогда  $x = [x_1, x_2, x_3]^T = [1/2, 1/2, -1]^T$ .

Легко проверить, что применение алгоритма 6.26 гл. I приводит к тем же результатам.

**7.6. Замечание.** В работе [Krishnamurthy и др., 1975a] показано, что при четном  $r$  и *положительном целом*  $\alpha \in \mathbb{F}_N$  в коде  $H(p, r, \alpha)$  последние  $r/2$  цифры нулевые. Например, если  $p = 5$ ,  $r = 8$  и  $\alpha = 199$ , то

$$H(5, 8, 199) = .44210000.$$

Подобным образом, у *отрицательного целого*  $\alpha$  последние  $r/2$  цифры кода  $H(p, r, \alpha)$  принимают значение  $p - 1$  (напомним дополнительное представление отрицательных чисел); например,

$$H(5, 8, -199) = .10234444.$$

Эти факты и очевидное равенство  $v(u/v) = u$  служат обоснованием для следующего метода восстановления дроби по ее коду. Пусть известен код Гензеля дроби  $u/v$ . Суммируем этот код сам с собой пока не получим целое число (т. е. не больше  $N - 1$  раз, поскольку  $v \leq N$ ). Это целое число есть числитель  $u$  дроби, а количество слагаемых в сумме — ее знаменатель  $v$ . Например, пусть  $H(5, 4, u/v) = .4131$ . образуем сумму

$$\begin{array}{r} .4131 \\ .4131 \\ .4131 \\ \hline .2000, \end{array}$$

равную  $u = 2$ , с  $v = 3$  слагаемыми. Следовательно,  $u/v = 2/3$ .

Мы включили описание этого метода из-за его простоты, хотя он совсем не так полезен, как два предыдущие. К примеру, в работе [Beiser, 1979] обнаружено, что может встретиться такой код Гензеля, который приведет к целой сумме при числе слагаемых, меньшем  $v$ . Другими словами, мы можем, не дойдя до  $u/v$ , «споткнуться» на рациональном числе из обобщенного класса вычетов, в который входит дробь  $u/v$ .

Четвертый метод заключается в непосредственном просмотре таблицы и предложен в [Rao, Gregory, 1981]. Однако этот метод непрактичен и не будет здесь описан. Тем не менее следует отметить, что в указанной работе приводятся некоторые таблицы, проливающие свет на особенности соответствия между дробями Фарея порядка  $N$  и их целыми представлениями.

### Упражнения II.7

1. Какие нормализованные коды Гензеля с плавающей точкой эквивалентны следующим данным ( $p = 5$ ,  $r = 4$ )?

- (i) (.0121, 0);      (iii) (.0140, 0);  
(ii) (.0231, 0);      (iv) (.0330, 0).

2. Какие нормализованные коды Гензеля с плавающей точкой эквивалентны следующим обычным кодам Гензеля?

- (i)  $H(5, 4, \alpha) = 4.200$ ;      (iii)  $H(5, 4, \alpha) = 2.322$ ;  
(ii)  $H(5, 4, \alpha) = 3.231$ ;      (iv)  $H(5, 4, \alpha) = 3.100$ .

3. Для каждого кода Гензеля из задачи 1 найти соответствующую дробь Фарея порядка 17.

4. Для каждого кода Гензеля из задачи 2 найти соответствующую дробь Фарея порядка 17.

5. Найти  $\alpha$  — дробь Фарея порядка 6, если задан код:

- (i)  $\hat{H}(3, 4, \alpha) = (.2101, 0)$ ;  
(ii)  $\hat{H}(3, 4, \alpha) = (.1202, 1)$ ;  
(iii)  $\hat{H}(3, 4, \alpha) = (.1111, -1)$ .

6. Решить систему линейных алгебраических уравнений  $Ax = b$ , где

$$A = \begin{bmatrix} 2 & -1 & -3 \\ 3 & 0 & 1 \\ -1 & 2 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1/2 \\ 2 \\ -3/2 \end{bmatrix},$$

используя конечноразрядную  $p$ -адическую арифметику. Показать, что допустимы значения  $p = 5$  и  $r = 4$ , т. е. что соответствующее неравенство в примере 7.5 при таких значениях выполняется.

# Глава III

## Точное вычисление обобщенных обратных матриц

### § 1. Введение

Хорошо известно, что всякая квадратная невырожденная вещественная или комплексная матрица  $A$  имеет единственную *обратную* матрицу, определяемую свойствами

$$(1.1) \quad AA^{-1} = A^{-1}A = I.$$

Наличие обратной матрицы гарантирует, что у системы линейных уравнений  $Ax = b$  существует единственное решение

$$(1.2) \quad x = A^{-1}b.$$

Обратную может иметь только квадратная матрица. Квадратная матрица  $A$  допускает обратную тогда и только тогда, когда  $A$  невырождена, другими словами, тогда и только тогда, когда (1)  $\det A \neq 0$ , или (2) столбцы  $A$  линейно независимы, или (3) строки  $A$  линейно независимы; каждое из этих свойств влечет за собой два других.

Если матрица прямоугольная или квадратная, но вырожденная, то у нее нет обратной, зато имеется *обобщенная обратная матрица* (называемая также *g-обратной*), обладающая следующими свойствами:

(1) *g-обратная* существует для более широкого класса матриц, чем класс невырожденных матриц;

(2) *g-обратная* разделяет некоторые свойства обычной обратной матрицы;

(3) *g-обратная* совпадает с обычной обратной матрицей, если  $A$  квадратная и невырожденная.

Если  $A$  —  $(m \times n)$ -матрица, то *g-обратной* называется  $(n \times m)$ -матрица  $G$ , определяемая следующим образом.

**1.3. Определение.** Рассмотрим матричные уравнения

$$(a) \quad AGA = A,$$

$$(b) \quad GAG = G,$$

$$(в) \quad (AG)^* = AG,$$

$$(г) \quad (GA)^* = GA,$$

где  $*$  означает комплексное сопряжение и транспонирование. Матрица  $G$  называется

- (1)  $g$ -обратной для  $A$  (обозначается через  $A^-$ ), если выполняется (а);  
 (2) *рефлексивной*  $g$ -обратной для  $A$  (обозначается через  $A_R^-$ ), если выполняются (а) и (б);  
 (3)  $g$ -обратной *со свойством наименьших квадратов* (обозначается через  $A_L^-$ ), если выполняются (а) и (в);  
 (4)  $g$ -обратной *со свойством минимальной нормы* (обозначается через  $A_M^-$ ), если выполняются (а) и (г);  
 (5)  $g$ -обратной *Мура — Пенроуза* для  $A$  (обозначается через  $A^+$ ), если выполняются все четыре условия (а) — (г).

## § 2. Свойства $g$ -обратных матриц

В этом параграфе мы покажем, что для произвольной  $(m \times n)$ -матрицы  $A$  существует  $g$ -обратная каждого типа, указанного в определении 1.3; кроме того, будет доказано, что  $g$ -обратная Мура — Пенроуза для  $A$  единственна.

**2.1. Теорема.** *Если для квадратной невырожденной матрицы  $A$  существует  $A^-$ , то  $A^- = A^{-1}$ .*

**Доказательство.** Пусть  $A$  — квадратная невырожденная матрица, и пусть существует  $A^-$ , такая, что

$$AA^-A = A.$$

Умножая обе части этого равенства слева и справа на  $A^{-1}$ , получим  $A^- = A^{-1}$ .  $\square$

Хорошо известно (см. [Boullion, Odell, 1971, с. 2]), что для произвольной матрицы  $A$  найдутся (квадратные) невырожденные матрицы  $P$  и  $Q$  (не обязательно единственные), такие, что

$$(2.2) \quad R = PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

Здесь  $r$  — ранг матрицы  $A$ , а  $I_r$  — единичная матрица порядка  $r$ . Преобразование, описываемое равенством (2.2), будем называть *диагональной редукцией*  $A$ . (Мы используем символ 0 для всех трех нулевых блоков в  $R$ , хотя их размеры могут быть неодинаковы.) Если  $r = m$ , то нижние два нулевых блока отсутствуют; если  $r = n$ , то отсутствуют два правых нулевых блока.

Если транспонировать  $R$  и заменить нулевые блоки произвольными матрицами  $U$ ,  $V$  и  $W$  (подходящих размеров), то получим

$$(2.3) \quad \hat{R} = \begin{bmatrix} I_r & U \\ V & W \end{bmatrix}.$$

**2.4. Теорема.** Матрица  $G = Q\hat{R}P$  является  $g$ -обратной для  $A$ .

**Доказательство.** Согласно (2.2),  $A = P^{-1}RQ^{-1}$  и, следовательно,

$$\begin{aligned}AGA &= (P^{-1}RQ^{-1})(Q\hat{R}P)(P^{-1}RQ^{-1}) = \\ &= P^{-1}R\hat{R}RQ^{-1} = P^{-1}RQ^{-1} = A.\end{aligned}$$

Таким образом, уравнение (а) в определении 1.3 выполнено.  $\square$

Если положить  $U = 0$ ,  $V = 0$  и  $W = 0$  (где символ 0 в каждом случае обозначает нулевую матрицу с соответствующими размерами), то  $\hat{R}$  превращается в  $R^T$ .

**2.5. Теорема.** Если  $G = QR^TP$ , то  $G$  является рефлексивной  $g$ -обратной для  $A$ .

**Доказательство.** Согласно теореме 2.4,  $AGA = A$ . Аналогично проверяется, что

$$\begin{aligned}GAG &= (QR^TP)(P^{-1}RQ^{-1})(QR^TP) = \\ &= QR^TRR^TP = QR^TP = G.\end{aligned}$$

Итак, уравнения (а) и (б) в определении 1.3 выполнены.  $\square$

**2.6. Теорема.** Если  $A$  — произвольная матрица и  $A^-$  —  $g$ -обратная для  $A$ , то

- (1)  $(A^-)^*$  —  $g$ -обратная для  $A^*$ ;
- (2) если  $\lambda \neq 0$ , то  $(1/\lambda)A^-$  —  $g$ -обратная для  $\lambda A$ ;
- (3)  $\text{rank } A^- \geq \text{rank } A$ .
- (4) матрицы  $AA^-$  и  $A^-A$  идемпотентны и имеют тот же ранг, что и  $A$ .

**Доказательство.** Из  $AA^-A = A$  следует, что  $A^*(A^-)^*A^* = A^*$ , чем доказано (1). Если  $\lambda \neq 0$ , то

$$(\lambda A)((1/\lambda)A^-)(\lambda A) = \lambda(AA^-A) = \lambda A,$$

и (2) доказано. Так как ранг произведения матриц не превосходит ранга каждого из сомножителей, то из равенства  $AA^-A = A$  выводим

$$\text{rank } A^- \geq \text{rank } A,$$

т. е. (3) доказано. Легко видеть, что

$$\begin{aligned}(AA^-)^2 &= (AA^-A)A^- = AA^-, \\ (A^-A)^2 &= A^-(AA^-A) = A^-A,\end{aligned}$$



Таким образом, обе матрицы  $AA^-$  и  $A^-A$  идемпотентны. Наконец, соотношения

$$\begin{aligned}\text{rank } AA^- &\leq \text{rank } A, \\ \text{rank } A &\leq \text{rank } AA^-\end{aligned}$$

(последнее следует из свойства  $(AA^-)A = A$ ) приводят к равенству

$$\text{rank } A = \text{rank } AA^-.$$

Аналогично можно показать, что

$$\text{rank } A = \text{rank } A^-A.$$

Этим установлено (4).  $\square$

Поскольку обобщенная обратная матрица каждого типа, перечисленного в определении 1.3, удовлетворяет условию (а) этого определения, а само условие (а) является единственным предположением теоремы 2.6, то имеем очевидное следствие.

**2.7. Следствие.** Теорема 2.6 верна для каждой из  $g$ -обратных матриц  $A_R^-$ ,  $A_L^-$ ,  $A_M^-$ ,  $A^+$ .

**2.8. Теорема.** Если  $Y$  и  $Z$  —  $g$ -обратные для матрицы  $A$ , то  $G = YAZ$  — рефлексивная  $g$ -обратная для  $A$ .

**Доказательство.** Так как  $AYA = A = AZA$ , можем написать

$$AGA = (A(YAZ))A = (AYA)ZA = AZA = A.$$

Аналогично

$$\begin{aligned}GAG &= YAZ)A(YAZ) = Y(AZA)(YAZ) = \\ &= Y(AYA)Z = G.\end{aligned}$$

**2.9. Теорема.** Для произвольной матрицы  $A$  матрица

$$G = (A^*A)^-A^*$$

является рефлексивной  $g$ -обратной для  $A$  со свойством наименьших квадратов, а матрица

$$\hat{G} = A^*(AA^*)^-$$

— рефлексивной  $g$ -обратной для  $A$  со свойством минимальной нормы.

**Доказательство** оставляется читателю в качестве упражнения.  $\square$

**2.10. Теорема.** Пусть  $A$  — произвольная матрица,  $X$  —  $g$ -обратная для  $A$  со свойством минимальной нормы,  $Y$  —  $g$ -об-

ратная для  $A$  со свойством наименьших квадратов. Тогда  $G = XAY$  есть  $g$ -обратная Мура — Пенроуза для  $A$ .

**Доказательство.** Согласно теореме 2.8,  $G$  — рефлексивная  $g$ -обратная для  $A$ , так что  $AGA = A$  и  $GAG = G$ . Тем самым условия (а) и (б) определения 1.3 выполнены.

Так как  $X$  и  $Y$  обладают соответственно свойством минимальности нормы и свойством наименьших квадратов, то

$$\begin{aligned} AXA &= A, & AYA &= A, \\ (XA)^* &= XA, & (AY)^* &= AY. \end{aligned}$$

Отсюда

$$\begin{aligned} AG &= A(XAY) = AY, \\ GA &= (XAY)A = XA, \end{aligned}$$

Следовательно,

$$\begin{aligned} (AG)^* &= (AY)^* = AY = AG, \\ (GA)^* &= (XA)^* = XA = GA. \end{aligned}$$

Итак, условия (в) и (г) в определении 1.3 удовлетворяются, а потому  $G = A^+$ .  $\square$

Посредством этих теорем установлено существование  $g$ -обратных всех типов, указанных в определении 1.3. Докажем теперь единственность  $g$ -обратной матрицы Мура — Пенроуза.

**2.11. Теорема.** Для произвольной матрицы  $A$   $g$ -обратная матрица Мура — Пенроуза единственна.

**Доказательство.** Пусть  $G$  и  $\underline{G}$  — две  $g$ -обратные Мура — Пенроуза для  $A$ . Имеем

$$\begin{aligned} AGA &= A, & \underline{A}\underline{G}\underline{A} &= A, \\ GAG &= G, & \underline{G}\underline{A}\underline{G} &= \underline{G}, \\ (AG)^* &= AG, & (\underline{A}\underline{G})^* &= \underline{A}\underline{G}, \\ (GA)^* &= GA, & (\underline{G}\underline{A})^* &= \underline{G}\underline{A}. \end{aligned}$$

Следовательно,

$$G = GAG = G(AG)^* = GG^*A^*.$$

Так как  $A^* = A^*G^*A^*$ , то

$$\begin{aligned} G &= GG^*(A^*G^*A^*) = G(AG)^*(\underline{A}\underline{G})^* = \\ &= GAG\underline{A}\underline{G} = G\underline{A}\underline{G}. \end{aligned}$$

Аналогично можем показать, что

$$\underline{G} = G\underline{A}\underline{G}.$$

Отсюда  $\underline{G} = G$ .  $\square$

**2.12. Теорема.** Для некоторых специальных классов матриц  $A$   $g$ -обратная Мура — Пенроуза может быть вычислена согласно следующим правилам:

(1) если  $A$  — квадратная и невырожденная, то  $A^+ = A^{-1}$ ;  
 (2) если  $A = 0$ , то и  $A^+ = 0$ ; эти две нулевые матрицы взаимно транспонированы;

(3) если  $A$  —  $(m \times n)$ -матрица ранга  $n$ , то

$$A^+ = (A^* A)^{-1} A^*;$$

(4) если  $A$  —  $(m \times n)$ -матрица ранга  $m$ , то

$$A^+ = A^* (A A^*)^{-1}.$$

**Доказательство** предоставляется читателю в качестве упражнения.  $\square$

**2.13. Теорема.** Если  $A$  —  $(m \times n)$ -матрица ранга  $r$ , то существуют  $(m \times r)$ -матрица  $B$  и  $(r \times n)$ -матрица  $C$ , обе ранга  $r$ , такие, что  $A = BC$ . (Это представление называется скелетным разложением  $A$ .) Кроме того,

$$A^+ = C^* (B^* A C^*)^{-1} B^*.$$

**Доказательство** можно найти в книге [Ben-Israel, Greville, 1974, с. 23].  $\square$

**2.14. Теорема.** Для произвольной матрицы  $A$  справедливы следующие соотношения:

- (1)  $(A^+)^+ = A$ ,
- (2)  $(A^*)^+ = (A^+)^*$ ,
- (3)  $(A^T)^+ = (A^+)^T$ ,
- (4)  $A^+ = (A^* A)^+ A^*$ ,
- (5)  $A^+ = A^* (A A^*)^+.$

**Доказательство** предоставляется читателю.  $\square$

### Упражнения III. 2

1. Доказать теорему 2.9.
2. Доказать теорему 2.12.
3. Доказать теорему 2.14.
4. Доказать, что если  $A$  — эрмитова и идемпотентная матрица, то  $A^+ = A$ .
5. Доказать, что если  $U$  и  $V$  — унитарные, то  $(UAV)^+ = V^* A^+ U^*$  для всякой матрицы  $A$ , такой, что произведение  $UAV$  имеет смысл.



**3.6. Теорема.** *Нормальное решение единственно (что может быть неверно в отношении  $A_M^-$ ).*

**3.7. Определение.** Пусть  $Ax = b$  — несовместная система линейных алгебраических уравнений. Вектор  $y$  называется *псевдорешением*<sup>1)</sup> этой системы, если  $\|Ay - b\|_2 \leq \|Ax - b\|_2$  для всех  $x$ .

**3.8. Теорема.** Пусть  $Ax = b$  — несовместная система линейных алгебраических уравнений, и пусть  $A_L^-$  —  $g$ -обратная для  $A$  со свойством наименьших квадратов. Тогда  $y = A_L^- b$  — псевдорешение системы  $Ax = b$ .

Теоремы 3.5 и 3.6 дают нам способ вычисления нормального решения совместной системы  $Ax = b$ , а теорема 3.8 — способ вычисления псевдорешения несовместной системы. Однако псевдорешение не является единственным, и мы хотели бы выбрать среди псевдорешений то, которое имеет наименьшую норму.

**3.9. Определение.** Пусть  $Ax = b$  — несовместная система линейных алгебраических уравнений. Пусть  $y$  — такое псевдорешение системы, что  $\|y\|_2 \leq \|x\|_2$  для всякого псевдорешения  $x$ . В этом случае  $y$  называют *нормальным псевдорешением*<sup>2)</sup> системы  $Ax = b$ .

**3.10. Теорема.** Пусть  $Ax = b$  — несовместная система линейных алгебраических уравнений, и пусть  $A^+$  —  $g$ -обратная Мура — Пенроуза для  $A$ . Тогда  $y = A^+ b$  — нормальное псевдорешение системы  $Ax = b$ .

**3.11. Теорема.** Нормальное псевдорешение системы  $Ax = b$  единственно.

## § 4. Точное вычисление $A^+$ в случае рациональной матрицы $A$

Мы выбрали построение  $g$ -обратной Мура — Пенроуза как пример приложения безошибочных вычислений, поскольку это наиболее полезный тип  $g$ -обратных (например, такая матрица единственна), а также потому, что в обычной машинной арифметике вычислять ее довольно трудно. Нужно заметить, что при вычислении  $A^+$  для некоторых специальных классов матриц может пригодиться теорема 2.12.

<sup>1)</sup> В оригинале least-squares solution. — Прим. перев.

<sup>2)</sup> В оригинале minimum-norm least-squares solution. — Прим. перев.

В случае произвольной матрицы  $A$  многие алгоритмы требуют умения распознавать значение численного ранга  $A$ , а это крайне трудная задача при наличии ошибок округления, так как нужно решать, можно ли считать малые машинные числа представляющими точные нули.

Если  $A$  — (квадратная) невырожденная матрица, то обратная матрица является непрерывной функцией элементов  $A$ , т. е.

$$(4.1) \quad \lim_{E \rightarrow 0} (A + E)^{-1} = A^{-1}.$$

С другой стороны, если  $A$  — произвольная матрица, то  $A^+$  уже не обязательно непрерывно зависит от элементов  $A$ ; см., например, [Stewart, 1969].

4.2. **Пример** (см. [Rao, 1975]). Пусть  $A$  имеет вид

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 + \varepsilon \end{bmatrix}.$$

Тогда  $A^+ = \frac{1}{4} A$ , если  $\varepsilon = 0$ ; однако при  $\varepsilon \neq 0$

$$A^+ = \begin{bmatrix} 1 + 1/\varepsilon & -1/\varepsilon \\ -1/\varepsilon & 1/\varepsilon \end{bmatrix} = A^{-1}.$$

Существование подобных разрывов еще более затрудняет вычисление  $A^+$ . Ясно поэтому, что крайне желательно было бы находить  $A^+$  методами безошибочных вычислений, например посредством арифметики вычетов либо арифметики конечно-разрядных  $p$ -адических чисел.

В этой главе мы применим арифметику вычетов. Для вычисления  $A^+$  будет употреблен алгоритм, опирающийся на диагональную редукцию квадратной матрицы  $M = (AA^T)^2$  к форме (2.2). В работе [Rao et al., 1976] ее называют канонической формой Эрмита<sup>1)</sup> и вследствие этого алгоритм — алгоритмом Эрмита. Он использует уравнение

$$(4.3) \quad A^+ = A^T M_R^{-1} A^T,$$

причем вычисление  $M_R^{-1}$  основывается на разложении (2.2) и теореме 2.5.

Мы будем считать матрицу  $A$  целочисленной, поскольку рациональные элементы подходящим масштабированием могут быть превращены в целые (см., например, теорему 2.6, часть (2) и следствие 2.7).

<sup>1)</sup> Другое определение канонической формы Эрмита см. [Ben-Israel, Greville, 1974] или [Rao, Mitra; 1971].

Вслед за обсуждением алгоритма Эрмита мы рассмотрим еще два алгоритма; один из них предложен Гревиллом [Greville, 1960], а другой, описанный в [Decell, 1965], опирается на метод Леверье (см. [Фаддеев, Фаддеева, 1963]).

### Алгоритм Эрмита

По заданной (произвольной) матрице  $A$  легко построить (квадратную) матрицу

$$(4.4) \quad M = (AA^T)^2.$$

Наиболее трудная часть алгоритма Эрмита связана с вычислением  $M_R^-$  при помощи (2.2) и теоремы 2.5. Как только  $M_R^-$  найдена, перемножение матриц, указанных в (4.3), дает  $A^+$ .

**4.5. Определение.** Говорят, что матрица имеет *строчную нормальную форму*<sup>1)</sup>, если она удовлетворяет следующим условиям:

- (1) ненулевые строки предшествуют нулевым;
- (2) первый ненулевой элемент каждой ненулевой строки равен 1;
- (3) в столбце, где стоит такой элемент, все остальные элементы равны нулю;
- (4) для любых двух ненулевых строк с номерами  $i$  и  $j$ ,  $i < j$ , первый ненулевой элемент строки  $i$  расположен левее первого ненулевого элемента строки  $j$ .

**4.6. Определение.** Про матрицу, удовлетворяющую всем вышеперечисленным условиям, кроме третьего, говорят, что она имеет *простую строчную нормальную форму*.

Аналогичным образом можно определить *столбцовую нормальную форму* и *простую столбцовую нормальную форму*.

**4.7. Пример.** Пусть

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

Эта матрица имеет простую строчную нормальную форму. Однако если переставить первые две строки, то полученная матрица потеряет даже и простую строчную нормальную форму, поскольку не будет выполнено (4).

<sup>1)</sup> В оригинале row-echelon form. — Прим. перев.

Вернемся теперь к матрице  $M = (AA^T)^2$ . Ясно, что найдется невырожденная матрица  $E$ , такая, что

$$(4.8) \quad EM = M_1,$$

и  $M_1$  имеет простую строчную нормальную форму<sup>1)</sup>. Очевидно, что у  $M_1^T$  будет простая столбцовая нормальная форма. Поэтому существует невырожденная матрица  $F$ , для которой

$$(4.9) \quad FM_1^T = R = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}.$$

Здесь  $k$  — ранг матрицы  $M$  и квадратная матрица  $R$  имеет одновременно строчную и столбцовую нормальную форму. Таким образом,

$$(4.10) \quad FM^T E^T = R,$$

или (поскольку  $R$  симметрична)

$$(4.11) \quad R = EMF^T.$$

Это равенство соответствует разложению (2.2), где  $E = P$ ,  $F^T = Q$ . Поэтому на основании теоремы 2.5 имеем

$$(4.12) \quad M_R^- = F^T R E.$$

**4.13. Замечание.** Если произведение в правой части (4.12) записать в блочном виде, то можно заметить, что

$$\begin{aligned} M_R^- &= \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} G & H \\ K & L \end{bmatrix} = \\ &= \begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix} \begin{bmatrix} G & H \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

Отсюда следует, что для построения рефлексивной  $g$ -обратной для матрицы  $M$  нужны только первые  $r$  столбцов  $F^T$  и первые  $r$  строк  $E$ .

Вычисление матриц  $E$  и  $F$  проводится следующим образом. Последовательное левое умножение окаймленной матрицы  $[M: I]$  на элементарные матрицы порождает окаймленную матрицу  $[M_1: E]$ :

$$(4.14) \quad E_s \dots E_2 E_1 [M: I] = [M_1: E].$$

<sup>1)</sup> Такую матрицу  $E$  можно построить, выполняя последовательность элементарных преобразований строк матрицы  $A$ ; см. ниже. — *Прим. перев.*



Используя аналогичную процедуру для  $M_1^T$ , получим

$$(4.15) \quad F_t \dots F_2 F_1 [M_1^T : I] = [R : F].$$

Здесь  $F_i$ ,  $i = 1, 2, \dots, t$ , — элементарные матрицы.

#### Одномодульная арифметика вычетов

Так как, согласно нашему предположению,  $A$  — целочисленная матрица, то и  $M = (AA^T)^2$  также имеет целые элементы. Однако  $M_R^-$  и  $A^+$  в общем случае уже не будут целочисленными. Если

$$(4.16) \quad A^+ = (\alpha_{ij}),$$

то матрица с целыми элементами

$$(4.17) \quad a_{ij} = |\alpha_{ij}|_p$$

будет обозначаться как

$$(4.18) \quad |A^+|_p = (a_{ij}).$$

4.19. **Пример** [Rao et al., 1976]. Пусть имеем вырожденную матрицу

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Тогда  $M = (AA^T)^2$  имеет вид

$$\begin{bmatrix} 9 & 6 & 9 \\ 6 & 6 & 6 \\ 9 & 6 & 9 \end{bmatrix}.$$

Прежде чем вычислять  $A^+$  посредством арифметики вычетов, продемонстрируем, как это делается в обычной рациональной арифметике. Можно следовать процедуре, сходной с той, что использовалась в примере 7.5 гл. II; тогда получим

$$\begin{aligned} \begin{bmatrix} 9 & 6 & 9 & 1 & 0 & 0 \\ 6 & 6 & 6 & 0 & 1 & 0 \\ 9 & 6 & 9 & 0 & 0 & 1 \end{bmatrix} &\rightarrow \begin{bmatrix} 1 & \frac{2}{3} & 1 & \frac{1}{9} & 0 & 0 \\ 6 & 6 & 6 & 0 & 1 & 0 \\ 9 & 6 & 9 & 0 & 0 & 1 \end{bmatrix} \rightarrow \\ \rightarrow \begin{bmatrix} 1 & \frac{2}{3} & 1 & \frac{1}{9} & 0 & 0 \\ 0 & 2 & 0 & -\frac{2}{3} & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix} &\rightarrow \begin{bmatrix} 1 & \frac{2}{3} & 1 & \frac{1}{9} & 0 & 0 \\ 0 & 1 & 0 & -\frac{1}{3} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Это означает

$$M_1 = \begin{bmatrix} 1 & 2/3 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} 1/9 & 0 & 0 \\ -1/3 & 1/2 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

Далее мы окаймляем  $M_1^T$  и получаем

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 2/3 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2/3 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}.$$

Следовательно,

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad F = \begin{bmatrix} 1 & 0 & 0 \\ -2/3 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

Если воспользоваться замечанием 4.13, то для построения  $M_R^-$  нам понадобятся только первые два столбца  $F^T$  и первые две строки  $E$ :

$$M_R^- = \begin{bmatrix} 1 & -2/3 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/9 & 0 & 0 \\ -1/3 & 1/2 & 0 \end{bmatrix} = \begin{bmatrix} 1/3 & -1/3 & 0 \\ -1/3 & 1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Согласно (4.3), имеем

$$A^+ = A^T M_R^- A A^T = (1/6) \begin{bmatrix} 1 & 2 & 1 \\ -1 & 4 & -1 \\ 2 & -2 & 2 \end{bmatrix}.$$

Теперь мы вычислим  $M_R^-$ , а затем  $A^+$  посредством одно-модульной арифметики вычетов. В [Rao et al., 1976] предлагается выбирать простой модуль  $p$ , удовлетворяющий неравенству

$$p > 2 \prod_{j=1}^m \|c_j\|_2.$$

Здесь  $\|c_j\|_2$  обозначает евклидову норму  $j$ -го столбца  $M$  (если  $M$  содержит нулевые столбцы, они не участвуют в произведении). Пользуясь этим критерием, можно взять  $p = 4357$ .

Как и прежде, преобразуем окаймленную матрицу  $[|M|_p : I]$  в матрицу  $[|M_1|_p : |E|_p]$ , выполняя арифметику по

модулю  $p$ :

$$\begin{aligned} & \begin{bmatrix} 9 & 6 & 9 & 1 & 0 & 0 \\ 6 & 6 & 6 & 0 & 1 & 0 \\ 9 & 6 & 9 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1453 & 1 & 3873 & 0 & 0 \\ 6 & 6 & 6 & 0 & 1 & 0 \\ 9 & 6 & 9 & 0 & 0 & 1 \end{bmatrix} \rightarrow \\ & \rightarrow \begin{bmatrix} 1 & 1453 & 1 & 3873 & 0 & 0 \\ 0 & 2 & 0 & 2904 & 1 & 0 \\ 0 & 0 & 0 & 4356 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1453 & 1 & 3873 & 0 & 0 \\ 0 & 1 & 0 & 1452 & 2179 & 0 \\ 0 & 0 & 0 & 4356 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Итак,

$$|M_1|_{4357} = \begin{bmatrix} 1 & 1453 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad |E|_{4357} = \begin{bmatrix} 3873 & 0 & 0 \\ 1452 & 2179 & 0 \\ 4356 & 0 & 1 \end{bmatrix}.$$

Далее преобразуем окаймленную матрицу  $[|M_1^T|_p : I]$  в  $[|R|_p : |F|_p]$ . В результате

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1453 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2904 & 1 & 0 \\ 0 & 0 & 0 & 4356 & 0 & 1 \end{bmatrix}.$$

Таким образом,

$$|R|_{4357} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad |F|_{4357} = \begin{bmatrix} 1 & 0 & 0 \\ 2904 & 1 & 0 \\ 4356 & 0 & 1 \end{bmatrix}.$$

В этом месте мы воспользуемся замечанием 4.13; тогда получим

$$|M_R^-|_{4357} = \begin{bmatrix} 1 & 2904 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3873 & 0 & 0 \\ 1452 & 2179 & 0 \end{bmatrix} = \begin{bmatrix} 2905 & 1452 & 0 \\ 1452 & 2179 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Согласно (4.3), имеем

$$|A^+|_p = |A^T|_p |M_R^-|_p |AA^T|_p;$$

следовательно,

$$|A^+|_{4357} = \begin{bmatrix} 3631 & 2905 & 3631 \\ 726 & 1453 & 726 \\ 2905 & 1452 & 2905 \end{bmatrix}.$$

Наконец, пользуясь алгоритмом 6.26 гл. I, отобразим целые числа 3631, 2905, 726, 1452 и 1453 в их рациональные эквива-

ленты:

	4357	0			4357	0
	3631	1			2905	1
1	726	-1		1	1452	-1
5	1	6		2	1	3
726	0	-4357		1452	0	-4357
	4357	0			4357	0
	726	1			1452	1
6	1	-6		3	1	-3
726	0	4357		1452	0	4357
	4357	0			4357	0
	1453	1				
2	1451	-2				
1	2	3				
725	1	-2177				
2	0	4357				

Следовательно,

$$A^+ = \begin{bmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{6} & \frac{2}{3} & -\frac{1}{6} \\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

### Многомодульная арифметика вычетов

В примере 4.19 мы применили для обращения матрицы порядка 3 единственный модуль  $p = 4357$ . Очевидно, что при увеличении порядка приходится увеличивать и значение  $p$ . Чтобы не допускать сильного роста промежуточных результатов, можно употребить многомодульную арифметику вычетов. Это иллюстрируется следующим примером.

4.20. **Пример** [Rao, 1975]. Пусть имеем вырожденную матрицу

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}.$$

Тогда

$$M = \begin{bmatrix} 20 & 40 \\ 40 & 80 \end{bmatrix}.$$

Критерий

$$\rho > 2 \sum_{j=1}^2 \|c_j\|_2$$

показывает, что при использовании только одного модуля он должен быть не меньше 8000. Мы предпочтем работу с двумя модулями, произведение которых превосходит 8000, именно  $p_1 = 101$  и  $p_2 = 103$ . В каждом случае  $|A|_{p_i} = A$  и  $|M|_{p_i} = M$ .

*Случай  $p_1 = 101$*

Приводим вычисления, дающие  $|M_1|_{101}$  и  $|E_1|_{101}$ :

$$\begin{bmatrix} 20 & 40 & 1 & 0 \\ 40 & 80 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 96 & 0 \\ 40 & 80 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 96 & 0 \\ 0 & 0 & 99 & 1 \end{bmatrix}.$$

Отсюда

$$|M_1|_{101} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}, \quad |E_1|_{101} = \begin{bmatrix} 96 & 0 \\ 99 & 1 \end{bmatrix}.$$

Теперь вычисляем  $|R|_{101}$  и  $|F|_{101}$ :

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 99 & 1 \end{bmatrix}.$$

Отсюда

$$|R|_{101} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad |F|_{101} = \begin{bmatrix} 1 & 0 \\ 99 & 1 \end{bmatrix}.$$

В этом месте мы воспользуемся замечанием 4.13; тогда получим

$$|M_R^-|_{101} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} [96 \ 0] = \begin{bmatrix} 96 & 0 \\ 0 & 0 \end{bmatrix}.$$

Согласно (4.3), имеем

$$|A^+|_{101} = |A^T|_{101} |M_R^-|_{101} |AA^T|_{101}.$$

Следовательно,

$$|A^+|_{101} = \begin{bmatrix} 91 & 81 \\ 91 & 81 \end{bmatrix}.$$

*Случай  $p_2 = 103$*

Повторяя для этого случая те же шаги, находим

$$|A^+|_{103} = \begin{bmatrix} 31 & 62 \\ 31 & 62 \end{bmatrix}.$$

*Объединение результатов*

Воспользуемся обозначениями § 3 гл. I, чтобы ввести векторное основание

$$\beta = [101, 103].$$

Тогда

$$|A^+|_{\beta} = \begin{bmatrix} [91, 31] & [81, 62] \\ [91, 31] & [81, 62] \end{bmatrix}.$$

Наша двухмодульная арифметика вычетов эквивалентна одномодульной с единственным (непростым) модулем  $p = p_1 p_2$ . Поскольку в данном случае  $p = 10403$ , вычислим  $|A^+|_{10403}$  посредством процедуры из § 4 гл. I.

Прежде всего преобразуем представление

$$|\alpha_{11}|_{\beta} = [91, 31]$$

в целое число  $|\alpha_{11}|_{10403}$ , имеющее единственное смешанное представление

$$|\alpha_{11}|_{10403} = y_0 + x_1(101),$$

где  $0 \leq x_0 < 101$  и  $0 \leq x_1 < 103$ . Согласно (4.15), гл. I первый разряд  $x_0$  в смешанном представлении совпадает с остатком по первому модулю. Итак,

$$x_0 = 91.$$

Используя метод задачи 4.39 гл. I, в качестве второго разряда получим

$$x_1 = 30.$$

Итак,

$$|\alpha_{11}|_{10403} = 3121.$$

Аналогично находим, что

$$|\alpha_{12}|_{10403} = 6242.$$

Поскольку  $|\alpha_{11}|_{10403} = |\alpha_{21}|_{10403}$ ,  $|\alpha_{12}|_{10403} = |\alpha_{22}|_{10403}$ , то

$$|A^+|_{10403} = \begin{bmatrix} 3121 & 6242 \\ 3121 & 6242 \end{bmatrix}.$$

Теперь с помощью алгоритма 6.26 гл. I отобразим целые числа 3121 и 6242 в их рациональные эквиваленты:

	10403	0		10403	0
	3121	1		6242	1
3	1040	—3	1	4161	—1
3	1	10	1	2081	2
1040	0	—10403	1	2080	—3
			1	1	5
			2080	0	—10403

Итак,

$$A^+ = \begin{bmatrix} 1/10 & 1/5 \\ 1/10 & 1/5 \end{bmatrix}.$$

#### Алгоритм Гревилла

Этот алгоритм, предложенный в [Greville, 1960], описан также в книге [Krishnamurthy, Sen, 1976]<sup>1)</sup>. Обобщенная обратная Мура — Пенроуза для  $(m \times n)$ -матрицы

$$(4.21) \quad A = [a_1, a_2, \dots, a_n]$$

строится путем рекурсивной процедуры обращения матриц

$$(4.22) \quad A_i = [a_1, a_2, \dots, a_i].$$

Процедура начинается с обращения первого столбца  $a_1$ ; на каждом шаге к обращаемой матрице приписывается очередной столбец, так что (4.22) можно записать в виде

$$(4.23) \quad A_i = [A_{i-1} : a_i], \quad i = 2, 3, \dots, n.$$

Основная теорема (см. [Rao et al., 1976, с. 161] или [Rao, Mitra, 1971, с. 64]<sup>2)</sup>) связывает  $A_i^+$  с  $A_{i-1}^+$ .

**4.24. Теорема.** Пусть  $A_{i-1}$  — матрица с размерами  $m \times (i-1)$ , а  $a_i$  —  $m$ -вектор. Тогда

$$A_i^+ = \begin{bmatrix} A_{i-1}^+ - d_i b_i^T \\ b_i^T \end{bmatrix},$$

<sup>1)</sup> А также в книге Гантмахер Ф. Р. Теория матриц. — М.: Наука, 1966, с. 39. — Прим. перев.

<sup>2)</sup> Или книгу Ф. Р. Гантмахера (см. предыдущую сноску). — Прим. перев.

где

$$\begin{aligned} d_i &= A_{i-1}^+ a_i, \\ c_i &= a_i - A_{i-1} d_i, \\ b_i &= \begin{cases} \frac{1}{c_i^T c_i} c_i, & c_i \neq 0, \\ \frac{1}{1 + d_i^T d_i} (A_{i-1}^+)^T d_i, & c_i = 0. \end{cases} \end{aligned}$$

**Доказательство** см. [Greville, 1960]. □

Для начала рекурсии воспользуемся формулами

$$(4.25) \quad A_1^+ = \begin{cases} a_1^T, & a_1 = 0, \\ \frac{1}{a_1^T a_1} a_1^T & a_1 \neq 0. \end{cases}$$

Для выбора способа вычисления  $b_i$  критически важно, будет ли нулевым вектор  $c_i$ . В обычной арифметике с плавающей точкой, где имеются ошибки округления, крайне трудно установить, являются ли компоненты вектора  $c_i$  точными нулями. Поэтому алгоритм подвержен численной неустойчивости. Понятно, что безошибочные вычисления могут быть очень полезны для преодоления этой трудности.

#### Алгоритм Диселла — Леверье

В работе [Stallings, Boullion, 1972] арифметика вычетов применена для вычисления  $A^+$  в соответствии с алгоритмом Диселла [Decell, 1965], который в свою очередь опирается на метод Леверье (см. описание метода в книге [Фаддеев, Фаддеева, 1963]). Основой алгоритма является следующая теорема.

**4.26. Теорема.** Пусть  $A$  — произвольная комплексная  $(m \times n)$ -матрица, и пусть

$$B(\lambda) = (-1)^m (a_0 \lambda^m + a_1 \lambda^{m-1} + \dots + a_m)$$

— характеристический многочлен матрицы  $B = AA^*$ ;  $a_0 = 1$ . Если  $k$  — наибольшее целое число, для которого  $a_k \neq 0$ , то

$$A^+ = \begin{cases} -\frac{1}{a_k} A^* (a_0 B^{k-1} + a_1 B^{k-2} + \dots + a_{k-1} I), & k > 0, \\ 0, & k = 0. \end{cases}$$

При этом ранг матрицы  $A$  равен  $k$ .



**Доказательство** см. [Decell, 1965].

Исходя из этой теоремы, Столлингс и Бульон описывают *точный* метод вычисления ранга  $A$  и матрицы  $A^+$  в частном случае целочисленной матрицы  $A$ .

*Шаг 1.* Вычислить  $B = AA^T$ .

*Шаг 2.* Найти характеристический многочлен  $B(\lambda)$ . Для этого с помощью разбираемого ниже метода Леверье вычисляются коэффициенты  $a_0, a_1, \dots, a_m$ .

Пусть  $\lambda_1, \lambda_2, \dots, \lambda_m$  — собственные значения матрицы  $B$ ; положим

$$(4.27) \quad s_k = \sum_{i=1}^m \lambda_i^k, \quad 1 \leq k \leq m.$$

Тогда

$$(4.28) \quad s_k = \text{tr}(B^k).$$

Если воспользоваться формулами Ньютона (см. Фаддеев, Фаддеева, 1963, с. 311), то можно написать

$$(4.29a) \quad \begin{bmatrix} 1 & & & & & \\ s_1 & 2 & & & & \\ s_2 & s_1 & 3 & & & \\ s_3 & s_2 & s_1 & 4 & & \\ \dots & \dots & \dots & \dots & \dots & \\ s_{m-1} & s_{m-2} & s_{m-3} & s_{m-4} & \dots & m \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_m \end{bmatrix} = - \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ \vdots \\ s_m \end{bmatrix},$$

или попросту

$$(4.29b) \quad La = s.$$

Вычисление коэффициентов  $a_k$ , согласно (4.29), а затем вычисление  $A^+$  происходит следующим образом.

(1) По формулам (4.28) определяем числа  $s_k$ ,  $1 \leq k \leq m$ . Это требует вычисления степеней матрицы  $B$  и значения следа каждой степени.

(2) Подставляя значения  $s_k$  в (4.29), решаем эту систему с нижней треугольной матрицей  $L$  прямым ходом. Треугольную систему (4.29) можно решить и путем предварительного обращения матрицы  $L$ , используя для вычисления  $L^{-1}$  параллельный метод Кришнамурти [Krishnamurthy, 1983] (прямой ход является существенно последовательной процедурой). Затем полагаем  $a = L^{-1}s$ .

*Шаг 3.* Пользуясь теоремой 4.26, находим  $A^+$ .

Однако Столлингс и Бульон применяют для вычисления коэффициентов  $a_k$  последовательный алгоритм, связанный с

построением такой цепочки матриц <sup>1)</sup>:

$$\begin{aligned} A_0 &= 0, & q_0 &= -1, & B_0 &= I, \\ A_1 &= AA^T, & q_1 &= \text{tr } A_1, & B_1 &= A_1 - q_1 I, \\ A_2 &= A_1 B_1, & q_2 &= \frac{1}{2} \text{tr } A_2, & B_2 &= A_2 - q_2 I, \\ A_3 &= A_1 B_2, & q_3 &= \frac{1}{3} \text{tr } A_3, & B_3 &= A_3 - q_3 I, \\ &\dots & & & & \dots \\ A_k &= A_1 B_{k-1}, & q_k &= \frac{1}{k} \text{tr } A_k, & B_k &= A_k - q_k I. \end{aligned}$$

Заметим, что  $q_i$  только знаком отличаются от чисел  $a_i$  в выражении для  $A^+$  из теоремы 4.26:  $q_i = -a_i$ . Следовательно, при  $k > 0$

$$(4.30) \quad A^+ = \frac{1}{q_k} A^T B_{k-1}.$$

Поскольку  $k$  заранее не известно, то итерации продолжают-ся, пока не будет

$$(4.31) \quad A_1 B_k = 0.$$

Далее,  $A^T$  и, оказывается,  $B_{k-1}$  будут целочисленными матрицами, поэтому в данном методе нет необходимости хранить общий знаменатель. (Общим знаменателем является константа  $q_k$ .)

Критическое место в алгоритме Диселла — это проверка условия  $A_1 B_k = 0$ . Ясно, что безошибочные вычисления полезны для преодоления численной неустойчивости этого алгоритма.

**4.32. Замечание.** Мы не привели численных иллюстраций использования безошибочной арифметики для проведения двух последних алгоритмов. Такие иллюстрации можно найти в работах [Stallings, Boullion, 1972], [Rao et al., 1976]. Во второй из них вдобавок показано превосходство алгоритма Эрмита над двумя другими алгоритмами.

### Упражнения III.4

1. Вычислить  $M = (AA^T)^2$  для матриц:

$$(a) \quad A = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 0 & 1 \end{bmatrix};$$

$$(b) \quad A = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix};$$

$$(c) \quad A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}.$$

<sup>1)</sup> То есть используют метод Леверье в форме, предложенной Д. К. Фаддеевым. — *Прим. перев.*

2. Вычислить  $M_R^-$  для каждой матрицы из задачи 1.
3. Вычислить  $A^+$  для каждой матрицы из задачи 1, пользуясь алгоритмом Эрмита и обычной рациональной арифметикой.
4. Вычислить  $A^+$  для каждой матрицы из задачи 1, пользуясь алгоритмом Эрмита и арифметикой вычетов.
5. Вычислить  $A^+$  для матрицы из задачи 1(а), пользуясь алгоритмом Гревилла и обычной рациональной арифметикой.
6. Вычислить  $A^+$  для матрицы из задачи 1(с), пользуясь алгоритмом Диселла — Леверье и обычной рациональной арифметикой.

### § 5. Неудачи при применении арифметики вычетов и предупредительные меры

Необходимо указать некоторые трудности, связанные с использованием арифметики вычетов, а также предупредительные меры, позволяющие избежать этих трудностей.

Для произвольной матрицы над полем комплексных чисел верны следующие утверждения:

- (а)  $\text{rank } A = \text{rank } AA^* = \text{rank } A^*A$ ;
- (б)  $\text{rank } A = \text{rank } A^2$ , если  $A$  — квадратная эрмитова матрица;
- (в) если  $AA^* = 0$ , то  $A = 0$ , и обратно;
- (г) если  $A^*A = 0$ , то  $A = 0$ , и обратно.

Если матрица  $A$  вещественна, то в этих формулировках нужно заменить  $A^*$  на  $A^T$ .

К сожалению, ни одно из этих утверждений не обязано выполняться для матрицы над конечным полем. Проиллюстрируем это следующими примерами.

5.1. Пример. Рассмотрим матрицу

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Ее элементы — числа конечного поля  $(\mathbb{I}_5, +, \cdot)$ . Легко проверить, что  $\text{rank } A = 1$ ,  $AA^T = 0$ ,  $\text{rank } AA^T = 0$ ,

$$A^T A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

и  $\text{rank } A^T A = 1$ .

## 5.2. Пример. Рассмотрим матрицу

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

с элементами из конечного поля  $(\mathbb{P}_3, +, \cdot)$ . Легко видеть, что  $\text{rank } A = 1$ ,  $AA^T = A^T A = A^2 = 0$ ,  $\text{rank } A^2 = 0$ .

Из этих примеров мы заключаем, что ранг матрицы  $AA^T$  над полем вещественных чисел может отличаться от ранга матрицы  $|AA^T|_p$ , а именно возможно, что

$$(5.3) \quad \text{rank } |AA^T|_p < \text{rank } AA^T.$$

Поскольку все формулы для вычисления  $A_L^-$ ,  $A_M^-$  и  $A^+$  основаны на предварительном вычислении  $AA^T$ ,  $A^T A$  или  $(AA^T)^2$ , то понятно, что между  $|A_L^-|_p$ ,  $|A_M^-|_p$ ,  $|A^+|_p$  и соответствующими вещественными матрицами может не быть однозначного соответствия.

Действительно, если используется многомодульная арифметика вычетов и при различных простых  $p_i$  вычисляются

$$A_L^-|_{p_i}, |A_M^-|_{p_i} \text{ и } |A^+|_{p_i}, \text{ то для восстановления } A_L^-, A_M^-, A^+$$

по китайской теореме об остатках требуется, чтобы матрицы по модулю  $p_i$  имели правильный ранг. Таким образом, необходимо обеспечить, чтобы ранг матрицы  $|AA^T|_{p_i}$  был одним и тем же для каждого модуля  $p_i$ .

Один из способов добиться равенства рангов состоит в следующем. При вычислении  $A^+$  воспользуемся формулой (4.12) для построения  $M_R^-$ . Итак, вычисляем

$$(5.4) \quad |M_R^-|_{p_i} = |F^T R E|_{p_i}$$

и для  $p_j \neq p_i$

$$(5.5) \quad |M_R^-|_{p_j} = |F^T R E|_{p_j},$$

а затем сравниваем их ранги. Если

$$(5.6) \quad \text{rank } |M_R^-|_{p_i} < \text{rank } |M_R^-|_{p_j},$$

то  $p_i$  следует заменить другим модулем. Кроме того, нужна проверка, что новый выбор  $p_i$  не приводит к увеличению

ранга по сравнению с его значением для модуля  $p_j$ . Если ранг увеличился, то все ранее выбранные модули должны быть отвергнуты, а процесс начат заново. Чтобы уменьшить вероятность такой ситуации, следует использовать очень большие простые модули.

В случае применения алгоритма Диселла — Леверье можно проводить такую проверку: убедиться, что  $|a_k|_{p_i} \neq 0$  для каждого простого модуля. Это обеспечивает равенство

$$\text{rank} | AA^T |_{p_i} = \text{rank} | AA^T |_{p_j}$$

для всех  $p_i, p_j$ .

# Глава IV

## Целочисленные решения линейных уравнений

### § 1. Введение

Рассмотрим систему линейных алгебраических уравнений  $Ax = b$ , в которой  $A$  — целочисленная  $(m \times n)$ -матрица, а  $b$  —  $m$ -вектор с целыми компонентами. В общем случае решение системы не обязано быть целочисленным. Однако существует много ситуаций, где приходится разыскивать именно целочисленные решения таких систем.

Примером может служить построение оптимальных решений в задачах целочисленного программирования; здесь требуется, чтобы часть или даже все неизвестные были целыми числами (см. [Руле, Cline, 1973] или [Zlobec, Ben-Israel, 1970]<sup>1)</sup>). Если для вычисления решений в таких задачах используется арифметика с плавающей точкой, то вследствие ошибок округления результаты, скорее всего, не будут целыми. В этом случае нет уверенности, что мы получим правильный ответ простым округлением результатов. Основания для сомнений следующие:

(1) существуют разные способы округления; например, мы можем округлять «вверх», «вниз» или до ближайшего целого числа;

(2) даже если принят определенный способ округления, полученный результат может удовлетворять не всем ограничениям исходной задачи;

(3) даже если все ограничения удовлетворены, округленный результат может не совпадать с оптимальным решением.

Ясно поэтому, что при решении данного класса задач применение безошибочных вычислений предпочтительней арифметики с плавающей точкой.

Другой пример ситуации, где нужны целые решения систем линейных алгебраических уравнений с целочисленными коэффициентами, дает отрасль математической химии, называемую стехиометрией. Здесь мы имеем дело с весовыми отношениями, определяемыми химическими уравнениями и формулами. Соответственно очень важным является уравнивание реакций (как окислительно-восстановительных, так и идущих

---

<sup>1)</sup> См. также Корбут А. А., Финкельштейн Ю. Ю. Дискретное программирование. — М.: Наука, 1969 или Ху Т. Целочисленное программирование и потоки в сетях. — Пер. с англ. — М.: Мир, 1974. — *Прим. перев.*

без изменения степеней окисления). Большинство методов уравнивания основано на подборе коэффициентов путем проб; при этом используются правила, определяющие изменение степеней окисления [Andrews, Kokes, 1963; Benson, 1962], уравнивание полуреакций или электронный баланс.

В этой главе будет развит детерминированный метод, свободный от проб и пригодный для автоматизации. Вычисления без ошибок будут применены для отыскания (нетривиальных) целых решений системы линейных однородных алгебраических уравнений с матрицей коэффициентов (называемой матрицей реакции) неполного ранга. Так как коэффициенты этой матрицы целые, то при построении решения используется рефлексивная  $g$ -обратная со специальными целочисленными свойствами (см. [Krishnamurthy, 1978]).

Поскольку данный метод не опирается на химическое понятие восстановления-окисления, мы не рекомендуем пользоваться им для уравнивания химических реакций на семинарских занятиях. Вообще же говоря, польза его в том, что он позволяет:

(1) автоматически отвергать невозможные реакции, если матрица реакции имеет полный ранг;

(2) классифицировать реакцию как единственную (в смысле относительных пропорций), если ранг матрицы реакции ровно на единицу меньше полного;

(3) установить неединственность реакции, если ранг матрицы реакции меньше полного по крайней мере на две единицы.

До описания алгоритма мы изложим в следующем параграфе необходимый подготовительный материал.

## § 2. Основы теории

Здесь будет дано резюме с. 93—96 книги [Ben-Israel, Greville, 1974], которые в свою очередь используют работу [Hurt, Waid, 1970]. Пусть  $\Pi$  обозначает множество целых чисел; напомним, что  $(\Pi, +, \cdot)$  есть кольцо. Введем следующие обозначения:

(1)  $\Pi^m$  — множество  $m$ -мерных векторов над  $\Pi$ ;

(2)  $\Pi^{mn}$  — множество  $(m \times n)$ -матриц над  $\Pi$ ;

(3)  $\Pi_r^{mn}$  — множество  $(m \times n)$ -матриц ранга  $r$  над  $\Pi$ .

**2.1. Определение** (см. [Маркус, Минк, 1972, с. 63]). Пусть  $A$  — невырожденная матрица из  $\Pi^{nn}$ . Если  $A^{-1} \in \Pi^{nn}$ , то  $A$  называется *матрицей-единицей*.

**2.2. Теорема.** Если  $A$  — матрица-единица, то  $A$  — унимодулярная матрица, т. е.  $|\det A| = 1$ .

**Доказательство** предоставляется читателю в качестве упражнения.  $\square$

Рассмотрим теперь систему линейных алгебраических уравнений

$$(2.3) \quad Ax = b,$$

где  $A \in \mathbb{I}^{mn}$ ,  $b \in \mathbb{I}^m$ . Чтобы определить условия, при которых решение системы, если оно существует, удовлетворяет требованию  $x \in \mathbb{I}^n$ , мы опишем алгоритм вычисления рефлексивной обратной для матрицы  $A$ , опирающийся на каноническую форму Смита этой матрицы. Существование канонической формы Смита утверждает приводимая ниже теорема 2.5.

**2.4. Определение.** Матрицы  $A$  и  $S$  из  $\mathbb{I}^{mn}$  эквивалентны над  $\mathbb{I}$ , если существуют матрицы-единицы  $P \in \mathbb{I}^{mm}$  и  $Q \in \mathbb{I}^{nn}$ , такие, что  $PAQ = S$ .

**2.5. Теорема (существование).** Пусть  $A \in \mathbb{I}^{m \times n}$ . Тогда  $A$  эквивалентна над  $\mathbb{I}$  единственной матрице  $S = (s_{ij}) \in \mathbb{I}_r^{mn}$ , такой, что

$$(1) \quad s_{ii} > 0, \quad i = 1, 2, \dots, r,$$

$$(2) \quad s_{ij} = 0 \quad \text{для прочих } i, j,$$

и  $s_{ii}$  делит  $s_{i+1, i+1}$ ,  $i = 1, \dots, r-1$ . Матрица  $S$  называется канонической формой Смита матрицы  $A$ .

**Доказательство** см. [Маркус, Минк, 1972, с. 66—69].  $\square$

Зная каноническую форму Смита матрицы  $A$ , можно построить для нее рефлексивную  $g$ -обратную со специальными целочисленными свойствами. Это показано в приведенных ниже следствиях.

**2.6. Следствие.** Пусть  $S = PAQ$  — каноническая форма Смита для  $A$ , описанная в теореме 2.5. Тогда  $G = QS^+P$  есть рефлексивная  $g$ -обратная для  $A$ .

**Доказательство.** Из равенств

$$\begin{aligned} PAQ &= S = SS^+S = (PAQ)S^+(PAQ) = \\ &= PA(QS^+P)AQ = P(AGA)Q \end{aligned}$$

следует, что  $AGA = A$ . Кроме того,

$$\begin{aligned} GAG &= (QS^+P)A(QS^+P) = QS^+(PAQ)S^+P = \\ &= Q(S^+SS^+)P = QS^+P = G. \end{aligned}$$

Итак,  $G$  — рефлексивная  $g$ -обратная для  $A$ .  $\square$



**2.7. Следствие.** Матрица  $G = QS^+P$  из следствия 2.6 имеет следующие свойства:

$$(1) \quad AG \in \mathbb{I}^{mm},$$

$$(2) \quad GA \in \mathbb{I}^{nn}.$$

**Доказательство.** Из равенств

$$PAG = PA(QS^+P) = SS^+P$$

следует, что

$$AG = P^{-1}SS^+P \in \mathbb{I}^{mm}.$$

Аналогично из

$$GAQ = (QS^+P)AQ = QS^+S$$

вытекает

$$GA = QS^+SQ^{-1} \in \mathbb{I}^{nn}.$$

□

Мы будем пользоваться обозначением

$$(2.8) \quad G = A_{\bar{I}},$$

чтобы указать, что  $G$  — рефлексивная  $g$ -обратная для  $A$  (следствие 2.6) с дополнительными целочисленными свойствами, описываемыми в следствии 2.7.

**2.9. Теорема.** Пусть  $A \in \mathbb{I}^{mn}$ ,  $b \in \mathbb{I}^m$ . Предположим, что система  $Ax = b$  совместна. Целое решение существует тогда и только тогда, когда  $A_{\bar{I}}b \in \mathbb{I}^n$ ; если это условие выполнено, то общая формула для целых решений имеет вид

$$x = A_{\bar{I}}b + (I - A_{\bar{I}}A)y.$$

Здесь  $y$  — произвольный вектор из  $\mathbb{I}^n$ .

**Доказательство** можно найти в [Hurt, Waid, 1970] или [Bowman, Burdet, 1974].

Эта теорема вместе с теоремой 3.3 гл. III приводит очевидным образом к следующему результату.

**2.10. Следствие.** Пусть  $A \in \mathbb{I}^{mn}$ ,  $b \in \mathbb{I}^m$ . Система линейных алгебраических уравнений  $Ax = b$  тогда и только тогда имеет целое решение, когда

$$(1) \quad A_{\bar{I}}b \in \mathbb{I}^n$$

и

$$(2) \quad AA_{\bar{I}}b = b.$$

Если эти условия выполнены, то общая формула для целых решений имеет вид

$$x = A_I^{-1}b + (I - A_I^{-1}A)y.$$

Здесь  $y$  — произвольный вектор из  $\Pi^n$ .

2.11. **Замечание.** Для однородной системы уравнений, т. е. при  $b = 0$ , формула общего решения принимает вид

$$x = (I - A_I^{-1}A)y,$$

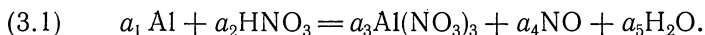
где  $y$  — произвольный вектор из  $\Pi^n$ . Отметим, что для квадратной невырожденной матрицы  $A$  справедливо  $A_I^{-1} = A^{-1}$ , и в этом специальном случае единственным решением будет  $x = 0$ .

### Упражнение IV.2

1. Доказать теорему 2.2.

## § 3. Матричная форма химических уравнений

Чтобы проиллюстрировать процедуру, которую мы собираемся описать, рассмотрим задачу уравнивания следующей химической реакции:



Алюминий реагирует с азотной кислотой, в результате чего получаются нитрат алюминия, окись азота и вода. Наша цель состоит в подборе  $a_1, a_2, \dots, a_5$ , удовлетворяющих закону сохранения атомов и закону сохранения электронов.

В данном конкретном случае мы будем рассматривать только неионизированные вещества. Это означает, что закон сохранения электронов автоматически будет выполнен, если выполнен закон сохранения атомов. (Если бы мы имели дело с ионизированными веществами, то нужно было бы учитывать электроны явным образом как часть системы, чтобы удовлетворить закон сохранения электронов.)

Пользуясь матричной формой химического уравнения, можно описать эту задачу математически. Это приводит к системе линейных однородных алгебраических уравнений относительно неизвестных  $a_1, a_2, \dots, a_5$ .

Предположим, что в химической реакции участвуют  $m$  различных химических элементов и  $n$  различных химических соединений (считая реагенты и продукты реакции). Обозначим через  $R = (r_{ij})$  ( $m \times n$ )-матрицу (матрицу реакции), кон-

струируемую следующим образом. Для  $1 \leq i \leq m$ ,  $1 \leq j \leq n$  целое число  $|r_{ij}|$  определяется как количество атомов типа  $i$  в соединении  $j$ . Само число  $r_{ij}$  положительно, если соединение  $j$  — реагент, и отрицательно, если это соединение — продукт реакции. Так, матрица реакции для (3.1) может быть построена исходя из табл. 3.2. Отметим, что химические со-

3.2. Таблица. Число атомов

	Al	HNO <sub>3</sub>	Al(NO <sub>3</sub> ) <sub>3</sub>	NO	H <sub>2</sub> O
Al	1	0	-1	0	0
H	0	1	0	0	-2
N	0	1	-3	-1	0
O	0	3	-9	-1	-1

единения в этой таблице упорядочены слева направо соответственно тому, как они появляются в (3.1); химические элементы упорядочены сверху вниз опять-таки согласно порядку появления в (3.1) при чтении слева направо. При таком соглашении

$$(3.3) \quad R = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -2 \\ 0 & 1 & -3 & -1 & 0 \\ 0 & 3 & -9 & -1 & -1 \end{bmatrix}.$$

Поэтому, если положить

$$(3.4) \quad a^T = [a_1 \ a_2 \ a_3 \ a_4 \ a_5],$$

то коэффициенты  $a_1, a_2, \dots, a_5$  в (3.1) можно получить, решая однородную систему линейных алгебраических уравнений

$$(3.5) \quad Ra = 0.$$

В общем случае  $R \in \mathbb{I}^{mn}$  при  $m \neq n$ . Однако и при  $m = n$  следует ожидать, что  $R$  вырождена; в противном случае мы имели бы только тривиальное решение  $a = 0$ , означающее, что никакая химическая реакция невозможна. Согласно замечанию 2.11, всякое целое решение системы (3.5) имеет вид

$$(3.6) \quad a = Wy, \quad y \in \mathbb{I}^n,$$

$$(3.7) \quad W = I - R_1^- R.$$

Ясно, что для этих вычислений нам нужна безошибочная арифметика.

### § 4. Решение однородной системы

Согласно (3.6) и (3.7), для решения однородной системы (3.5) требуется матрица  $W \in \mathbb{I}^{nn}$ . Очень важен следующий результат.

**4.1. Теорема.** Пусть  $\text{rank } R = r$ ,  $\text{rank } W = k$ . Тогда  $k$  есть дефект  $R$ , т. е.

$$k = n - r.$$

**Доказательство** предоставляется читателю в качестве упражнения.  $\square$

Ясно, что либо  $k = 0$ , либо  $k > 0$ . Это мотивирует рассмотрение таких двух случаев.

(1) Если  $r = n \leq m$ , то  $k = 0$ .

(2) Если  $r < n \leq m$  либо  $r \leq m < n$ , то  $k > 0$ .

Если  $R$  — матрица химической реакции типа (3.1), то можно классифицировать реакции, основываясь на значении  $r$ . Тогда возможны три случая.

(1) Если  $r = n$ , то  $k = 0$  и пространство решений системы  $Ra = 0$  нульмерно. Следовательно,  $a = 0$  — единственное решение (3.5), что означает: реакция не может быть уравнена; никакая химическая реакция в данном случае невозможна.

(2) Если  $r = n - 1$ , то  $k = 1$  и пространство решений системы  $Ra = 0$  одномерно. Это означает, что имеется единственное (с точностью до числового множителя) ненулевое решение. Таким образом, химическая реакция единственна, если говорить об относительных пропорциях реагентов и продуктов реакции.

(3) Если  $r \leq n - 2$ , то  $k \geq 2$  и пространство решений системы имеет размерность 2 или большую. Это означает, что возможны два или более линейно независимых решений, и химическая реакция неединственна даже в смысле относительных пропорций. Таким образом, реакцию можно уравнивать при различных (линейно независимых) относительных пропорциях реагентов и продуктов реакции.

#### Вычисление $R$

Согласно следствию 2.6, процедура построения  $R_l^-$  включает вычисление канонической формы Смита для  $R$ :

$$(4.2) \quad S = PRQ,$$

после чего находим

$$(4.3) \quad R_l^- = QS^+P.$$

Чтобы найти матрицы-единицы  $P$  и  $Q$ , действуем следующим образом. Приводим  $R$  к канонической форме Смита, пользуясь элементарными операциями трех типов:

- (1) переставить две строки (два столбца)  $R$ ;
- (2) умножить строку (столбец)  $R$  на ненулевое целое число и прибавить к другой строке (другому столбцу);
- (3) умножить строку (столбец)  $R$  на минус единицу.

Каждая из этих элементарных строчных (столбцовых) операций эквивалентна умножению  $R$  слева (справа) на элементарную матрицу. Элементарные матрицы, отвечающие указанным выше трем типам элементарных операций, имеют целые элементы. Так как  $P$  и  $Q$  суть произведения элементарных матриц, то  $P \in \mathbb{I}^{mm}$ ,  $Q \in \mathbb{I}^{nn}$ . Точно так же, поскольку обратные к этим элементарным строчным (столбцовым) операциям сохраняют тип, то  $P^{-1} \in \mathbb{I}^{mm}$ ,  $Q^{-1} \in \mathbb{I}^{nn}$ . Это гарантирует, что  $P$  и  $Q$  являются матрицами-единицами (см. определение 2.1).

Матрицы  $P$  и  $Q$  могут быть получены, если начать с двух единичных матриц соответствующих порядков и выполнять над одной из них (той, что представляет  $P$ ) те же *строчные*, а над другой (той, что представляет  $Q$ ) те же *столбцовые* операции, какие выполняются над  $R$ .

Для приведения мы применим модифицированный вариант алгоритма Эрмита из § 4 гл. III. Так как, пользуясь матрицами-единицами, в общем случае нельзя сделать ненулевые диагональные элементы равными 1, то мы удовлетворимся их приведением к целым значениям, каждое из которых делит последующее (каноническая форма Смита). Вначале мы исключим внедиагональные элементы первого столбца и первой строки в указанном порядке. Далее исключим внедиагональные элементы второго столбца и второй строки снова в указанном порядке. Процесс продолжается, пока не будет получена каноническая форма Смита.

**4.4. Пример.** Возьмем  $R$  из (3.3). Это матрица реакции для химического уравнения, выписанного в (3.1). Начинаем с  $R$  и двух единичных матриц:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -2 \\ 0 & 1 & -3 & -1 & 0 \\ 0 & 3 & -9 & -1 & -1 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Все внедиагональные элементы первого столбца  $R$  — нули, поэтому переключим внимание на первую строку. В ней имеется только один ненулевой внедиагональный элемент, который мы исключим прибавлением столбца 1 к столбцу 3:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -2 \\ 0 & 1 & -3 & -1 & 0 \\ 0 & 3 & -9 & -1 & -1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Теперь исключаем внедиагональные элементы второго столбца, прибавляя к строкам 3 и 4 строку 2, умноженную соответственно на  $-1$  и  $-3$ :

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -3 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -2 \\ 0 & 0 & -3 & -1 & 2 \\ 0 & 0 & -9 & -1 & 5 \end{bmatrix}.$$

Вслед за этим исключаем внедиагональный элемент второй строки, прибавляя столбец 2, умноженный на 2, к столбцу 5:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -3 & -1 & 2 \\ 0 & 0 & -9 & -1 & 5 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Этот процесс продолжается следующим образом. Умножаем строку 3 на  $-1$  и прибавляем к строке 4 умноженную на 3 новую строку 3:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -3 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & -2 \\ 0 & 0 & 0 & 2 & -1 \end{bmatrix}.$$

Переставляем столбцы 3 и 4:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 & -2 \\ 0 & 0 & 2 & 0 & -1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Прибавляем к строке 4 строку 3, умноженную на  $-2$ :

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & -2 & -1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & -6 & 3 \end{bmatrix}.$$

Прибавляем к столбцам 4 и 5 столбец 3, умноженный соответственно на  $-3$  и  $2$ :

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -6 & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -3 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Переставляем столбцы 4 и 5, после чего прибавляем к столбцу 5 столбец 4, умноженный на 2:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix}.$$

Из этих результатов следует, что

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -2 & -1 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix}$$

и

$$S^+ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

(относительно последнего равенства см. задачу 6 из упражнений III. 2). Таким образом,

$$R_I^- = QS^+P = \frac{1}{3} \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & -1 & -2 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & -5 & 2 \\ 0 & -2 & -1 & 1 \end{bmatrix}$$

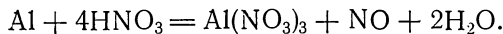
откуда

$$W = I - R_I^- R = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \end{bmatrix}.$$

Согласно (3.6), общая формула для целых решений системы  $Ra = 0$  имеет вид  $a = Wy$ , где  $y$  — произвольный вектор из  $\mathbb{I}^n$ . Поэтому

$$a = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = y_3 \begin{bmatrix} 1 \\ 4 \\ 1 \\ 1 \\ 2 \end{bmatrix}.$$

Например, при  $y_3 = 1$  получим уравненную химическую реакцию



**4.5. Замечание.** Вычисленная химическая реакция единственна в смысле относительных пропорций реагентов и продуктов; эти пропорции указаны выше. Единственность реакции согласуется с нашей теорией: очевидно, что ранг  $R$  равен  $n - 1$  и, следовательно, пространство решений системы



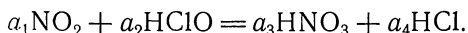
$Ra = 0$  одномерно (см. теорему 4.1 и сопровождающие ее замечания).

**4.6. Замечание.** Из следствия 2.7 мы знаем, что матрица  $W = I - R_l^- R$  имеет целые элементы, даже если не все элементы матриц  $S^+$  и  $R_l^- = QS^+P$  целые. Поэтому для вычисления  $W$  нужна безошибочная арифметика гл. I либо гл. II. В примере 4.19 гл. III использовалась арифметика вычетов, а в примере 7.5 гл. II — конечноразрядная  $p$ -адическая арифметика. И та и другая могут быть применены в примере 4.4.

### Дополнительные примеры

Следующие два примера иллюстрируют случаи, когда химическую реакцию нельзя уравнивать или же можно уравнивать, но относительные пропорции реагентов и продуктов определяются неединственным образом.

**4.7. Пример.** Рассмотрим химическое уравнение



Для него

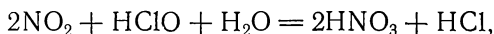
$$R = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 2 & 1 & -3 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 1 & 0 & -1 \end{bmatrix}.$$

Легко проверить, что  $r = 4$ ; так как  $m = n = 4$ , то  $R$  невырожденна. Можно вычислить матрицу

$$R^{-1} = \begin{bmatrix} 1 & 0 & -1 & 1 \\ -2 & 1 & -1 & 1 \\ 0 & 0 & -1 & 1 \\ -2 & 1 & -1 & 0 \end{bmatrix},$$

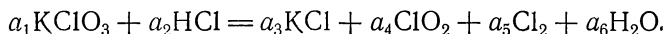
что дает  $W = I - R^{-1}R = 0$ , откуда  $a = 0$ . *Эту химическую реакцию уравнивать нельзя.*

С другой стороны, если добавить реагент  $\text{H}_2\text{O}$ , то получим уравненную реакцию



и эта реакция в смысле относительных пропорций единственна (см. [Krishnamurthy, 1978]).

**4.8. Пример.** Рассмотрим химическое уравнение



Для него

$$R = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & -2 & 0 \\ 3 & 0 & 0 & -2 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & -2 \end{bmatrix},$$

$$R_I^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -22 & 4 & 6 & -3 \\ 0 & 0 & 0 & 0 \\ 7 & -1 & -2 & 1 \\ -14 & 2 & 4 & -2 \\ -11 & 2 & 3 & -2 \end{bmatrix}.$$

и

$$W = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -18 & 16 & 8 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 6 & -4 & -2 & 0 \\ 0 & 0 & -12 & 10 & 5 & 0 \\ 0 & 0 & -9 & 8 & 4 & 0 \end{bmatrix}.$$

Общая формула для целых решений системы  $Ra = 0$  имеет вид  $a = Wy$ , где  $y$  — произвольный вектор из  $\mathbb{P}^n$ . Так как четвертый столбец  $W$  равен удвоенному пятому (т. е. в  $W$  только два линейно независимых столбца), то общее решение можно представить в виде линейной комбинации

$$a = k_1 \begin{bmatrix} 1 \\ -18 \\ 1 \\ 6 \\ -12 \\ -9 \end{bmatrix} + k_2 \begin{bmatrix} 0 \\ 8 \\ 0 \\ -2 \\ 5 \\ 4 \end{bmatrix}.$$

С математической точки зрения любые два целых числа  $k_1$  и  $k_2$  порождают решение системы  $Ra = 0$ . Так, например, при  $k_1 = 1$ ,  $k_2 = 3$  получим

$$a^T = [1 \ 6 \ 1 \ 0 \ 3 \ 3],$$

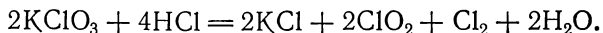
что дает возможную химическую реакцию



Аналогичным образом при  $k_1 = 2$ ,  $k_2 = 5$  находим

$$a^T = [2 \ 4 \ 2 \ 2 \ 1 \ 2],$$

что дает возможную химическую реакцию



Этот случай мы классифицируем как случай неединственной реакции, так как полученные уравнения не являются пропорциональными: они линейно независимы. Ясно, что существует бесконечно много выборов  $k_1$  и  $k_2$ , которые порождают возможные химические реакции.

**4.9. Замечание.** Результаты примера 4.8 подтверждают нашу теорию. Легко проверить, что ранг  $R$  равен  $n - 2$ , откуда следует, что пространство решений системы  $Ra = 0$  двумерно.

**4.10. Замечания.** К материалу § 3 и 4 добавим следующий комментарий.

(1) Представленная здесь матричная модель аналогична замкнутой статической модели «вход — выход» Леонтьева, используемой при исследовании некоторых аспектов экономического равновесия (см., например, [Куппе, 1963]).

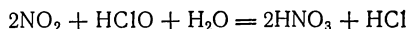
(2) Если химическая реакция неединственна, то для выбора одного из возможных решений можно применить термодинамические критерии (см. [Krishnamurthy, Adegbeyeni, 1977] или [Van Zeggeren, Storey, 1970]). Это соответствует оптимизации при наличии ограничений.

(3) Элементы  $R_I^-$  не обязаны быть непрерывными функциями элементов  $R$ . Поэтому требование, что для осуществимости химической реакции ранг матрицы  $R \in \mathbb{P}^{mn}$  должен быть меньше  $n$ , согласуется с известным фактом, что даже малые изменения  $R$  могут сопровождаться чрезвычайно большими, разрывными изменениями химической реакции.

#### Упражнения IV. 4

1. Доказать теорему 4.1.

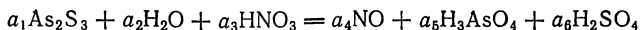
2. Проверить, что если в примере 4.7 добавить в качестве реагента  $\text{H}_2\text{O}$ , то уравненная химическая реакция



является единственно возможной (в смысле относительных пропорций).

3. Проверить вычисление  $R_I^-$  и  $W$  в примере 4.8 или, поскольку  $R_I^-$  определяется неединственным образом, найти другую матрицу  $R_I^-$  и соответствующую ей  $W$ . В последнем случае показать, что конечные результаты будут теми же самыми.

4. Уравнять реакцию



и показать, что химическая реакция единственна (в смысле относительных пропорций).

### § 5. Решение неоднородной системы

В этом параграфе будет рассматриваться не какое-либо конкретное приложение, а попросту задача решения неоднородной системы  $Ax = b$ ,  $b \neq 0$ .

5.1. **Пример.** Пусть имеем вырожденную матрицу [Hurt, Waid, 1970]

$$A = \begin{bmatrix} -9 & -8 & -5 \\ 6 & 5 & 2 \\ 3 & 2 & -1 \end{bmatrix}.$$

Канонической формой Смита этой матрицы будет

$$S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

а рефлексивной  $g$ -обратной со специальными целочисленными свойствами (следствие 2.7) — матрица

$$A_I^- = \frac{1}{3} \begin{bmatrix} 0 & 1 & 2 \\ 0 & -1 & -2 \\ 0 & 1 & -1 \end{bmatrix}.$$

Рассмотрим теперь систему  $Ax = b$  при нескольких выборах вектора  $b$ .

1. Пусть

$$b = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}.$$

Легко проверить, что

$$AA_I^-b = \begin{bmatrix} -4 \\ 1 \\ -2 \end{bmatrix},$$

т. е. условие (2) следствия 2.10 не выполнено. Поэтому система несовместна.

2. Пусть

$$b = \begin{bmatrix} 10 \\ -5 \\ 0 \end{bmatrix}.$$

Легко проверить, что  $AA_I^-b = b$ ; стало быть, система совместна. Однако

$$A_I^-b = \frac{5}{3} \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix},$$

т. е. условие (1) следствия 2.10 не выполнено. Таким образом, система не имеет целых решений, хотя и совместна.

3. Пусть

$$b = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

Поскольку  $AA_I^-b = b$  и

$$A_I^-b = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix},$$

то оба условия следствия 2.10 выполнены, и система имеет целые решения. Так как, согласно общей формуле для целых решений,

$$x = A_I^-b + (I - A_I^-A)y$$

( $y$  — произвольный вектор из  $\Pi^n$ ), то можем написать

$$x = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \begin{bmatrix} -3 & -3 & 0 \\ 4 & 4 & 0 \\ -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + k \begin{bmatrix} -3 \\ 4 \\ -1 \end{bmatrix}.$$

Здесь  $k = y_1 + y_2$  — произвольное целое число. Итак, общее решение имеет вид

$$\begin{aligned} x_1 &= 1 - 3k, \\ x_2 &= -1 + 4k, \\ x_3 &= -k. \end{aligned}$$

#### Упражнения IV.5

1. Проверить вычисление  $A_I^-$  в примере 5.1 либо, пользуясь неединственностью  $A_I^-$ , найти другую матрицу  $A_I^-$  и соответствующую матрицу

$I - A^{-1}A$ , а затем показать, что конечные результаты будут теми же самыми:

2. Показать, что система

$$\begin{bmatrix} 33 & 16 & 72 \\ -24 & -10 & -57 \\ 9 & 6 & 15 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 23 \\ -23 \\ 0 \end{bmatrix}$$

совместна, и найти целые решения, пользуясь методом данного параграфа.

## § 6. Решение интервальных задач линейного программирования

Имеется класс задач, называемых *задачами оптимизации*, в которых разыскивается максимум или минимум функции нескольких переменных, причем на значения этих переменных могут быть наложены некоторые ограничения. Более чем столетие для решения таких задач развивались классические методы оптимизации; они с успехом применялись к задачам физического, технического и экономического происхождения.

В годы второй мировой войны и последующие годы в области экономики появился новый большой класс оптимизационных задач; эти задачи объединяют под общим названием *задач программирования*. К ним относятся практические задачи государственного планирования, промышленности, военного дела. К сожалению, для задач этого типа классические методы оказались неэффективны, а потому пришлось разрабатывать новые методы их решения.

Типичным примером задачи программирования является задача об оптимальном размещении ограниченных ресурсов, где нужно достигнуть некоторой указанной цели при удовлетворении имеющихся ограничений (связей). Если все соотношения между переменными линейны (т. е. линейны и ограничения, и оптимизируемая функция), то говорят о *задаче линейного программирования*.

Специальный класс задач линейного программирования, называемых обычно *интервальными задачами линейного программирования*, может быть описан следующим образом (см., например, [Murthy, 1976] или [Rao, Mitra, 1971]).

6.1. Задача. Найти максимум целевой функции

$$\sum_{j=1}^n b_j x_j = b^T x$$

при интервальных ограничениях

$$c_i \leq \sum_{j=1}^n a_{ij} x_j \leq d_i, \quad i = 1, 2, \dots, m,$$

На матричном языке: разыскивается максимум функции  $b^T x$  при ограничениях

$$c \leq Ax \leq d.$$

Здесь  $b$  и  $x$  —  $n$ -мерные, а  $c$  и  $d$  —  $m$ -мерные векторы;  $A$  — матрица размером  $m \times n$ .

Алгоритм решения этой задачи основан на нижеследующих определениях и теоремах.

**6.2. Определение.** Интервальная задача линейного программирования (сокращенно ИЗ) называется *совместной*, если множество

$$S = \{x \in \mathbb{R}^n: c \leq Ax \leq d\}$$

непусто; в этом случае элементы  $S$  называются допустимыми решениями ИЗ.

**6.3. Определение.** Совместная ИЗ называется *ограниченной*, если величина

$$\max \{b^T x: x \in S\}$$

конечна. В этом случае *оптимальными решениями* ИЗ будут допустимые решения  $x_0$ , для которых

$$b^T x_0 = \max \{b^T x: x \in S\}.$$

**6.4. Лемма.** Совместная ИЗ будет ограниченной, если  $b$  принадлежит пространству столбцов  $A^T$ , или, что то же, если

$$b^T A^- A = b^T$$

для произвольной  $g$ -обратной  $A^-$ .

Доказательство дано на с. 193 книги [Rao, Mitra, 1971].

**6.5. Теорема.** Пусть имеем совместную и ограниченную ИЗ, причем  $\text{rank } A = m$ . Оптимальные решения ИЗ описываются формулой

$$x = A^- e + (I - A^- A) z,$$

где  $z$  — произвольный вектор из  $\mathbb{R}^n$ , а вектор  $e \in \mathbb{R}^m$  определяется так: для  $i = 1, 2, \dots, m$

$$e_i = \begin{cases} c_i, & (b^T A^-)_i < 0, \\ d_i, & (b^T A^-)_i > 0, \\ \text{произвольное число из } [c_i, d_i], & (b^T A^-)_i = 0. \end{cases}$$

Доказательство дано на с. 193 книги [Rao, Mitra, 1971].

Случай  $\text{rank } A < m$  рассматривается в книгах [Rao, Mitra, 1971; Ben-Israel, Greville, 1974].

Сформулируем теперь алгоритм решения задачи 6.1.

**6.6. Алгоритм.** Пусть  $A = (a_{ij})$ ,  $b^T x$ ,  $c$  и  $d$  суть величины, определенные в постановке задачи 6.1.

Этап 1. Вычислить вспомогательные матрицы  $E$  и  $F$  (см. формулы (4.8)–(4.9) гл. III).

Этап 2. Вычислить  $A_R^-$  в соответствии с формулой (4.12) гл. III.

Этап 3. Вычислить  $b^T A_R^-$ .

Этап 4. Вычислить  $b^T A_R^- A$ .

Этап 5. Если  $b^T = b^T A_R^- A$ , то выдать на печать сообщение «Задача ограничена» и перейти к этапу 6; в противном случае перейти к этапу 9.

Этап 6. Определить компоненты вектора  $e$  согласно формулам теоремы 6.5. Если  $(b^T A_R^-)_i = 0$ , полагать  $e_i = d_i$ .

Этап 7. Вычислить вектор  $x = A_R^- e + (I - A_R^- A)z$ , выбирая произвольный вектор  $z$  (мы полагаем  $z_i = 1$ ,  $i = 1, 2, \dots, n$ ).

Этап 8. Стоп.

Этап 9. Выдать на печать сообщение «Задача не ограничена» и перейти к этапу 8.

**6.7. Пример.** Предположим, что нужно найти максимум целевой функции

$$f(x_1, x_2, x_3) = 4x_1 + 5x_2 + 10x_3$$

при интервальных ограничениях

$$0 \leq 2x_1 + x_2 + 2x_3 \leq 140,$$

$$0 \leq 3x_1 + 4x_2 + 8x_3 \leq 360.$$

На матричном языке это эквивалентно максимизации функции  $b^T x$  при интервальных ограничениях  $c \leq Ax \leq d$ , где

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ 5 \\ 10 \end{bmatrix}, \quad c = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad d = \begin{bmatrix} 140 \\ 360 \end{bmatrix},$$

$$A = \begin{bmatrix} 2 & 1 & 2 \\ 3 & 4 & 8 \end{bmatrix}.$$

Если воспользоваться процедурой, описываемой формулами (4.8), (4.9) и (4.12) гл. III, то окажется, что  $\text{rank } A = 2$  и

$$A_R^- = \frac{1}{5} \begin{bmatrix} 4 & -1 \\ -3 & 2 \\ 0 & 0 \end{bmatrix}.$$



Следовательно,

$$AA_{\bar{R}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A_{\bar{R}}A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix},$$

$$I - A_{\bar{R}}A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -2 \\ 0 & 0 & 1 \end{bmatrix}.$$

Кроме того, найдем, что

$$b^T A_{\bar{R}}A = [4 \ 5 \ 10] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} = [4, 5, 10] = b^T,$$

т. е., согласно лемме 6.4, задача ограничена. Так как к тому же

$$b^T A_{\bar{R}} = (1/5)[1, 6],$$

то

$$(b^T A_{\bar{R}})_i > 0$$

для  $i = 1, 2$ . Поскольку  $\text{rang } A = 2$ , применима теорема 6.5, и

$$x = A_{\bar{R}}d + (I - A_{\bar{R}}A)z,$$

где  $z$  — произвольный вектор. Полагая

$$z = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

получим

$$x = \begin{bmatrix} 40 \\ 58 \\ 1 \end{bmatrix}$$

Заметим, что

$$Ax = \begin{bmatrix} 2 & 1 & 2 \\ 3 & 4 & 8 \end{bmatrix} \begin{bmatrix} 40 \\ 58 \\ 1 \end{bmatrix} = \begin{bmatrix} 140 \\ 360 \end{bmatrix}$$

и  $b^T x = 460$ .

6.8. **Пример** (с невырожденной матрицей  $A$ ). Пусть заданы

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

$$b = \begin{bmatrix} -4 \\ 1 \\ -3 \\ 2 \\ -2 \\ 3 \\ -1 \\ 4 \end{bmatrix}, \quad c = \begin{bmatrix} -9 \\ -9 \\ -9 \\ -9 \\ -9 \\ -9 \\ -9 \\ -9 \end{bmatrix}, \quad d = \begin{bmatrix} 9 \\ 9 \\ 9 \\ 9 \\ 9 \\ 9 \\ 9 \\ 9 \end{bmatrix}.$$

Матрица  $A$  — это хорошо известная невырожденная матрица; ее обратную можно указать в явном виде<sup>1)</sup>

$$A^{-1} = \frac{1}{n+1} B,$$

где для  $(n \times n)$ -матрицы  $B$

$$b_{ij} = \begin{cases} i(n-i+1), & i=j, \\ b_{i,j-1} - i, & j > i, \\ b_{ji}, & j < i; \end{cases}$$

см., например, с. 48 книги [Gregory, Karney, 1978]. Следовательно,

$$A^{-1} = \frac{1}{9} \begin{bmatrix} 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 7 & 14 & 12 & 10 & 8 & 6 & 4 & 2 \\ 6 & 12 & 18 & 15 & 12 & 9 & 6 & 3 \\ 5 & 10 & 15 & 20 & 16 & 12 & 8 & 4 \\ 4 & 8 & 12 & 16 & 20 & 15 & 10 & 5 \\ 3 & 6 & 9 & 12 & 15 & 18 & 12 & 6 \\ 2 & 4 & 6 & 8 & 10 & 12 & 14 & 7 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{bmatrix}.$$

<sup>1)</sup> Для произвольного порядка  $n$ . — Прим. перев.

Поскольку  $A$  не вырождена, то  $I - A^{-1}A = 0$ . Так как к тому же  $b^T A^{-1}A = b^T$ , то (по лемме 6.4) задача ограничена. Прямое вычисление дает

$$b^T A^{-1} = (1/9)[-30, -24, -27, -3, 3, 27, 24, 30],$$

откуда

$$e^T = [-9, -9, -9, -9, 9, 9, 9, 9].$$

Согласно теореме 6.4,

$$x^T = (A^{-1}e)^T = [-16, -23, -21, -10, 10, 21, 23, 16].$$

Заметим, что

$$(Ax)^T = [-9, -9, -9, -9, 9, 9, 9, 9]$$

и  $b^T x = 168$ .

#### Упражнение IV.6

1. Найти максимум целевой функции

$$f(x_1, x_2, x_3) = 3x_1 - x_2 - 2x_3$$

при интервальных ограничениях

$$-10 \leq 2x_1 + x_2 + 2x_3 \leq 10,$$

$$-20 \leq 3x_1 + 4x_2 + 8x_3 \leq 20.$$

### § 7. Решение полуцелых систем линейных уравнений

В этом параграфе мы опишем алгоритм для решения полуцелых систем линейных уравнений вида <sup>1)</sup>

$$(7.1) \quad Ax + By = c,$$

где  $A$  и  $B$  — матрицы с рациональными элементами, а  $c$  — вектор, компоненты которого также рациональны. Вектор  $x$  должен иметь целые компоненты, но у вектора  $y$  компоненты могут быть рациональными.

Заметим, что при  $B = 0$  мы получаем диофантову систему <sup>2)</sup>

$$(7.2) \quad Ax = c,$$

а при  $A = 0$  — стандартную систему линейных уравнений

$$(7.3) \quad By = c,$$

краткое обсуждение которой проводится в § 3 гл. III.

<sup>1)</sup> В оригинале systems of mixed-integer linear equations. — Прим. перев.

<sup>2)</sup> Имеется в виду, что разыскивается целое решение. — Прим. перев.

Пусть  $B_R^-$  — рефлексивная  $g$ -обратная для  $B$ . По теореме 3.3 гл. III система (7.3) разрешима тогда и только тогда, когда

$$(7.4) \quad BB_R^-c = c.$$

Если это условие выполнено, то общее решение имеет вид

$$(7.5) \quad y = B_R^-c + (I - B_R^-B)z,$$

где  $z$  — произвольный вектор с рациональными компонентами. Формула (7.5) легко проверяется посредством левого умножения на  $B$ .

Исследуем теперь возможность решения диофантовой системы (7.2) с помощью  $g$ -обратных аналогично тому, как это делалось в § 2 и 5, с тем отличием, что сейчас мы не требуем, чтобы  $A$  имела целые элементы.

В работе [Hurt, Waid, 1970] для этой системы использована  $g$ -обратная специального вида. Мы, однако, употребим класс  $g$ -обратных, предложенный в [Bowman, Burdet, 1974], поскольку он приводит к классическим формам Смита и Эрмита.

Аналогично теореме 2.9 и следствию 2.10 можно показать, что (7.2) разрешима тогда и только тогда, когда

$$(7.6) \quad A_I^-c \in \mathbb{I}^n$$

и

$$(7.7) \quad AA_I^-c = c.$$

Если эти условия выполнены, то общая формула для целых решений имеет вид

$$(7.8) \quad x = A_I^-c + (I - A_I^-A)w,$$

где  $w$  — произвольный вектор из  $\mathbb{I}^n$ .

Заметим, что задача (7.1) эквивалентна одновременному решению двух систем

$$(7.9) \quad \begin{aligned} Ax &= q, \\ By &= c - q, \end{aligned}$$

из которых первая — диофантова система, а вторая — стандартная линейная система. Можно объединить решения обеих систем в (7.9) и получить решение полуцелой системы (7.1).

В применении к (7.9) условие (7.4) принимает вид

$$(7.10) \quad BB_R^-(c - q) = c - q.$$

Таким образом, система  $By = c - q$  разрешима тогда и только тогда, когда выполнено (7.10). Можно придать этому

условию форму

$$(7.11) \quad (I - BB_R^-)q = (I - BB_R^-)c.$$

Итак, приемлемы только значения  $q$ , удовлетворяющие (7.11). (По определению  $q$  есть целочисленная комбинация столбцов  $A$ .) Если заменить  $q$  на  $Ax$ , то разрешимость полуцелой системы (7.1) или пары систем (7.9) равносильна разрешимости диофантовой системы

$$(7.12) \quad [(I - BB_R^-)A]x = (I - BB_R^-)c.$$

Чтобы упростить обозначения, положим

$$(7.13) \quad (I - BB_R^-)A = D,$$

$$(7.14) \quad (I - BB_R^-)c = d.$$

Теперь (7.12) превращается в

$$(7.15) \quad Dx = d.$$

Согласно теореме 2.9, система  $Dx = d$  тогда и только тогда имеет целое решение, когда

$$(7.16) \quad D_I^- d \in \Pi^n$$

и

$$(7.17) \quad DD_I^- d = d.$$

Если эти условия выполнены, то общая формула для целых решений такова:

$$(7.18) \quad x = D_I^- d + (I - D_I^- D)\omega,$$

где  $\omega$  — произвольный вектор из  $\Pi^n$ . Суммируя все сказанное, приходим к следующей теореме.

**7.19. Теорема.** *Полуцелая система линейных уравнений  $Ax + By = c$  разрешима тогда и только тогда, когда*

$$(1) \quad D_I^- d = [(I - BB_R^-)A]_I^- (I - BB_R^-)c \in \Pi^n,$$

$$(2) \quad DD_I^- d = [(I - BB_R^-)A][I - BB_R^-]_I^- (I - BB_R^-)c = \\ = (I - BB_R^-)c = d.$$

*Если эти условия выполнены, то общее решение имеет вид*

$$x = D_I^- d + (I - D_I^- D)\omega,$$

$$y = B_R^- c - B_R^- A D_I^- d - B_R^- A (I - D_I^- D)\omega + (I - B_R^- B)z,$$

*где  $z$  и  $\omega$  — произвольные векторы соответственно с рациональными и целыми компонентами.*

Этот вывод заимствован нами из работы [Bowman, Burdet, 1974]. Заметим, что если  $B = 0$ , а  $A$  квадратная и невырожденная, то

$$(7.20) \quad x = A^{-1}c;$$

если же  $A = 0$ , а  $B$  квадратная и невырожденная, то

$$(7.21) \quad y = B^{-1}c.$$

### Алгоритм

Перечислим теперь основные этапы построения векторов  $x$  и  $y$ , решающих полуцелую линейную систему  $Ax + By = c$ .

*Этап 1.* Найти какую-либо рефлексивную  $g$ -обратную  $B_R^-$  для матрицы  $B$  (пользуясь, например, формулой (4.12) и теоремой 2.5 гл. III). Затем вычислить

$$D = (I - BB_R^-)A, \quad d = (I - BB_R^-)c.$$

*Этап 2.* Вычислить  $D_I^-$ , следуя указаниям § 4.

*Этап 3.* Вычислить  $D_I^-d$ . Если все компоненты этого вектора целые, то перейти к этапу 4; в противном случае перейти к этапу 6.

*Этап 4.* Проверить условие  $DD_I^-d = d$ . Если оно выполнено, перейти к этапу 5; в противном случае перейти к этапу 6.

*Этап 5.* Вычислить векторы <sup>1)</sup>,

$$x = D_I^-d + (I - D_I^-D)w,$$

$$y = B_R^-c - B_R^-AD_I^-d - B_R^-A(I - D_I^-D)w + (I - B_R^-B)z$$

и перейти к этапу 7.

*Этап 6.* Выдать на печать сообщение «Решения нет»; перейти к этапу 7.

*Этап 7.* Стоп.

**7.22. Пример.** Рассмотрим полуцелую линейную систему

$$\begin{bmatrix} 3 & 0 & 2 \\ -2 & -1 & 1 \\ 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 1 & -1 & 5 & 4 \\ -1 & -3 & 3 & 0 \\ 2 & 4 & -2 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \\ 7 \end{bmatrix},$$

записанную в виде  $Ax + By = c$ . Ищем решение, в котором компоненты  $x$  — целые, а компоненты  $y$  — рациональные числа.

<sup>1)</sup> Следует задать векторы  $w$  и  $z$ . — Прим. перев.

Одной из рефлексивных  $g$ -обратных для  $B$  будет матрица

$$B_R^- = \begin{bmatrix} 3/4 & -1/4 & 0 \\ -1/4 & -1/4 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix};$$

для нее

$$BB_R^- = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1/2 & -3/2 & 0 \end{bmatrix}, \quad B_R^- B = \begin{bmatrix} 1 & 0 & 3 & 3 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Следовательно,

$$I - BB_R^- = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1/2 & 3/2 & 1 \end{bmatrix}, \quad I - B_R^- B = \begin{bmatrix} 0 & 0 & -3 & -3 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Далее находим

$$B_R^- A = \frac{1}{4} \begin{bmatrix} 11 & 1 & 5 \\ -1 & 1 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ 0 \\ 14 \end{bmatrix},$$

$$D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -7/2 & 1/2 & 1/2 \end{bmatrix}.$$

Полагая

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 7 & 0 \\ 0 & 0 & 7 \end{bmatrix},$$

получим

$$S = PDQ = \begin{bmatrix} -7/2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix};$$

отсюда

$$S^+ = \begin{bmatrix} -2/7 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Поэтому

$$D_I^- = QS^+P = \begin{bmatrix} 0 & 0 & -2/7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad D_I^- D = \begin{bmatrix} 1 & -1/7 & -1/7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$DD_I^- = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$d = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1/2 & 3/2 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ 7 \\ 7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 14 \end{bmatrix}.$$

Замечаем, что вектор

$$D_I^- d = \begin{bmatrix} -4 \\ 0 \\ 0 \end{bmatrix}$$

имеет целые компоненты и

$$DD_I^- d = \begin{bmatrix} 0 \\ 0 \\ 14 \end{bmatrix} = d.$$

Если взять

$$z = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad w = \begin{bmatrix} 7 \\ 7 \\ 7 \\ 7 \end{bmatrix},$$

то для указанных выше матриц и векторов получим

$$x = \begin{bmatrix} -4 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 1/7 & 1/7 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ 7 \\ 7 \end{bmatrix} = \begin{bmatrix} -2 \\ 7 \\ 7 \end{bmatrix},$$

$$y = \begin{bmatrix} 7/2 \\ -7/2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 11 \\ -1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 16 \\ -4 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -6 \\ 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -15/2 \\ 5/2 \\ 1 \\ 1 \end{bmatrix}.$$



Выбирая различные  $z$  и  $w$ , можем построить другие векторы  $x$  (с целыми компонентами) и  $y$ , удовлетворяющие заданной полуцелой линейной системе.

*Упражнение IV.7*

1. Решить полуцелую линейную систему

$$\begin{bmatrix} 1 & 0 & 1 \\ -1 & 2 & -2 \\ 0 & 3 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 1 & -1 \\ 0 & -1 & 0 & 2 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \\ 6 \end{bmatrix}.$$

## Глава V

# Итерационные методы обращения матриц и решения систем линейных уравнений

### § 1. Введение

В § 5 гл. II было показано, как по заданному коду Гензеля  $H(p, r, \alpha)$  вычислить код  $H(p, r, 1/\alpha)$ , используя быстрый итерационный алгоритм, основанный на методе Ньютона. Хорошо известно, что метод Ньютона для обращения вещественного числа можно обобщить на случаи

(1) вычисления обратной для невырожденной матрицы (см., например, [Stoer, Bulirsch, 1980, с. 310]<sup>1)</sup>), где соответствующий метод называется методом Шульца);

(2) вычисления  $g$ -обратной Мура — Пенроуза для произвольной матрицы (см., например, [Ben-Israel, Greville, 1974, с. 300]).

В этой главе мы опишем способы применения метода Ньютона — Шульца к решению задач (1) и (2) для матриц с целыми (или рациональными) элементами в условиях, когда используются арифметика вычетов и конечноразрядная  $p$ -адическая арифметика.

Некоторыми важными особенностями этого подхода являются следующие.

(1) *Детерминированный характер.* Начальное приближение к обратной (или к  $g$ -обратной) матрице выбирается вполне детерминированно как обратная (или  $g$ -обратная) по модулю  $p$  для заданной матрицы;  $p$  — простое число. На  $i$ -м итерационном шаге генерируются элементы обратной матрицы по модулю  $p^{2^i}$ . Итерационная процедура заканчивается, когда  $i$  «достаточно велико», т. е. если результаты выражаются дробями Фарея порядка  $N$ , где  $2N^2 + 1 \leq p^r$ , то при  $2^i \geq r$ .

(2) *Скорость вычислений.* Рациональные элементы обратной матрицы можно определить с квадратичной (или даже более высокой) скоростью сходимости.

---

<sup>1)</sup> Или Воеводин В. В. Численные методы алгебры. — М.: Наука, 1966, с. 102. — *Прим. перев.*

## § 2. Метод Ньютона — Шульца для обращения матрицы

Прежде чем описывать метод Ньютона — Шульца, введем некоторые определения и теоремы, связанные с обращением целочисленной матрицы по модулю  $m$ . Нас будет интересовать случай  $m = p^r$ , где  $p$  — простое число. Поскольку все нужные нам результаты содержатся в книге [Young, Gregory, с. 853—858], то мы не приводим здесь доказательств. Считаем, что  $A$ ,  $C$  и  $E$  —  $(n \times n)$ -матрицы с целыми элементами.

**2.1. Определение.** Если  $A = (a_{ij})$ , то

$$|A|_m = (|a_{ij}|_m).$$

**2.2. Определение.** Если  $|AC|_m = |CA|_m = I$  и  $|C|_m = C$ , то  $C$  называется обратной для  $A$  по модулю  $m$  и обозначается  $C = A^{-1}(m)$ .

**2.3. Теорема.** Если матрица  $A^{-1}(m)$  существует, то она единственна.

Хотя  $A^{-1}(m)$  в случае существования определена однозначно, сама она может быть обратной по модулю  $m$  к многим матрицам.

**2.4. Теорема.** Если  $|A|_m = |E|_m$  и  $A^{-1}(m)$  существует, то существует и  $E^{-1}(m)$ , причем  $A^{-1}(m) = E^{-1}(m)$ .

**2.5. Определение.** Говорят, что  $A$  невырождена по модулю  $m$ , если  $|\det A|_m \neq 0$  и  $(\det A, m) = 1$ . В противном случае говорят, что  $A$  вырождена по модулю  $m$ .

**2.6. Теорема.** Для существования  $A^{-1}(m)$  необходимо и достаточно, чтобы  $A$  была невырождена по модулю  $m$ .

**2.7. Теорема.**  $|\det A|_m = |\det A|_m$ .

*Алгоритм вычисления матрицы  $A^{-1}$*

Пусть  $A$  —  $(n \times n)$ -матрица с целыми элементами, причем  $A$  невырождена по модулю  $m$ . В случае  $m = p^r$ , где  $p$  — простое число, легко показать, что  $A$  невырождена по модулю  $m$  тогда и только тогда, когда она невырождена по модулю  $p$  (см. задачу 1 в упражнениях).

Первый шаг в алгоритме Ньютона — Шульца — это построение (произвольным методом) матрицы  $A^{-1}(p)$ ; она рассматривается как начальное приближение к  $C = A^{-1}(m)$ .

Если ввести обозначения по аналогии с формулами (5.32) и (5.33) гл. II, то можно положить

$$(2.8) \quad B_1 = A^{-1}(p)$$

и генерировать последующие приближения итерационным процессом

$$(2.9) \quad B_{2^k} = [B_{2^{k-1}}(2I - AB_{2^{k-1}})]_{p^{2^k}},$$

$k = 1, 2, \dots, i$ . Здесь  $r = 2^i$  должно быть достаточно велико, чтобы охватить рациональные элементы обратной матрицы. Другими словами, если  $\mathbb{F}_N$  — множество дробей Фарея порядка  $N$ , которому принадлежат элементы обратной матрицы, то  $i$  должно удовлетворять неравенству

$$(2.10) \quad p^{2^i} \geq 2N^2 + 1.$$

Будем обозначать через  $H(p, r, A^{-1})$  матрицу, элементы которой суть коды Гензеля, представляющие элементы  $A^{-1}$  (т. е. дроби Фарея порядка  $N$ ). Тогда для  $r = 2^i$  матрица  $B_{2^i}$  (с целыми элементами) соответствует матрице кодов Гензеля  $H(p, 2^i, A^{-1})$ , и соответствие между одноименными элементами обеих матриц такое же, как между  $b_{2^k}$  и  $H(p, 2^k, 1/\alpha)$  в формуле (5.32) гл. II. Как только  $i$  удовлетворяет условию (2.10), итерации можно закончить и перевести элементы матрицы в соответствующие дроби Фарея порядка  $N$ , пользуясь обратным отображением из гл. I, II.

Прежде чем доказывать корректность процесса (2.9), проиллюстрируем его численным примером.

### 2.11. Пример. Пусть

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 3 & 2 & 4 \\ 0 & 1 & -2 \end{bmatrix}.$$

Возьмем  $p = 3$ . Тогда

$$|A|_3 = \begin{bmatrix} 1 & 2 & 2 \\ 0 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Вычисление  $A^{-1}(3)$  проведем методом Гаусса — Жордана в  $(\mathbb{I}_3, +, \cdot)$ ; получим

$$\begin{bmatrix} 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & 2 & 2 \end{bmatrix}.$$

Следовательно,

$$B_1 = A^{-1}(3) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 2 & 2 \end{bmatrix}.$$

По формуле (2.9) последовательно находим

$$B_2 = A^{-1}(9) = \begin{bmatrix} 1 & 0 & 1 \\ 6 & 7 & 2 \\ 3 & 8 & 5 \end{bmatrix},$$

$$B_4 = A^{-1}(81) = \begin{bmatrix} 1 & 0 & 1 \\ 60 & 61 & 20 \\ 30 & 71 & 50 \end{bmatrix},$$

$$B_8 = A^{-1}(6561) = \begin{bmatrix} 1 & 0 & 1 \\ 4920 & 4921 & 1640 \\ 2460 & 5741 & 4100 \end{bmatrix},$$

$$B_{16} = A^{-1}(43046721) = \begin{bmatrix} 1 & 0 & 1 \\ 32285040 & 32285041 & 10761680 \\ 16142520 & 37665881 & 26904200 \end{bmatrix}.$$

Используя обратное отображение из гл. I, замечаем, что  $B_8$  и  $B_{16}$  порождают одну и ту же матрицу. Поэтому здесь итерации можно прекратить.

Матрицам  $B_1, B_2, B_4, B_8$  соответствуют следующие матрицы из кодов Гензеля:

$$H(3, 1, B_1) = \begin{bmatrix} .1 & .0 & .1 \\ .0 & .1 & .2 \\ .0 & .2 & .2 \end{bmatrix},$$

$$H(3, 2, B_2) = \begin{bmatrix} .10 & .00 & .10 \\ .02 & .12 & .20 \\ .01 & .22 & .21 \end{bmatrix},$$

$$H(3, 4, B_4) = \begin{bmatrix} .1000 & .0000 & .1000 \\ .0202 & .1202 & .2020 \\ .0101 & .2212 & .2121 \end{bmatrix},$$

$$H(3, 8, B_8) = \begin{bmatrix} .10000000 & .00000000 & .10000000 \\ .02020202 & .12020202 & .20202020 \\ .01010101 & .22121212 & .21212121 \end{bmatrix}.$$

Чтобы проиллюстрировать применение обратного отображения, вычислим элемент (2,1) матрицы  $A^{-1}$ . Код Гензеля .02020202 соответствует целому числу

$$20202020_{\text{три}} = 4920_{\text{десять}}$$

в позиции (2, 1) матрицы  $B_8$ . Далее:

	6561	0
	4920	1
1	1641	-1
2	1638	3
1	3	-4
546	0	2187

Полученная дробь Фарея порядка 57 равна  $-3/4$ . После того как все элементы будут вычислены, окажется, что

$$A^{-1} = \begin{bmatrix} 1 & 0 & 1 \\ -3/4 & 1/4 & -1/4 \\ -3/8 & 1/8 & -5/8 \end{bmatrix}.$$

Корректность алгоритма Ньютона — Шульца устанавливает следующая теорема.

**2.12. Теорема.** Пусть  $(n \times n)$ -матрица  $A$  невырождена по модулю  $p$ . Последовательность матриц  $\{B_1, B_2, B_4, \dots, B_r\}$ , построенная согласно (2.9), такова, что

$$|AB_{2^k}|_{p^{2^k}} = I, \quad k = 0, 1, \dots, i \quad (2^i = r),$$

и, более того,

$$B_{2^k} = A^{-1}(p^{2^k}).$$

**Доказательство.** Докажем по индукции, что для  $k \geq 0$

$$|AB_{2^k}|_{p^{2^k}} = I.$$

Согласно (2.8), первым членом последовательности является  $B_1 = A^{-1}(p)$ . Поэтому по определению  $|AB_1|_p = I$ . Напомним, что  $B_1$  можно вычислить методом Гаусса — Жордана.

Пусть теперь (индуктивное предположение)

$$|AB_{2^{k-1}}|_{p^{2^{k-1}}} = I.$$

Используя (2.9), получим

$$\begin{aligned} |AB_{2k}|_{p^{2k}} &= |A|_{B_{2k-1}}(2I - AB_{2k-1})|_{p^{2k}}|_{p^{2k}} = \\ &= |AB_{2k-1}(2I - AB_{2k-1})|_{p^{2k}}. \end{aligned}$$

Так как из индуктивного предположения вытекает представление

$$AB_{2k-1} = I + p^{2k-1}E_{k-1}$$

для некоторой матрицы  $E_{k-1}$ , то можно написать

$$\begin{aligned} |AB_{2k}|_{p^{2k}} &= |(I + p^{2k-1}E_{k-1})[2I - (I + p^{2k-1}E_{k-1})]|_{p^{2k}} = \\ &= |(I + p^{2k-1}E_{k-1})(I - p^{2k-1}E_{k-1})|_{p^{2k}} = I. \end{aligned}$$

Поскольку теорема справедлива при  $k=0$  (по самому способу построения  $B_1$ ) и верен индуктивный переход (справедливость результата для  $k-1$  влечет его справедливость для  $k$ ), то она справедлива при всех  $k \geq 0$ .  $\square$

### Число итераций

Пусть выполнено  $i$  итераций, где  $i$  — наименьшее целое число, удовлетворяющее неравенству

$$(2.13) \quad p^{2^i} > 2 \prod_{j=1}^n \|c_j\|_2.$$

Здесь  $\|c_j\|_2$  — евклидова норма  $j$ -го столбца  $A$ ; если  $\|c_j\|_2 = 0$ , то  $j$ -й столбец в произведении опускается<sup>1)</sup>. Пусть  $N$  — наибольшее целое число, для которого

$$(2.14) \quad 2N^2 + 1 \leq p^{2^i}.$$

Ясно<sup>2)</sup>, что все рациональные элементы  $A^{-1}$  принадлежат  $\mathbb{F}_N$ . Это значит, что если к целочисленным элементам  $B_{2^i}$  или к кодам Гензеля из  $H(p, r, B_{2^i})$  применить обратное отображение, то псевдопереполнений не будет.

Так как неравенство (2.13) весьма консервативно, то число  $i$  в (2.13) обычно много больше необходимого минимума. Поэтому в  $i$  итерациях может быть немало лишней работы. Более практично было бы выбрать достаточно большое значение  $i$ , исходя из интуитивных соображений, а затем све-

<sup>1)</sup> Однако при  $c_j = 0$  бессмысленно говорить об обращении  $A$ . — Прим. перев.

<sup>2)</sup> См. [Rao et al., 1976], а также пример 4.19 гл. III.

речь результаты после  $i$ -й и  $(i+1)$ -й итераций. Если они порождают одни и те же рациональные элементы для  $A^{-1}$ , то  $i$  итераций достаточно и  $B_{2i}$ , или, что все равно,  $H(p, 2^i, B_{2i})$  однозначно определяет  $A^{-1}$ .

**2.15. Замечание.** Каждая итерация порождает однотипное  $p$ -адическое представление сразу для всех элементов матрицы  $H(p, 2^k, B_{2k})$ . (См. пример 2.11, а также § 5 гл. II.)

### *Сходимость более высокого порядка*

Скорость сходимости описанного нами алгоритма квадратична (отметим, что число  $p$ -адических разрядов в кодах Гензеля примера 2.11 на каждом шаге удваивается). Совершенно таким же образом, как в гл. II для повышения порядка сходимости (5.33) было заменено на (5.39), можно модифицировать и (2.9). Например, по аналогии с (5.39) имеем

$$(2.16) \quad B_{qk} = |B_{qk-1} [I + D_{k-1} (I + D_{k-1} (I + \dots))] |_{p^q k}.$$

В этом выражении уровень вложенности равен  $q-1$  и

$$(2.17) \quad D_{k-1} = |I - AB_{qk-1} |_{p^q k}.$$

Так, для кубической сходимости полагаем

$$(2.18) \quad \begin{aligned} B_{3k} &= |B_{3k-1} [I + D_{k-1} (I + D_{k-1})] |_{p^3 k} = \\ &= |B_{3k-1} [I + (I - AB_{3k-1}) (2I - AB_{3k-1})] |_{p^3 k}. \end{aligned}$$

### *Выбор простого $p$*

Мы предполагаем, что  $A$  невырождена по модулю  $p$ , откуда следовало, что  $A$  невырождена и по модулю  $p'$ . Если на первом этапе (когда вычисляется  $A^{-1}(p)$ ) обнаруживается, что  $A^{-1}(p)$  не существует, то приходится выбирать иное значение  $p$  и начинать заново.

Другая возможность состоит в использовании следующего метода одноранговой модификации. Пусть  $a$  и  $b$  — произвольные  $n$ -мерные векторы; сформируем с их помощью матрицу

$$(2.19) \quad V = ab^T.$$

Применим наш алгоритм к  $A + V$  вместо  $A^1$ ; для вычисления  $A^{-1}$  воспользуемся формулой

$$(2.20) \quad A^{-1} = (A + V)^{-1} \left[ I + \frac{1}{s} V (A + V)^{-1} \right],$$

<sup>1)</sup> Предполагая, что  $A + V$  невырождена. — *Прим. перев.*



где

$$(2.21) \quad s = 1 - b^T (A + V)^{-1} a.$$

Это вариант хорошо известной формулы Шермана — Моррисона (см., например, [Householder, 1964, с. 123] или [Dahlquist, Björck, 1974, с. 161])<sup>1)</sup>.

### Упражнения V.2

1. Пусть  $m = p^r$ , где  $p$  простое. Доказать, что  $A$  невырождена по модулю  $m$  тогда и только тогда, когда она невырождена по модулю  $p$ .

2. Проверить, что в примере 2.11 элементы матриц  $B_8$  и  $B_{16}$  отображаются на одно и то же множество рациональных чисел, т. е. на множество элементов  $A^{-1}$ .

3. Положим

$$A = \begin{bmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}.$$

Используя метод Ньютона — Шульца (вариант с квадратичной сходимостью), найти  $A^{-1}$ . *Указание:* взять  $p = 3$ .

4. Найти  $A^{-1}$  для матрицы задачи 3, используя вариант с кубической сходимостью. *Указание:* взять  $p = 2$ .

5. Обратить следующие матрицы обоими вариантами метода Ньютона — Шульца (с квадратичной и кубической сходимостью):

$$(a) \begin{bmatrix} 1 & 0 & -1 & 0 \\ 2 & 1 & -3 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 1 & 0 & -1 \end{bmatrix}; \quad (b) \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}; \quad (c) \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix}.$$

## § 3. Итерационное решение линейной системы

Займемся теперь итерационным решением системы линейных алгебраических уравнений

$$(3.1) \quad Ax = b,$$

где  $A$  — невырожденная  $(n \times n)$ -матрица,  $b$  —  $n$ -мерный вектор (и то и другое задано), а  $x$  — неизвестный  $n$ -мерный вектор, который мы называем решением.

Не будет потери общности в предположении, что  $A$  и  $b$  имеют целочисленные элементы (компоненты): если бы они были рациональными, то подходящим строчным масштабированием их можно было бы превратить в целые. (Мы не рассматриваем системы уравнений с иррациональными коэффициентами, поскольку такие системы не могут быть записаны

<sup>1)</sup> См. также Муртаф Б. Современное программирование. — Перев. с англ. — М.: Мир, 1984, с. 17. — *Прим. перев.*

в память вычислительной машины; иррациональные числа машинно непредставимы.) Пусть

$$(3.2) \quad A_1 = |A|_p,$$

$$(3.3) \quad b_1 = |b|_p,$$

и пусть  $A$  невырождена по модулю  $p$ , т. е.<sup>1)</sup>

$$(3.4) \quad |\det A|_p \neq 0,$$

$$(3.5) \quad (\det A, p) = 1.$$

Ясно (см. теорему 2.4), что  $A_1$  также невырождена по модулю  $p$ . Положим

$$(3.6) \quad x_k = |x|_{p^k}, \quad k = 1, 2, \dots$$

Первым этапом будет вычисление  $A^{-1}(p)$ , а затем вектора  $x_1$  путем решения системы

$$(3.7) \quad |A_1 x_1 - b_1|_p = |Ax_1 - b|_p = 0.$$

После этого проводим для  $k = 1, 2, \dots, i$  итерационный процесс

$$(3.8) \quad x_{k+1} = |(I - A_1^{-1}A)x_k + A_1^{-1}b|_{p^{k+1}}.$$

Здесь для упрощения обозначений мы пишем  $A_1^{-1}$  вместо  $A_1^{-1}(p)$ . Если  $\mathbb{F}_N$  — множество дробей Фарея порядка  $N$ , которому принадлежат компоненты  $x$ , то  $i$  должно быть настолько велико, чтобы удовлетворялось неравенство

$$(3.9) \quad p^i \geq 2N^2 + 1.$$

На практике мы отображаем (целочисленные) компоненты  $x_i$  и  $x_{i+1}$  в их рациональные эквиваленты и, если будет получено одно и то же множество рациональных чисел, прекращаем итерации.

Прежде чем доказывать корректность процесса (3.8), рассмотрим численный пример.

**3.10. Пример.** Пусть дана система линейных алгебраических уравнений

$$\begin{bmatrix} 1 & 5 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \end{bmatrix}.$$

Беря  $p = 3$ , имеем

$$A_1 = |A|_3 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}.$$

---

<sup>1)</sup> Если  $p$  — простое, то предположение (3.5) повторяет (3.4). — *Прим. перев.*

Проводя исключение в  $(\Pi_3, +, \cdot)$ , получаем

$$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

Итак,

$$A_1^{-1}(3) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Поскольку

$$b_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

то

$$x_1 = A_1^{-1}b_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

и начальный этап закончен.

Вектор  $x_2 = |x|_9$  найдем следующим образом. Вначале вычисляем

$$I - A_1^{-1}A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 5 \\ 6 & 4 \end{bmatrix} = \begin{bmatrix} -6 & -9 \\ -6 & -3 \end{bmatrix},$$

$$A_1^{-1}b = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \\ 3 \end{bmatrix},$$

а затем

$$\begin{aligned} x_2 &= |I - A_1^{-1}A|_9 x_1 + |A_1^{-1}b|_9|_9 = \\ &= \left| \begin{bmatrix} 3 & 0 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right|_9 = \begin{bmatrix} 4 \\ 6 \end{bmatrix}. \end{aligned}$$

Аналогично

$$\begin{aligned} x_3 &= |I - A_1^{-1}A|_{27} x_2 + |A_1^{-1}b|_{27}|_{27} = \\ &= \left| \begin{bmatrix} 21 & 18 \\ 21 & 24 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \end{bmatrix} + \begin{bmatrix} 10 \\ 3 \end{bmatrix} \right|_{27} = \begin{bmatrix} 13 \\ 15 \end{bmatrix}, \\ x_4 &= |I - A_1^{-1}A|_{81} x_3 + |A_1^{-1}b|_{81}|_{81} = \\ &= \left| \begin{bmatrix} 75 & 72 \\ 75 & 78 \end{bmatrix} \begin{bmatrix} 13 \\ 15 \end{bmatrix} + \begin{bmatrix} 10 \\ 3 \end{bmatrix} \right|_{81} = \begin{bmatrix} 40 \\ 42 \end{bmatrix}. \end{aligned}$$

В этот момент можно прекратить итерации, так как

$$\begin{array}{c|cc} & 27 & 0 \\ & 13 & 1 \\ \hline 2 & 1 & -2 \\ \hline 13 & 0 & 27 \end{array} \qquad \begin{array}{c|cc} & 81 & 0 \\ & 40 & 1 \\ \hline 2 & 1 & -2 \\ \hline 40 & 0 & 81 \end{array}$$

$$\begin{array}{c|cc} & 27 & 0 \\ & 15 & 1 \\ \hline 1 & 12 & -1 \\ 1 & 3 & 2 \\ \hline 4 & 0 & -9 \end{array} \qquad \begin{array}{c|cc} & 81 & 0 \\ & 42 & 1 \\ \hline 1 & 39 & -1 \\ 1 & 3 & 2 \\ \hline 13 & 0 & -27 \end{array}$$

Тем самым и  $x_3$ , и  $x_4$  отображаются на вектор

$$x = \begin{bmatrix} -1/2 \\ 3/2 \end{bmatrix}.$$

Вернемся теперь к проверке корректности процесса (3.8).

3.11. **Теорема.** Пусть

$$x_{k+1} = |(I - A_1^{-1}A)x_k + A_1^{-1}b|_{p^{k+1}}.$$

Тогда из  $|Ax_k - b|_{p^k} = 0$  следует

$$|Ax_{k+1} - b|_{p^{k+1}} = 0.$$

**Доказательство.**

$$\begin{aligned}
 |Ax_{k+1} - b|_{p^{k+1}} &= |A|(I - A_1^{-1}A)x_k + A_1^{-1}b|_{p^{k+1}} - b|_{p^{k+1}} = \\
 &= |A[(I - A_1^{-1}A)x_k + A_1^{-1}b] - b|_{p^{k+1}} = \\
 &= |Ax_k - AA_1^{-1}Ax_k + AA_1^{-1}b - b|_{p^{k+1}} = \\
 &= |(Ax_k - b) - AA_1^{-1}(Ax_k - b)|_{p^{k+1}} = \\
 &= |(I - AA_1^{-1})(Ax_k - b)|_{p^{k+1}}.
 \end{aligned}$$

По построению  $A_1 = |A|_p$ . Следовательно,

$$A_1 = A + pD$$

для некоторой матрицы  $D$ . Поэтому<sup>1)</sup>

$$AA_1^{-1} = (A_1 - pD)A_1^{-1} = I - pDA_1^{-1},$$

<sup>1)</sup> Здесь авторов подводят собственные обозначения. Матрица  $A_1^{-1}(p)$  обратна к  $A_1$ , вообще говоря, только по модулю  $p$ , т. е. не обязана сов-

или

$$I - AA_1^{-1} = pDA_1^{-1}.$$

Так как по предположению  $|Ax_k - b|_{p^k} = 0$ , то

$$Ax_k - b = p^k c$$

для некоторого вектора  $c$ . Отсюда

$$|Ax_{k+1} - b|_{p^{k+1}} = |(pDA_1^{-1})(p^k c)|_{p^{k+1}} = 0. \quad \square$$

Согласно (3.7),  $|A_1 x_1 - b_1|_p = 0$ . Вместе с теоремой 3.11 это позволяет доказать по индукции, что процесс (3.8) последовательно генерирует векторы  $x_2, x_3, \dots, x_k, \dots$ , для которых при всех  $k \geq 2$

$$(3.12) \quad |Ax_k - b|_{p^k} = 0.$$

### 3.13. Замечания

(1) Изложенный метод имеет линейную скорость сходимости. Другими словами, каждая итерация вычисляет очередной разряд  $p$ -адического представления решения.

(2) Итерационная формула (3.8) требует<sup>1)</sup> выполнять только умножения типа «матрица — вектор». Поэтому она более экономична, чем вычисление решения путем предварительного обращения матрицы процессом (2.9), где необходимы умножения матриц на матрицы.

(3) В статье [Dixon, 1982] предложен алгоритм, сходный с описанным в данном параграфе.

### Упражнения V.3

1. Решить систему линейных алгебраических уравнений

$$\begin{bmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix},$$

пользуясь методом данного параграфа.

падать с обычной обратной  $A_1^{-1}$ . Поэтому  $A_1 A_1^{-1}(p)$  может отличаться от единичной матрицы матричным кратным числа  $p$ . Однако все равно  $AA_1^{-1}$  представима в виде  $I - pF$ , где  $F \in \mathbb{P}^{n \times n}$ , и последующие выкладки сохраняют силу. — Прим. перев.

<sup>1)</sup> При вычислении векторов  $x_2, x_3, \dots$ . Однако вначале приходится находить  $A_1^{-1}A$ . — Прим. перев.

2. То же, что в задаче 1, для следующих систем:

$$(a) \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 12 \\ 14 \\ 6 \end{bmatrix};$$

$$(b) \begin{bmatrix} 5 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 9 \\ 4 \\ -6 \end{bmatrix};$$

$$(c) \begin{bmatrix} 4 & -1 & 0 & 0 & 0 \\ -1 & 4 & -1 & 0 & 0 \\ 0 & -1 & 4 & -1 & 0 \\ 0 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & -1 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ u \\ v \end{bmatrix} = \begin{bmatrix} 100 \\ 200 \\ 200 \\ 200 \\ 100 \end{bmatrix}.$$

#### § 4. Итерационное вычисление $g$ -обратных

Как отмечено во введении к этой главе, метод Ньютона — Шульца можно обобщить на задачу вычисления  $g$ -обратной Мура — Пенроуза для матрицы над вещественным или комплексным полем. Спрашивается, можно ли описанный в § 2 вариант метода Ньютона — Шульца, опирающийся на конечно-разрядную  $p$ -адическую арифметику, также перенести на вычисление  $g$ -обратной Мура — Пенроуза для целочисленной (или рациональной)  $(m \times n)$ -матрицы  $A = (a_{ij})$ ? (Необходимые определения и обозначения введены в гл. III.)

Мы покажем сейчас, что такое обобщение возможно.

Пусть  $A$  —  $(m \times n)$ -матрица; положим

$$(4.1) \quad M = (AA^T)^2.$$

В обозначениях гл. III  $M_R^- = (r_{ij})$  есть рефлексивная  $g$ -обратная к матрице  $M$ :

$$(4.2) \quad M_R^- = [(AA^T)^2]_R^-.$$

Пусть все  $r_{ij}$  принадлежат множеству  $\mathbb{F}_N$  дробей Фарея порядка  $N$ , и пусть для целого числа  $k$  выполнено неравенство

$$(4.3) \quad p^{2k} \geq 2N^2 + 1,$$

где  $p$  — выбранное для вычислений простое число. Если преобразовать элементы  $r_{ij}$  (дроби Фарея порядка  $N$ ) в  $\mathbb{I}_{p^{2k}}$ , пользуясь прямым отображением из гл. I, то  $M_R^-$  можно

представить в виде конечного  $p$ -адического матричного степенного ряда. Именно

$$(4.4) \quad \begin{aligned} |M_R^-|_{p^{2k}} = B_{2k} = B_{2^0} + (B_{2^1} - B_{2^0}) + \\ + (B_{2^2} - B_{2^1}) + \dots + (B_{2^k} - B_{2^{k-1}}). \end{aligned}$$

Здесь

$$(4.5) \quad \begin{aligned} B_{2^0} &= |M_R^-|_p, \\ B_{2^1} &= |M_R^-|_{p^2}, \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ B_{2^k} &= |M_R^-|_{p^{2^k}}. \end{aligned}$$

Итак, если справедливо (4.3), то  $B_{2^k}$  соответствует конечно-разрядному  $p$ -адическому представлению  $M_R^-$  в том смысле, что  $B_{2^k}$  (матрица с целыми элементами) соответствует матрице кодов Гензеля

$$(4.6) \quad H(p, 2^k, B_{2^k}) = H(p, 2^k, M_R^-).$$

Поэлементное соответствие между обеими матрицами описывается формулами (5.32) гл. II и иллюстрируется примером 2.11. Чтобы воспользоваться равенством (4.4), нужно уметь

(а) вычислять  $|M_R^-|_p = B_1$  и для  $1 < i \leq k$ ,

(б) вычислять  $|M_R^-|_{p^{2^i}} = B_{2^i}$ , если матрица  $B_{2^{i-1}}$  известна.

Матрицу  $B_1$  мы легко найдем методом гл. III; обобщение процедуры Ньютона — Шульца из § 2 позволит нам по  $B_{2^{i-1}}$  строить  $B_{2^i}$ .

#### 4.7. Алгоритм

*Этап 1. Пользуясь методом гл. III, найти  $|M_R^-|_p = B_1$ . Для этой матрицы справедливы соотношения  $B_1 = |B_1 M B_1|_p$ ,  $|M|_p = |M B_1 M|_p$ .*

*Этап 2. Для  $i = 1, 2, \dots, k$  вычислить*

$$B_{2^i} = |B_{2^{i-1}}(2I - M B_{2^{i-1}})|_{p^{2^i}}.$$

*Этап 3. В этот момент  $|M_R^-|_{p^{2^k}} = B_{2^k}$ ; вычислить матрицу  $A^+$ ,  $g$ -обратную Мура — Пенроуза для  $A$ , по формуле (4.3) гл. III<sup>1)</sup>.*

<sup>1)</sup> Имеется в виду — после восстановления самой матрицы  $M_R^-$  из  $B_{2^k}$ . — Прим. перев.

4.8. **Замечание.** Дадим доказательство корректности алгоритма. На этапе 1 будет построена матрица  $B_1$ , такая, что

$$\begin{aligned} B_1 &= |B_1 M B_1|_p = |M_R^-|_p, \\ |M|_p &= |M B_1 M|_p. \end{aligned}$$

На этапе 2 последовательные итерации удовлетворяют соотношению

$$(1) \quad B_{2i} = |B_{2i-1}(2I - M B_{2i-1})|_{p^{2i}},$$

$i = 1, 2, \dots, k$ . Нужно доказать, что если

$$\begin{aligned} |M B_{2i-1} M|_{p^{2i-1}} &= |M|_{p^{2i-1}}, \\ |B_{2i-1} M B_{2i-1}|_{p^{2i-1}} &= |B_{2i-1}|_{p^{2i-1}}. \end{aligned}$$

то и

$$\begin{aligned} |M B_{2i} M|_{p^{2i}} &= |M|_{p^{2i}}, \\ |B_{2i} M B_{2i}|_{p^{2i}} &= |B_{2i}|_{p^{2i}}. \end{aligned}$$

Из первого индуктивного предположения следует, что для некоторой матрицы  $K$  над  $\mathbb{Q}$  справедливо равенство <sup>1)</sup>

$$(2) \quad M B_{2i-1} M = (I + p^{2i-1} K) M.$$

Используя (1), имеем

$$\begin{aligned} |M B_{2i} M|_{p^{2i}} &= |M [B_{2i-1}(2I - M B_{2i-1})] M|_{p^{2i}} = \\ &= |2M B_{2i-1} M - M B_{2i-1} M B_{2i-1} M|_{p^{2i}}. \end{aligned}$$

Подставляя сюда (2), находим <sup>2)</sup>

$$\begin{aligned} &|2(I + p^{2i-1} K) M - (I + p^{2i-1} K)(I + p^{2i-1} K) M|_{p^{2i}} = \\ &= |(I + p^{2i-1} K)(2M - M - p^{2i-1} K M)|_{p^{2i}} = \\ &= |(I + p^{2i-1} K)(I - p^{2i-1} K) M|_{p^{2i}} = |M|_{p^{2i}}. \end{aligned}$$

<sup>1)</sup> Второй автор (в письме редактору русского перевода) предлагает в качестве  $K$  матрицу  $(M B_{2i-1} - I)/p^{2i-1}$  — Прим. перев.

<sup>2)</sup> По поводу последнего перехода в этой цепочке заметим: авторы полагают, что  $|(I - p^{2i} K^2) M|_{p^{2i}} = |M|_{p^{2i}}$ , что верно, вообще говоря, лишь если элементы матрицы  $K$  суть  $p$ -адические целые числа. Неявно предполагается также, что  $M$  — целочисленная матрица. — Прим. перев.



Аналогичным образом для некоторых матриц  $K_1, K_2$  над  $\mathbb{Q}$  справедливы представления <sup>1)</sup>

$$(3) \quad B_{2i-1}MB_{2i-1} = B_{2i-1}(I + p^{2i-1}K_1),$$

$$(4) \quad B_{2i-1}MB_{2i-1} = (I + p^{2i-1}K_2)B_{2i-1}.$$

Используя (1), получаем

$$\begin{aligned} |B_{2i}MB_{2i}|_{p^{2i}} &= |B_{2i-1}(2I - MB_{2i-1})MB_{2i-1}(2I - MB_{2i-1})|_{p^{2i}} = \\ &= |(2B_{2i-1} - B_{2i-1}MB_{2i-1})M(2B_{2i-1} - B_{2i-1}MB_{2i-1})|_{p^{2i}}. \end{aligned}$$

Опираясь на (3) и (4), выводим

$$\begin{aligned} &(2B_{2i-1} - B_{2i-1} - p^{2i-1}K_2B_{2i-1})M(2B_{2i-1} - B_{2i-1} - \\ &- p^{2i-1}B_{2i-1}K_1)|_{p^{2i}} = |(I - p^{2i-1}K_2)B_{2i-1}MB_{2i-1}(I - p^{2i-1}K_1)|_{p^{2i}} = \\ &= |(I - p^{2i-1}K_2)B_{2i-1}(I + p^{2i-1}K_1)(I - p^{2i-1}K_1)|_{p^{2i}} = \\ &= |(I - p^{2i-1}K_2)B_{2i-1}|_{p^{2i}} = |[2I - (I + p^{2i-1}K_2)]B_{2i-1}|_{p^{2i}} = \\ &= |(2B_{2i-1} - B_{2i-1}MB_{2i-1})|_{p^{2i}} = |B_{2i-1}(2I - MB_{2i-1})|_{p^{2i}} = \\ &= |B_{2i}|_{p^{2i}}. \end{aligned}$$

В последних переходах использованы (4) и (1).

**4.9. Пример.** Возьмем вырожденную матрицу  $A$  из примера 4.19 гл. III:

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Берем  $p=5$  и по методике, примененной в указанном примере, строим матрицы

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} 4 & 0 & 0 \\ 3 & 3 & 0 \\ 4 & 0 & 1 \end{bmatrix}.$$

<sup>1)</sup> В качестве  $K_1$  можно взять прежнюю матрицу  $K$  (см. примечание 1 на предыдущей странице). — *Прим. перев.*

Следовательно,

$$B_1 = |M_R^-|_5 = |F^T R E|_5 = \begin{bmatrix} 2 & 3 & 0 \\ 3 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

По итерационной формуле этапа 2 имеем

$$B_2 = |M_R^-|_{25} = \begin{bmatrix} 17 & 8 & 0 \\ 8 & 13 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$B_4 = |M_R^-|_{625} = \begin{bmatrix} 417 & 208 & 0 \\ 208 & 313 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Выполняя для элементов  $B_2$  и  $B_5$  обратное отображение из гл. I, получим в обоих случаях одинаковые результаты. Поэтому итерации можно закончить. Например, для элемента (1,1) мы нашли бы

	25	0		625	0
	17	1		417	1
1	8	-1	1	208	-1
2	1	3	2	1	3
8	0	-25	208	0	-625

Когда отображение проделано для всех элементов, будет получена матрица

$$M_R^- = \begin{bmatrix} 1/3 & -1/3 & 0 \\ -1/3 & 1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

а затем матрица

$$A^+ = \frac{1}{6} \begin{bmatrix} 1 & 2 & 1 \\ -1 & 4 & -1 \\ 2 & -2 & 2 \end{bmatrix},$$

что согласуется с результатами примера 4.19 гл. III.

**4.10. Замечание.** Совместную систему линейных алгебраических уравнений  $Ax = b$  с вырожденной матрицей  $A$  можно решать, заменяя в итерационной формуле (3.8)  $A_1^{-1}$  на  $B_1 = |M_R^-|_p$ .

## Упражнения V.4

1. Вычислить рефлексивную  $g$ -обратную для каждой из матриц

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

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

Использовать при этом метод данного параграфа.

2. Найти  $A^+$  для матрицы

$$A = \begin{bmatrix} 5 & 0 & 6 \\ 3 & 3 & 0 \\ 5 & 0 & 6 \end{bmatrix}.$$

# Глава VI

## Точное вычисление характеристического многочлена матрицы

### § 1. Введение

В общем случае не рекомендуется вычислять коэффициенты характеристического многочлена матрицы в качестве первого шага процедуры, использующей для отыскания собственных значений методы вычисления корней многочленов. Дело в том, что вследствие ошибок округления, неизбежных в обычной арифметике с плавающей точкой, будут получены лишь приближенные значения коэффициентов. Если многочлен плохо обусловлен, то корни «приближенного характеристического уравнения» могут быть плохими приближениями к корням истинного уравнения. Обсуждение понятия *обусловленности* алгебраического уравнения по отношению к вычислению его корней можно найти в гл. 2 книги [Wilkinson, 1963].

Тем не менее на с. 397 учебника [Pennigton, 1970] все же рекомендуется вычисление характеристического многочлена, а затем его корней как средство найти все собственные значения матрицы. Вычисление коэффициентов характеристического многочлена предлагается проводить методом Леверье — Фаддеева (см., например, [Фаддеев, Фаддеева, 1963]); этот метод упоминался в § 4 гл. III в связи с алгоритмом Диселла — Леверье. К сожалению, автор не предупреждает читателей о потенциальных опасностях, сопряженных с таким подходом. Мы настоятельно советуем: если уж необходимо вычислять коэффициенты характеристического многочлена (с целью получить собственные значения или по какой-то другой причине), то это нужно делать с помощью безошибочных вычислений.

Не будет потери общности, если мы ограничимся целочисленными матрицами, поскольку случай матрицы с рациональными элементами подходящим масштабированием можно свести к целочисленному. Итак, в данной главе будет рассмотрена задача точного вычисления коэффициентов характеристического многочлена целочисленной матрицы посредством арифметики вычетов. Соответствующий алгоритм был предложен в работе [Rao, 1978].

## § 2. Алгоритм вычисления характеристического многочлена нижней матрицы Хессенберга

Пусть дана нижняя матрица Хессенберга

$$(2.1) \quad A = \begin{bmatrix} a_{11} & a_{12} & & & \\ a_{21} & a_{22} & a_{23} & & \\ a_{31} & a_{32} & a_{33} & a_{34} & \\ \vdots & \vdots & \vdots & \vdots & a_{n-1,n} \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} & \dots & a_{nn} \end{bmatrix},$$

причем все элементы первой наддиагонали ненулевые<sup>1)</sup>, а все элементы над ними равны нулю (в формуле они не выписаны).

Пусть  $x$  — собственный вектор матрицы  $A$ , соответствующий собственному значению  $\lambda$ . Тогда  $Ax = \lambda x$ , или

$$(2.2) \quad \begin{bmatrix} (a_{11} - \lambda) & a_{12} & & & \\ a_{21} & (a_{22} - \lambda) & a_{23} & & \\ a_{31} & a_{32} & (a_{33} - \lambda) & a_{34} & \\ \vdots & \vdots & \vdots & \vdots & \dots & a_{n-1,n} \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} & \dots & (a_{nn} - \lambda) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Без потери общности можем считать, что  $x_1 = 1$ <sup>2)</sup>. Тогда

$$(2.3) \quad \begin{aligned} (a_{11} - \lambda) + a_{12}x_2 &= 0, \\ a_{21} + (a_{22} - \lambda)x_2 + a_{23}x_3 &= 0, \\ a_{31} + a_{32}x_2 + (a_{33} - \lambda)x_3 + a_{34}x_4 &= 0, \\ \dots &\dots \\ a_{n1} + a_{n2}x_2 + a_{n3}x_3 + \dots + (a_{nn} - \lambda)x_n &= 0. \end{aligned}$$

Решая первое уравнение относительно  $x_2$ , получим

$$(2.4) \quad x_2 = \frac{1}{a_{12}} \lambda - \frac{a_{11}}{a_{12}},$$

т. е. многочлен от  $\lambda$  *первой степени*. Подставляя это выражение для  $x_2$  во второе уравнение и решая его относительно

<sup>1)</sup> Если  $a_{i, i+1} = 0$  для некоторого  $i$ , то матрица расщепляется, и мы можем обрабатывать диагональные подматрицы независимо друг от друга.

<sup>2)</sup> Для матрицы Хессенберга с ненулевой наддиагональю первая компонента *любого* собственного вектора не может быть равна нулю. — *Прим. перев.*

$x_3$ , получаем

$$(2.5) \quad x_3 = \frac{1}{a_{12}a_{23}} \lambda^2 - \frac{a_{22} + a_{11}}{a_{12}a_{23}} \lambda + \frac{a_{11}a_{22} - a_{21}a_{12}}{a_{12}a_{23}},$$

т. е. многочлен от  $\lambda$  степени два. Продолжаем этот процесс, пока из  $(n-1)$ -го уравнения не будет найдено  $x_n$ . Подставляя выражения для  $x_2, x_3, \dots, x_n$  в  $n$ -е уравнение, приходим к многочлену от  $\lambda$  степени  $n$ ; он обращается в нуль тогда и только тогда, когда  $\lambda$  — собственное значение  $A$ . Следовательно, этот многочлен лишь скалярным множителем может отличаться от характеристического многочлена  $A$ .

Пусть  $P_{i+1}(\lambda)$  обозначает многочлен от  $\lambda$  степени  $i$ , а  $P_{i+1}$  —  $(n+1)$ -мерный строчный вектор, компонентами которого являются коэффициенты  $P_{i+1}(\lambda)$ . Если положить

$$(2.6) \quad P_1 = [0, 0, \dots, 0, 0, 1],$$

то можем написать

$$(2.7) \quad P_2 = \left[ 0, 0, \dots, 0, \frac{1}{a_{12}}, -\frac{a_{11}}{a_{12}} \right],$$

$$(2.8) \quad P_3 = \left[ 0, \dots, \frac{1}{a_{12}a_{23}}, -\frac{a_{11} + a_{22}}{a_{12}a_{23}}, \frac{a_{11}a_{22} - a_{21}a_{12}}{a_{12}a_{23}} \right]$$

и т. д. Заметим, что

$$(2.9) \quad P_2 = -\frac{a_{11}}{a_{12}} P_1 + \frac{1}{a_{12}} \tilde{P}_1,$$

где  $\tilde{P}_1 = [0, 0, \dots, 0, 1, 0]$ , т. е.  $\tilde{P}_1$  получен из  $P_1$  путем циклического левого сдвига компонент на одну позицию. Аналогично

$$(2.10) \quad P_3 = -\frac{a_{21}}{a_{23}} P_1 - \frac{a_{22}}{a_{23}} P_2 + \frac{1}{a_{23}} \tilde{P}_2,$$

где  $\tilde{P}_2$  есть результат циклического сдвига компонент  $P_2$  на одну позицию влево. В общем случае

$$(2.11) \quad P_{i+1} = \sum_{j=1}^i \frac{-a_{ij}}{a_{i, i+1}} P_j + \frac{1}{a_{i, i+1}} \tilde{P}_i, \quad i = 1, 2, \dots, n-1,$$

а при  $i = n$  получим

$$(2.12) \quad P_{n+1} = \sum_{j=1}^n a_{nj} P_j - \tilde{P}_n.$$

По вектору  $P_{n+1}$  легко восстанавливается характеристический многочлен.

**2.13. Пример** [Rao, 1978]. Пусть имеем целочисленную матрицу

$$B = \begin{bmatrix} 4 & 3 & 2 \\ 2 & 3 & 4 \\ 4 & 3 & 5 \end{bmatrix}.$$

Приведем ее подобным преобразованием к нижней форме Хессенберга:

$$\begin{aligned} A = SBS^{-1} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2/3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 3 & 2 \\ 2 & 3 & 4 \\ 4 & 3 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2/3 \\ 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 4 & 3 & 0 \\ 14/3 & 5 & 4 \\ 4 & 3 & 3 \end{bmatrix}. \end{aligned}$$

В данном случае:

$$P_1 = [0, 0, 0, 1],$$

$$P_2 = -\frac{4}{3}[0, 0, 0, 1] + \frac{1}{3}[0, 0, 1, 0] = [0, 0, \frac{1}{3}, -\frac{4}{3}],$$

$$\begin{aligned} P_3 &= -\frac{7}{6}[0, 0, 0, 1] - \frac{5}{4}[0, 0, \frac{1}{3}, -\frac{4}{3}] + \frac{1}{4}[0, \frac{1}{3}, -\frac{4}{3}, 0] = \\ &= [0, \frac{1}{12}, -\frac{3}{4}, \frac{1}{2}], \end{aligned}$$

$$\begin{aligned} P_4 &= 4[0, 0, 0, 1] + 3[0, 0, \frac{1}{3}, -\frac{4}{3}] + 3[0, \frac{1}{12}, -\frac{3}{4}, \frac{1}{2}] - \\ &- [\frac{1}{12}, -\frac{3}{4}, \frac{1}{2}, 0] = [-\frac{1}{12}, 1, -\frac{7}{4}, \frac{3}{2}]. \end{aligned}$$

Вектору  $P_4$  соответствует многочлен

$$P_4(\lambda) = -\frac{1}{12}\lambda^3 + \lambda^2 - \frac{7}{4}\lambda + \frac{3}{2}.$$

Так как  $B$  — целочисленная матрица, ее характеристический многочлен должен иметь целые коэффициенты. В частности, коэффициент при  $\lambda^3$  должен быть равен единице. Подобные матрицы  $B$  и  $A$  имеют один и тот же характеристический многочлен. Поэтому характеристический многочлен  $B$  получается из  $P_4(\lambda)$  умножением на  $(-12)$ . Итак, многочлен

$$(-12)P_4(\lambda) = \lambda^3 - 12\lambda^2 + 21\lambda - 18$$

является характеристическим многочленом  $B$ .

**2.14. Замечание.** Если мы хотим провести эти вычисления в конечном поле  $(\mathbb{P}_p, +, \cdot)$ , то нужна процедура выбора достаточно большого простого модуля  $p$ . Очень консервативное

неравенство для  $p$  дано в [Рао, 1978]: рекомендуется брать  $p$  настолько большим, чтобы выполнялось неравенство

$$p \geq 2 \max [m^n, n(n+1)m^{n-1}],$$

$$m = \max [\|AA^T\|, \operatorname{tr}(AA^T)].$$

Матричную норму можно брать любую.

**2.15. Пример.** Возьмем ту же матрицу  $B$ , что в примере 2.13. Мы увидим вскоре, что можно использовать  $p = 101$ , хотя неравенство из замечания 2.14 и не будет удовлетворено. Прежде всего вычисляем матрицу

$$|B|_{101} = \begin{bmatrix} 4 & 3 & 2 \\ 2 & 3 & 4 \\ 4 & 3 & 5 \end{bmatrix}.$$

Затем прибавляем к 3-му столбцу 2-й, умноженный на 33 (в результате элемент  $b_{13} = 2$  обратится в нуль), и завершаем подобное преобразование, прибавляя ко 2-й строке 3-ю, умноженную на 68 (число 68 является аддитивным обратным для числа 33). Матрица

$$|A|_{101} = |SBS^{-1}|_{101} = \left| \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 68 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 3 & 2 \\ 2 & 3 & 4 \\ 4 & 3 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 33 \\ 0 & 0 & 1 \end{bmatrix} \right|_{101} =$$

$$= \begin{bmatrix} 4 & 3 & 0 \\ 72 & 5 & 4 \\ 4 & 3 & 3 \end{bmatrix}$$

имеет нижнюю форму Хессенберга.

Начиная с вектора  $|P_1|_{101} = [0, 0, 0, 1]$ , последовательно находим

$$|P_2|_{101} = [0, 0, 34, 66],$$

$$|P_3|_{101} = [0, 59, 75, 51],$$

$$|P_4|_{101} = [42, 1, 74, 52].$$

Так как  $42^{-1}(101) = 89$ , то

$$|89 \cdot P_4|_{101} = [1, 89, 21, 83].$$

Правильные алгебраические знаки можно определить с помощью симметричных вычетов:

$$/89 \cdot P_4 /_{101} = [1, -12, 21, -18].$$



Итак, характеристическим многочленом  $B$  будет

$$\lambda^3 - 12\lambda^2 + 21\lambda - 18.$$

**2.16. Замечание.** В § 1 мы отметили, что матрицу с рациональными элементами еще до начала вычислений можно преобразовать в целочисленную посредством подходящего масштабирования. Нужно сказать, что это вовсе не является необходимым. Используя методы, описанные в § 5 гл. I, мы можем начинать прямо с матрицы с рациональными элементами.

### Упражнения VI.2

1. Найти характеристический многочлен матрицы

$$A = \begin{bmatrix} 3 & 2 & 5 \\ 6 & -5 & 3 \\ -24 & 38 & 2 \end{bmatrix}.$$

2. Найти характеристический многочлен матрицы

$$A = \begin{bmatrix} 5/2 & 1 & 1/2 \\ 1/2 & 2 & -1/2 \\ -1/2 & -1 & 3/2 \end{bmatrix}.$$

# Литература

Alparslan E. Finite  $p$ -adic number systems with possible applications. Ph. D. Dissertation. Department of Electrical Engineering, University of Maryland, College Park, 1975.

Andrews D. H., Kokes R. J. Fundamental Chemistry. John Wiley, New York, 1963.

Bachman G. Introduction to  $p$ -adic Numbers and Valuation Theory. Academic Press, New York, 1964.

Beiser P. S. An examination of finite-segment  $p$ -adic number systems as an alternative methodology for performing exact arithmetic. M. S. Thesis. Department of Applied Mathematics and Computer Science, University of Virginia, Charlottesville, 1979.

Ben-Israel A., Greville T. N. C. Generalized Inverses: Theory and Applications. Wiley-Interscience, New York, 1974, Reprinted with corrections by Robert E. Krieger Pub. Co., Melbourne, Florida, 1980.

Benson S. W. Chemical Calculations. John Wiley, New York, 1962.  
Bhimasankaram P. Some contributions to the theory, applications, and computation of generalized inverses of matrices. Ph. D. Dissertation. Indian Statistical Institute, Calcutta, 1971.

Boullion T. L., Odell P. L. Generalized Inverse Matrices. Wiley-Interscience, New York, 1971.

Bowman V. J., Burdet C. A. On the general solution of systems of mixed-integer linear equations. SIAM J. Appl. Math., 26, 1974, 120—125.

Dahlquist G., Björk A. Numerical Methods. (translated by N. Anderson). Prentice-Hall, Englewood Cliffs, N. J., 1974.

Decell H. P., Jr. An application of the Cayley—Hamilton theorem to generalized matrix inversion. SIAM Review, 7, 1965, 526—528.

Dixon J. D. Exact solution of linear equations using  $p$ -adic expansions. Numer. Math., 40, 1982, 137—141.

Farinnade J. A. Fast parallel exact matrix computations using  $p$ -adic arithmetic. M. S. Thesis. Department of Computer Sciences, University of Lagos, Nigeria, 1976.

Gregory R. T. The use of Finite-segment  $p$ -adic arithmetic for exact computation. BIT, 18, 1978, 282—300.

Gregory R. T. Error-Free Computation. Robert E. Krieger Pub. Co., Melbourne, Florida, 1980.

Gregory R. T. Residue arithmetic with rational operands. Proceedings 5th Symposium on Computer Arithmetic. IEEE Computer Society, Ann Arbor, Michigan, 144—145, 1981a.

Gregory R. T. Error-free computation with rational numbers. BIT, 21, 1981b, 194—202.

Gregory R. T. A method for and an application of error-free computations. Proceedings of the AFCET Symposium «Mathematics for Computer Science». Paris, 152—158, 1982.

Gregory R. T. Exact computation with order- $N$  Farey fractions. Computer Science and Statistics: Proceedings of the 15th Symposium on the Interface. J. E. Gentle Editor, North Holland, Amsterdam, 1983.

Gregory R. T., Karney D. L. A Collection of Matrices for Testing Computational Algorithms. Robert E. Krieger Pub. Co., Melbourne, Florida, 1978.

Greville T. N. E. Some applications of the pseudo-inverse of matrix. *SIAM Review*, 2, 1960, 15—22.

Hardy G. M., Wright E. M. An Introduction to the Theory of Numbers (4th ed.). Clarendon Press, Oxford, 1960.

Hehner E. C. R., Horspool R. N. S. A new representation of the rational numbers for fast easy arithmetic. *SIAM J. Comp.*, 8, 1979, 124—134.

Hensel K. *Theorie der Algebraischen Zahlen*, Teubner, Leipzig, 1908.

Householder A. S. *The Theory of Matrices in Numerical Analysis*. Blaisdell Pub. Co., New York, 1964.

Howell J. A. On the reduction of a matrix to Frobenius form using residue arithmetic. Ph. D. Dissertation. Department of Computer Sciences, University of Texas, Austin, 1971.

Howell J. A., Gregory R. T. An algorithm for solving linear algebraic equations using residue arithmetic, Parts I and II. *BIT*, 9, 1969, 220—224; and 324—337.

Howell J. A., Gregory R. T. Solving linear equations using residue arithmetic — Algorithm II, *BIT*, 10, 1970, 23—37.

Hurt M. F., Maid C. A generalized inverse which gives all the integral solutions to a system of linear equations, *SIAM J. Appl. Math.*, 10, 1970, 547—550.

Hwang Shu-Hwa. Computation in a finite field using rational operands. M. S. Thesis. Department of Computer Science, University of Tennessee, Knoxville, 1981.

Kornerup P., Gregory R. T. Mapping integers and Hensel codes onto Farey fractions, *BIT*, 23, 1983, 9—20.

Krishnamurthy E. V. On optimal iterative schemes for high speed division. *IEEE Transactions on Computers*, C-19, 1970, 227—231.

Krishnamurthy E. V. Economical iterative and range transformation schemes for division. *IEEE Transactions on Computers*, C-20, 1971, 470—472.

Krishnamurthy E. V. Matrix processors using  $p$ -adic arithmetic for exact linear computations, *IEEE Transactions on Computers*, C-26, 1977, 633—639.

Krishnamurthy E. V. Generalized matrix inverse approach for automatic balancing of chemical equations. *Inter. J. Math. Educ. Sci. Technol.*, 9, 1978, 323—328.

Krishnamurthy E. V. Fast parallel exact computation of the Moore—Penrose inverse and rank of a matrix. *Elektronische Informationsverarbeitung und Kybernetik*, 19, 1983, 95—98.

Krishnamurthy E. V., Adegbeyeni E. O. Finite field computational techniques for the exact solutions of mixed-integer linear equations. *Inter. J. Syst. Sci.*, 8, 1977, 1181—1192.

Krishnamurthy E. V., Klette R. Fast parallel realization of matrix multiplication. *Elektronische Informationsverarbeitung und Kybernetik*, 17, 1981, 279—292.

Krishnamurthy E. V., Murthy V. K. Fast iterative division of  $p$ -adic numbers. *IEEE Transactions on Computers*, C-32, 1983, 396—398.

Krishnamurthy E. V., Rao T. M., Subramanian K. Finite-segment  $p$ -adic number systems with applications to exact computation. *Proc. Indian Acad. Sci.*, 81A, 1975a, 58—79.

Krishnamurthy E. V., Rao T. M., Subramanian K.  $p$ -adic arithmetic procedures for exact matrix computations, *Proc. Indian Acad. Sci.*, 82A, 1975b, 165—175.

Krishnamurthy E. V., Sen S. K. *Computer-based Numerical Algorithms*. East-West Press, New Delhi, 1976.

- Kuene R. E. The Theory of General Economic Equilibrium. Princeton University, Princeton, N. J., 1963.
- Lewis Ruth Ann.  $p$ -adic number systems for error-free computation. Ph. D. Dissertation. Department of Mathematics, University of Tennessee, Knoxville, 1979.
- MacDuffee C. C. The  $p$ -adic numbers of Hensel. Amer. Math. Monthly, 45, 1938, 500—508.
- Mahler K. Introduction to  $p$ -adic Numbers and Their Functions. Cambridge University Press, Cambridge, 1973.
- McCoy N. H. Rings and Ideals. Carus Monography #8. The Mathematical Association of America, Washington, D. C., 1948.
- Miola A. M. The conversion of Hensel codes to their rational equivalents. ACM Sigsam Bulletin, Nov. 1982, 24—26.
- Murthy K. G. Linear and Combinatorial Programming. John Wiley, New York, 1976.
- Pennington R. H. Introductory Computer Methods and Numerical Analysis (2nd ed.). Macmillan, Toronto, 1970.
- Pettoufrezzo A. J., and Byrkit D. R. Elements of Number Theory. Prentice-Hall, Englewood Cliffs, N. J., 1970.
- Pyle L. D., and Cline R. E. The generalized inverse in linear programming—interior gradient projection methods, SIAM J. Appl. Math., 24, 1973, 511—534.
- Rao C. R., and Mitra S. K. Generalized Inverse of Matrices and its Applications. John Wiley, New York, 1971.
- Rao T. M. Finite field computational techniques for exact solution of numerical problems. Ph. D. Dissertation. Department of Applied Mathematics, Indian Institute of Sciences, Bangalore, 1975.
- Rao T. M. Error-free computation of characteristic polynomial of a matrix. Comp. and Math. with Appl., 4, 1978, 61—65.
- Rao T. M., and Gregory R. T. The conversion of Hensel codes to rational numbers. Proceedings 5th Symposium on Computer Arithmetic. IEEE Computer Society, Ann Arbor, Michigan, 1981, 10—20.
- Rao T. M., Subramanian K., and Krishnamurthy E. V. Residue arithmetic algorithms for exact computation of  $q$ -inverses of matrices, SIAM J. Numer. Anal., 13, 1976, 155—171.
- Smyre J. S. Exact computation using extended-precision single-modulus residue arithmetic. M. S. Thesis. Department of Computer Science, University of Tennessee, Knoxville, 1983.
- Stallings W. I., and Boullion T. L. Computation of pseudoinverse matrices using residue arithmetic. SIAM Review, 14, 1972, 152—163.
- Stewart G. W. On the continuity of the generalized inverse. SIAM J. Appl. Math., 17, 1969, 33—45.
- Stoer J., Bulirsch R. A. An Introduction to Numerical Analysis. Springer-Verlag, New York, 1980.
- Subramanian K. Symbolic processing of polynomial matrices using finite-field transforms. Ph. D. Dissertation. School of Automation, Indian Institute of Science, Bangalore, 1977.
- Szabó N. S., Tanaka R. I. Residue Arithmetic and its Applications to Computer Technology. McGraw-Hill, New York, 1967.
- Van Zeggeren F., Storey S. H. The Computation of Chemical Equilibria. Cambridge University Press, Cambridge, 1970.
- Wilkinson J. H. Rounding Errors in Algebraic Processes. Prentice-Hall, Englewood Cliffs, N. J., 1963.
- Young D. M., Gregory R. T. A Survey of Numerical Mathematics, vol. I. Addison-Wesley, Reading, Massachusetts, 1972.
- Young D. M., Gregory R. T. A Survey of Numerical Mathematics, vol. II. Addison-Wesley, Reading, Massachusetts, 1973.

Zassenhaus H. On Hensel Factorization. J. of Number Theory, 1, 1969, 291—311.

Zlobec S., Ben-Israel A. On explicit solutions of interval linear programs. Israel J. Math., 8, 1970, 12—22.

#### Издания на русском языке

Кнут Д. Искусство программирования для ЭВМ. Т. 1. Пер. с англ. — М.: Мир, 1976.

Коблиц Н.  $p$ -адические числа,  $p$ -адический анализ и дзета-функции. Пер. с англ. — М.: Мир, 1982.

Маркус М., Минк Х. Обзор по теории матриц и матричных неравенств. Пер. с англ. — М.: Наука, 1972.

Фаддеев Д. К., Фаддеева В. Н. Вычислительные методы линейной алгебры. 2-е изд. — М.: Физматгиз, 1963.

Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений. Пер. с англ. — М.: Мир, 1980.

# Именной указатель

- Ахо** (Acho A.) 27
- Буллон** (Boullion T. L.) 143
- Воеводин** В. В. 177
- Гензель** (Henzel K.) 75
- Грегори** (Gregory C.) 67, 74
- Евклид** (Euclides) 42
- Кнут** (Knut D.) 27
- Коблиц** (Koblitz N.) 76, 78, 79, 82
- Корбут** А. А. 148
- Корнеруп** (Kornerup P.) 42, 47, 53, 56
- Коши** (Cauchy A. L.) 77
- Кришнамурти** (Krishnamurthy E. V.) 42, 53
- Льюис** (Lewis Ruth Ann) 102
- Малькольм** (Malcolm M. A.) 11
- Маркус** (Marcus M.) 150
- Матула** (Matula D.) 64, 67, 74
- Минк** (Mink H.) 150
- Моулер** (Moler C. B.) 11
- Муртаф** (Murtagh B.) 183
- Понтрягин** Л. С. 16, 75
- Рао** (Rao T. M.) 20
- Столингс** (Stallings W. I.) 143
- Ульман** (Ulman J.) 27
- Фаддеев** Д. К. 142, 143, 194
- Фаддеева** В. Н. 142, 143, 194
- Финкельштейн** Ю. Ю. 148
- Форсайт** (Forsythe G. E.) 11
- Хинчин** А. Я. 55
- Хопкрофт** (Hopcroft J.) 27

# Предметный указатель

- Алгоритм Гревилла (Grevill algorithm) 141
- Дисела — Леверрье (Decel—Le-verrier algorithm) 134, 142
- Евклида (Euclidean algorithm) 42
- — расширенный (extended) 45, 48, 50
- — свойства 48—53
- Матулы — Грегори (Gregory — Matula algorithm) 56
- Эрмита (Hermite algorithm) 133
- — каноническая форма (canonical form) 118
- Анализ ошибок обратный (backward error analysis) 12
- Арифметика в одномодульной системе вычетов над рациональными числами (single modulus residue arithmetic with rational numbers) 34, 135
- — — — — целыми числами (with integers) 13, 15
- вычетов многомодульная над рациональными числами (multiple modulus residue arithmetic with rational numbers) 60, 63, 138
- — — — — целыми числами (with integers) 21, 23
- конечноразрядная  $p$ -адическая (finite-segment  $p$ -adic arithmetic) 74, 104
- $p$ -адическая ( $p$ -adic arithmetic) 86
- Векторное основание в многомодульной арифметике (base vector in multiple modulus residue arithmetic) 21, 60, 63, 138
- Вычеты симметричные (symmetric residues) 19
- Вычитание  $p$ -адическое ( $p$ -adic subtraction) 87
- Деление  $p$ -адическое ( $p$ -adic division) 90
- Диагональная редукция (diagonal reduction) 125
- Дробь Фарея (Farey fraction) 38
- — порядка  $N$  (order  $N$  Farey fraction) 39
- Единица  $p$ -адическая ( $p$ -adic unit) 88
- Задача линейного программирования (linear programming problem) 164
- — — интервальная (interval) 164, 165
- — — ограниченная (bounded) 165
- — — совместная (feasible, consistent) 165
- Задачи оптимизации (optimization problems) 164
- плохо обусловленные (ill-conditioned problems) 12
- Интервальные ограничения (interval constraints) 164, 165
- Каноническая форма Смита (Smith canonical form) 150
- — Эрмита (canonical hermitian form) 132
- Китайская теорема об остатках (Chinese remainder theorem) 27
- Класс эквивалентности (equivalence classe) 79
- Коды Гензеля (Genzel codes) 75, 93
- — вычисление обратного (computing a reciprocal) 108
- — вычитание (subtraction) 105
- — деление (division) 106
- — с плавающей точкой (floating point Genzel codes) 98
- — — — — ненормализованные (unnormalized) 115
- — — — — нормализованные (normalized) 99
- — сложение (addition) 104
- — таблица (table) 100
- — умножение (multiplication) 105

Конечно-разностная система  $p$ -адических чисел (finite segment  $p$ -adic number system) 93

Конечный отрезок  $p$ -адического разложения (finite segment of  $p$ -adic expansion) 93—94

Мантисса (mantissa) 99

Матрица начальная (seed matrix) 45  
— невырожденная по модулю  $m$  (non singular modulo  $m$  matrix) 177

— обобщенная обратная ( $g$ -обратная) (generalized inverse [ $g$ -inverse]) 124

— обратная (inverse of a matrix) 124 См.  $g$ -обратная матрица

— унимодулярная (unimodular matrix) 149

— *Гессенберга* (Hessenberg matrix) 181

—  $g$ -обратная ( $g$ -inverse of a matrix) 124

— *Мура — Пенроуза* (Moore — Penrose matrix) 125

— — рефлексивная (reflexive  $g$ -inverse matrix) 125

— — со свойством минимальной нормы (minimum norm  $g$ -inverse matrix) 125

— — — наименьших квадратов (least square  $g$ -inverse matrix) 125

Матрица-единица (unit matrix) 149

Метод Гаусса — Жордана (Gauss — Jordan method) 178

— *Леверье* (Leverrier method) 133, 142

— *Фаддеева* (Leverrier — Faddeev method) 194

— *Ньютона* (Newton method) 108—109

— — *Шульца* (Newton — Schultz method) 177

— общего знаменателя (common denominator method) 57

Метрика (metric) 64

—  $p$ -адическая ( $p$ -adic metric) 78

Модуль в одномодульной арифметике (module) 13, 14, 135, 138

Наибольший общий делитель (greatest common divisor) 17, 42

Наилучшая рациональная аппроксимация (best rational approximation) 55

Наименьший неотрицательный вычет (least non-negative residue) 15

Норма (norm, valuation) 75, 76  
—  $p$ -адическая ( $p$ -adic norm) 76

Обобщенный класс вычетов (generalized residue class) 35, 61

Обратный по модулю  $m$  элемент (inverse modulo  $m$ ) 16, 46

Оптимальные решения (optimal solutions) 163

Основание (radix) 27

Основная дробь (radix fraction) 82

Отношение эквивалентности (equivalence relation) 78

Отображение кода Гензеля (mapping for Genzel code) 115—117

— рациональных чисел (mapping for rational numbers) 41, 42

— — обратное (inverse) 42, 53

Ошибка округления (rounding error) 12

Плавающая точка (floating point) 98

Показатель (exponent) 99

Поле вещественных чисел (real field) 79

— Галуа (Galois field) 17

— — конечное (finite) 17, 18

—  $p$ -адических чисел (field of  $p$ -adic numbers) 75, 79, 80

Полное метрическое пространство (complete metric space) 77

Пополнение метрического пространства (completion of a metric space) 78

Последовательность Коши (Cauchy sequence) 77

Предел последовательности (limit of a convergent sequence) 77

Представление дополнительное (complement representation) 81

— симметричное (symmetric representation) 23

— стандартное обратного числа (multiple-modulus residue representation of an inverse) 25

— — целого числа (of an integer) 21

— целых чисел в системе счисления со смешанным основанием (mixed-radix number representation) 27

— числа в системе с фиксированным основанием (fixed-radix number representation) 28

—  $p$ -адическое ( $p$ -adic representation) 80



- Псевдопереполнение (over-flow) 20, 40
- Разложение  $p$ -адическое ( $p$ -adic expansion) 79, 81, 84
- — периодическое (periodic) 81
- Свойство деления (division property) 43
- Система вычетов симметричная (symmetric residue number system) 23
- конечноразрядных  $p$ -адических чисел (finite segment  $p$ -adic number system) 75
- линейных уравнений полуцелая (mixed-integer linear equations) 169
- счисления стандартная (standard mixed-radix system) 28
- Скелетное разложение (full-rank factorization) 129
- Сложение  $p$ -адических чисел (addition of  $p$ -adic numbers) 86
- Совместная ИЗ (feasible IP) 165
- — ограниченная (bounded) 165
- Сравнение (congruence) 14
- Стандартное представление цифры (standard residue digit) 21
- Статическая модель *Леонтьева* (Leontif static model) 161
- Стехиометрия (stoichiometry) 148
- Столбцовая нормальная форма (column-echelon form) 133
- — — простая (simple) 133
- Строчная нормальная форма (row-echelon form) 133
- — — простая (simple) 133
- Сходящаяся последовательность (convergent sequence) 77
- Теорема о разложении  $p$ -адических чисел ( $p$ -adic number expansion theorem) 79
- Точка  $p$ -адическая ( $p$ -adic point) 80
- Умножение  $p$ -адическое ( $p$ -adic multiplication) 88
- Упорядоченный набор цифр (digit sequence) 28
- Уравнивание химических уравнений (balancing chemical equations) 149, 152
- Формула *Шермана — Моррисона* (Sherman—Morrison formula) 183
- Формулы *Ньютона* (Newton formulas) 143
- Функция расстояния (distance function) 76
- Целевая функция (objective function) 164
- Целое  $p$ -адическое ( $p$ -adic integer) 88
- Цифра симметричного представления (symmetric residue digit) 23
- $p$ -адического разложения (digit of  $p$ -adic expansion) 81
- Цифры стандартного представления для смешанного основания (mixed-radix standard digits) 18
- Числа взаимно простые (relatively primes) 17
- машинные (computer-representable numbers) 11
- с плавающей точкой (floating-point numbers) 11
- $p$ -адические ( $p$ -adic numbers) 75, 79

# Оглавление

Предисловие редактора перевода . . . . .	5
Предисловие . . . . .	7
Список обозначений . . . . .	9
<i>Глава I. Арифметика вычетов или модульная арифметика . . . . .</i>	<i>11</i>
§ 1. Введение . . . . .	11
§ 2. Арифметика в одномодульной системе вычетов . . . . .	13
§ 3. Многомодульная арифметика вычетов . . . . .	21
§ 4. Отображение стандартных представлений вычетами в целые числа . . . . .	27
§ 5. Одномодульная арифметика вычетов для рациональных чисел . . . . .	34
§ 6. Прямое и обратное отображения . . . . .	42
§ 7. Многомодульная арифметика вычетов для рациональных чисел . . . . .	60
<i>Глава II. Конечноразрядная <math>p</math>-адическая арифметика . . . . .</i>	<i>75</i>
§ 1. Введение . . . . .	75
§ 2. Поле $p$ -адических чисел . . . . .	75
§ 3. Арифметика в $\mathbb{Q}_p$ . . . . .	86
§ 4. Конечноразрядная система $p$ -адических чисел . . . . .	93
§ 5. Арифметические операции над кодами Гензеля . . . . .	104
§ 6. Удаление первого нуля в коде Гензеля . . . . .	114
§ 7. Отображение кода Гензеля в единственную дробь Фарей порядка $N$ . . . . .	115
<i>Глава III. Точное вычисление обобщенных обратных матриц . . . . .</i>	<i>124</i>
§ 1. Введение . . . . .	124
§ 2. Свойства $g$ -обратных матриц . . . . .	125
§ 3. Приложения $g$ -обратных матриц . . . . .	130
§ 4. Точное вычисление $A^+$ в случае рациональной матрицы $A$ . . . . .	131
§ 5. Неудачи при применении арифметики вычетов и предупредительные меры . . . . .	145
<i>Глава IV. Целочисленные решения линейных уравнений . . . . .</i>	<i>148</i>
§ 1. Введение . . . . .	148
§ 2. Основы теории . . . . .	149
§ 3. Матричная форма химических уравнений . . . . .	152
§ 4. Решение однородной системы . . . . .	154
§ 5. Решение неоднородной системы . . . . .	162
§ 6. Решение интервальных задач линейного программирования . . . . .	164
§ 7. Решение полунелых систем линейных уравнений . . . . .	169
<i>Глава V. Итерационные методы обращения матриц и решения систем линейных уравнений . . . . .</i>	<i>176</i>
§ 1. Введение . . . . .	176
§ 2. Метод Ньютона — Шульца для обращения матрицы . . . . .	177
§ 3. Итерационное решение линейной системы . . . . .	183
§ 4. Итерационное вычисление $g$ -обратных . . . . .	188
<i>Глава VI. Точное вычисление характеристического многочлена матрицы . . . . .</i>	<i>194</i>
§ 1. Введение . . . . .	194
§ 2. Алгоритм вычисления характеристического многочлена нижней матрицы Хессенберга . . . . .	195
Литература . . . . .	200
Именной указатель . . . . .	204
Предметный указатель . . . . .	205



1 р. 60 к.

---

Теория безошибочных вычислений имеет дело с задачами, для которых входная информация представима набором целых чисел (или многочленов с целыми коэффициентами), а решение является рациональной функцией от этих чисел (или многочленов). К задачам такого типа относятся обращение и построение характеристического многочлена целочисленной матрицы, а также решение линейной системы с целыми коэффициентами.

*Из предисловия редактора перевода*

---

ISBN 5—03—001145—5 (русск.)

ISBN 0—387—90967—2 (англ.)