

ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ ЛИНЕЙНОЙ АЛГЕБРЫ**ОГЛАВЛЕНИЕ**

Предисловие	6
Глава I. Основные сведения из линейной алгебры	7
§ 1. Матрицы	7
§ 2. Матрицы специального вида	33
§ 3. Аксиомы линейного пространства	41
§ 4. Базис и координаты	45
§ 5. Подпространства	50
§ 6. Линейные операторы	58
§ 7. Каноническая форма Жордана	71
§ 8. Строение инвариантных подпространств	85
§ 9. Ортогональность векторов и подпространств	87
§ 10. Линейные операторы в унитарном пространстве и евклидовом пространстве	94
§ 11. Самосопряженный оператор	99
§ 12. Квадратичные формы	111
§ 13. Понятие предела в линейной алгебре	117
§ 14. Градиент функционала	134
Глава II. Точные методы решения систем линейных уравнений	137
§ 15. Обусловленность матриц	138
§ 16. Метод Гаусса	147
§ 17. Вычисление определителей	157
§ 18. Компактные схемы для решения неоднородной линейной системы	160
§ 19. Связь метода Гаусса с разложением матрицы на множители	162
§ 20. Метод квадратных корней	165
§ 21. Обращение матрицы	168
§ 22. Задача исключения	172
§ 23. Исправление элементов обратной матрицы	182
§ 24. Обращение матрицы при помощи разбиения на клетки	184
§ 25. Метод окаймления	187
§ 26. Эскалаторный метод	192
§ 27. Метод Перселла	195
§ 28. Метод пополнения для обращения матрицы	198
Глава III. Итерационные методы решения Систем линейных уравнений	204
§ 29. Принципы построения итерационных процессов	204
§ 30. Метод последовательных приближений	207
§ 31. Подготовка системы линейных уравнений к виду, удобному для применения метода последовательных приближений. Метод простой итерации	214
§ 32. Одношаговый циклический процесс	220

§ 33. Метод П. А. Некрасова	226
§ 34. Методы полной релаксации	230
§ 35. Неполная релаксация	232
§ 36. Исследование итерационных методов для систем с квазитрехдиагональными матрицами	237
§ 37. Теорема сходимости	244
§ 38. Управление релаксацией	248
§ 39. Релаксация по длине вектора невязки	253
§ 40. Групповая релаксация	254
Глава IV. Полная проблема собственных значений	257
§ 41. Устойчивость проблемы собственных значений	259
§ 42. Метод А. Н. Крылова	263
§ 43. Определение собственных векторов по методу А. Н. Крылова	271
§ 44. Метод Хессенберга	273
§ 45. Метод Самуэльсона	280
§ 46. Метод А. М. Данилевского	285
§ 47. Метод Леверье и видоизменение Д. К. Фаддеева	295
§ 48. Эскалаторный метод	300
§ 49. Метод интерполяции	308
§ 50. Метод ортогонализации последовательных итераций	314
§ 51. Преобразование симметричной матрицы к трехдиагональному виду посредством вращений	317
§ 52. Уточнение полной проблемы собственных значений	324
Глава V. Частичная проблема собственных значений	328
§ 53. Определение наибольшего по модулю собственного значения матрицы при помощи последовательных итераций	329
§ 54. Ускорение сходимости степенного метода	346
§ 55. Модификации степенного метода	352
§ 56. Применение степенного метода к отысканию нескольких собственных значений	355
§ 57. Ступенчатый степенной метод	358
§ 58. Метод λ -разности	367
§ 59. Метод исчерпывания	370
§ 60. Метод понижения	375
§ 61. Координатная релаксация	378
§ 62. Уточнение отдельного собственного значения и принадлежащего ему собственного вектора	386
Глава VI. Метод минимальных итераций и другие методы, основанные на идее ортогонализации	392
§ 63. Метод минимальных итераций	392
§ 64. Биортогональный алгорифм	404
§ 65. Метод A -минимальных итераций	416
§ 66. A -биортогональный алгорифм	425

§ 67. Двучленные формулы метода минимальных итераций и биортогонального алгорифма	427
§ 68. Методы сопряженных направлений и их общие свойства	433
§ 69. Некоторые методы сопряженных направлений	437
Глава VII. Градиентные итерационные методы	455
§ 70. Метод наискорейшего спуска для решения линейных систем	456
§ 71. Градиентный метод с минимальными невязками	465
§ 72. Градиентные методы с неполной релаксацией	466
§ 73. s -шаговые градиентные методы наискорейшего спуска	472
§ 74. Определение алгебраически наибольшего собственного значения симметричной матрицы и принадлежащего ему собственного вектора градиентными методами	480
§ 75. Решение частичной проблемы собственных значений с помощью полиномов Ланцоша	494
§ 76. s -шаговый метод наискорейшего спуска	498
Глава VIII. Итерационные методы для решения полной проблемы собственных значений	508
§ 77. Алгорифм деления и вычитания	508
§ 78. Треугольный степенной метод	524
§ 79. LR-алгорифм	530
§ 80. AP-алгорифм	533
§ 81. Итерационные процессы, основанные на применении вращений	536
§ 82. Решение полной проблемы собственных значений при помощи спектрального анализа последовательных итераций	547
Глава IX. Универсальные алгорифмы	553
§ 83. Общая идея подавления компонент	554
§ 84. Прием Л. А. Люстерника для ускорения сходимости метода последовательных приближений при решении системы линейных уравнений	557
§ 85. Подавление компонент при помощи полиномов низших степеней	559
§ 86. Различные формы проведения универсальных алгорифмов	563
§ 87. Универсальный алгорифм, наилучший в смысле первого критерия	567
§ 88. Универсальный алгорифм, наилучший в смысле второго критерия	570
§ 89. Прием А. А. Абрамова для ускорения сходимости метода последовательных приближений при решении систем линейных уравнений	572
§ 90. BT-процессы	574
§ 91. Общие трехчленные итерационные процессы	577
§ 92. Универсальный алгорифм Ланцоша	582
§ 93. Универсальные алгорифмы, наилучшие в среднем	586
§ 94. Метод подавления компонент в комплексной области	589
§ 95. Применение конформного отображения к решению линейных систем	591
§ 96. Примеры s -универсальных алгорифмов	599

§ 97. Метод конформного отображения в применении к неподготовленной системе	603
§ 98. Применение идеи подавления компонент к решению частичной проблемы собственных значений	609
§ 99. Применение конформного отображения к решению частичной проблемы собственных значений	610
Заключение	612
Дополнение	615
Литература	617
Дополнительная литература	654

ПРЕДИСЛОВИЕ

Настоящая книга посвящена изложению вычислительных методов для решения основных задач линейной алгебры.

Этими задачами являются решение системы линейных уравнений, обращение матрицы, решение полной и частичной проблем собственных значений.

Огромное количество численных методов решения этих задач, появившихся главным образом в последние годы, поставило авторов перед необходимостью попытки их систематизации и изложения с некоторых общих точек зрения. При этом авторы старались строить изложение не выходя за области понятий линейной алгебры в той мере, в какой это было возможно. Так, например, авторы сознательно исключили использование теории непрерывных дробей, заменив ее теорией ортогональных полиномов, в которой, в свою очередь, ортогональность понимается в линейно-алгебраическом смысле.

В книге почти не затрагивается важный вопрос о влиянии ошибок округления на результат вычислений.

Первая глава книги носит вводный характер. Остальные восемь глав посвящены изложению вычислительных методов. Материал этих глав частично был освещен в книге В. Н. Фаддеевой, вышедшей в 1950 г. под тем же названием.

В конце книги приложены библиография по вычислительным методам линейной алгебры и вопросам оценки и распределения собственных значений матрицы. При ее составлении существенную помощь оказали авторам И. А. Лифшиц и Р. С. Александрова. Авторы приносят им свою благодарность.

Рукопись книги была прочитана В. Н. Кублановской, сделавшей ряд ценных замечаний. Авторы приносят ей глубокую благодарность. Авторы благодарят также редактора книги Г. П. Акилова и всех своих товарищей, проявивших интерес к их работе.

ГЛАВА I

ОСНОВНЫЕ СВЕДЕНИЯ ИЗ ЛИНЕЙНОЙ АЛГЕБРЫ

§ 1. Матрицы

1. Определения. Прямоугольной матрицей называется совокупность чисел, вообще говоря, комплексных, расположенных в виде прямоугольной таблицы, содержащей n строк и m столбцов.

Такая матрица записывается в виде:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \quad (1)$$

или, сокращенно, в виде:

$$A = (a_{ij}), \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m.$$

Две матрицы называются равными, если равны их соответствующие элементы.

Матрица, состоящая из одной строки, называется просто строкой; матрица, состоящая из одного столбца, — столбцом; матрица $A = (a)$, состоящая из одного числа, отождествляется с этим числом. Если число n строк матрицы равно числу ее столбцов, то матрица называется квадратной. В этом случае число n называется порядком матрицы.

Среди квадратных матриц важную роль играют так называемые диагональные матрицы, т. е. матрицы, у которых отличны от нуля лишь элементы, стоящие вдоль диагонали. Диагональные матрицы обозначаются $[\alpha_1, \alpha_2, \dots, \alpha_n]$, так что

$$[\alpha_1, \alpha_2, \dots, \alpha_n] = \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_n \end{bmatrix}. \quad (2)$$

Если все числа a_i при этом равны между собой, матрица называется скалярной:

$$[\alpha] = \begin{bmatrix} \alpha & 0 & \dots & 0 \\ 0 & \alpha & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \alpha \end{bmatrix} \quad (3)$$

и в случае, если $\alpha = 1$, — единичной:

$$E = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix} = (\delta_{ij}), \quad (4)$$

где δ_{ij} так называемый символ Кронекера, т. е. $\delta_{ij} = 0$ при $i \neq j$, $\delta_{ii} = 1$.

Наконец, матрица, все элементы которой равны нулю, называется нулевой. Мы будем обозначать ее символом 0.

Переставив в матрице

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

строки со столбцами, мы получим так называемую транспонированную матрицу

$$A' = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{bmatrix}. \quad (5)$$

Квадратная матрица A равна транспонированной A' в том и только в том случае, когда она симметрична, т. е. если $a_{ij} = a_{ji}$.

Очевидно, что матрица, транспонированная со строкой, есть столбец, составленный из тех же элементов. Это обстоятельство мы часто будем использовать для удобства записи столбцов. Так, вместо столбца

$$\begin{bmatrix} 4 \\ 2 \\ 3 \\ 5 \end{bmatrix}$$

мы будем писать $(4, 2, 3, 5)'$.

Посредством замены элементов матрицы на комплексно-сопряженные числа мы приходим к комплексно-сопряженной матрице \bar{A} . Если элементы матрицы A вещественны, то $\bar{A} = A$.

Матрица $A^* = \bar{A}'$, комплексно-сопряженная с транспонированной, называется матрицей сопряженной с матрицей A . Очевидно, что

$$(A^*)^* = A.$$

Если матрица A вещественна, то сопряженная с ней матрица совпадает с транспонированной.

Определителем квадратной матрицы называется определитель, элементы которого равны элементам матрицы. Определитель матрицы A обозначается через $|A|$.

Далее, любой определитель, строки и столбцы которого „укладываются“ в строки и столбцы матрицы называется минором этой матрицы. Подробнее, минор порядка k матрицы A есть определитель k -го порядка, составленный из элементов, находящихся на пересечении некоторых k строк и k столбцов матрицы A в их естественном расположении.

Рангом матрицы A называется максимальный порядок отличных от нуля миноров матрицы. Иначе, рангом матрицы называется такое число r , что среди миноров матрицы существует минор порядка r , неравный нулю, а все миноры порядка $r+1$ и выше равны нулю или не могут быть составлены.

2. Умножение матриц на число и сложение матриц. Произведением матрицы $A = (a_{ij})$ на число α называется матрица, элементы которой получены из элементов матрицы A умножением на число α :

$$\alpha A = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \dots & \alpha a_{1m} \\ \vdots & \ddots & \ddots & \vdots \\ \alpha a_{n1} & \alpha a_{n2} & \dots & \alpha a_{nm} \end{bmatrix}. \quad (6)$$

Суммой двух прямоугольных матриц $A = (a_{ij})$ и $B = (b_{ij})$, имеющих одинаковое число как строк, так и столбцов, называется матрица C , элементы которой равны суммам соответствующих элементов матриц A и B , т. е.

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2m} + b_{2m} \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \dots & a_{nm} + b_{nm} \end{bmatrix}. \quad (7)$$

Введенные выше операции, как это нетрудно видеть, обладают свойствами:

1. $A + (B + C) = (A + B) + C$.
2. $A + B = B + A$.
3. $A + 0 = A$.

4. $(\alpha + \beta)A = \alpha A + \beta A.$
5. $\alpha(A+B) = \alpha A + \alpha B.$
6. $1 \cdot A = A.$
7. $\alpha(\beta A) = \alpha\beta A.$

Здесь A , B и C матрицы, а α и β числа.

3. Умножение матриц. Умножение матриц A и B определяется только в предположении, что число столбцов матрицы A равно числу строк матрицы B . В этом предположении элементы произведения C определяются следующим образом: элемент i -й строки j -го столбца матрицы C равен сумме произведений элементов i -й строки матрицы A на соответствующие элементы j -го столбца матрицы B . Таким образом,

$$AB = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \dots & \dots & \dots & \dots \\ b_{m1} & b_{m2} & \dots & b_{mp} \end{bmatrix} = \\ = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & c_{np} \end{bmatrix}, \quad (8)$$

где

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{im}b_{mj} \quad (i = 1, \dots, n; j = 1, \dots, p). \quad (9)$$

Заметим, что произведение двух прямоугольных матриц есть снова прямоугольная матрица, число строк которой равно числу строк первой матрицы, а число столбцов равно числу столбцов второй матрицы. Так, например, произведение квадратной матрицы на матрицу, состоящую из одного столбца, есть матрица из одного столбца.

Перестановочный закон при умножении матриц, вообще говоря, не имеет места. Легко видеть, что сама постановка вопроса о равенстве матриц AB и BA имеет смысл только для квадратных матриц A и B одинакового порядка. Действительно, матрицы AB и BA имеют смысл одновременно только в случае, если число строк первой матрицы равно числу столбцов второй, а число столбцов первой матрицы равно числу строк второй. При выполнении этих условий матрицы AB и BA обе будут квадратными, но разных порядков, если A и B не квадратные. Но даже и для квадратных матриц одинакового порядка, вообще говоря, $AB \neq BA$.

Например,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} -5 & 3 \\ -9 & 7 \end{bmatrix}$$

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

В отдельных случаях умножение может быть коммутативно — в таком случае матрицы называются перестановочными. Так, например, скалярные матрицы перестановочны с любыми квадратными матрицами того же порядка, ибо

$$\begin{aligned} & \begin{bmatrix} \alpha & 0 & \dots & 0 \\ 0 & \alpha & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = \\ & = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} \alpha & 0 & \dots & 0 \\ 0 & \alpha & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha \end{bmatrix} = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \dots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \dots & \alpha a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha a_{n1} & \alpha a_{n2} & \dots & \alpha a_{nn} \end{bmatrix}. \end{aligned}$$

Из последней формулы следует особая роль единичной матрицы при умножении матриц. Именно, единичная матрица среди всех квадратных матриц данного порядка играет такую же роль, как число единица среди чисел. Действительно,

$$AE = EA = A.$$

Можно доказать, что умножение матриц ассоциативно, именно, если AB и $(AB)C$ имеют смысл, то имеют смысл BC и $A(BC)$ и

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

Действительно, элемент i -й строки и j -го столбца $(AB)C$ равен

$$\sum_{\beta} \left[\sum_{\alpha} a_{i\alpha} b_{\alpha\beta} \right] c_{\beta j} = \sum_{\alpha} \sum_{\beta} a_{i\alpha} b_{\alpha\beta} c_{\beta j},$$

а элемент i -й строки и j -го столбца матрицы $A(BC)$ равен

$$\sum_{\alpha} a_{i\alpha} \left[\sum_{\beta} b_{\alpha\beta} c_{\beta j} \right] = \sum_{\alpha} \sum_{\beta} a_{i\alpha} b_{\alpha\beta} c_{\beta j}.$$

Таким образом, соответствующие элементы матриц $(AB)C$ и $A(BC)$ равны, следовательно, равны и сами матрицы.

Произведение матриц обладает также свойствами:

$$\begin{aligned} \alpha(AB) &= (\alpha A)B = A(\alpha B) \\ (A+B)C &= AC + BC \\ C(A+B) &= CA + CB, \end{aligned}$$

где A, B, C матрицы, а α число.

Имеет место следующее правило транспонирования произведения:

$$(AB)' = B'A'. \quad (10)$$

Действительно, элемент i -й строки и j -го столбца матрицы $(AB)'$ равен элементу j -й строки и i -го столбца матрицы AB , т. е. равен

$$a_{j1}b_{1i} + a_{j2}b_{2i} + \dots + a_{jm}b_{mi}.$$

Последнее выражение, очевидно, равно сумме произведений элементов i -й строки матрицы B' на соответствующие элементы j -го столбца матрицы A' , т. е. равно элементу i -й строки и j -го столбца матрицы $B'A'$.

Ясно также, что

$$\overline{AB} = \overline{A}\overline{B}$$

и

$$(AB)^* = B^*A^*. \quad (11)$$

Как уже говорилось выше, матрица AB будет квадратной, если число n строк матрицы A равно числу столбцов матрицы B . Обозначим через m число столбцов матрицы A (оно равно также числу строк матрицы B , так как только при этом условии произведение AB имеет смысл). Из теории определителей известно, что определитель матрицы AB равен нулю, если $n > m$, и равен сумме произведений всех миноров порядка n , составленных из матрицы A , на соответствующие миноры, составленные из матрицы B , если $n \leq m$. Точнее, если

$$A = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \dots & b_{mn} \end{bmatrix}$$

и $n \leq m$, то

$$|AB| = \sum_{i_1 < i_2 < \dots < i_n} \begin{vmatrix} a_{1i_1} & \dots & a_{1i_n} \\ \vdots & \ddots & \vdots \\ a_{ni_1} & \dots & a_{ni_n} \end{vmatrix} \cdot \begin{vmatrix} b_{i_11} & \dots & b_{i_1n} \\ \vdots & \ddots & \vdots \\ b_{i_n1} & \dots & b_{i_nn} \end{vmatrix}.$$

В частности, при $m = n$

$$|AB| = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} \cdot \begin{vmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nn} \end{vmatrix},$$

т. е. определитель произведения двух квадратных матриц равен произведению определителей перемножаемых матриц.

4. Разбиение матриц на клетки. Иногда бывает целесообразно свести вычисления над матрицами высоких порядков к вычислениям с матрицами меньших порядков. Такое сведение осуществляется при помощи разбиения данных матриц на так называемые клетки. Именно,

каждую матрицу можно представить и притом многими способами, как состоящую из нескольких матриц меньшего порядка. Например,

$$\begin{aligned} & \left[\begin{array}{cccc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{array} \right] = \\ & = \left[\begin{array}{c|cc|c} a_{11} & a_{12} & a_{13} & a_{14} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{array} \right] = \left[\begin{array}{cc|cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \end{array} \right]. \end{aligned}$$

Матрицы, на которые разбивается данная матрица, называются клетками. При клеточном разбиении предполагается, что горизонтальные и вертикальные разделяющие линии пересекают всю матрицу.

Мы не будем останавливаться на общем случае разбиения матриц на клетки, а рассмотрим лишь такие разбиения квадратных матриц, при которых диагональные клетки квадратны.

Основные действия над матрицами с диагональными клетками одинаковых порядков естественным образом связываются с действиями над самими клетками.

Именно, если

$$A = \left[\begin{array}{cccc} A_{11} & A_{12} & \dots & A_{1k} \\ A_{21} & A_{22} & \dots & A_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ A_{k1} & A_{k2} & \dots & A_{kk} \end{array} \right]$$

$$B = \left[\begin{array}{cccc} B_{11} & B_{12} & \dots & B_{1k} \\ B_{21} & B_{22} & \dots & B_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ B_{k1} & B_{k2} & \dots & B_{kk} \end{array} \right],$$

причем A_{ii} и B_{ii} — квадратные матрицы одинаковых порядков, то

$$A + B = \left[\begin{array}{cccc} A_{11} + B_{11} & A_{12} + B_{12} & \dots & A_{1k} + B_{1k} \\ A_{21} + B_{21} & A_{22} + B_{22} & \dots & A_{2k} + B_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ A_{k1} + B_{k1} & A_{k2} + B_{k2} & \dots & A_{kk} + B_{kk} \end{array} \right] \quad (12)$$

$$AB = \left[\begin{array}{cccc} C_{11} & C_{12} & \dots & C_{1k} \\ C_{21} & C_{22} & \dots & C_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ C_{k1} & C_{k2} & \dots & C_{kk} \end{array} \right], \quad (13)$$

где

$$C_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j} + \dots + A_{ik}B_{kj} \quad i, j = 1, \dots, k.$$

Действительно, все произведения $A_{ii}B_{ij}$ имеют смысл, так как число столбцов матрицы A_{ii} всегда равно числу строк матрицы B_{ij} . Сумма $A_{ii}B_{ij} + \dots + A_{ik}B_{kj}$ имеет смысл, так как все слагаемые матрицы имеют одинаковое строение. Далее, пусть $c_{\alpha\beta}$ есть некоторый элемент клетки C_{ij} . Тогда

$$\begin{aligned} c_{\alpha\beta} = & (a_{\alpha 1}b_{1\beta} + \dots + a_{\alpha s_1}b_{s_1\beta}) + \\ & + \dots + \dots + \\ & + (a_{\alpha s_{k-1}+1}b_{s_{k-1}+1\beta} + \dots + a_{\alpha s_k}b_{s_k\beta}). \end{aligned}$$

Здесь $s_1, s_2 - s_1, \dots, s_k - s_{k-1}$ обозначают порядки матриц $A_{11}, A_{22}, \dots, A_{kk}$. Ясно, что заключенные в скобки слагаемые, на которые мы разбили $c_{\alpha\beta}$, являются элементами матриц $A_{11}B_{1j}, \dots, A_{kk}B_{kj}$, занимающими в этих матрицах такое же положение, какое занимает $c_{\alpha\beta}$ в матрице C_{ij} . Следовательно,

$$C_{ij} = A_{ii}B_{ij} + \dots + A_{ik}B_{kj}.$$

Формулы (12) и (13) показывают, что действия с матрицами, разбитыми на клетки указанного вида, производятся так, как будто на месте клеток находятся числа.

Важным частным случаем клеточных матриц являются окаймленные матрицы. Именно, пусть мы имеем квадратную матрицу A_{n-1} порядка $n-1$:

$$A_{n-1} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1, n-1} \\ a_{21} & a_{22} & \dots & a_{2, n-1} \\ \vdots & \ddots & \ddots & \vdots \\ a_{n-1, 1} & a_{n-1, 2} & \dots & a_{n-1, n-1} \end{bmatrix}.$$

Образуем квадратную матрицу n -го порядка A_n , присоединяя к матрице A_{n-1} строку $v_{n-1} = (a_{n1} \dots a_{n, n-1})'$, столбец $u_{n-1} = (a_{1n} a_{2n} \dots a_{n-1, n})'$ и число a_{nn} :

$$A_n = \begin{bmatrix} & a_{1n} \\ A_{n-1} & a_{2n} \\ & \vdots \\ & a_{n-1, n} \\ a_{n1} \dots a_{n, n-1} & a_{nn} \end{bmatrix} = \begin{bmatrix} A_{n-1} & u_{n-1} \\ v_{n-1} & a_{nn} \end{bmatrix}. \quad (14)$$

Будем говорить, что матрица A_n получена окаймлением матрицы A_{n-1} . Матрица A_n естественным образом разделена на клетки.

Действия над окаймленными матрицами производятся согласно общим правилам действий над клеточными матрицами.

Пусть

$$A = \begin{bmatrix} M & u \\ v & a \end{bmatrix}, \quad B = \begin{bmatrix} P & y \\ x & b \end{bmatrix}$$

две окаймленные матрицы порядка n . Смысл обозначений M , v , u , a и P , x , y , b тот же, что и в определении.

Тогда справедливы следующие равенства:

$$\begin{aligned} \alpha A &= \begin{bmatrix} \alpha M & ux \\ av & \alpha a \end{bmatrix} \\ A + B &= \begin{bmatrix} M + P & u + y \\ v + x & a + b \end{bmatrix} \\ AB &= \begin{bmatrix} MP + ux & My + ub \\ vP + ax & vy + ab \end{bmatrix}. \end{aligned} \quad (15)$$

Здесь MP и ux — матрицы $n \times n$ 1-го порядка; My и ub — столбцы, состоящие из n — 1-го элемента, vP и ax — аналогичные строки и, наконец, $vy + ab$ — число.

5. Квазидиагональные матрицы. Рассмотрим еще один частный случай клеточных матриц, именно так называемые квазидиагональные матрицы. Квазидиагональной матрицей называется квадратная матрица, у которой вдоль главной диагонали расположены квадратные клетки, а остальные элементы равны нулю. Например, матрица 7-го порядка

$$\left[\begin{array}{cc|ccccc} a_{11} & a_{12} & 0 & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & b_{11} & b_{12} & b_{13} & 0 & 0 \\ 0 & 0 & b_{21} & b_{22} & b_{23} & 0 & 0 \\ 0 & 0 & b_{31} & b_{32} & b_{33} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & c_{11} & c_{12} \\ 0 & 0 & 0 & 0 & 0 & c_{21} & c_{22} \end{array} \right]$$

квазидиагональна. Клетками этой матрицы будут, очевидно, матрицы

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix};$$

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}; \quad C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

и шесть нулевых матриц.

Если строение двух квазидиагональных матриц одинаково, то произведение таких матриц будет также квазидиагональной матрицей того же строения, диагональные клетки которой равны произведениям соответствующих клеток перемножаемых матриц.

В силу известной теоремы Лапласа определитель квазidiагональной матрицы равен произведению определителей диагональных клеток.

6. Обратная и союзная матрицы. Квадратная матрица $A = (a_{ij})$ называется неособенной или невырожденной, если ее определитель не равен нулю; в противном случае матрица называется особенной.

Введем теперь важное понятие обратной матрицы. Матрицу B назовем обратной к квадратной матрице A , если

$$AB = E. \quad (16)$$

Докажем, что необходимым и достаточным условием существования обратной матрицы является неособенность матрицы A .

Необходимость сразу вытекает из теоремы об определителе произведения двух матриц. Действительно, если $AB = E$, то $|A||B| = 1$ и, следовательно, $|A| \neq 0$.

Допустим теперь, что $|A| \neq 0$. Для построения обратной матрицы рассмотрим предварительно так называемую союзную матрицу, т. е. матрицу

$$C = \begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix}. \quad (17)$$

Здесь A_{ij} — алгебраическое дополнение элемента a_{ij} в определителе матрицы A .

Докажем, что союзная матрица обладает следующим свойством:

$$AC = |A|E. \quad (18)$$

Действительно, вычисляя общий элемент матрицы AC по правилам умножения матриц, получим, что он равен

$$a_{i1}A_{j1} + a_{i2}A_{j2} + \dots + a_{in}A_{jn},$$

т. е. равен нулю при $i \neq j$ и определителю матрицы A при $i = j$, в силу известной теоремы о разложении определителя.

Таким же образом устанавливается равенство

$$CA = |A|E. \quad (18')$$

Союзная матрица имеет смысл для любой квадратной матрицы A . Из равенства $AC = |A|E$ следует, что матрица

$$B = \frac{1}{|A|} C \quad (19)$$

есть обратная для неособенной матрицы A .

Действительно,

$$AB = A \frac{1}{|A|} C = \frac{1}{|A|} AC = E.$$

Построенная матрица обладает также свойством

$$BA = E, \quad (20)$$

что следует из равенства $CA = |A|E$.

Докажем, наконец, единственность обратной матрицы. Положим, что существует такая матрица X , что $AX = E$. Умножая это равенство на B слева, получим, что $X = B$. Если положить, что $YA = E$, то, умножая справа на B , получим $Y = B$.

Матрица, обратная к матрице A , обозначается A^{-1} . Очевидно, что $|A^{-1}| = |A|^{-1}$. Далее $(A^{-1})^{-1} = A$. Это следует непосредственно из равенства $A^{-1}A = E$.

Заметим еще, что матрица, обратная к произведению двух матриц, равна произведению обратных матриц, взятых в обратном порядке, т. е.

$$(A_1 A_2)^{-1} = A_2^{-1} A_1^{-1}. \quad (21)$$

Действительно,

$$A_1 A_2 A_2^{-1} A_1^{-1} = A_1 A_1^{-1} = E.$$

Равенство (19) дает возможность вычисления обратной матрицы. Однако вычисление союзной матрицы настолько трудоемко, что упомянутое равенство важно лишь в теоретическом отношении.

Задача численного нахождения обратной матрицы есть одна из важнейших вычислительных задач линейной алгебры, и мы к ней будем неоднократно возвращаться в последующих главах.

7. Полиномы от матриц. Определим теперь целую положительную степень квадратной матрицы, полагая

$$A^n = \overbrace{A \dots A}^{\text{n раз}}. \quad (22)$$

В силу ассоциативного закона безразлично, как в этом произведении расставить скобки, и потому мы их опускаем. Из определения ясно, что

$$\left. \begin{aligned} A^n A^m &= A^{n+m} \\ (A^n)^m &= A^{nm} \end{aligned} \right\}. \quad (23)$$

Отсюда следует, что степени одной и той же матрицы *перестановочны*.

Выражение вида

$$\alpha_0 A^n + \alpha_1 A^{n-1} + \dots + \alpha_n E,$$

где $\alpha_0, \alpha_1, \dots, \alpha_n$ комплексные числа, называется *полиномом от матрицы* или *матричным полиномом*. Полином от матрицы можно рассматривать как результат подстановки матрицы A вместо переменной t в алгебраический полином

$$f(t) = \alpha_0 t^n + \alpha_1 t^{n-1} + \dots + \alpha_n. \quad (24)$$

Важно отметить, что правила действий над полиномами от матриц ничем не отличаются от правил действий над алгебраическими полиномами.

Именно, если

$$\left. \begin{array}{l} \rho(t) = \psi(t) \pm \chi(t) \\ \omega(t) = \psi(t) \chi(t), \\ \rho(A) = \psi(A) \pm \chi(A) \\ \omega(A) = \psi(A) \chi(A). \end{array} \right\} \quad (25)$$

то

Это следует из перестановочности степеней матрицы.

8. Характеристический полином. Соотношение Кели—Гамильтона. Минимальный полином. Вековым, или характеристическим, уравнением матрицы $A = (a_{ij})$ называется уравнение

$$\left| \begin{array}{cccc} a_{11} - t & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - t & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - t \end{array} \right| = 0. \quad (26)$$

Левая часть этого уравнения, которую можно записать сокращенно в виде $|A - tE|$, носит название характеристического полинома матрицы. Вековые уравнения часто встречаются в приложениях.

Непосредственное вычисление характеристического полинома представляет значительные технические трудности. Действительно, если

$$\varphi(t) = |A - tE| = (-1)^n [t^n - p_1 t^{n-1} - p_2 t^{n-2} - \dots - p_n], \quad (27)$$

то

$$\left. \begin{array}{l} p_1 = a_{11} + a_{22} + \dots + a_{nn} \\ p_n = (-1)^{n-1} |A|. \end{array} \right\} \quad (28)$$

а остальные коэффициенты p_k суть взятые со знаком $(-1)^{k-1}$ суммы всех миноров определителя матрицы A порядка k , опирающихся на главную диагональ. Число таких миноров равно числу сочетаний из n по k .

Корни λ_i характеристического полинома называются собственными значениями или характеристическими числами матрицы. Из известной теоремы Виета, дающей связь корней уравнения с его коэффициентами, следует, что

$$\left. \begin{array}{l} \lambda_1 + \lambda_2 + \dots + \lambda_n = p_1 = a_{11} + a_{22} + \dots + a_{nn} \\ \lambda_1 \lambda_2 \dots \lambda_n = (-1)^{n-1} p_n = |A|. \end{array} \right\} \quad (29)$$

Величина $p_1 = a_{11} + \dots + a_{nn}$ называется следом матрицы A и обозначается $\text{Sp } A$.

Задача численного нахождения корней характеристического полинома является одной из важнейших задач линейной алгебры. Практически удобные методы определения коэффициентов и корней характеристического полинома будут разобраны в дальнейшем.

Старший коэффициент характеристического полинома равен $(-1)^n$. Иногда вместо характеристического полинома рассматривается нормированный характеристический полином, отличающийся от обычного множителем $(-1)^n$. Старший коэффициент нормированного характеристического полинома равен единице.

Для любой матрицы имеет место следующее замечательное соотношение, известное под названием соотношения Кели — Гамильтона: если $\varphi(t)$ есть характеристический полином матрицы A , то $\varphi(A) = 0$, т. е. говоря условно, матрица является корнем своего характеристического полинома.

Для доказательства рассмотрим матрицу B , союзную с матрицей $A - tE$. Так как каждое алгебраическое дополнение в определителе $|A - tE|$ является полиномом от t степени не выше $n-1$ -й, то союзную матрицу можно представить в виде

$$B = B_{n-1} + B_{n-2}t + \dots + B_0 t^{n-1},$$

где B_{n-1}, \dots, B_0 некоторые матрицы, не зависящие от t . В силу основного свойства союзной матрицы имеем:

$$(B_{n-1} + B_{n-2}t + \dots + B_0 t^{n-1})(A - tE) = |A - tE| E = \\ = (-1)^n (t^n - p_1 t^{n-1} - \dots - p_n) E.$$

Это равенство равносильно системе равенств

$$\begin{aligned} B_{n-1}A &= (-1)^{n+1} p_n E \\ B_{n-2}A - B_{n-1} &= (-1)^{n+1} p_{n-1} E \\ &\dots \\ B_0A - B_1 &= (-1)^{n+1} p_1 E \\ -B_0 &= (-1)^n E. \end{aligned}$$

Умножив эти равенства справа на E , A , A^2, \dots, A^{n-1} , A^n и сложив, получим в левой части нулевую матрицу, а в правой части

$$(-1)^n [-p_n E - p_{n-1} A - p_{n-2} A^2 - \dots + A^n] = \varphi(A). \quad (30)$$

Итак, $\varphi(A) = 0$, что и требовалось доказать.

Соотношение Кели — Гамильтона показывает, что для данной матрицы существует полином, корнем которого она является.

Очевидно, что такой полином не единственный, ибо если $\psi(t)$ обладает этим свойством, то им обладает и всякий полином, делящийся на $\psi(t)$. Полином наименьшей степени, обладающий тем свойством, что матрица A является его корнем, называется **минимальным полиномом матрицы**.

Докажем, что характеристический полином делится на минимальный полином.

Пусть $q(t)$ и $r(t)$ целая часть и остаток при делении характеристического полинома $\varphi(t)$ на минимальный полином $\psi(t)$. Тогда

$$\varphi(t) = \psi(t)q(t) + r(t),$$

причем степень $r(t)$ меньше степени $\psi(t)$.

Подставив в это равенство A вместо t , получим $r(A) = \varphi(A) - \psi(A)q(A) = 0$. Таким образом, матрица A оказывается „корнем“ полинома $r(t)$. Отсюда следует, что $r(t) = 0$, так как иначе $\varphi(t)$ не был бы минимальным полиномом. Следовательно, $\varphi(t)$ действительно делится на $\psi(t)$.

Точно так же доказывается, что любой полином $\omega(t)$, аннулирующий матрицу A , т. е. удовлетворяющий требованию $\omega(A) = 0$, делится на минимальный полином.

9. Подобные матрицы. Матрица B называется подобной матрице A , если существует такая неособенная матрица C , что $B = C^{-1}AC$. Говорят при этом, что матрица B получена из матрицы A преобразованием подобия.

Преобразование подобия обладает следующими свойствами:

$$C^{-1}A_1C + C^{-1}A_2C + \dots + C^{-1}A_nC = C^{-1}(A_1 + A_2 + \dots + A_n)C$$

$$C^{-1}A_1C \cdot C^{-1}A_2C \dots C^{-1}A_nC = C^{-1}(A_1A_2 \dots A_n)C. \quad (31)$$

В частности $(C^{-1}AC)^n = C^{-1}A^nC$. Далее

$$f(C^{-1}AC) = C^{-1}f(A)C$$

для любого полинома $f(t)$.

Из последнего свойства сразу вытекает, что **минимальные полиномы подобных матриц одинаковы**.

Покажем, что **подобные матрицы имеют также одинаковые характеристические полиномы**.

Действительно,

$$|B - tE| = |C^{-1}AC - tE| = |C^{-1}AC - tC^{-1}EC| =$$

$$= |C|^{-1}|A - tE||C| = |A - tE|.$$

10. Собственные значения полинома от матриц. Пусть A — матрица с собственными значениями $\lambda_1, \dots, \lambda_n$, среди которых могут быть равные, и пусть $f(t) = a_0 + a_1t + \dots + a_mt^m$ данный полином.

Покажем, что собственными значениями матрицы $f(A)$ будут числа $f(\lambda_1), \dots, f(\lambda_n)$.

Предварительно вычислим определитель матрицы $f(A)$. С этой целью разложим полином $f(t)$ на линейные множители

$$f(t) = a_m(t - \mu_1) \dots (t - \mu_m),$$

где μ_1, \dots, μ_m корни полинома $f(t)$.

Тогда

$$f(A) = a_m(A - \mu_1 E) \dots (A - \mu_m E)$$

и, следовательно,

$$|f(A)| = a_m^n |A - \mu_1 E| \dots |A - \mu_m E| = a_m^n \varphi(\mu_1) \dots \varphi(\mu_m),$$

где $\varphi(t)$ есть характеристический полином матрицы A . Но

$$\varphi(t) = (\lambda_1 - t) \dots (\lambda_n - t).$$

Поэтому

$$|f(A)| = a_m^n \prod_{i=1}^n \prod_{j=1}^m (\lambda_i - \mu_j) = \prod_{i=1}^n \left(a_m \prod_{j=1}^m (\lambda_i - \mu_j) \right) = \prod_{i=1}^n f(\lambda_i).$$

Равенство

$$|f(A)| = f(\lambda_1) \dots f(\lambda_n)$$

есть тождество относительно коэффициентов полинома $f(t)$. Применив это тождество к полиному $f(t) - u$, получим

$$|f(A) - uE| = (f(\lambda_1) - u) \dots (f(\lambda_n) - u).$$

Это и значит, что собственные значения матрицы $f(A)$ суть числа $f(\lambda_1), \dots, f(\lambda_n)$.

Отметим, в частности, что собственные значения матрицы A^k суть $\lambda_1^k, \dots, \lambda_n^k$.

11. Элементарные преобразования. Часто над матрицами нужно совершать следующие операции:

а) умножение элементов какой-либо строки на число;

б') добавление к элементам какой-либо строки чисел, пропорциональных элементам какой-либо предыдущей строки;

б'') добавление к элементам какой-либо строки чисел, пропорциональных элементам какой-либо последующей строки.

Иногда такие преобразования приходится делать над столбцами. Преобразования указанного вида будем называть элементарными преобразованиями матриц.

Нетрудно проверить, что всякое элементарное преобразование над строками равносильно умножению матрицы слева на некоторую

неособенную матрицу специального вида. Именно, операция а) равносильна умножению слева на матрицу

$$\begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \alpha \\ & & & & 1 \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix}. \quad (32)$$

операция б') равносильна умножению слева на матрицу

$$\begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \vdots \\ & & & & 1 \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix}. \quad (33)$$

операция б'') равносильна умножению слева на матрицу

$$\begin{bmatrix} 1 & & & \\ & & \ddots & \\ & & & 1 & \dots & \alpha \\ & & & & \ddots & \\ & & & & & 1 \\ & & & & & & \ddots \\ & & & & & & & 1 \end{bmatrix}. \quad (34)$$

Например,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ x & y & z \\ u & v & w \end{bmatrix} = \begin{bmatrix} a & b & c \\ \alpha x & \alpha y & \alpha z \\ u & v & w \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ x & y & z \\ u & v & w \end{bmatrix} = \begin{bmatrix} a & b & c \\ x & y & z \\ u + \alpha x & v + \alpha y & w + \alpha z \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ x & y & z \\ u & v & w \end{bmatrix} = \begin{bmatrix} a & b & c \\ x + zu & y + xv & z + aw \\ u & v & w \end{bmatrix}.$$

Операции а), б''), б') над столбцами осуществляются посредством умножения на те же матрицы справа.

В дальнейшем нам придется часто производить над матрицами преобразования вида а) и б').

Результат нескольких преобразований этого вида равносителен умножению матрицы слева на некоторую „левую треугольную“ матрицу, т. е. матрицу вида

$$\begin{bmatrix} \gamma_{11} & 0 & \dots & 0 \\ \gamma_{21} & \gamma_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} \end{bmatrix} \quad (35)$$

с ненулевыми диагональными элементами γ_{ii} .

Действительно, каждое отдельное преобразование вида а) и б') равносильно умножению слева на некоторую треугольную матрицу указанного вида, а произведение двух или нескольких треугольных матриц одинакового строения есть снова треугольная матрица.

Аналогично, результат нескольких преобразований вида а) и б'') равносителен умножению слева на „правую треугольную“ матрицу, т. е. матрицу вида

$$\begin{bmatrix} -b_{11} & b_{12} & \dots & b_{1n} \\ 0 & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & b_{nn} \end{bmatrix}. \quad (36)$$

Далее, результат нескольких преобразований вида а) и б') над столбцами равносителен умножению справа на правую треугольную

матрицу, а результат нескольких преобразований вида а) и б'') над столбцами равносителен умножению справа на левую треугольную матрицу.

Отметим еще, что результат нескольких преобразований вида б'') и б''), произведенных над столбцами, таких, что ко всем столбцам добавляются лишь элементы i -го столбца (который сам остается без изменения), равносителен умножению данной матрицы справа на матрицу вида:

$$\left[\begin{array}{cccccc|c} 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ m_{i1} & m_{i2} & \dots & m_{i, i-1} & 1 & m_{i, i+1} & \dots & m_{in} \\ \dots & \dots \\ 0 & 0 & \dots & \dots & \dots & \dots & \dots & 1 \end{array} \right]. \quad (37)$$

12. Разложение матрицы в произведение двух треугольных матриц. Треугольные матрицы обладают рядом удобных свойств. Так, определитель треугольной матрицы равен произведению элементов главной диагонали; произведение двух треугольных матриц одинакового строения есть снова треугольная матрица того же строения; неособенная треугольная матрица легко обращается, и ее обратная матрица имеет одинаковое с ней строение и т. д.

Поэтому интересна следующая теорема.

Теорема 1.1. Матрицу

$$A = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{array} \right]$$

можно представить в виде произведения левой и правой треугольных матриц, при условии, что

$$a_{11} \neq 0, \quad \left| \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right| \neq 0, \quad \dots, \quad |A| \neq 0.$$

Доказательство проведем методом математической индукции. Для $n=1$ утверждение очевидно: $(a_{11}) = (b_{11})(c_{11})$, причем один из множителей может быть взят произвольно. Пусть теорема верна для матриц $n-1$ -го порядка. Покажем, что она верна для матриц n -го порядка.

Разобьем матрицу A на клетки следующим образом:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} =$$

$$= \begin{bmatrix} & & a_{1n} & \\ & A_{n-1} & \vdots & \\ & & \vdots & \\ & a_{n-1, n} & & \end{bmatrix} = \begin{bmatrix} A_{n-1} & u \\ v & a_{nn} \end{bmatrix}$$

$$\begin{bmatrix} a_{n1} & a_{n2} & \dots & a_{n, n-1} & a_{nn} \end{bmatrix}$$

и будем искать разложение $A = CB$ матрицы A в произведение матриц B и C требуемого вида, разбив эти матрицы на клетки подобно A :

$$C = \begin{bmatrix} C_{n-1} & 0 \\ x & c_{nn} \end{bmatrix}, \quad B = \begin{bmatrix} B_{n-1} & y \\ 0 & b_{nn} \end{bmatrix}.$$

По правилу умножения для клеточных матриц:

$$CB = \begin{bmatrix} C_{n-1} & 0 \\ x & c_{nn} \end{bmatrix} \begin{bmatrix} B_{n-1} & y \\ 0 & b_{nn} \end{bmatrix} =$$

$$= \begin{bmatrix} C_{n-1}B_{n-1} & C_{n-1}y \\ xB_{n-1} & xy + c_{nn}b_{nn} \end{bmatrix} = A.$$

Поставленная цель будет достигнута, если мы определим матрицы C_{n-1} , B_{n-1} , x , y и числа c_{nn} , b_{nn} так, что

$$C_{n-1}B_{n-1} = A_{n-1}$$

$$C_{n-1}y = u$$

$$xB_{n-1} = v$$

$$xy + c_{nn}b_{nn} = a_{nn}.$$

Первому требованию можно удовлетворить в силу индукционного предположения. При этом из предположения, что $|A_{n-1}| \neq 0$, следует, что $|C_{n-1}| \neq 0$ и $|B_{n-1}| \neq 0$.

Далее, y и x однозначно определяются по формулам

$$y = C_{n-1}^{-1}u, \quad x = vB_{n-1}^{-1}.$$

Таким образом, нам остается определить только диагональные элементы c_{nn} и b_{nn} из равенства

$$c_{nn}b_{nn} = a_{nn} - xy.$$

Это, очевидно, всегда возможно, причем одному из чисел c_{nn} или b_{nn} можно приписать произвольное отличное от нуля значение, и тогда второе число определится однозначно. Тем самым теорема доказана.

Из хода доказательства следует, что разложение матрицы в произведение двух треугольных будет единственным, если мы заранее предпишем диагональным элементам одной из треугольных матриц фиксированные значения.

Удобно считать, например, что $b_{ii} = 1$, $i = 1, 2, \dots, n$. Если при этом из строк матрицы C вынести диагональные элементы, то мы придем к разложению

$$A = \tilde{C} \Lambda B,$$

где $\Lambda = [\alpha_1, \dots, \alpha_n]$, α_i — диагональные элементы матрицы C , и $\tilde{c}_{ii} = 1$.

Легко проверить, что

$$\alpha_i = \frac{|A_i|}{|A_{i-1}|}, \quad i = 1, \dots, n.$$

Действительно, из проведенной конструкции следует, что

$$|A_i| = |C_i| \cdot |B_i| = c_{ii} |C_{i-1}| b_{ii} |B_{i-1}| = \alpha_i |C_{i-1}| |B_{i-1}| = \alpha_i |A_{i-1}|.$$

Нетрудно дать явные формулы для разложения матрицы в произведение двух треугольных.

Обозначим через β_{ik} , $i \leq k$, минор матрицы A , составленный из первых i строк, первых $i-1$ столбцов и k -го столбца. Соответственно через γ_{ik} обозначим минор матрицы A , составленный из первых i столбцов, первых $i-1$ строк и k -й строки. В этих обозначениях $\beta_{kk} = \gamma_{kk} = |A_k|$.

Далее, обозначим через $\bar{\beta}_{ik}$, $i \leq k$, алгебраическое дополнение i -го элемента последней строки матрицы A_k , через $\bar{\gamma}_{ki}$ алгебраическое дополнение i -го элемента последнего столбца матрицы A_k . В этих обозначениях $\bar{\beta}_{kk} = \bar{\gamma}_{kk} = |A_{k-1}|$.

Построим следующие матрицы:

$$B_1 = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1n} \\ 0 & \beta_{22} & \dots & \beta_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \beta_{nn} \end{bmatrix}$$

$$B_2 = \begin{bmatrix} \bar{\beta}_{11} & \bar{\beta}_{12} & \dots & \bar{\beta}_{1n} \\ 0 & \bar{\beta}_{22} & \dots & \bar{\beta}_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \bar{\beta}_{nn} \end{bmatrix}$$

$$C_1 = \begin{bmatrix} \gamma_{11} & 0 & \dots & 0 \\ \gamma_{21} & \gamma_{22} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} \end{bmatrix}$$

$$C_2 = \begin{bmatrix} \bar{\gamma}_{11} & 0 & \dots & 0 \\ \bar{\gamma}_{21} & \bar{\gamma}_{22} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \bar{\gamma}_{n1} & \bar{\gamma}_{n2} & \dots & \bar{\gamma}_{nn} \end{bmatrix}.$$

Из элементарных свойств определителей следует, что

$$AB_2 = C_1; \quad C_2 A = B_1. \quad (38)$$

Далее, введем в рассмотрение диагональные матрицы

$$S_1 = [\beta_{11}, \beta_{22}, \dots, \beta_{nn}] = [\gamma_{11}, \gamma_{22}, \dots, \gamma_{nn}] = [|\mathcal{A}_1|, |\mathcal{A}_2|, \dots, |\mathcal{A}_n|]$$

и

$$S_2 = [\bar{\beta}_{11}, \bar{\beta}_{22}, \dots, \bar{\beta}_{nn}] = [\bar{\gamma}_{11}, \bar{\gamma}_{22}, \dots, \bar{\gamma}_{nn}] = [1, |\mathcal{A}_1|, \dots, |\mathcal{A}_{n-1}|].$$

Введем обозначения

$$\tilde{B} = S_1^{-1} B_1$$

$$\tilde{C} = C_1 S_1^{-1}$$

$$S = S_1 S_2^{-1} = S_2^{-1} S_1$$

$$B_0 = B_2 S_2^{-1}$$

$$C_0 = S_2^{-1} C_2.$$

Очевидно, матрицы \tilde{B} , \tilde{C} , B_0 и C_0 имеют единичные диагональные элементы.

Из равенств (38) следует

$$AB_0 = \tilde{C}S; \quad C_0 A = S\tilde{B},$$

откуда

$$A = \tilde{C}S B_0^{-1} = C_0^{-1} S \tilde{B}.$$

Матрицы \tilde{C} и C_0^{-1} являются левыми треугольными матрицами с единичными диагональными элементами, матрицы же $S\tilde{B}$ и $S B_0^{-1}$ быть правые треугольные матрицы. В силу однозначности разложения матрицы в произведение треугольных матриц с предписанными диагональными элементами для одного из сомножителей заключаем, что $\tilde{C} = C_0^{-1}$ и $S B_0^{-1} = S\tilde{B}$, откуда $\tilde{B} = B_0^{-1}$.

Таким образом,

$$A = \tilde{C}S\tilde{B}. \quad (39)$$

Ясно, что

$$\tilde{C} = \begin{bmatrix} 1 & & & \\ \frac{\gamma_{21}}{\gamma_{11}} & 1 & & 0 \\ \cdot & \cdot & \cdot & \\ \frac{\gamma_{n1}}{\gamma_{11}} & \frac{\gamma_{n2}}{\gamma_{22}} & \cdots & 1 \end{bmatrix}, \quad (40)$$

$$\tilde{B} = \begin{bmatrix} 1 & \frac{\beta_{12}}{\beta_{11}} & \cdots & \frac{\beta_{1n}}{\beta_{11}} \\ 0 & 1 & \cdots & \frac{\beta_{2n}}{\beta_{22}} \\ \cdot & \cdot & \cdot & \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

13. Разложение клеточной матрицы в произведении двух квазитреугольных. Клеточная матрица с квадратными диагональными клетками называется правой квазитреугольной, если она имеет вид

$$\begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ 0 & B_{22} & \cdots & B_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & B_{nn} \end{bmatrix},$$

и левой квазитреугольной, если она имеет вид

$$\begin{bmatrix} C_{11} & 0 & \cdots & 0 \\ C_{21} & C_{22} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}.$$

Очевидно, что определитель квазитреугольной матрицы равен произведению определителей ее диагональных клеток.

Имеет место следующая теорема.

Теорема 1.2 Если матрицы

$$A_1 = A_{11}, \quad A_2 = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad \dots, \quad A_{n-1} = \begin{bmatrix} A_{11} & \cdots & A_{1n-1} \\ \cdot & \cdots & \cdot \\ A_{n-11} & \cdots & A_{n-1n-1} \end{bmatrix}$$

неособенные, то матрица

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix}$$

может быть представлена в виде произведения левой и правой квазитреугольных матриц, причем диагональные клетки одной из них можно взять равными любым наперед заданным неособенным матрицам соответствующих порядков.

Доказательство проводится аналогично доказательству теоремы 1.1. Допустим, что для матрицы A_{n-1} уже получено искомое разложение $A_{n-1} = C_{n-1}B_{n-1}$. Так как $|A_{n-1}| \neq 0$, то $|C_{n-1}| \neq 0$ и $|B_{n-1}| \neq 0$. Далее положим

$$A = \begin{bmatrix} A_{n-1} & U \\ V & A_{nn} \end{bmatrix}$$

и будем искать квазитреугольные матрицы B и C в виде

$$B = \begin{bmatrix} B_{n-1} & Y \\ 0 & B_{nn} \end{bmatrix} \quad \text{и} \quad C = \begin{bmatrix} C_{n-1} & 0 \\ X & C_{nn} \end{bmatrix}.$$

Матричное равенство $CB = A$ распадается на равенства

$$\begin{aligned} C_{n-1}B_{n-1} &= A_{n-1} \\ C_{n-1}Y &= U \\ XB_{n-1} &= V \\ XY + C_{nn}B_{nn} &= A_{nn}. \end{aligned}$$

Первое из этих равенств выполняется автоматически, из второго и третьего находим $Y = C_{n-1}^{-1}U$, $X = VB_{n-1}^{-1}$, из последнего находим

$$B_{nn} = C_{nn}^{-1}(A_{nn} - XY) \quad (\text{или } C_{nn} = (A_{nn} - XY)B_{nn}^{-1}),$$

взяв в качестве C_{nn} (или B_{nn}) любую неособенную матрицу.

14. Матричная запись системы линейных уравнений. Рассмотрим систему n линейных уравнений с n неизвестными:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= f_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= f_2 \\ \vdots &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= f_n. \end{aligned} \tag{41}$$

Воспользовавшись матричными обозначениями, в частности правилом умножения матриц, мы можем записать систему (41) в виде одного матричного равенства

$$Ax = f. \tag{41'}$$

Здесь A обозначает матрицу из коэффициентов системы, f — столбец свободных членов, x — столбец, элементами которого являются значения неизвестных.

Если матрица A неособенная, мы сразу получаем решение системы умножением равенства (41') на A^{-1} слева. Именно,

$$x = A^{-1}f = \frac{1}{|A|} Bf, \quad (42)$$

где B матрица союзная с A .

Покажем, что последняя формула представляет собой матричную запись известных формул Крамера

$$x_i = \frac{|A_i|}{|A|}, \quad (43)$$

где A_i матрица, которая получается из матрицы A путем замены элементов i -го столбца через свободные члены.

Действительно, матричное равенство (42) равносильно n равенствам

$$x_i = \frac{A_{1i}f_1 + A_{2i}f_2 + \dots + A_{ni}f_n}{|A|} \quad i = 1, \dots, n.$$

Так как A_{ki} есть алгебраические дополнения элемента a_{ki} в определителе матрицы A , мы получаем, очевидно, что

$$A_{1i}f_1 + A_{2i}f_2 + \dots + A_{ni}f_n = |A_i|,$$

что доказывает наше утверждение.

15. Матричная запись квадратичной формы. Во многих разделах математики существенную роль играют квадратичные формы, т. е. однородные полиномы второй степени от нескольких переменных. Ясно, что квадратичная форма состоит из слагаемых двух видов, именно: квадратов переменных и попарных произведений переменных, взятых с некоторыми коэффициентами. Разобъем каждое слагаемое, содержащее попарное произведение переменных на две равные части, расположив перемножаемые переменные в обоих возможных порядках. При таком соглашении квадратичная форма будет записываться в виде следующей квадратной схемы:

Матрица

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

называется матрицей квадратичной формы. Из самого ее определения следует, что она симметрична, т. е. $a_{ij} = a_{ji}$. Таким образом, каждая

квадратичная форма естественно связывается с некоторой симметричной матрицей и обратно, каждой симметричной матрице может быть соотнесена некоторая квадратичная форма.

Квадратичная форма может быть коротко записана в матричных обозначениях. Действительно,

$$\begin{aligned}
 F(x_1, x_2, \dots, x_n) &= x_1(a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n) + \\
 &\quad + x_2(a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n) + \\
 &\quad \dots \dots \dots \dots \dots \dots \\
 &\quad + x_n(a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n) = \\
 &= (x_1, x_2, \dots, x_n) \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \dots \dots \dots \dots \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n \end{bmatrix} = \\
 &= (x_1, x_2, \dots, x_n) \begin{bmatrix} a_{11} & a_{12} \dots & a_{1n} \\ a_{21} & a_{22} \dots & a_{2n} \\ \dots & \dots \dots & \dots \\ a_{n1} & a_{n2} \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x'Ax, \quad (45) \\
 x' &= (x_1, x_2, \dots, x_n).
 \end{aligned}$$

Вещественная квадратичная форма называется положительно-определенной, если все ее значения при вещественных значениях переменных положительны, за исключением значения при $x_1 = x_2 = \dots = x_n = 0$.

Примером положительно-определенной формы может служить форма $x_1^2 + x_2^2 + \dots + x_n^2$.

Термин „положительно-определенный“ распространяется также на симметричные матрицы. Именно, вещественная симметричная матрица $A = (a_{ij})$ называется положительно-определенной, если квадратичная форма $F(x_1, x_2, \dots, x_n) = \sum_{i,j=1}^n a_{ij}x_i x_j$ положительно определена. Так, например, единичная матрица является положительно-определенной, ибо соответствующая ей квадратичная форма положительно определена. Очевидно далее, что квазидиагональная матрица, составленная из положительно-определенных ящиков, положительно определена.

Понятие положительной определенности может быть распространено на комплексные матрицы специального вида — так называемые Эрмитовы матрицы, которые связываются с формами Эрмита. Этим формам посвящен п. 6 § 12.

Выясним, как изменяются коэффициенты квадратичной формы при линейном преобразовании переменных. Пусть

$$\begin{aligned}x_1 &= b_{11}y_1 + b_{12}y_2 + \dots + b_{1n}y_n \\x_2 &= b_{21}y_1 + b_{22}y_2 + \dots + b_{2n}y_n \\&\vdots \\x_n &= b_{n1}y_1 + b_{n2}y_2 + \dots + b_{nn}y_n\end{aligned}$$

или, в матричной записи, $x = By$. Тогда

$$x'Ax = (By)'A(By) = y'B'ABy = y'Cy,$$

где $C = B'AB$. Легко видеть, что матрица C симметрична. Действительно, $C' = (B'AB)' = B'A'(B')' = B'AB = C$, ибо $A' = A$, $(B')' = B$.

Таким образом, при линейном преобразовании переменных с матрицей B квадратичная форма превращается в квадратичную же форму, причем ее матрица коэффициентов заменяется на матрицу $B'AB$.

Отметим, что при преобразовании переменных с неособенной матрицей B положительно-определенная квадратичная форма остается положительно-определенной. Действительно, если предположить, что при некоторой системе значений y_0 преобразованная квадратичная форма примет отрицательное или нулевое значение, то исходная квадратичная форма примет то же значение при системе значений переменных $x_0 = By_0$, что возможно только при $x_0 = 0$, а следовательно, и $y_0 = 0$.

16. Трансформации Гаусса. Решение системы линейных уравнений

$$Ax = f$$

с невырожденной матрицей A всегда может быть приведено к решению системы с симметричной и даже положительно-определенной матрицей. Это сведение основывается на следующей теореме.

Теорема 1.3. Если A невырожденная матрица, то матрицы $A'A$ и AA' положительно определены.

Доказательство. Если в квадратичной форме с единичной матрицей E сделать замену переменных с матрицей A (соответственно с матрицей A'), то получится квадратичная форма с матрицей $A'EA = A'A$ (соответственно $AEA' = AA'$). Так как E положительно-определенна, положительно-определенными будут и матрицы $A'A$ и AA' .

Если систему уравнений

$$Ax = f$$

умножить слева на матрицу A' , мы получим равносильную систему

$$A'Ax = A'f.$$

с положительно-определенной матрицей $A'A$. Такое преобразование будем называть первой (левой) трансформацией Гаусса.

Вторая (правая) трансформация Гаусса заключается в том, что вместо системы

$$Ax = f$$

рассматривается вспомогательная система

$$AA'y = f$$

с положительно-определенной матрицей AA' . Найдя решение у вспомогательной системы, мы найдем решение исходной системы по формуле

$$x = A'y.$$

§ 2. Матрицы специального вида

1. Симметричные матрицы. Симметричные матрицы обладают рядом замечательных свойств, о которых речь будет впереди. Пока мы лишь отметим, что произведение двух симметричных матриц не всегда есть матрица симметричная. Точнее, произведение симметричных матриц есть симметричная матрица в том и только в том случае, когда они перестановочны. Действительно, $(AB)' = B'A' = BA$, так что $AB = (AB)'$, только если $BA = AB$.

2. Ортогональные матрицы. Вещественная матрица называется ортогональной, если сумма квадратов элементов каждого столбца равна единице и суммы произведений соответствующих элементов из двух различных столбцов равны нулю. Ортогональность матрицы может быть охарактеризована одним матричным равенством, именно

$$A'A = E. \quad (1)$$

Действительно, диагональные элементы матрицы $A'A$ являются суммами квадратов элементов столбцов матрицы A , а недиагональные элементы равны суммам произведений соответствующих элементов из двух различных столбцов. Ортогональные матрицы обладают следующими свойствами.

1. Единичная матрица ортогональна.

2. Если A ортогональна, то $A^{-1} = A'$. Это следует из равенства $A'A = E$.

3. Если A ортогональна, то A' тоже ортогональна. Другими словами, из выполнения условий ортогональности для столбцов матрицы A следует выполнение тех же условий для строк матрицы A . В самом деле,

$$(A')' A' = AA' = AA^{-1} = E.$$

4. Произведение двух ортогональных матриц есть ортогональная матрица. Действительно, если A и B ортогональны, то

$$(AB)' AB = B'A'AB = B'E B = E.$$

5. Определитель ортогональной матрицы равен ± 1 . Действительно, из $A'A = E$ следует, что

$$|A'A| = |A'| \cdot |A| = |A|^2 = 1.$$

Последнее обстоятельство определяет естественное разбиение ортогональных матриц на два класса — собственно ортогональные с определителем $+1$ и несобственно ортогональные с определителем -1 .

К классу ортогональных матриц принадлежат элементарные матрицы вращения вида

$$T_{ij} = \begin{bmatrix} 1 & & & \\ & c \dots -s & & \\ & \vdots & \ddots & \\ & s \dots & c & \\ & & & 1 \end{bmatrix}. \quad (2)$$

где $c^2 + s^2 = 1$. Последнее соотношение показывает, что существует такой угол φ , что $c = \cos \varphi$, $s = \sin \varphi$.

Элементарные матрицы вращения отличаются от единичной матрицы лишь четырьмя элементами, находящимися на пересечении двух строк и двух столбцов с номерами i и j , $i < j$. Эти четыре элемента составляют матрицу $\begin{bmatrix} c, -s \\ s, c \end{bmatrix} = \begin{bmatrix} \cos \varphi, -\sin \varphi \\ \sin \varphi, \cos \varphi \end{bmatrix}$, совпадающую с матрицей преобразования декартовых координат на плоскости при повороте осей на угол φ .

В дальнейшем элементарные матрицы вращений будут неоднократно использоваться как вспомогательные для преобразования данной матрицы A посредством цепочки умножений слева или справа, или с обоих сторон на эти матрицы.

Очевидно, что при левом умножении матрицы $A = (a_{\alpha\beta})$ на матрицу T_{ij} изменяются лишь i -я и j -я строки матрицы A , именно, для матрицы $A^{(1)} = T_{ij}A$ будем иметь

$$\begin{aligned} a_{i\beta}^{(1)} &= ca_{i\beta} - sa_{j\beta} \\ a_{j\beta}^{(1)} &= sa_{i\beta} + ca_{j\beta} \end{aligned} \quad (\beta = 1, 2, \dots, n). \quad (3)$$

Соответственно, при умножении матрицы $A = (a_{ij})$ справа на матрицу T_{ij} изменяются только i -й и j -й столбцы по формулам

$$\begin{aligned} a_{\alpha i}^{(1)} &= ca_{\alpha i} + sa_{\alpha j} \\ a_{\alpha j}^{(1)} &= -sa_{\alpha i} + ca_{\alpha j} \end{aligned} \quad (\alpha = 1, 2, \dots, n). \quad (4)$$

Ясно, что если хотя бы один из двух элементов $a_{i\beta}$ и $a_{j\beta}$ отличен от нуля, то можно подобрать c и s так, чтобы для матрицы $A^{(1)} = T_{ij}A$ элемент $a_{j\beta}^{(1)}$ оказался равным нулю. Для этого нужно взять

$$s = -\frac{a_{j\beta}}{\sqrt{a_{i\beta}^2 + a_{j\beta}^2}}, \quad c := \frac{a_{i\beta}}{\sqrt{a_{i\beta}^2 + a_{j\beta}^2}}. \quad (5)$$

При таком выборе s и c получим

$$a_{j\beta}^{(1)} = \sqrt{a_{i\beta}^2 + a_{j\beta}^2} > 0, \quad a_{i\beta}^{(1)} = 0.$$

Теорема 2.1. Любая вещественная невырожденная матрица посредством цепочки умножений слева на элементарные матрицы вращений может быть преобразована в правую треугольную матрицу, все диагональные элементы которой, кроме, может быть, последнего, положительны.

Доказательство. Пусть

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

невырожденная вещественная матрица.

Допустим сначала, что $a_{11} \neq 0$. Умножим матрицу A слева по-очередно на матрицы $T_{12}, T_{13}, \dots, T_{1n}$, выбирая их так, чтобы последовательно аннулировать все элементы первого столбца, кроме верхнего. Этот же последний сделаем на первом шагу положительным и далее будем сохранять его положительность.

Если же $a_{11} = 0$, начнем преобразования с умножения на T_{1j_0} , где j_0 — наименьший номер, при котором $a_{j_0 1} \neq 0$. В силу предположения о невырожденности матрицы A хотя бы один элемент первого столбца отличен от нуля, так что такой номер j_0 найдется.

После описанных преобразований мы придем к матрице

$$A^{(1)} = T_{1n} T_{1n-1} \dots T_{12} A = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(1)} & \dots & a_{2n}^{(1)} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & a_{n2}^{(1)} & \dots & a_{nn}^{(1)} \end{bmatrix},$$

причем $a_{11}^{(1)} > 0$.

В силу невырожденности матрицы $A^{(1)}$ хотя бы один из элементов $a_{22}^{(1)}, \dots, a_{2n}^{(1)}$ отличен от нуля. Теперь подбираем элементарные матрицы вращений $T_{23}, T_{24}, \dots, T_{2n}$ так, чтобы после цепочки умножений на эти матрицы все элементы второго столбца ниже диагонали обратились в нуль, а диагональный стал положительным. Затем переходим к аннулированию поддиагональных элементов третьего столбца и т. д. В конце процесса придем к матрице

$$A^{(n-1)} = T_{n-1n} \cdots T_{12} A = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn}^{(n-1)} \end{bmatrix},$$

в которой все диагональные элементы, кроме, может быть, последнего $a_{nn}^{(n-1)}$, положительны. Знак $a_{nn}^{(n-1)}$ совпадает, очевидно, со знаком определителя матрицы A . Теорема доказана.

Отметим, что общее число матричных умножений для получения матрицы $A^{(n-1)}$ не превосходит числа поддиагональных элементов, равного $\frac{n(n-1)}{2}$.

Следствие 1. Всякая невырожденная вещественная матрица есть произведение собственно ортогональной на правую треугольную.

Действительно, $A = PA^{(n-1)}$, где $P = (T_{n-1n} \cdots T_{12})^{-1}$ есть собственно ортогональная матрица.

Следствие 2. Всякая собственно ортогональная матрица есть произведение элементарных матриц вращения.

Действительно, пусть A собственно ортогональная матрица. Тогда

$$A = T_{12}^{-1} \cdots T_{n-1n}^{-1} A^{(n-1)},$$

причем все диагональные элементы матрицы $A^{(n-1)}$ положительны. Легко видеть, что матрица $A^{(n-1)}$ единичная. Действительно, сумма квадратов элементов первого столбца равна единице, откуда следует, что $a_{11}^{(1)} = 1$. Из того, что сумма произведений элементов первого столбца на элементы каждого другого столбца равна нулю, заключаем, что $a_{12}^{(1)} = \dots = a_{1n}^{(1)} = 0$. Далее, сумма квадратов элементов второго столбца равна единице. Следовательно, $a_{22}^{(2)} = 1$ и т. д. Так последовательно заключаем, что все недиагональные элементы матрицы $A^{(n-1)}$ равны нулю, а все диагональные элементы равны единице.

Симметричные матрицы, так же как и ортогональные матрицы, входят в более общий класс так называемых нормальных матриц. Вещественная матрица называется нормальной, если она перестановочна со своей транспонированной, т. е. если

$$A' A = AA'.$$

3. Эрмитовы матрицы. Матрица с комплексными элементами называется эрмитовой, если

$$a_{ij} = \bar{a}_{ji}, \quad (6)$$

или в сокращенной записи

$$A = A^*.$$

Из определения следует, что диагональные элементы эрмитовой матрицы вещественны. Вещественные симметричные матрицы являются частным случаем эрмитовых. Многие свойства вещественных симметричных матриц сохраняются почти без изменений для матриц эрмитовых. В частности, произведение эрмитовых матриц будет эрмитовой матрицей тогда и только тогда, когда они перестановочны.

Далее, для любой матрицы с комплексными элементами матрица A^*A будет эрмитовой.

4. Унитарные матрицы. Матрица с комплексными элементами называется унитарной, если суммы квадратов модулей элементов столбцов равны единице и суммы произведений элементов одного столбца на числа комплексно-сопряженные к элементам другого столбца равны нулю. Унитарные матрицы могут быть охарактеризованы матричным равенством

$$A^*A = E. \quad (7)$$

Ортогональные матрицы являются, очевидно, частным случаем унитарных.

Свойства 1—4 ортогональных матриц сохраняются и для матриц унитарных. Определитель унитарной матрицы есть комплексное число, по модулю равное единице. Как эрмитовы, так и унитарные матрицы входят в более общий класс комплексных нормальных матриц, характеризуемых тем, что они перестановочны со своей сопряженной матрицей.

К унитарным матрицам относятся элементарные унитарные матрицы вида

$$\begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & ce^{i\varphi_1} & \dots & -se^{i\varphi_p} \\ & & \vdots & \ddots & \vdots \\ & & se^{i\varphi_p} & \dots & ce^{i\varphi_1} \\ & & & & 1 \end{bmatrix} \quad (8)$$

при $c > 0$, $s > 0$, $c^2 + s^2 = 1$, $\varphi_1 - \varphi_2 = \varphi_3 - \varphi_4$. Определитель такой матрицы равен $e^{i(\varphi_1 + \varphi_4)}$. Он равен единице в том и только в том случае, когда $\varphi_4 = -\varphi_1$ и, следовательно, $\varphi_3 = -\varphi_2$ (с точностью до кратных 2π).

Если даны два комплексных не равных одновременно нулю числа a и b , то всегда можно подобрать такие c , s , φ_1 и φ_2 , что

$$\begin{aligned} ace^{i\varphi_1} - bse^{i\varphi_2} &> 0 \\ ase^{-i\varphi_2} + bce^{-i\varphi_1} &= 0. \end{aligned}$$

Для этого достаточно взять

$$\begin{aligned} c &= \frac{|b|}{\sqrt{|a|^2 + |b|^2}}; \\ s &= \frac{|a|}{\sqrt{|a|^2 + |b|^2}}; \quad \varphi_1 = -\arg a; \quad \varphi_2 = \pi - \arg b. \end{aligned} \quad (9)$$

Это замечание позволяет доказать теорему:

Теорема 2.2. Всякая невырожденная комплексная матрица A преобразуется посредством умножения слева на цепочку элементарных унитарных матриц с определителями, равными единице, в правую треугольную матрицу, все диагональные элементы которой, кроме, может быть, последнего, положительны.

Ясно, что аргумент этого последнего элемента совпадает с аргументом определителя матрицы A .

Из теоремы вытекают следующие следствия.

Следствие 1. Всякая невырожденная комплексная матрица представима в виде произведения унитарной матрицы с определителем, равным единице, на правую треугольную, все диагональные элементы которой, кроме, может быть, последнего, положительны.

Следствие 2. Всякая унитарная матрица с определителем, равным единице, есть произведение элементарных унитарных матриц с определителями, равными единице.

5. Трехдиагональные матрицы. Трехдиагональной матрицей называется матрица вида

$$\left[\begin{array}{cccccc} a_1 & b_1 & 0 & \dots & 0 & 0 \\ c_1 & a_2 & b_2 & \dots & 0 & 0 \\ 0 & c_2 & a_3 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a_{n-1} & b_{n-1} \\ 0 & 0 & 0 & \dots & c_{n-1} & a_n \end{array} \right]. \quad (10)$$

Вещественная трехдиагональная матрица называется якобиевой, если $b_i c_i > 0$ при $i = 1, 2, \dots, n-1$. Любая симметричная

трехдиагональная матрица с ненулевыми недиагональными элементами автоматически будет якобиевой.

Трехдиагональные матрицы замечательны тем, что их характеристические полиномы вычисляются по несложным рекуррентным формулам. Пусть $\varphi_k(t)$ есть нормированный характеристический полином укороченной матрицы, т. е. матрицы

$$A_k = \begin{bmatrix} a_1 & b_1 & & & \\ c_1 & a_2 & b_2 & & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & & c_{k-1} \quad a_k \end{bmatrix}.$$

Тогда

$$\begin{aligned} \varphi_k(t) &= (-1)^k |A_k - tE| = \\ &= (-1)^k \begin{vmatrix} a_1 - t & b_1 & & & \\ c_1 & a_2 - t & b_2 & & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & & a_{k-1} - t \quad b_{k-1} \\ & & & & c_{k-1} \quad a_k - t \end{vmatrix}. \end{aligned}$$

Раскладывая этот определитель по элементам последнего столбца, получим, при $k \geq 3$,

$$\begin{aligned} \varphi_k(t) &= -(a_k - t) \varphi_{k-1}(t) - (-1)^k b_{k-1} \begin{vmatrix} a_1 - t & b_1 & & & \\ & \ddots & & & \\ & & a_{k-2} - t & b_{k-2} & \\ 0 & & & & c_{k-1} \end{vmatrix} = \\ &= (t - a_k) \varphi_{k-1}(t) - b_{k-1} c_{k-1} \varphi_{k-2}(t). \end{aligned} \tag{11}$$

Последняя формула остается верной и для $k = 2$, если положить $\varphi_0 = 1$. Ясно, что $\varphi_1(t) = t - a_1$. Полагая $k = 2, 3, \dots, n$, мы последовательно определим полиномы $\varphi_2(t), \varphi_3(t), \dots, \varphi_n(t)$, где $\varphi_n(t)$ характеристический полином данной трехдиагональной матрицы.

Для матрицы Якоби последовательность полиномов $\varphi_0, \varphi_1(t), \dots, \varphi_n(t)$ является последовательностью Штурма¹⁾. Так как старшие коэффициенты всех этих полиномов положительны, то, согласно теореме Штурма, все корни этих полиномов вещественны, причем корни двух соседних полиномов разделяются. Итак, все собственные значения якобиевой матрицы вещественны и различны.

¹⁾ А. Г. Курош. Курс высшей алгебры, 1949, стр. 269.

Изложим один прием для решения системы линейных уравнений с трехдиагональной матрицей. Пусть

$$a_1x_1 + b_1x_2 = f_1$$

$$c_1x_1 + a_2x_2 + b_2x_3 = f_2$$

$$c_2x_2 + a_3x_3 + b_3x_4 = f_3$$

· · · · ·

$$c_{n-2}x_{n-2} + a_{n-1}x_{n-1} + b_{n-1}x_n = f_{n-1}$$

$$c_{n-1}x_{n-1} + a_nx_n = f_n$$

Допустим, что $b_i \neq 0$, $i = 1, 2, \dots, n-1$. Если это условие не выполнено, то из системы выделяется система с меньшим числом неизвестных.

Отбросим последнее уравнение и найдем два решения $x^{(0)}$ и $x^{(1)}$ оставшейся системы из $n-1$ уравнений, положив $x_1^{(0)} = 0$, $x_1^{(1)} = 1$. Для этого придется дважды решить систему с треугольной матрицей. Очевидно, что столбец $x^{(0)} + t(x^{(1)} - x^{(0)})$ при любом t будет давать решение срезанной системы. Найдем t так, чтобы удовлетворялось и отброшенное ранее последнее уравнение.

Для этого нужно решить уравнение

$$c_{n-1}x_{n-1}^{(0)} + c_{n-1}t(x_{n-1}^{(1)} - x_{n-1}^{(0)}) + a_nx_n^{(0)} + a_nt(x_n^{(1)} - x_n^{(0)}) = f_n,$$

откуда

$$t = \frac{r_n^{(0)}}{r_n^{(0)} - r_n^{(1)}},$$

где $r_n^{(0)}$ и $r_n^{(1)}$ невязки последнего уравнения при подстановке решений $x^{(0)}$ и $x^{(1)}$.

6. Почти треугольные матрицы. Матрица называется (правой) почти треугольной, если она имеет вид

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n-1} & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n-1} & a_{2n} \\ 0 & a_{32} & a_{33} & \dots & a_{3n-1} & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn-1} & a_{nn} \end{bmatrix}. \quad (12)$$

Определитель почти треугольной матрицы связан простым рекуррентным соотношением со своими главными минорами. Именно, если положить $\Delta_0 = 1$ и обозначить

$$\Delta_k = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{kk} \end{vmatrix} \quad (k = 1, 2, \dots, n),$$

то

$$\Delta_k = a_{kk} \Delta_{k-1} - a_{kk-1} a_{k-1k} \Delta_{k-2} + a_{kk-1} a_{k-1k-2} a_{k-2k} \Delta_{k-3} + \\ + \dots + (-1)^{k-1} a_{kk-1} a_{k-1k-2} \dots a_{21} a_{1k} \Delta_0. \quad (13)$$

В этом легко убедиться посредством разложения определителя Δ_k по элементам последнего столбца.

Применение этой формулы к вычислению определителя требует выполнения примерно $\frac{n^3}{6}$ умножений, что, как мы увидим далее, несколько меньше, чем число операций для вычислений определителя общего вида.

Формула (13) может быть применена к рекуррентному построению характеристического полинома почти треугольной матрицы, именно

$$\varphi_k(t) = (t - a_{kk}) \varphi_{k-1}(t) - a_{kk-1} a_{k-1k} \varphi_{k-2}(t) - \\ - a_{kk-1} a_{k-1k-2} a_{k-2k} \varphi_{k-3}(t) - \dots - a_{kk-1} a_{k-1k-2} \dots a_{21} a_{1k}. \quad (14)$$

Здесь

$$\varphi_k(t) = \begin{vmatrix} t - a_{11} & -a_{12} & \dots & -a_{1k} \\ -a_{21} & t - a_{22} & \dots & -a_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & t - a_{kk} \end{vmatrix} \quad (k = 1, 2, \dots, n).$$

§ 3. Аксиомы линейного пространства

Как известно, методы аналитической геометрии дают возможность сопоставлять одному из важнейших геометрических объектов — вектору пространства — тройку вещественных чисел — проекций вектора на выбранные оси координат. Такое сопоставление делает возможным, с одной стороны, исследовать свойства геометрических объектов средствами алгебры и, с другой стороны, интерпретировать геометрическими образами некоторые алгебраические задачи. Например, совместное решение системы трех линейных уравнений с тремя неизвестными интерпретируется как задача о пересечении трех плоскостей в пространстве.

Это обстоятельство делает разумным и целесообразным введение геометрической терминологии в линейную алгебру.

Ясно, что для возможности геометрической трактовки линейной алгебры понятие вектора должно быть надлежащим образом обобщено. Это обобщение осуществляется посредством введения так называемого линейного пространства. Мы введем это понятие аксиоматически.

Линейным пространством называется совокупность математических (или физических) объектов, для которых определены два действия — сложение и умножение на любые вещественные или любые

комплексные числа, причем эти действия удовлетворяют следующим требованиям (аксиомам):

- 1) $X + Y = Y + X$ (переместительный закон);
- 2) $(X + Y) + Z = X + (Y + Z)$ (сочетательный закон);
- 3) существует „ $\mathbf{0}$ “, т. е. такой элемент, что $X + \mathbf{0} = X$ при любом X ;
- 4) для любого X существует противоположный элемент „ $-X$ “, такой, что $X + (-X) = \mathbf{0}$;
- 5) $1 \cdot X = X$;
- 6) $(a + b)X = aX + bX$;
- 7) $a(X + Y) = aX + aY$;
- 8) $a(bX) = abX$.

Элементы линейного пространства называются векторами.

Из перечисленных аксиом легко выводятся единственность нулевого элемента, единственность противоположного элемента, равенства $0X = a\mathbf{0} = \mathbf{0}$, $(-X) = (-1)X$. Мы не будем останавливаться на доказательстве этих утверждений.

Пространство называется вещественным, если для его векторов определено умножение только на вещественные числа, и комплексным, если определено умножение на комплексные числа.

Пространство называется конечно-мерным, если выполнена следующая аксиома:

9) Существует конечное число векторов X_1, \dots, X_N таких, что любой вектор пространства представляется в виде

$$c_1X_1 + \dots + c_NX_N.$$

Размерностью конечно-мерного пространства называется наименьшее число векторов, удовлетворяющих требованию аксиомы 9. Если же аксиома 9 не выполнена, то пространство называется бесконечно-мерным. Изучение бесконечно-мерных пространств выходит за рамки линейной алгебры, и, при тех или других дополнительных ограничениях, бесконечно-мерные пространства исследуются в одной из важнейших математических дисциплин — в функциональном анализе.

Как в вещественном, так и в комплексном линейном пространстве, может быть введено понятие скалярного произведения следующим образом. Каждой паре векторов X, Y сопоставляется число (X, Y) (вещественное в вещественном пространстве, комплексное в комплексном), причем должны быть удовлетворены следующие аксиомы:

- 10) $(X, X) > 0$, если $X \neq \mathbf{0}$; $(X, X) = 0$, если $X = \mathbf{0}$;
- 11) $(X, Y) = (\overline{Y}, \overline{X})$;
- 12) $(X_1 + X_2, Y) = (X_1, Y) + (X_2, Y)$;
- 13) $(aX, Y) = a(X, Y)$.

Для вещественного пространства одиннадцатая аксиома выглядит проще, именно $(X, Y) = (Y, X)$.

Вещественное линейное пространство с так определенным скалярным произведением называется евклидовым пространством, комплексное — унитарным пространством.

Число $\sqrt{(\mathbf{X}, \mathbf{X})}$ называется длиной вектора и обозначается $|\mathbf{X}|$. Скалярное произведение двух векторов удовлетворяет следующему важному неравенству:

$$|(\mathbf{X}, \mathbf{Y})| \leq |\mathbf{X}| \cdot |\mathbf{Y}|, \quad (1)$$

которое называется неравенством Коши—Буняковского. Докажем его. При $\mathbf{X} = \mathbf{0}$ неравенство очевидно. Пусть $\mathbf{X} \neq \mathbf{0}$. Введем вектор $\mathbf{Z} = \mathbf{Y} - \alpha \mathbf{X}$, где $\alpha = \frac{(\mathbf{Y}, \mathbf{X})}{(\mathbf{X}, \mathbf{X})}$, и подсчитаем квадрат его длины, принимая во внимание, что

$$(\mathbf{Z}, \mathbf{X}) = (\mathbf{Y}, \mathbf{X}) - \alpha (\mathbf{X}, \mathbf{X}) = 0.$$

Имеем

$$\begin{aligned} |\mathbf{Z}|^2 &= (\mathbf{Z}, \mathbf{Z}) = (\mathbf{Z}, \mathbf{Y} - \alpha \mathbf{X}) = (\mathbf{Z}, \mathbf{Y}) - \alpha (\mathbf{Z}, \mathbf{X}) = \\ &= (\mathbf{Y}, \mathbf{Y}) - \frac{(\mathbf{Y}, \mathbf{X})(\mathbf{X}, \mathbf{Y})}{(\mathbf{X}, \mathbf{X})} = \frac{|\mathbf{Y}|^2 \cdot |\mathbf{X}|^2 - |(\mathbf{X}, \mathbf{Y})|^2}{|\mathbf{X}|^2}. \end{aligned}$$

Следовательно, $|\mathbf{X}|^2 \cdot |\mathbf{Y}|^2 - |(\mathbf{X}, \mathbf{Y})|^2 = |\mathbf{X}|^2 \cdot |\mathbf{Z}|^2 \geq 0$, откуда непосредственно следует неравенство (1).

Для пары ненулевых векторов \mathbf{X} и \mathbf{Y} в евклидовом пространстве естественным образом вводится понятие угла по формуле

$$\cos \varphi = \frac{(\mathbf{X}, \mathbf{Y})}{|\mathbf{X}| \cdot |\mathbf{Y}|}. \quad (2)$$

Это определение всегда имеет смысл, ибо число $\frac{(\mathbf{X}, \mathbf{Y})}{|\mathbf{X}| \cdot |\mathbf{Y}|}$ по абсолютной величине не превосходит единицы в силу неравенства Коши—Буняковского.

Важнейшим примером линейного пространства является так называемое арифметическое пространство. Векторами этого пространства являются упорядоченные совокупности из n вещественных (комплексных) чисел, которые называются компонентами. Два вектора арифметического пространства считаются равными в том и только в том случае, если равны их соответствующие компоненты. Действия сложения и умножения на вещественное (комплексное) число определяются покомпонентно. Тем самым эти действия для векторов арифметического пространства ничем не отличаются от тех же действий для строк (т. е. односторонних матриц). Следовательно, все формальные законы этих действий, установленные на стр. 9 для произвольных матриц, верны и для векторов, так что аксиомы 1—8 линейного пространства оказываются выполненными. Роль нулевого вектора играет вектор, все компоненты которого равны нулю. Противоположным вектором — \mathbf{X} для вектора \mathbf{X} является $(-1)\mathbf{X}$.

Выполнение аксиомы 9-й вытекает из того, что любой вектор арифметического пространства допускает представление в виде

$$\mathbf{X} = x_1 e_1 + \dots + x_n e_n,$$

где e_i вектор, i -я компонента которого равна 1, а все остальные равны нулю. Ниже будет установлено, что не существует системы векторов, удовлетворяющих требованию аксиомы 9-й и состоящей из меньшего чем n числа векторов. Тем самым арифметическое пространство, векторы которого составлены из n чисел, оказывается n -мерным линейным пространством.

Пространство, составленное из векторов с вещественными компонентами, является вещественным линейным пространством; пространство, составленное из векторов с комплексными компонентами, является комплексным пространством.

Скалярное произведение вводится по формуле

$$(\mathbf{X}, \mathbf{Y}) = x_1 \bar{y}_1 + \dots + x_n \bar{y}_n,$$

которая в вещественном пространстве принимает более простой вид

$$(\mathbf{X}, \mathbf{Y}) = x_1 y_1 + \dots + x_n y_n.$$

Легко проверяется, что скалярное произведение удовлетворяет аксиомам 10—13.

Длина вектора равна, очевидно, $\sqrt{|x_1|^2 + \dots + |x_n|^2}$.

Скалярное произведение векторов арифметического пространства может быть выражено и в терминах матриц. Именно,

$$(\mathbf{X}, \mathbf{Y}) = x_1 \bar{y}_1 + \dots + x_n \bar{y}_n =$$

$$= (\bar{y}_1, \dots, \bar{y}_n) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = (x_1, \dots, x_n) \begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_n \end{bmatrix},$$

где $(\bar{y}_1, \dots, \bar{y}_n)$ строка из чисел, комплексно-сопряженных с компонентами вектора \mathbf{Y} , а (x_1, \dots, x_n) строка, элементы которой равны компонентам вектора \mathbf{X} .

Другими примерами линейных конечно-мерных пространств являются: совокупность всех матриц данного строения, совокупность полиномов от одной переменной, степень каждого из которых не превосходит данного числа, совокупность решений линейного однородного дифференциального уравнения и т. д.

В некоторых из этих пространств естественным образом вводится скалярное умножение. Так, в пространстве полиномов ограниченной

степени с вещественными коэффициентами скалярное произведение полиномов $f_1(t)$ и $f_2(t)$ может быть определено как $\int_a^b f_1(t) f_2(t) dt$.

Легко видеть, что при этом все аксиомы 10—13 выполнены.

В линейных пространствах, возникающих в связи с изучением конкретных задач, скалярное произведение обычно вводится так, чтобы оно находилось в естественной связи со специфическими особенностями элементов изучаемого пространства.

В некоторых задачах, исследуемых методами линейной алгебры, вообще не возникает необходимости во введении скалярного произведения.

Линейное пространство, в котором не вводится действие скалярного умножения, называется афинным пространством, и, вообще, те свойства пространства, которые не связаны с понятием скалярного умножения, называются афинными свойствами.

С описания афинных свойств мы и начнем систематическое изложение. При этом мы, вообще говоря, не будем оговаривать, о каком пространстве идет речь — о вещественном или комплексном, ввиду полного параллелизма формулировок и доказательства результатов.

§ 4. Базис и координаты

1. Линейная зависимость. Вектор $Y = c_1 X_1 + c_2 X_2 + \dots + c_m X_m$ называется линейной комбинацией векторов X_1, X_2, \dots, X_m .

Легко видеть, что если векторы Y_1, Y_2, \dots, Y_k являются линейными комбинациями векторов X_1, X_2, \dots, X_m , то любая их линейная комбинация $\gamma_1 Y_1 + \gamma_2 Y_2 + \dots + \gamma_k Y_k$ является также линейной комбинацией векторов X_1, X_2, \dots, X_m .

Векторы X_1, X_2, \dots, X_m называются линейно-зависимыми, если существуют такие числа c_1, c_2, \dots, c_m , не равные нулю одновременно, что имеет место равенство

$$c_1 X_1 + c_2 X_2 + \dots + c_m X_m = 0. \quad (1)$$

Если же это равенство имеет место только тогда, когда все постоянные c_1, c_2, \dots, c_m равны нулю, то векторы X_1, X_2, \dots, X_m называются линейно-независимыми.

Если векторы X_1, X_2, \dots, X_m линейно-зависимы, то по крайней мере один из них является линейной комбинацией остальных. Действительно, если, например, $c_m \neq 0$, то из (1) находим:

$$X_m = -\frac{c_1}{c_m} X_1 - \frac{c_2}{c_m} X_2 - \dots - \frac{c_{m-1}}{c_m} X_{m-1}. \quad (2)$$

Теорема 4. I. Если векторы Y_1, Y_2, \dots, Y_k являются линейными комбинациями векторов X_1, X_2, \dots, X_m и $k > m$, то они линейно-зависимы.

Доказательство проведем методом математической индукции.

Для $m = 1$ теорема очевидна. Допустим, что теорема верна в предположении, что число комбинируемых векторов равно $m - 1$, и в этом предположении докажем ее для m комбинируемых векторов. Пусть

$$Y_1 = c_{11}X_1 + \dots + c_{1m}X_m$$

$$\dots \dots \dots \dots$$

$$Y_k = c_{k1}X_1 + \dots + c_{km}X_m.$$

Могут представиться два случая.

1. Все коэффициенты $c_{11}, c_{21}, \dots, c_{k1}$ равны нулю. Тогда Y_1, Y_2, \dots, Y_m фактически являются линейными комбинациями только $m - 1$ векторов X_2, \dots, X_m . В силу индукционного предположения Y_1, \dots, Y_k будут линейно-зависимы.

2. Хотя бы один коэффициент при X_1 отличен от нуля. Без нарушения общности можно считать, что $c_{11} \neq 0$.

Рассмотрим систему векторов

$$Y'_2 = Y_2 - \frac{c_{21}}{c_{11}} Y_1$$

$$\dots \dots \dots \dots$$

$$Y'_k = Y_k - \frac{c_{k1}}{c_{11}} Y_1.$$

Построенные векторы, очевидно, являются линейными комбинациями векторов X_2, \dots, X_m , и число их $k - 1$ больше $m - 1$. В силу индукционного предположения, они линейно-зависимы, т. е. найдутся числа $\gamma_2, \dots, \gamma_k$, не равные нулю одновременно, такие, что

$$\gamma_2 Y'_2 + \dots + \gamma_k Y'_k = 0.$$

Подставляя вместо Y'_2, \dots, Y'_k их выражение через Y_1, \dots, Y_k , получим

$$\gamma_1 Y_1 + \gamma_2 Y_2 + \dots + \gamma_k Y_k = 0,$$

где $\gamma_1 = -\frac{c_{21}}{c_{11}} \gamma_2 - \dots - \frac{c_{k1}}{c_{11}} \gamma_k$. Числа $\gamma_1, \gamma_2, \dots, \gamma_k$ не равны нулю одновременно и, следовательно, Y_1, Y_2, \dots, Y_k линейно-зависимы.

2. Базис пространства. Система линейно-независимых векторов называется базисом пространства, если любой вектор пространства является линейной комбинацией векторов этой системы.

Например, в арифметическом пространстве векторы e_1, \dots, e_n образуют базис. Действительно, они линейно-независимы, так как компонентами вектора $c_1 e_1 + \dots + c_n e_n$ являются c_1, \dots, c_n и потому из равенства $c_1 e_1 + \dots + c_n e_n = 0$ следует, что $c_1 =$

$= c_2 = \dots = c_n = 0$. Далее для любого вектора X с компонентами x_1, \dots, x_n имеем

$$X = x_1e_1 + \dots + x_ne_n.$$

Базис e_1, \dots, e_n будем называть естественным базисом арифметического пространства.

Установим теперь, что в общем линейном конечно-мерном пространстве всегда существует базис. Пусть n — размерность пространства. В силу определения размерности в пространстве существует система n векторов U_1, \dots, U_n таких, что все векторы пространства являются их линейными комбинациями и не существует системы из меньшего числа векторов, обладающих тем же свойством.

Покажем, что векторы U_1, \dots, U_n образуют базис пространства, для чего достаточно установить их линейную независимость. Но она почти очевидна. Действительно, если допустить, что векторы U_1, \dots, U_n линейно-зависимы, то хотя бы один из них, например U_n , был бы линейной комбинацией остальных U_1, \dots, U_{n-1} . Тогда все векторы пространства оказались бы линейными комбинациями векторов U_1, \dots, U_{n-1} , что противоречит тому, что n есть размерность пространства.

Как мы увидим ниже, базис пространства не единственен, и в выборе базиса имеется широкий произвол.

Теорема 4.2. В n -мерном пространстве не существует более чем n линейно-независимых векторов.

Доказательство. Действительно, в n -мерном пространстве существует базис, состоящий из n векторов U_1, \dots, U_n . Любая система из $m > n$ векторов будет состоять из линейно- зависимых векторов в силу теоремы 4.1.

Следствие. Число векторов базиса пространства не зависит от выбора базиса и совпадает с размерностью пространства.

Действительно, базис n -мерного пространства не может состоять меньше чем из n векторов по определению размерности и больше чем из n векторов в силу теоремы 4.2.

Из теоремы 4.2 следует, что размерность арифметического пространства, составленного из n -компонентных векторов, действительно равна n , ибо в этом пространстве существует естественный базис e_1, \dots, e_n , состоящий ровно из n векторов.

Произвол в выборе базиса выясняет следующая теорема.

Теорема 4.3. Любая система из n линейно-независимых векторов образует базис n -мерного пространства.

Доказательство. Пусть U_1, \dots, U_n система из n линейно-независимых векторов и X — любой вектор пространства. По теореме 4.1 векторы U_1, \dots, U_n, X линейно-зависимы, ибо каждый из них является линейной комбинацией базисных векторов, число которых равно n . Следовательно, найдутся не равные одновременно нулю числа c_0, c_1, \dots, c_n такие, что $c_0X + c_1U_1 + \dots + c_nU_n = 0$.

При этом $c_0 \neq 0$, ибо если $c_0 = 0$, то $c_1\mathbf{U}_1 + \dots + c_n\mathbf{U}_n = \mathbf{0}$, что противоречит линейной независимости векторов $\mathbf{U}_1, \dots, \mathbf{U}_n$. Следовательно,

$$\mathbf{X} = -\frac{c_1}{c_0}\mathbf{U}_1 - \dots - \frac{c_n}{c_0}\mathbf{U}_n.$$

Итак, мы доказали, что любой вектор пространства есть линейная комбинация векторов $\mathbf{U}_1, \dots, \mathbf{U}_n$, откуда следует, что векторы $\mathbf{U}_1, \dots, \mathbf{U}_n$ образуют базис.

Доказанная теорема позволяет указать следующую конструкцию для построения базиса. Возьмем вектор \mathbf{U}_1 произвольным, только отличным от нуля. Вектор \mathbf{U}_2 возьмем произвольно, но неравным линейной комбинации вектора \mathbf{U}_1 (такой вектор найдется, если $n > 1$). Далее, за вектор \mathbf{U}_3 примем произвольный вектор, не являющийся линейной комбинацией первых двух и т. д. В силу определения размерности эта конструкция позволит нам построить систему из n векторов $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_n$, которые будут линейно-независимыми в силу самой конструкции. Из описанной конструкции также следует, что любую систему линейно-независимых векторов можно дополнить до базиса пространства.

8. Координаты векторов. Пусть $\mathbf{U}_1, \dots, \mathbf{U}_n$ какой-либо базис пространства. Тогда каждый вектор \mathbf{X} является линейной комбинацией векторов $\mathbf{U}_1, \dots, \mathbf{U}_n$, именно

$$\mathbf{X} = x_1\mathbf{U}_1 + \dots + x_n\mathbf{U}_n. \quad (3)$$

Коэффициенты в этом разложении однозначно определяются вектором \mathbf{X} , ибо если

$$\mathbf{X} = x_1\mathbf{U}_1 + \dots + x_n\mathbf{U}_n = x'_1\mathbf{U}_1 + \dots + x'_n\mathbf{U}_n,$$

то

$$(x_1 - x'_1)\mathbf{U}_1 + \dots + (x_n - x'_n)\mathbf{U}_n = \mathbf{0},$$

и, следовательно,

$$x_1 - x'_1 = \dots = x_n - x'_n = 0$$

в силу линейной независимости векторов $\mathbf{U}_1, \dots, \mathbf{U}_n$.

Коэффициенты x_1, \dots, x_n называются координатами вектора \mathbf{X} в базисе $\mathbf{U}_1, \dots, \mathbf{U}_n$.

В арифметическом пространстве компоненты вектора x_1, \dots, x_n являются, очевидно, его координатами в естественном базисе.

Введение координат дает возможность каждому вектору \mathbf{X} общего линейного n -мерного пространства сопоставить столбец $X = (x_1, \dots, x_n)'$ из его координат в выбранном базисе $\mathbf{U}_1, \dots, \mathbf{U}_n$.

При этом любой столбец $X = (x_1, \dots, x_n)'$ окажется сопоставлен некоторому вектору, именно вектору $\mathbf{X} = x_1\mathbf{U}_1 + \dots + x_n\mathbf{U}_n$. Если вектору \mathbf{X} сопоставлен столбец X , то вектору $c\mathbf{X}$ будет сопоставлен столбец cX , если векторам \mathbf{X} и \mathbf{Y} сопоставлены столбцы X и Y , то вектору $\mathbf{X} + \mathbf{Y}$ будет сопоставлен столбец $X + Y$. Построенное соответствие, очевидно, взаимно однозначное.

Таким образом, каждый выбор базиса определяет представление векторов n -мерного пространства в виде столбцов из их координат. Каждое такое представление взаимно однозначно. Произведение числа на вектор представляется произведением того же числа на столбец, представляющий вектор. Сумма векторов представляется суммой столбцов, представляющих слагаемые, иначе говоря, эти представления являются изоморфными — действиям над векторами соответствуют одноименные действия над представляющими их столбцами.

Для арифметического пространства в частности, естественный базис порождает представление векторов арифметического пространства в виде столбцов из их компонент. Такое представление мы будем называть естественным. Однако другие выборы базиса дают другие представления векторов арифметического пространства в виде столбцов.

Из приведенных рассуждений ясно, что общее линейное пространство размерности n изоморфно арифметическому пространству той же размерности.

В вычислительных задачах линейной алгебры подлежащие определению совокупности неизвестных и совокупности чисел, входящих в исходные данные, следует объединять в столбцы. Оказывается целесообразным рассматривать их как векторы арифметического пространства в их естественном представлении, а при решении задач переходить к другим представлениям, находящимся в той или другой связи со спецификой задачи.

4. Преобразование координат. Выясним, как изменятся координаты вектора при изменении базиса. Пусть U_1, U_2, \dots, U_n и U'_1, U'_2, \dots, U'_n два базиса и пусть

$$\begin{aligned} U'_1 &= a_{11} U_1 + a_{12} U_2 + \dots + a_{1n} U_n \\ U'_2 &= a_{21} U_1 + a_{22} U_2 + \dots + a_{2n} U_n \\ &\vdots \\ U'_n &= a_{n1} U_1 + a_{n2} U_2 + \dots + a_{nn} U_n. \end{aligned} \quad (4)$$

Связем с преобразованием координат матрицу, столбцы которой состоят из координат векторов U'_1, U'_2, \dots, U'_n в базисе U_1, U_2, \dots, U_n , т. е. матрицу

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}. \quad (5)$$

Матрица A неособенная, ибо она имеет обратную матрицу, посредством которой векторы U_1, U_2, \dots, U_n выражаются через векторы U'_1, U'_2, \dots, U'_n .

Обозначим через x_1, x_2, \dots, x_n координаты вектора \mathbf{X} в базисе $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_n$, через x'_1, x'_2, \dots, x'_n координаты в базисе $\mathbf{U}'_1, \mathbf{U}'_2, \dots, \mathbf{U}'_n$. Найдем зависимость между старыми и новыми координатами. Имеем

$$\begin{aligned}\mathbf{X} &= x_1 \mathbf{U}_1 + x_2 \mathbf{U}_2 + \dots + x_n \mathbf{U}_n = \\ &= x'_1 \mathbf{U}'_1 + x'_2 \mathbf{U}'_2 + \dots + x'_n \mathbf{U}'_n = \\ &= x'_1 (a_{11} \mathbf{U}_1 + a_{21} \mathbf{U}_2 + \dots + a_{n1} \mathbf{U}_n) + \\ &+ x'_2 (a_{12} \mathbf{U}_1 + a_{22} \mathbf{U}_2 + \dots + a_{n2} \mathbf{U}_n) + \\ &+ \dots \dots \dots \dots \dots \dots + \\ &+ x'_n (a_{1n} \mathbf{U}_1 + a_{2n} \mathbf{U}_2 + \dots + a_{nn} \mathbf{U}_n) = \\ &= (a_{11} x'_1 + a_{12} x'_2 + \dots + a_{1n} x'_n) \mathbf{U}_1 + \\ &+ (a_{21} x'_1 + a_{22} x'_2 + \dots + a_{2n} x'_n) \mathbf{U}_2 + \\ &+ \dots \dots \dots \dots \dots \dots + \\ &+ (a_{n1} x'_1 + a_{n2} x'_2 + \dots + a_{nn} x'_n) \mathbf{U}_n.\end{aligned}$$

Отсюда, в силу линейной независимости векторов $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_n$,

$$\begin{aligned}x_1 &= a_{11} x'_1 + a_{12} x'_2 + \dots + a_{1n} x'_n \\ x_2 &= a_{21} x'_1 + a_{22} x'_2 + \dots + a_{2n} x'_n \\ &\dots \dots \dots \dots \dots \dots \\ x_n &= a_{n1} x'_1 + a_{n2} x'_2 + \dots + a_{nn} x'_n.\end{aligned}\quad (6)$$

Последние равенства можно записать в матричной форме

$$\mathbf{X} = A \mathbf{X}',$$

где

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{и} \quad \mathbf{X}' = \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix}$$

суть столбцы, составленные из координат вектора \mathbf{X} в базисах $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_n$ и $\mathbf{U}'_1, \mathbf{U}'_2, \dots, \mathbf{U}'_n$.

§ 5. Подпространства

1. Определение подпространства, размерность, базис. Подпространством пространства \mathbf{R} называется множество векторов $\mathbf{X} \in \mathbf{R}$ такое, что любая линейная комбинация векторов множества снова

является вектором множества. Очевидно, что множество, состоящее только из нулевого вектора, так же как и все пространство, являются подпространствами в смысле этого определения. Мы будем называть их тривиальными подпространствами.

Ясно, что совокупность векторов, образующих подпространство, удовлетворяет аксиомам 1—8 линейного пространства, так что подпространство, рассматриваемое само по себе, является линейным пространством. Нетрудно убедиться в том, что подпространство n -мерного пространства будет конечно-мерным и его размерность не превосходит числа n . Действительно, в подпространстве (как и во всем пространстве) может существовать не более чем n линейно-независимых векторов.

В каждом подпространстве, очевидно, существует свой базис, причем число векторов базиса (размерность подпространства) не более чем n .

Если размерность подпространства равна n , то оно совпадает со всем пространством. В самом деле, базис U_1, \dots, U_n подпространства состоит из линейно-независимых векторов в количестве, равном размерности всего пространства и, следовательно, является базисом всего пространства.

Теорема 5.1. Любой базис U_1, \dots, U_m подпространства может быть дополнен до базиса всего пространства.

Доказательство. За первые m базисных векторов можно взять векторы U_1, \dots, U_m , ибо они линейно-независимы, а, как мы видели выше, любая система линейно-независимых векторов может быть дополнена до базиса всего пространства.

2. Подпространство, натянутое на данную систему векторов. Если дана совокупность векторов X_1, \dots, X_m линейно-независимых или даже линейно-зависимых, то, очевидно, множество всевозможных их линейных комбинаций образует подпространство. Так построенное подпространство называют подпространством, натянутым на систему векторов X_1, \dots, X_m .

Теорема 5.2. Размерность подпространства, натянутого на данную систему векторов X_1, \dots, X_m , равна рангу матрицы, составленной из координат этих векторов по отношению к любому базису.

Доказательство. Пусть

$$B = \begin{bmatrix} x_{11} & \dots & x_{1m} \\ x_{21} & \dots & x_{2m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nm} \end{bmatrix}$$

матрица, столбцами которой являются, соответственно, координаты данных векторов X_1, \dots, X_m относительно некоторого базиса. Пусть ранг этой матрицы равен r . Тогда, по определению ранга, существует отличный от нуля минор порядка r матрицы A , а все миноры

порядков $r+1$ и выше равны нулю или не могут быть составлены. Без нарушения общности (в случае необходимости можно изменить нумерацию данных векторов X_1, \dots, X_m и базисных векторов, что равносильно изменению нумерации столбцов и строк матрицы \mathbf{B}) можно принять, что отличным от нуля минором является

$$\delta = \begin{vmatrix} x_{11}, & \dots, & x_{1r} \\ \dots & \dots & \dots \\ x_{r1}, & \dots, & x_{rr} \end{vmatrix}.$$

Докажем, что векторы X_1, \dots, X_r образуют базис подпространства P , натянутого на векторы X_1, \dots, X_m . Установим, прежде всего, линейную независимость векторов X_1, \dots, X_r . Пусть

$$c_1X_1 + \dots + c_rX_r = 0.$$

Записывая это равенство в координатах, получим

$$\begin{aligned} c_1x_{11} + \dots + c_rx_{1r} &= 0 \\ \dots &\dots \dots \dots \\ c_1x_{r1} + \dots + c_rx_{rr} &= 0 \\ c_1x_{r+11} + \dots + c_rx_{r+1r} &= 0 \\ \dots &\dots \dots \dots \\ c_1x_{n1} + \dots + c_rx_{nr} &= 0. \end{aligned}$$

Первые r равенств представляют собой систему r линейных однородных уравнений относительно c_1, \dots, c_r , определитель из коэффициентов которой равен $\delta \neq 0$. Следовательно, эта система имеет единственное решение, именно $c_1 = \dots = c_r = 0$. Тем самым линейная независимость векторов X_1, \dots, X_r установлена.

Покажем теперь, что все данные векторы $X_1, \dots, X_r, X_{r+1}, \dots, X_m$ являются линейными комбинациями X_1, \dots, X_r . Для векторов X_1, \dots, X_r это тривиально, так что нужно рассмотреть только векторы X_s , $s = r+1, \dots, m$.

Рассмотрим определитель

$$\Delta_s = \begin{vmatrix} x_{11} & \dots & x_{1r} & x_{1s} \\ \dots & \dots & \dots & \dots \\ x_{r1} & \dots & x_{rr} & x_{rs} \\ z_1 & \dots & z_r & z \end{vmatrix},$$

где z_1, \dots, z_r, z некоторые числа, значения которых нам безразличны. Через M_1, M_2, \dots, M_r, M обозначим алгебраические дополнения элементов последней строки в определителе Δ_s .

Рассмотрим вектор $\mathbf{Y} = M_1 \mathbf{X}_1 + \dots + M_r \mathbf{X}_r + M \mathbf{X}_s$. Его координатами будут

$$\begin{aligned} y_1 &= M_1 x_{11} + \dots + M_r x_{1r} + M x_{1s} \\ &\dots \dots \dots \dots \dots \dots \dots \\ y_r &= M_1 x_{r1} + \dots + M_r x_{rr} + M x_{rs} \\ y_{r+1} &= M_1 x_{r+11} + \dots + M_r x_{r+1r} + M x_{r+1s} \\ &\dots \dots \dots \dots \dots \dots \dots \\ y_n &= M_1 x_{n1} + \dots + M_r x_{nr} + M x_{ns}. \end{aligned}$$

Первые r координат y_1, \dots, y_r равны нулю, так как они представляют собой суммы произведений алгебраических дополнений элементов последней строки определителя Δ_s на соответствующие элементы других строк. Остальные координаты y_{r+1}, \dots, y_n тоже равны нулю. Действительно,

$$\begin{aligned} y_{r+1} &= \begin{vmatrix} x_{11} & \dots & x_{1r} & x_{1s} \\ \dots & \dots & \dots & \dots \\ x_{r1} & \dots & x_{rr} & x_{rs} \\ x_{r+11} & \dots & x_{r+1r} & x_{r+1s} \\ \dots & \dots & \dots & \dots \end{vmatrix} \\ y_n &= \begin{vmatrix} x_{11} & \dots & x_{1r} & x_{1s} \\ \dots & \dots & \dots & \dots \\ x_{r1} & \dots & x_{rr} & x_{rs} \\ x_{n1} & \dots & x_{nr} & x_{ns} \end{vmatrix} \end{aligned}$$

и они равны нулю как миноры $r+1$ -го порядка, составленные из матрицы $\mathbf{\Xi}$ ранга r . Следовательно, $M_1 \mathbf{X}_1 + \dots + M_r \mathbf{X}_r + M \mathbf{X}_s = \mathbf{0}$. Так как $M = \delta \neq 0$, то

$$\mathbf{X}_s = -\frac{M_1}{M} \mathbf{X}_1 - \dots - \frac{M_r}{M} \mathbf{X}_r.$$

Итак, мы доказали, что векторы $\mathbf{X}_1, \dots, \mathbf{X}_r$ линейно-независимы и все векторы $\mathbf{X}_1, \dots, \mathbf{X}_m$ являются их линейными комбинациями. Следовательно, векторы $\mathbf{X}_1, \dots, \mathbf{X}_r$ образуют базис подпространства P , ибо любая линейная комбинация векторов $\mathbf{X}_1, \dots, \mathbf{X}_m$ является и линейной комбинацией векторов $\mathbf{X}_1, \dots, \mathbf{X}_r$. Тем самым доказано, что размерность подпространства равна r , что и требовалось доказать.

В терминах теории матриц теорема 5.2 может быть переформулирована следующим образом: *максимальное число линейно-независимых столбцов матрицы, так же как и максимальное число линейно-независимых строк, совпадает с рангом матрицы.*

3. Относительный базис. Пусть P есть подпространство размерности m в n -мерном пространстве R . Векторы V_1, \dots, V_k называются линейно-независимыми относительно P , если никакая их линейная комбинация, кроме нулевой, не принадлежит P , иными словами, если из $c_1V_1 + \dots + c_kV_k \in P$ следует, что $c_1 = \dots = c_k = 0$.

Система векторов V_1, \dots, V_k называется базисом R относительно P , если векторы V_1, \dots, V_k линейно-независимы относительно P и если всякий вектор из R представляется в виде суммы некоторого вектора из P и линейной комбинации векторов V_1, \dots, V_k .

Теорема 5.3. Пусть U_1, \dots, U_m базис подпространства P . Векторы V_1, \dots, V_k линейно-независимы относительно подпространства P в том и только в том случае, если векторы $U_1, \dots, U_m, V_1, \dots, V_k$ линейно-независимы.

Доказательство. Пусть V_1, \dots, V_k линейно-независимы относительно P и пусть $c_1V_1 + \dots + c_kV_k + d_1U_1 + \dots + d_mU_m = 0$. Тогда $c_1V_1 + \dots + c_kV_k = -d_1U_1 - \dots - d_mU_m \in P$. Следовательно, $c_1 = \dots = c_k = 0$ по определению линейной независимости относительно P . Поэтому $d_1U_1 + \dots + d_mU_m = 0$, откуда $d_1 = \dots = d_m = 0$. Таким образом, векторы $V_1, \dots, V_k, U_1, \dots, U_m$ линейно-независимы, так что необходимость сформулированного условия доказана.

Допустим теперь, что векторы $V_1, \dots, V_k, U_1, \dots, U_m$ линейно-независимы. Пусть $Y = c_1V_1 + \dots + c_kV_k \in P$. Тогда $Y = d_1U_1 + \dots + d_mU_m$, откуда $c_1V_1 + \dots + c_kV_k = -d_1U_1 - \dots - d_mU_m = 0$. В силу линейной независимости векторов $V_1, \dots, V_k, U_1, \dots, U_m$ все коэффициенты в последнем равенстве равны нулю. В частности, $c_1 = \dots = c_k = 0$, так что V_1, \dots, V_k линейно-независимы относительно P . Тем самым доказана и достаточность условия.

Теорема 5.4. Для того чтобы векторы V_1, \dots, V_k составляли базис пространства R относительно подпространства P , необходимо и достаточно, чтобы векторы $V_1, \dots, V_k, U_1, \dots, U_m$ составляли базис пространства R . Здесь векторы U_1, \dots, U_m составляют базис P .

Доказательство. Пусть V_1, \dots, V_k базис R относительно P . Тогда, в силу теоремы 5.3, векторы $V_1, \dots, V_k, U_1, \dots, U_m$ линейно-независимы. Далее, для любого вектора X из R имеем

$$X = c_1V_1 + \dots + c_kV_k + Y,$$

где $Y \in P$, и потому

$$X = c_1V_1 + \dots + c_kV_k + d_1U_1 + \dots + d_mU_m.$$

Необходимость доказана.

Пусть теперь векторы $V_1, \dots, V_k, U_1, \dots, U_m$ составляют базис R . Тогда, в силу теоремы 5.3, векторы V_1, \dots, V_k линейно-независимы

относительно P . Далее, для любого вектора $X \in R$ имеем

$$\begin{aligned} X = c_1 V_1 + \dots + c_k V_k + d_1 U_1 + \dots + d_m U_m = \\ = c_1 V_1 + \dots + c_k V_k + Y, \end{aligned}$$

где $Y \in P$. Достаточность доказана.

Следствие 1. Относительный базис всегда существует и число составляющих его векторов равно разности размерностей R и P .

Действительно, как мы видели, любой базис U_1, \dots, U_m подпространства P может быть дополнен до базиса пространства R . Совокупность дополнительных векторов V_1, \dots, V_k есть базис R относительно P , и их число равно разности размерностей R и P .

Следствие 2. Пусть $P_k \supset P_{k-1} \supset \dots \supset P_1$ убывающая цепочка подпространств. Тогда объединение базиса P_1 , базиса P_2 относительно P_1, \dots , базиса P_k относительно P_{k-1} образует базис P_k .

Теорема 5.5. Любая система V_1, \dots, V_s векторов, линейно-независимых относительно P , может быть дополнена до базиса R относительно P .

Доказательство. Векторы $U_1, \dots, U_m, V_1, \dots, V_s$ линейно-независимы в силу теоремы 5.3. Эта система может быть дополнена до системы $U_1, \dots, U_m, V_1, \dots, V_s, V_{s+1}, \dots, V_k$, образующей базис R . Тогда векторы V_1, \dots, V_k и составят базис R относительно P в силу теоремы 5.4.

4. Векторная сумма и пересечение подпространств. Пусть P и Q два подпространства пространства R . Векторной суммой подпространств P и Q называется совокупность всех векторов $Z = X + Y$, где $X \in P$, $Y \in Q$. Очевидно, что векторная сумма двух подпространств в свою очередь есть подпространство. Оно может быть охарактеризовано как наименьшее подпространство, содержащее подпространства P и Q . Будем обозначать векторную сумму подпространств P и Q через (P, Q) .

Пересечением подпространств P и Q называется совокупность всех векторов, принадлежащих как подпространству P , так и подпространству Q . Ясно, что пересечение двух подпространств есть в свою очередь подпространство. Оно может быть охарактеризовано как наибольшее подпространство, содержащееся в P и Q . Пересечение подпространств P и Q обозначается через $P \cap Q$.

Теорема 5.6. Пусть s — размерность (P, Q) , t — размерность $P \cap Q$. Тогда $s + t = p + q$, где p — размерность P , q — размерность Q .

Доказательство. Ясно, что $t \leq p \leq s$, $t \leq q \leq s$. Пусть U_1, \dots, U_t базис $P \cap Q$. Включим его в базисы $U_1, \dots, U_t, V_1, \dots, V_{p-t}$ и $U_1, \dots, U_t, W_1, \dots, W_{q-t}$ подпространств P и Q . Докажем, что векторы $U_1, \dots, U_t, V_1, \dots, V_{p-t}, W_1, \dots, W_{q-t}$ образуют базис

(P, Q). Установим сначала их линейную независимость. Пусть

$$c_1 \mathbf{U}_1 + \dots + c_t \mathbf{U}_t + d_1 \mathbf{V}_1 + \dots + d_{p-t} \mathbf{V}_{p-t} + \\ + d'_1 \mathbf{W}_1 + \dots + d'_{q-t} \mathbf{W}_{q-t} = \mathbf{0}. \quad (1)$$

Положим

$$\mathbf{Z} = c_1 \mathbf{U}_1 + \dots + c_t \mathbf{U}_t + d_1 \mathbf{V}_1 + \dots + d_{p-t} \mathbf{V}_{p-t}.$$

Ясно, что $\mathbf{Z} \in P$. С другой стороны, из равенства (1) заключаем, что

$$\mathbf{Z} = -d'_1 \mathbf{W}_1 - \dots - d'_{q-t} \mathbf{W}_{q-t},$$

откуда $\mathbf{Z} \in Q$. Следовательно, $\mathbf{Z} \in P \cap Q$, и потому

$$\mathbf{Z} = v_1 \mathbf{U}_1 + \dots + v_t \mathbf{U}_t$$

при некоторых v_1, \dots, v_t . Сравнивая второе и третье представления вектора \mathbf{Z} , получим, что

$$v_1 \mathbf{U}_1 + \dots + v_t \mathbf{U}_t + d'_1 \mathbf{W}_1 + \dots + d'_{q-t} \mathbf{W}_{q-t} = \mathbf{0},$$

откуда заключаем, что $v_1 = \dots = v_t = 0$, $d'_1 = \dots = d'_{q-t} = 0$ в силу линейной независимости векторов $\mathbf{U}_1, \dots, \mathbf{U}_t, \mathbf{W}_1, \dots, \mathbf{W}_{q-t}$. Тем самым равенство (1) превращается в равенство

$$c_1 \mathbf{U}_1 + \dots + c_t \mathbf{U}_t + d_1 \mathbf{V}_1 + \dots + d_{p-t} \mathbf{V}_{p-t} = \mathbf{0}.$$

откуда $c_1 = \dots = c_t = 0$, $d_1 = \dots = d_{p-t} = 0$.

Итак, все коэффициенты в равенстве (1) оказались нулями и, следовательно, векторы $\mathbf{U}_1, \dots, \mathbf{U}_t, \mathbf{V}_1, \dots, \mathbf{V}_{p-t}, \mathbf{W}_1, \dots, \mathbf{W}_{q-t}$ линейно-независимы. Остается доказать, что любой вектор из (P, Q) является их линейной комбинацией. Пусть $\mathbf{Z} \in (P, Q)$. Тогда $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$, где $\mathbf{X} \in P$, $\mathbf{Y} \in Q$. Представляя \mathbf{X} и \mathbf{Y} через базисные векторы подпространств P и Q , получим

$$\mathbf{X} = c_1 \mathbf{U}_1 + \dots + c_t \mathbf{U}_t + d_1 \mathbf{V}_1 + \dots + d_{p-t} \mathbf{V}_{p-t}$$

$$\mathbf{Y} = c'_1 \mathbf{U}_1 + \dots + c'_t \mathbf{U}_t + d'_1 \mathbf{W}_1 + \dots + d'_{q-t} \mathbf{W}_{q-t},$$

откуда

$$\mathbf{Z} = (c_1 + c'_1) \mathbf{U}_1 + \dots + (c_t + c'_t) \mathbf{U}_t + \\ + d_1 \mathbf{V}_1 + \dots + d_{p-t} \mathbf{V}_{p-t} + d'_1 \mathbf{W}_1 + \dots + d'_{q-t} \mathbf{W}_{q-t}.$$

Тем самым мы доказали, что векторы $\mathbf{U}_1, \dots, \mathbf{U}_t, \mathbf{V}_1, \dots, \mathbf{V}_{p-t}, \mathbf{W}_1, \dots, \mathbf{W}_{q-t}$ образуют базис подпространства (P, Q). Таким образом, размерность s подпространства (P, Q) равна $t + p - t + q - t = = p + q - t$, откуда следует, что

$$s + t = p + q.$$

5. Прямая сумма. Если любой вектор X пространства R представляется в виде суммы векторов Y_1, \dots, Y_k из подпространств P_1, \dots, P_k , то говорят, что R есть векторная сумма подпространств P_1, \dots, P_k . Если при этом представление

$$X = Y_1 + \dots + Y_k, \quad Y_i \in P_i \quad (i = 1, \dots, k)$$

однозначно, то R есть прямая сумма подпространств P_1, \dots, P_k .

Теорема 5.7. Для того чтобы пространство R было прямой суммой своих подпространств P_1, \dots, P_k , необходимо и достаточно, чтобы объединение базисов этих подпространств составляло базис всего пространства.

Доказательство. Пусть R есть прямая сумма подпространств P_1, \dots, P_k и пусть векторы $U_1, \dots, U_{s_1}; \dots; U_{s_{k-1}+1}, \dots, U_{s_k}$ составляют базисы этих подпространств. Тогда для любого вектора из R имеем

$$X = Y_1 + \dots + Y_k,$$

где $Y_i \in P_i$, и потому

$$X = c_1 U_1 + \dots + c_{s_1} U_{s_1} + \dots + c_{s_{k-1}+1} U_{s_{k-1}+1} + \dots + c_{s_k} U_{s_k}.$$

Остается доказать линейную независимость векторов U_1, \dots, U_{s_k} . Пусть

$$c_1 U_1 + \dots + c_{s_1} U_{s_1} + \dots + c_{s_{k-1}+1} U_{s_{k-1}+1} + \dots + c_{s_k} U_{s_k} = 0.$$

Введем обозначения:

$$\begin{aligned} c_1 U_1 + \dots + c_{s_1} U_{s_1} &= Y_1 \\ &\dots \\ c_{s_{k-1}+1} U_{s_{k-1}+1} + \dots + c_{s_k} U_{s_k} &= Y_k. \end{aligned}$$

Тогда $Y_i \in P_i$ и $0 = Y_1 + \dots + Y_k$. Но все подпространства P_i содержат нулевой вектор и $0 = 0 + \dots + 0$. В силу единственности разложения векторов из R по подпространствам P_1, \dots, P_k , мы заключаем, что $Y_1 = \dots = Y_k = 0$. Следовательно, и все коэффициенты $c_1, \dots, c_{s_1}, \dots, c_{s_{k-1}+1}, \dots, c_{s_k}$ равны нулю. Линейная независимость векторов U_1, \dots, U_{s_k} доказана. Тем самым доказана необходимость условия.

Предположим теперь, что векторы $U_1, \dots, U_{s_1}; \dots; U_{s_{k-1}+1}, \dots, U_{s_k}$, составляющие базисы подпространств P_1, \dots, P_k , образуют базис R . Тогда для любого вектора $X \in R$ имеем

$$\begin{aligned} X = c_1 U_1 + \dots + c_{s_1} U_{s_1} + \dots + c_{s_{k-1}+1} U_{s_{k-1}+1} + \dots + c_{s_k} U_{s_k} &= \\ &= Y_1 + \dots + Y_k, \end{aligned}$$

где

$$Y_1 = c_1 U_1 + \dots + c_{s_1} U_{s_1} \in P_1$$

$$\dots \dots \dots \dots \dots \dots \dots$$

$$Y_k = c_{s_{k-1}+1} U_{s_{k-1}+1} + \dots + c_{s_k} U_{s_k} \in P_k.$$

Это представление однозначно, ибо если

$$X = Y'_1 + \dots + Y'_k$$

при $Y'_i \in P_i$, то

$$X = c'_1 U_1 + \dots + c'_{s_1} U_{s_1} + \dots + c'_{s_{k-1}+1} U_{s_{k-1}+1} + \dots + c'_{s_k} U_{s_k},$$

где

$$Y'_1 = c'_1 U_1 + \dots + c'_{s_1} U_{s_1}$$

$$\dots \dots \dots \dots \dots \dots$$

$$Y'_k = c'_{s_{k-1}+1} U_{s_{k-1}+1} + \dots + c'_{s_k} U_{s_k}.$$

В силу единственности разложения вектора X по базисным векторам, заключаем, что $c_1 = c'_1, \dots, c_{s_1} = c'_{s_1}, \dots, c_{s_{k-1}+1} = c'_{s_{k-1}+1}, \dots, c_{s_k} = c'_{s_k}$ и потому $Y_1 = Y'_1, \dots, Y_k = Y'_k$. Теорема доказана и в части достаточности.

Теорема 5.8. Если пространство R есть векторная сумма подпространств P_1, \dots, P_k и размерность R равна сумме размерностей P_1, \dots, P_k , то R есть прямая сумма P_1, \dots, P_k .

Доказательство. Векторы пространства R являются линейными комбинациями базисных векторов всех подпространств P_1, \dots, P_k . Следовательно, размерность R не превосходит суммы размерностей подпространств P_1, \dots, P_k и равна этой сумме только если совокупность всех базисных векторов всех P_i линейно-независима. Но в этом случае, в силу теоремы 5.7, пространство R есть прямая сумма подпространств P_1, \dots, P_k .

Из последней теоремы следует, в частности, что векторная сумма двух подпространств будет прямой суммой в том и только в том случае, если пересечение этих подпространств имеет размерность 0, т. е. состоит только из нулевого вектора. Это последнее утверждение легко доказывается и непосредственно.

§ 6. Линейные операторы

1. Функция от векторного аргумента. Функцией от векторного аргумента с областью значений Ω называется закон сопоставления каждому вектору пространства R (или некоторого его подмножества) элемента из Ω .

Если областью значений Ω является совокупность чисел, то функция от векторного аргумента называется функционалом, если

областью значений Ω является совокупность векторов того же пространства, то функция от векторного аргумента называется преобразованием или оператором.

Примерами функционалов могут служить скалярное произведение (X, Y_0) при фиксированном векторе Y_0 , длина вектора X , квадратичная форма от координат вектора в некотором базисе. Вообще, если в пространстве зафиксирован базис, то функционалом будет любая функция от n переменных, именно от n координат переменного вектора в этом выбранном базисе. Очевидно, что при изменении базиса функция от координат вектора, задающая функционал, должна быть подвергнута соответствующему преобразованию переменных.

Функционал Φ называется линейным, если

$$\Phi(c_1X_1 + c_2X_2) = c_1\Phi(X_1) + c_2\Phi(X_2). \quad (1)$$

Ясно, что общий вид линейного функционала есть

$$\Phi(X) = \Phi(x_1U_1 + \dots + x_nU_n) = a_1x_1 + \dots + a_nx_n,$$

где a_1, \dots, a_n — числа, именно $a_1 = \Phi(U_1), \dots, a_n = \Phi(U_n)$; здесь U_1, \dots, U_n выбранный базис, x_1, x_2, \dots, x_n координаты вектора X в этом базисе. В дальнейшем важную роль будут играть квадратичные функционалы и некоторые другие.

2. Линейные операторы. Оператор называется линейным, если он удовлетворяет следующим условиям линейности.

1. $A(\alpha X) = \alpha AX$ при любом комплексном числе α .
2. $A(X_1 + X_2) = AX_1 + AX_2$.

Здесь через AX обозначен результат применения оператора A к вектору X .

Определим действия над линейными операторами. Произведением AB линейных операторов A и B назовем оператор C , состоящий в последовательном применении сначала линейного оператора B , а затем линейного оператора A .

Произведение линейных операторов A и B , как легко видеть, снова есть линейный оператор.

Действительно,

$$\begin{aligned} C(X_1 + X_2) &= A(B(X_1 + X_2)) = A(BX_1 + BX_2) = \\ &= A(BX_1) + A(BX_2) = CX_1 + CX_2. \end{aligned}$$

Умножение операторов ассоциативно, что непосредственно следует из определения.

Оператор E , сопоставляющий каждому вектору X этот же вектор, называется единичным оператором. Ясно, что единичный оператор линеен и $EA = AE = A$ при любом операторе A .

Суммой $A + B$ линейных операторов A и B назовем оператор C , сопоставляющий вектору X вектор $AX + BX$.

Оператор **0**, отображающий все векторы пространства на нулевой вектор, называется **нулевым оператором**. Ясно, что нулевой оператор линеен и $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$ при любом операторе \mathbf{A} .

Произведением $\alpha\mathbf{A}$ линейного оператора \mathbf{A} на число α назовем оператор, сопоставляющий вектору \mathbf{X} вектор $\alpha(\mathbf{AX})$.

Очевидно, что сумма линейных операторов, а также произведение оператора на число есть линейные операторы. Данные определения действий позволяют естественным образом определить степень оператора как произведение равных сомножителей и полином от оператора согласно формуле

$$f(\mathbf{A}) = a_0\mathbf{A}^n + a_1\mathbf{A}^{n-1} + \dots + a_{n-1}\mathbf{A} + a_n\mathbf{E},$$

где

$$f(t) = a_0t^n + a_1t^{n-1} + \dots + a_{n-1}t + a_n.$$

В дальнейшем мы будем иметь дело только с линейными операторами и потому слово линейный будем опускать.

3. Представление оператора матрицей. Выберем в пространстве \mathbb{R} некоторый базис $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_n$. Оператор \mathbf{A} соотносит векторам базиса векторы $\mathbf{AU}_1, \mathbf{AU}_2, \dots, \mathbf{AU}_n$.

Пусть векторы $\mathbf{AU}_1, \mathbf{AU}_2, \dots, \mathbf{AU}_n$ заданы своими координатами в базисе $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_n$, т. е. пусть

$$\begin{aligned} \mathbf{AU}_1 &= a_{11}\mathbf{U}_1 + a_{21}\mathbf{U}_2 + \dots + a_{n1}\mathbf{U}_n \\ \mathbf{AU}_2 &= a_{12}\mathbf{U}_1 + a_{22}\mathbf{U}_2 + \dots + a_{n2}\mathbf{U}_n \\ &\vdots \\ \mathbf{AU}_n &= a_{1n}\mathbf{U}_1 + a_{2n}\mathbf{U}_2 + \dots + a_{nn}\mathbf{U}_n. \end{aligned} \quad (2)$$

Рассмотрим матрицу A , столбцы которой состоят из координат векторов $\mathbf{AU}_1, \mathbf{AU}_2, \dots, \mathbf{AU}_n$:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}. \quad (3)$$

Покажем, что матрица A ¹⁾ вполне определяет оператор \mathbf{A} .

Действительно, если для оператора \mathbf{A} известна матрица A , то тем самым известны значения оператора \mathbf{A} на базисных векторах $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_n$ и, в силу линейности оператора, легко определить его значение для любого вектора. Именно, если $\mathbf{X} = x_1\mathbf{U}_1 + \dots + x_n\mathbf{U}_n$, то

$$\mathbf{AX} = x_1\mathbf{AU}_1 + \dots + x_n\mathbf{AU}_n.$$

1) Отметим, что матрица коэффициентов в соотношениях (2) образует матрицу, транспонированную к той, которую мы связываем с оператором.

Отсюда легко находятся координаты преобразованного вектора. Именно,

$$Y = AX = \sum_{k=1}^n y_k U_k = \sum_{i=1}^n x_i A U_i = \sum_{i=1}^n \sum_{k=1}^n a_{ki} x_i U_k,$$

откуда

$$y_k = \sum_{i=1}^n a_{ki} x_i$$

или, в матричной записи,

$$Y = AX \quad (4)$$

где Y и X суть столбцы из координат векторов Y и X .

Обратно, произвольная матрица A может быть связана с некоторым оператором.

Действительно, преобразование, задаваемое формулой

$$Y = AX,$$

где Y и X по-прежнему столбцы из координат векторов Y и X , линейно при любой матрице A .

Установленное одно-однозначное соответствие между операторами и матрицами сохраняется при действиях над операторами. Именно, матрица суммы операторов равна сумме матриц слагаемых, матрица произведения операторов равна произведению матриц, соответствующих сомножителям.

Короче, совокупность операторов n -мерного пространства изоморфна совокупности матриц порядка n , и такой изоморфизм осуществляется посредством сопоставления каждому оператору соответствующей ему матрицы в некотором фиксированном базисе пространства.

В тех рассуждениях, в которых базис пространства заранее фиксируется, имеет смысл отождествление оператора с соответствующей ему матрицей, подобно указанному выше отождествлению вектора и столбца из его координат. При таком отождествлении результат воздействия оператора на вектор совпадает с результатом умножения матрицы на столбец.

4. Связь между матрицами оператора в различных базисах. Выясним теперь, как изменится матрица оператора при изменении базиса пространства.

Положим, что от базиса U_1, U_2, \dots, U_n мы перешли к базису U'_1, U'_2, \dots, U'_n . Координаты любого вектора пространства при этом изменяются по формулам

$$X = CX',$$

где X столбец из координат вектора X в базисе U_1, \dots, U_n , X' — столбец из координат в базисе U'_1, \dots, U'_n , C — матрица преобразования координат.

Рассмотрим теперь оператор A , и пусть в базисе U_1, \dots, U_n ему соответствует матрица A , а в базисе U'_1, \dots, U'_n матрица B .

Пусть $Y = AX$, Y и Y' столбцы из координат вектора Y в базисах U_1, \dots, U_n и U'_1, \dots, U'_n соответственно.

Тогда

$$Y = AX$$

$$Y' = BX'.$$

Но $X = CX'$, $Y = CY'$ и потому

$$CY' = ACX',$$

откуда

$$Y' = C^{-1}ACX'$$

и, следовательно,

$$B = C^{-1}AC. \quad (5)$$

Таким образом, одному и тому же оператору в различных базисах соответствуют подобные матрицы. При этом матрица, посредством которой осуществляется преобразование подобия, совпадает с матрицей преобразования координат.

5. Ранг оператора. Совокупность AR векторов AX , где A данный оператор, а X вектор, пробегающий некоторое подпространство R n -мерного пространства R , образует подпространство. Действительно, если $Y_1 \in AR$, $Y_2 \in AR$, то $Y_1 = AX_1$, $Y_2 = AX_2$, где X_1 и X_2 некоторые векторы из R , и, следовательно, $c_1Y_1 + c_2Y_2 = A(c_1X_1 + c_2X_2) \in AR$, ибо $c_1X_1 + c_2X_2 \in R$.

В частности, AR является подпространством R . Это подпространство называется образом оператора A . Размерность этого подпространства называется рангом оператора A . Очевидно, что AR есть подпространство, натянутое на векторы AU_1, \dots, AU_n , где U_1, \dots, U_n базис R . Поэтому, согласно теореме 5.2, ранг оператора равен рангу матрицы, сопоставляемой оператору в базисе U_1, \dots, U_n .

Заметим, что, так как размерность подпространства AR не зависит от выбора базиса, ранги всех матриц, сопоставляемых оператору A в различных базисах, равны между собой. Следовательно, ранги подобных матриц равны.

Образ AR совпадает со всем пространством в том и только в том случае, когда ранг оператора A равен n , т. е. когда определитель его матрицы не равен нулю. В этом случае оператор называется невырожденным. Оператор, ранг которого меньше размерности пространства, называется вырожденным.

Совокупность Q векторов $Y \in R$, таких, что $AY = 0$, есть также подпространство. Действительно, если $Y_1 \in Q$ и $Y_2 \in Q$, то $AY_1 = AY_2 = 0$ и $A(c_1Y_1 + c_2Y_2) = c_1AY_1 + c_2AY_2 = 0$ и, следовательно, $c_1Y_1 + c_2Y_2 \in Q$. Подпространство Q называется ядром оператора A .

Теорема 6.1. Сумма размерностей ядра оператора и его образа равна размерности всего пространства.

Доказательство. Пусть U_1, \dots, U_m базис ядра \mathbf{Q} оператора \mathbf{A} . Дополним его до базиса пространства R векторами V_1, \dots, V_{n-m} . Покажем, что векторы AV_1, \dots, AV_{n-m} составляют базис образа AR оператора A . Докажем сначала линейную независимость этих векторов. Пусть

$$c_1AV_1 + \dots + c_{n-m}AV_{n-m} = \mathbf{0}.$$

Тогда $A(c_1V_1 + \dots + c_{n-m}V_{n-m}) = \mathbf{0}$, т. е. $c_1V_1 + \dots + c_{n-m}V_{n-m} \in Q$. Но это возможно только при $c_1 = \dots = c_{n-m} = 0$, ибо векторы V_1, \dots, V_{n-m} линейно-независимы относительно Q .

Пусть теперь $Y \in AR$. Тогда $Y = AX$. Разложим X по векторам выбранного базиса

$$X = c_1U_1 + \dots + c_mU_m + d_1V_1 + \dots + d_{n-m}V_{n-m}.$$

Следовательно,

$$Y = d_1AV_1 + \dots + d_{n-m}AV_{n-m},$$

ибо $AU_1 = \dots = AU_m = \mathbf{0}$. Итак, мы доказали, что размерность образа AR оператора A равна $n - m$, где m — размерность ядра. Теорема доказана.

Из доказанной теоремы следует, что ядро состоит из нулевого вектора в том и только в том случае, когда размерность AR равна n , т. е. когда оператор невырожденный. Заметим, что если ядро и образ оператора имеют нулевое пересечение, то все пространство является их прямой суммой. Однако это обстоятельство выполняется далеко не всегда.

В терминах теории матриц содержание теоремы может быть сформулировано следующим образом. *Максимальное число линейно-независимых решений системы n линейных однородных уравнений с n неизвестными равно $n - r$, где r ранг матрицы, составленной из коэффициентов системы.*

Действительно, пусть дана система

$$\begin{aligned} a_{11}y_1 + \dots + a_{1n}y_n &= 0 \\ \vdots &\quad \vdots \\ a_{n1}y_1 + \dots + a_{nn}y_n &= 0. \end{aligned} \tag{6}$$

Эта система равносильна векторному равенству

$$AY = \mathbf{0},$$

где A — оператор с матрицей

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix},$$

\mathbf{Y} — вектор с координатами y_1, \dots, y_n . Поэтому каждое решение системы (6) есть вектор из ядра \mathbf{Q} оператора \mathbf{A} , и обратно, координаты любого вектора $\mathbf{Y} \in \mathbf{Q}$ образуют решение системы (6), так что максимальное число линейно-независимых решений равно размерности \mathbf{Q} . По теореме 6.1 размерность \mathbf{Q} равна $n - r$, где r есть размерность образа оператора \mathbf{A} , т. е. ранг матрицы A .

6. Обратный оператор. Как мы видели, образ невырожденного оператора есть все пространство, так что невырожденный оператор осуществляет отображение пространства на себя. Это отображение взаимно однозначно. Действительно, если $\mathbf{AX} = \mathbf{Z}$ и $\mathbf{AY} = \mathbf{Z}$, то $\mathbf{A}(\mathbf{X} - \mathbf{Y}) = \mathbf{0}$, откуда следует $\mathbf{X} = \mathbf{Y}$, так как ядро невырожденного оператора состоит только из нулевого вектора. Поэтому для каждого невырожденного оператора \mathbf{A} существует обратный оператор \mathbf{A}^{-1} , сопоставляющий каждому вектору $\mathbf{Z} \in \mathbf{R}$ однозначно определенный вектор \mathbf{X} , такой, что $\mathbf{AX} = \mathbf{Z}$. Линейность оператора \mathbf{A}^{-1} очевидна.

Из определения обратного оператора следует, что $\mathbf{A}^{-1}\mathbf{A} = \mathbf{AA}^{-1} = \mathbf{E}$.

В любом базисе взаимно обратным операторам \mathbf{A} и \mathbf{A}^{-1} соответствуют взаимно обратные матрицы.

7. Собственные векторы и собственные значения оператора. Собственным значением (или собственным числом) оператора \mathbf{A} называется такое число λ , что для некоторого ненулевого вектора \mathbf{X} имеет место равенство

$$\mathbf{AX} = \lambda \mathbf{X}. \quad (7)$$

Любой ненулевой вектор \mathbf{X} , удовлетворяющий равенству (7), называется собственным вектором оператора \mathbf{A} , соответствующим (или принадлежащим) собственному значению λ .

Спектром оператора называется совокупность всех его собственных значений.

Собственные векторы и собственные значения оператора находятся следующим образом.

Пусть оператор \mathbf{A} в каком-либо базисе представляется матрицей $A = (a_{ik})$; пусть координаты собственного вектора в этом базисе суть x_1, \dots, x_n .

Тогда координаты вектора \mathbf{AX} будут

$$\sum_{k=1}^n a_{1k}x_k, \dots, \sum_{k=1}^n a_{nk}x_k,$$

и потому для определения координат x_1, \dots, x_n и собственного значения λ мы получим систему уравнений:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= \lambda x_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= \lambda x_2 \\ \vdots &\quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= \lambda x_n \end{aligned} \quad (8)$$

или

$$\begin{aligned} (a_{11} - \lambda)x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= 0 \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + \dots + a_{2n}x_n &= 0 \\ \vdots &\quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + (a_{nn} - \lambda)x_n &= 0. \end{aligned} \tag{9}$$

Эта система однородных относительно x_1, x_2, \dots, x_n уравнений будет иметь ненулевое решение в том и только в том случае, если

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{vmatrix} = 0, \tag{10}$$

т. е. если λ будет корнем характеристического полинома матрицы. Таким образом, верна следующая

Теорема 6.2. Собственные значения оператора совпадают с корнями характеристического полинома матрицы, представляющей оператор.

Вспомним, что матрицы, представляющие один и тот же оператор в различных базисах, подобны между собой, и, следовательно, их характеристические полиномы совпадают. Это дает основание назвать характеристический полином любой матрицы, представляющей оператор, характеристическим полиномом оператора.

Если $\varphi(t)$ характеристический полином оператора A и A какая-либо представляющая его матрица, то, в силу соотношения Кели—Гамильтона, $\varphi(A) = 0$. Следовательно, и $\varphi(A) = 0$, так как $\varphi(A)$ представляется нулевой матрицей $\varphi(A)$.

С оператором естественно связывается также и минимальный полином, который определяется как полином наименьшей степени среди полиномов, анулирующих оператор. Ясно, что минимальный полином оператора является также и минимальным полиномом для матрицы, сопоставляемой оператору в произвольном базисе. Характеристический полином делится на минимальный.

Поэтому каждый корень минимального полинома является корнем характеристического полинома. Справедливо и обратное утверждение. Именно, каждое собственное значение оператора, т. е. каждый корень характеристического полинома, является также и корнем минимального полинома оператора.

Действительно, пусть λ собственное значение оператора, X соответствующий ему собственный вектор, $\psi(t)$ минимальный полином оператора. По теореме Безу $\psi(t) = p(t)(t - \lambda) + \psi(\lambda)$. Следовательно, $\psi(A)X = p(A)(A - \lambda E)X + \psi(\lambda)X$. Но $(A - \lambda E)X = 0$ и $\psi(A)X = 0X = 0$. Поэтому $\psi(\lambda)X = 0$ и $\psi(\lambda) = 0$.

Таким образом, корни характеристического и минимального полиномов оператора совпадают в совокупности и могут отличаться лишь кратностями.

На основании так называемой основной теоремы высшей алгебры нам известно, что каждый полином имеет хотя бы один корень. Следовательно, оператор имеет по крайней мере одно собственное значение, которое может быть комплексным, даже если матрица оператора вещественна.

Для каждого собственного значения соответствующий собственный вектор (точнее векторы) определяется из системы (9) после подстановки в нее вместо буквы λ найденного численного значения. Собственных векторов, отвечающих собственному значению λ , бесконечно много, и они образуют подпространство пространства R .

Действительно, все собственные векторы, отвечающие собственному значению λ , образуют ядро оператора $A - \lambda E$. Размерность l этого подпространства, т. е. число линейно-независимых собственных векторов, соответствующих собственному значению λ , равно $n - r$, где r ранг оператора $A - \lambda E$.

Покажем, что l не превосходит кратности k числа λ как корня характеристического полинома оператора. Действительно, пусть X_1, \dots, X_l линейно-независимые собственные векторы, соответствующие собственному значению λ . Построим базис пространства X_1, \dots, X_n , взяв за первые l векторов векторы X_1, \dots, X_l . В этом базисе рассматриваемый оператор представляется матрицей, первые l столбцов которой имеют вид:

$$\begin{matrix} \lambda & 0 & \dots & 0 \\ 0 & \lambda & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{matrix}$$

ибо $AX_1 = \lambda X_1, \dots, AX_l = \lambda X_l$. Характеристический полином этой матрицы делится на $(t - \lambda)^l$, и, следовательно, λ имеет кратность k , не меньшую чем l , т. е. $l \leq k$. Естественно было бы предполагать, что $l = k$, т. е. что кратным корням характеристического полинома соответствуют k линейно-независимых собственных векторов. Однако на самом деле это не так. Именно, число линейно-независимых векторов может быть меньше, чем кратность собственного числа.

Подтвердим сказанное примером. Рассмотрим оператор с матрицей

$$A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}.$$

Тогда $|A - tE| = (t - 3)^2$, следовательно, $\lambda = 3$ является двойным корнем характеристического полинома.

Система уравнений для определения координат собственного вектора оператора A будет

$$\begin{aligned} 3x_1 + x_2 &= 3x_1 \\ 3x_2 &= 3x_2, \end{aligned}$$

откуда $x_2 = 0$, и потому все собственные векторы рассматриваемого оператора суть $(x_1, 0) = x_1(1, 0)$. Таким образом, двойному собственному числу в данном примере соответствует только один линейно-независимый собственный вектор, так что здесь l строго меньше k .

Важно отметить, что если $k = 1$, т. е. если λ есть простой корень характеристического полинома, то $l = k = 1$. В самом деле, $l \leq 1$ и $l > 0$, ибо хоть один собственный вектор, принадлежащий собственному значению λ , существует.

8. Собственные векторы матрицы. Собственным вектором матрицы A называется ненулевой столбец, удовлетворяющий условию

$$AX = \lambda X, \quad (11)$$

где λ — некоторое число. Ясно, что собственный вектор матрицы A есть столбец из координат собственного вектора оператора A , которому сопоставлена в избранном базисе данная матрица A .

Заметим, что если собственное значение вещественной матрицы комплексно, координаты собственного вектора также будут комплексными. Вектор, координаты которого комплексно сопряжены с координатами собственного вектора вещественной матрицы, тоже является собственным вектором этой матрицы, принадлежащим комплексно-сопряженному собственному значению.

Для того чтобы в этом убедиться, достаточно в равенстве $AX = \lambda X$ заменить все числа комплексно-сопряженными.

Выше было установлено, что подобные матрицы имеют одинаковые характеристические полиномы и, следовательно, одинаковые спектры собственных чисел.

Мы выяснили геометрическую причину этого обстоятельства, именно, подобные матрицы можно рассматривать как матрицы одного и того же оператора, отнесенного к различным базисам. Поэтому «собственные векторы» подобных матриц являются столбцами из координат собственных векторов рассматриваемого оператора в различных базисах и, следовательно, связаны соотношением

$$X' = C^{-1}X, \quad (12)$$

где C есть матрица преобразования координат.

Это обстоятельство проверяется и формально: если

$$AX = \lambda X, \text{ то } (C^{-1}AC)(C^{-1}X) = \lambda C^{-1}X.$$

9. Собственные векторы треугольной матрицы. Пусть

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & \dots & \dots & b_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & & & b_{nn} \end{bmatrix}$$

— правая треугольная матрица, диагональные элементы которой попарно различны. Очевидно, что эти диагональные элементы будут собственными значениями матрицы B . Найдем соответствующие им собственные векторы. Пусть $X_i = (x_{1i}, \dots, x_{ni})'$ есть собственный вектор, принадлежащий собственному значению b_{ii} . Для определения компонент вектора X_i прежде всего сравним в равенстве

$$BX_i = b_{ii}X_i \quad (13)$$

компоненты, начиная с $i+1$ -й. Это дает

$$\begin{aligned} (b_{i+1i+1} - b_{ii})x_{i+1i} + b_{i+1i+2}x_{i+2i} + \dots + b_{i+1n}x_{ni} &= 0 \\ (b_{i+2i+2} - b_{ii})x_{i+2i} + \dots + b_{i+2n}x_{ni} &= 0 \\ \vdots &\vdots \\ (b_{nn} - b_{ii})x_{ni} &= 0, \end{aligned}$$

откуда находим последовательно, что $x_{ni} = 0, x_{n-1i} = 0, \dots, x_{i+1i} = 0$.

Сравнение i -х компонент в равенстве (13) дает тождество

$$(b_{ii} - b_{ii})x_{ii} = 0,$$

которое показывает, что компоненту x_{ii} можно взять произвольной. Возьмем для определенности $x_{ii} = 1$. Тогда первые $i-1$ компонент вектора X_i определяются из треугольной системы

$$\begin{aligned} (b_{11} - b_{ii})x_{1i} + b_{12}x_{2i} + \dots + b_{1i-1}x_{i-1i} &= -b_{1i} \\ (b_{22} - b_{ii})x_{2i} + \dots + b_{2i-1}x_{i-1i} &= -b_{2i} \\ \vdots &\vdots \\ (b_{i-1i-1} - b_{ii})x_{i-1i} &= -b_{i-1i}. \end{aligned}$$

Таким образом, собственный вектор X_i , принадлежащий собственному значению b_{ii} , имеет все компоненты, начиная с $i+1$ -й, равными нулю. Поэтому матрица Ξ , столбцы которой состоят из компонент собственных векторов X_1, \dots, X_n , будет правой треугольной матрицей.

Аналогично устанавливается, что компоненты собственных векторов левой треугольной матрицы с попарно различными диагональными элементами составляют левую треугольную матрицу.

10. Приведение матрицы оператора к диагональной форме. Выясним вопрос об условиях, которым должен удовлетворять оператор для того, чтобы в пространстве существовал базис, состоящий из его собственных векторов. Это обстоятельство не всегда имеет место, как показывает пример, рассмотренный в п. 7.

Базис из собственных векторов замечателен тем, что в нем матрица оператора имеет диагональную форму $[\lambda_1, \lambda_2, \dots, \lambda_n]$. Действительно, если X_1, \dots, X_n базис, состоящий из собственных векторов оператора A , и $\lambda_1, \lambda_2, \dots, \lambda_n$ соответствующие собственные значения (среди которых могут быть равные), то $AX_1 = \lambda_1 X_1, \dots, AX_n = \lambda_n X_n$, так что в этом базисе i -я координата вектора AX_i равна λ_i , а все остальные координаты равны нулю. Следовательно, матрица оператора A в базисе X_1, \dots, X_n будет

$$\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

Обратно, если оператор A имеет в некотором базисе U_1, \dots, U_n диагональную матрицу, то векторы этого базиса являются линейно-независимыми собственными векторами оператора A . Действительно, в столбце координат вектора AU_i ненулевой будет лишь i -я координата и потому $AU_i = \lambda_i U_i$.

Теорема 6.3. *Собственные векторы, соответствующие попарно различным собственным значениям, линейно-независимы.*

Доказательство. Пусть $\lambda_1, \dots, \lambda_s$ попарно различные собственные значения оператора A , X_1, \dots, X_s какие-либо соответствующие им собственные векторы. Допустим, что они линейно-зависимы. Без нарушения общности можно считать, что векторы X_1, \dots, X_j , где $j < s$, линейно-независимы, а векторы X_{j+1}, \dots, X_s являются их линейными комбинациями. В частности, пусть

$$X_s = \sum_{i=1}^j c_i X_i.$$

Тогда

$$AX_s = \sum_{i=1}^j c_i AX_i = \sum_{i=1}^j c_i \lambda_i X_i.$$

С другой стороны,

$$AX_s = \lambda_s X_s = \sum_{i=1}^j c_i \lambda_s X_i.$$

Отсюда

$$\sum_{i=1}^j (\lambda_s - \lambda_i) c_i X_i = 0.$$

В силу линейной независимости векторов X_i все коэффициенты $(\lambda_s - \lambda_i)c_i = 0$ и, так как по предположению $\lambda_s \neq \lambda_i$ при $i = 1, 2, \dots, j$, все c_i равны нулю и потому $X_s = 0$, что противоречит тому, что X_s собственный вектор. Итак, векторы X_1, \dots, X_s линейно-независимы.

Теорема 6.4. Если характеристический полином оператора имеет только простые корни, то в пространстве существует базис, состоящий из собственных векторов оператора.

Доказательство. По условию теоремы оператор имеет n различных собственных значений, где n размерность пространства. Соответствующие им собственные векторы X_1, \dots, X_n , в силу теоремы 6.3, линейно-независимы, и значит их можно принять за базис.

В терминах теории матриц эта теорема может быть перефразирована следующим образом. Если собственные значения матрицы A попарно различны, то существует неособенная матрица C такая, что $C^{-1}AC = \Lambda$, где $\Lambda = [\lambda_1, \dots, \lambda_n]$.

Действительно, рассмотрим оператор A с матрицей A относительно какого-либо базиса. Для оператора A существует базис из собственных векторов. В этом базисе оператору A соотносится диагональная матрица $\Lambda = [\lambda_1, \dots, \lambda_n]$, причем $\Lambda = C^{-1}AC$. Здесь C — матрица преобразования координат при переходе от исходного базиса к базису, состоящему из собственных векторов. Следовательно, ее столбцы состоят из координат собственных векторов по отношению к исходному базису.

Теорема 6.5. Для того чтобы существовал базис из собственных векторов оператора A , необходимо и достаточно, чтобы каждому собственному значению соответствовало столько линейно-независимых векторов, какова его кратность.

Доказательство. Пусть $\lambda_1, \dots, \lambda_s$ все различные собственные значения оператора A и пусть k_1, \dots, k_s их кратности как корней характеристического полинома, $k_1 + \dots + k_s = n$. Обозначим через l_i число линейно-независимых собственных векторов, отвечающих собственному числу λ_i . Пусть X_{11}, \dots, X_{il_i} эти собственные векторы. Покажем, что векторы

$$X_{11}, \dots, X_{il_i}, \dots, X_{s1}, \dots, X_{sl_s}$$

линейно-независимы.

Положим

$$c_{11}X_{11} + \dots + c_{1l_i}X_{il_i} + \dots + c_{s1}X_{s1} + \dots + c_{sl_s}X_{sl_s} = 0. \quad (14)$$

Обозначим

$$Y_1 = c_{11}X_{11} + \dots + c_{1l_i}X_{il_i}; \dots; Y_s = c_{s1}X_{s1} + \dots + c_{sl_s}X_{sl_s}. \quad (14')$$

Тогда

$$Y_1 + \dots + Y_s = 0. \quad (14'')$$

Каждый из векторов Y_i есть или нулевой вектор, или собственный вектор, принадлежащий собственному значению λ_i . В силу теоремы 6.3, равенство (14'') возможно, только если все векторы Y_1, \dots, Y_s равны нулевому вектору. Но тогда, в силу (14') и линейной независимости векторов X_{ii}, \dots, X_{ii} при каждом i , заключаем, что

$$c_{11} = \dots = c_{1l_1} = \dots = c_{sl_1} = \dots = c_{sl_s} = 0,$$

так что равенство (14) возможно только при нулевых коэффициентах. Тём самым доказана линейная независимость векторов $X_{11}, \dots, X_{1l_1}, \dots, X_{sl_1}, \dots, X_{sl_s}$. Таким образом, максимальное число линейно-независимых векторов, соответствующих всем собственным значениям, равно $l_1 + \dots + l_s$. Поэтому для существования базиса из собственных векторов необходимо и достаточно, чтобы $l_1 + \dots + l_s = n$, что будет выполнено, только если все $l_i = k_i$.

§ 7. Каноническая форма Жордана

В предыдущем параграфе мы видели, что если матрица не имеет кратных собственных значений, то она всегда может быть приведена к диагональной форме преобразованием подобия. Однако, если кратные значения имеются, то преобразования к диагональному виду может не существовать. Это обстоятельство является исключительным в том смысле, что многообразие матриц, имеющих кратные собственные значения, имеет меньшую размерность, чем пространство всех матриц. Тем не менее исследование строения таких матриц представляет очень большой интерес для приложений как теоретического, так и практического характера. В вычислительной математике, в обстоятельствах, когда элементы матриц задаются неточно, резкая грань между случаями простых и кратных собственных значений стирается, так как при малых деформациях элементов матрицы всегда можно перейти от матрицы с кратными собственными значениями к матрице с простыми собственными значениями. Поэтому в вычислительной алгебре исследование матриц с кратными собственными значениями важно преимущественно для правильной ориентировки в структуре матриц, имеющих очень близкие, но различные собственные значения. С такими же матрицами приходится встречаться очень часто в приложениях.

Настоящий параграф посвящен изучению структуры матриц, не приводящихся к диагональной форме, и в частности, установлению некоторой простейшей канонической формы, обобщающей диагональную, к которой может быть приведена преобразованием подобия уже совершенно произвольная матрица.

1. Инвариантные подпространства. Пусть A оператор, действующий в n -мерном пространстве R . Подпространство P пространства R называется инвариантным по отношению к оператору A , если векторы из P преобразуются оператором снова в векторы из P , т. е. из $X \in P$ следует, что и $AX \in P$ (или в сокращенной записи $AP \subset P$).

Из данного определения следует, что если P инвариантное подпространство для A , то оно будет инвариантным и для оператора $f(A)$, где $f(t)$ любой полином. Действительно, если $X \in P$ и P инвариантно, то $AX \in P$, $A^2X \in P$, ..., и, следовательно, $f(A)X \in P$.

Отметим, в частности, что подпространство, инвариантное относительно оператора A , инвариантно и относительно оператора $A - \mu E$ при любом числе μ . Верно и обратное утверждение: если подпространство инвариантно относительно оператора $A - \mu E$, то оно инвариантно относительно A , ибо $A = A - \mu E + \mu E$.

Очевидно, что все пространство и пространство, состоящее из нулевого вектора, суть инвариантные подпространства. Нетривиальными примерами инвариантных подпространств могут служить, например, подпространства, натянутые на один или несколько собственных векторов оператора A . Действительно, пусть X_1, \dots, X_k собственные векторы оператора A и P натянутое на них подпространство. Тогда любой вектор X , принадлежащий P , может быть представлен в виде

$$X = c_1X_1 + \dots + c_kX_k,$$

и потому $AX = c_1AX_1 + \dots + c_kAX_k = c_1\lambda_1X_1 + \dots + c_k\lambda_kX_k$ (среди чисел $\lambda_1, \dots, \lambda_k$ могут быть равные). В случае, если все собственные числа оператора A различны, указанными подпространствами, как мы увидим далее, исчерпываются все инвариантные подпространства оператора.

Другим важным типом инвариантных подпространств являются циклические подпространства. Для определения этого понятия рассмотрим следующую конструкцию. Пусть дан вектор X_0 . Построим систему векторов X_0, AX_0, A^2X_0, \dots . Ясно, что в этой системе векторов когда-то в первый раз встретится вектор A^qX_0 , являющийся линейной комбинацией предыдущих $X_0, AX_0, \dots, A^{q-1}X_0$.

Циклическим подпространством Q , порожденным вектором X_0 , называется подпространство, натянутое на векторы $X_0, AX_0, \dots, A^{q-1}X_0$. Так как векторы $X_0, AX_0, \dots, A^{q-1}X_0$ линейно-независимы, то они образуют базис циклического подпространства Q и потому размерность Q равна показателю степени q .

Докажем теперь, что циклическое подпространство, порожденное вектором X_0 , есть наименьшее инвариантное подпространство, содержащее X_0 , т. е. что оно само инвариантно и что оно содержится во всяком инвариантном подпространстве, содержащем X_0 .

Действительно, пусть $A^q X_0 = \gamma_0 X_0 + \dots + \gamma_{q-1} A^{q-1} X_0$ и пусть $Y \in Q$. Тогда

$$Y = c_0 X_0 + c_1 A X_0 + \dots + c_{q-2} A^{q-2} X_0 + c_{q-1} A^{q-1} X_0;$$

$$\begin{aligned} AY &= c_0 A X_0 + c_1 A^2 X_0 + \dots + c_{q-2} A^{q-1} X_0 + c_{q-1} A^q X_0 = \\ &= c_0 A X_0 + c_1 A^2 X_0 + \dots + c_{q-2} A^{q-1} X_0 + \\ &\quad + c_{q-1} (\gamma_0 X_0 + \gamma_1 A X_0 + \dots + \gamma_{q-1} A^{q-1} X_0) = \\ &= c'_0 X_0 + c'_1 A X_0 + \dots + c'_{q-1} A^{q-1} X_0 \in Q. \end{aligned}$$

Тем самым инвариантность Q доказана.

Далее, пусть Q' какое-либо инвариантное подпространство, содержащее X_0 . Тогда $X_0 \in Q'$; $AX_0 \in Q'$, ..., $A^{q-1} X_0 \in Q'$ и, следовательно, $Q \subset Q'$, что доказывает минимальность Q среди инвариантных подпространств, содержащих X_0 .

Отметим для дальнейшего, что любое циклическое по отношению к оператору A подпространство, порожденное вектором X_0 , будет циклическим и по отношению к оператору $A - \mu E$ при любом численном значении для μ . Действительно, каждое инвариантное по отношению к A подпространство будет инвариантным и для $A - \mu E$ и обратно, и, следовательно, минимальные инвариантные подпространства, содержащие X_0 , должны совпадать.

2. Минимальный аннулирующий вектор X_0 полином. Размерность q циклического подпространства связана также со следующим важным понятием. Пусть A данный оператор. Мы будем называть полином $\chi(t)$ аннулирующим вектора X_0 , если $\chi(A)X_0 = 0$. Среди полиномов, аннулирующих вектор X_0 , существует полином $\theta(t)$ наименьшей степени, называемый минимальным аннулирующим полиномом для вектора X_0 .

Степень минимального аннулирующего полинома равна размерности циклического подпространства, порожденного вектором X_0 .

Действительно, пусть q размерность циклического подпространства, порожденного вектором X_0 , и пусть $A^q X_0 = \gamma_0 X_0 + \dots + \gamma_{q-1} A^{q-1} X_0$.

Положив $\theta(t) = t^q - \gamma_{q-1} t^{q-1} - \dots - \gamma_0$, получим $\theta(A)X_0 = 0$, т. е. $\theta(t)$ есть аннулирующий вектор X_0 полином. С другой стороны, если $\chi(t)$ полином меньшей степени чем q , то $\chi(A)X_0 \neq 0$ в силу линейной независимости векторов $X_0, AX_0, \dots, A^{q-1} X_0$. Следовательно, полином q -й степени $\theta(t)$ есть минимальный аннулирующий полином для вектора X_0 .

Легко доказать, что всякий полином, аннулирующий X_0 , делится на минимальный аннулирующий X_0 полином. Действительно, пусть $\chi(A)X_0 = 0$. Поделив полином $\chi(t)$ на полином $\theta(t)$ с остатком, получим $\chi(t) = p(t)\theta(t) + r(t)$, где остаток $r(t)$ имеет степень,

меньшую, чем q . Следовательно, $\mathbf{0} = \theta(\mathbf{A})\mathbf{X}_0 = p(\mathbf{A})\theta(\mathbf{A})\mathbf{X}_0 + r(\mathbf{A})\mathbf{X}_0 = r(\mathbf{A})\mathbf{X}_0$, откуда $r(t) = 0$, ибо иначе полином $\theta(t)$ не был бы минимальным аннулирующим \mathbf{X}_0 полиномом. В частности, минимальный полином оператора (а следовательно, и характеристический) делится на минимальный аннулирующий \mathbf{X}_0 полином. Поэтому размерность любого циклического подпространства не превосходит степени минимального полинома оператора.

Теорема 7.1. Если минимальный аннулирующий полином $\theta(t)$ для вектора \mathbf{X} раскладывается в произведение попарно взаимно простых множителей

$$\theta(t) = \theta_1(t)\theta_2(t)\dots\theta_s(t),$$

то вектор \mathbf{X} может быть представлен в виде суммы векторов $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_s$, аннулируемых соответственно полиномами $\theta_1(t), \theta_2(t), \dots, \theta_s(t)$. При этом слагаемые $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_s$ можно взять принадлежащими любому инвариантному подпространству, содержащему \mathbf{X} .

Доказательство. Очевидно, что теорему достаточно доказать для $s = 2$, ибо переход к общему случаю осуществляется методом математической индукции. Так как $\theta_1(t)$ и $\theta_2(t)$ взаимно просты, то найдутся два полинома $p_1(t)$ и $p_2(t)$ такие, что $\theta_1(t)p_1(t) + \theta_2(t)p_2(t) = 1^1$). Это равенство влечет за собой операторное равенство

$$\theta_1(\mathbf{A})p_1(\mathbf{A}) + \theta_2(\mathbf{A})p_2(\mathbf{A}) = \mathbf{E},$$

и потому верно векторное равенство

$$\mathbf{X} = \theta_1(\mathbf{A})p_1(\mathbf{A})\mathbf{X} + \theta_2(\mathbf{A})p_2(\mathbf{A})\mathbf{X}.$$

Положим

$$\mathbf{X}_1 = \theta_2(\mathbf{A})p_2(\mathbf{A})\mathbf{X}, \quad \mathbf{X}_2 = \theta_1(\mathbf{A})p_1(\mathbf{A})\mathbf{X}.$$

Тогда

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2,$$

причем

$$\theta_1(\mathbf{A})\mathbf{X}_1 = \theta_1(\mathbf{A})\theta_2(\mathbf{A})p_2(\mathbf{A})\mathbf{X} = \theta(\mathbf{A})p_2(\mathbf{A})\mathbf{X} =$$

$$= p_2(\mathbf{A})\theta(\mathbf{A})\mathbf{X} = p_2(\mathbf{A})\mathbf{0} = \mathbf{0},$$

и аналогично $\theta_2(\mathbf{A})\mathbf{X}_2 = \mathbf{0}$.

Построенные векторы \mathbf{X}_1 и \mathbf{X}_2 принадлежат к любому инвариантному подпространству, содержащему \mathbf{X} , ибо если $\mathbf{X} \in \mathbf{P}$ и \mathbf{P} инвариантно, то

$$\mathbf{X}_1 = \theta_2(\mathbf{A})p_2(\mathbf{A})\mathbf{X} \in \mathbf{P}, \quad \mathbf{X}_2 = \theta_1(\mathbf{A})p_1(\mathbf{A})\mathbf{X} \in \mathbf{P}.$$

Замечание. Нетрудно показать, что полиномы $\theta_1(t), \dots, \theta_s(t)$ будут минимальными аннулирующими полиномами для векторов $\mathbf{X}_1, \dots, \mathbf{X}_s$ соответственно.

¹) А. Г. Курош. Курс высшей алгебры, 1940, стр. 175.

3. Индуцированный оператор. Пусть оператор A действует в n -мерном пространстве R и пусть P инвариантное подпространство для этого оператора. Тогда оператор A сопоставляет каждому вектору из P вектор из P , т. е. определяет некоторое преобразование подпространства P . Очевидно, что это преобразование является линейным оператором, определенным на P . Этот оператор имеет название оператора, индуцированного оператором A на подпространстве P . Индуцированный оператор отличается от оператора A только областью определения.

Пусть P инвариантное относительно оператора A подпространство, U_1, \dots, U_m — базис P , $U_1, \dots, U_m, V_1, \dots, V_{n-m}$ — базис всего пространства. Выясним вид, который имеет в этом базисе матрица оператора A . Так как векторы AU_1, \dots, AU_m принадлежат P , т. е. являются линейными комбинациями лишь векторов U_1, \dots, U_m , то их координаты в выбранном базисе, начиная с $m+1$ -й, равны нулю. Следовательно, матрица оператора A имеет вид:

$$\begin{bmatrix} a_{11} & \dots & a_{1m} & a_{1m+1} & \dots & a_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mm} & a_{mm+1} & \dots & a_{mn} \\ 0 & \dots & 0 & a_{m+1, m+1} & \dots & a_{m+1, n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nm+1} & \dots & a_{nn} \end{bmatrix},$$

или сокращенно

$$\begin{bmatrix} A_P & B \\ 0 & \tilde{A}_{P^\perp} \end{bmatrix}.$$

Здесь A_P есть квадратная матрица порядка m , \tilde{A}_P квадратная матрица порядка $n-m$, B прямоугольная матрица из m строк и $n-m$ столбцов, 0 нулевая прямоугольная матрица. Ясно, что A_P есть матрица оператора, индуцированного на P .

Матрица оператора еще более упрощается, если пространство R есть прямая сумма двух инвариантных подпространств. Действительно, пусть $R = P_1 + P_2$. Возьмем за базис R объединение базисов P_1 и P_2 . В этом базисе матрица оператора A , очевидно, примет вид

$$\begin{bmatrix} A_{P_1} & 0 \\ 0 & A_{P_2} \end{bmatrix},$$

где A_{P_1} и A_{P_2} матрицы операторов, индуцированных оператором A на P_1 и P_2 . Если пространство разбивается в прямую сумму k инвариантных подпространств, то в базисе, составленном из объединения

базисов этих подпространств, матрица оператора \mathbf{A} принимает квазидиагональный вид

$$\begin{bmatrix} -A_{P_1} & & & 0 \\ & A_{P_2} & & \\ & & \ddots & \\ 0 & & & A_{P_k} \end{bmatrix}, \quad (1)$$

где $A_{P_1}, A_{P_2}, \dots, A_{P_k}$ — матрицы операторов, индуцированных на P_1, P_2, \dots, P_k .

Из последнего разложения вытекает следующая

Теорема 7.2. *Если пространство R есть прямая сумма инвариантных относительно оператора \mathbf{A} подпространств P_1, P_2, \dots, P_k , то характеристический полином оператора \mathbf{A} равен произведению характеристических полиномов операторов A_1, A_2, \dots, A_k , индуцированных оператором \mathbf{A} на подпространствах P_1, P_2, \dots, P_k .*

Для доказательства достаточно отнять t от элементов главной диагонали матрицы (1) и воспользоваться теоремой о том, что определитель квазидиагональной матрицы есть произведение определителей, составленных из ее диагональных клеток.

Если в разложении пространства в прямую сумму инвариантных подпространств входят одномерные инвариантные подпространства, т. е. подпространства, натянутые на собственные векторы, то соответствующие диагональные клетки будут клетками первого порядка, именно диагональными элементами матрицы. Очевидно, что эти диагональные элементы будут собственными значениями матрицы.

В дальнейшем мы часто будем заменять выражение „оператор, индуцированный оператором \mathbf{A} на P “ выражением „оператор \mathbf{A} на P “.

4. Корневые подпространства. Среди инвариантных подпространств особо важную роль играют так называемые корневые подпространства. Корневым вектором для оператора \mathbf{A} , соответствующим числу μ , называется вектор \mathbf{X} такой, что $(\mathbf{A} - \mu E)^m \mathbf{X} = \mathbf{0}$ при некотором целом $m > 0$. Ясно, что совокупность корневых векторов, отвечающих заданному числу μ , образует подпространство. Действительно, если $(\mathbf{A} - \mu E)^{m_1} \mathbf{X}_1 = \mathbf{0}$ и $(\mathbf{A} - \mu E)^{m_2} \mathbf{X}_2 = \mathbf{0}$, то $(\mathbf{A} - \mu E)^m (c_1 \mathbf{X}_1 + c_2 \mathbf{X}_2) = \mathbf{0}$, где $m = \max(m_1, m_2)$. Это подпространство называется корневым подпространством, соответствующим числу μ . Покажем, что оно инвариантное. Действительно, если $(\mathbf{A} - \mu E)^m \mathbf{X} = \mathbf{0}$, то $(\mathbf{A} - \mu E)^m \mathbf{AX} = \mathbf{A}(\mathbf{A} - \mu E)^m \mathbf{X} = \mathbf{0}$.

Понятие корневого вектора обобщает понятие собственного вектора, именно, каждый собственный вектор \mathbf{X} , принадлежащий собст-

венному значению λ , является и корневым для того же числа λ , ибо $(A - \mu E)X = 0$.

Высотой ненулевого корневого вектора называется наименьшее число из таких показателей m , что $(A - \mu E)^m X = 0$. Иными словами, высота корневого вектора есть такое число k , что $(A - \mu E)^k X = 0$, но $(A - \mu E)^{k-1} X \neq 0$. Высота нулевого вектора считается равной нулю по определению. Собственные векторы являются корневыми векторами высоты единица.

Полином $(t - \mu)^k$ есть минимальный аннулирующий полином для корневого вектора высоты k . Действительно, $(A - \mu E)^k X = 0$, и, следовательно, минимальный аннулирующий X полином является делителем $(t - \mu)^k$. Но делителями $(t - \mu)^k$ являются только полиномы $(t - \mu)^j$ при $j \leq k$. Полиномы же $(t - \mu)^j$ при $j < k$ не аннулируют вектор X , ибо $(A - \mu E)^j X \neq 0$.

Теорема 7.3. Для того чтобы для числа μ существовал ненулевой корневой вектор, необходимо и достаточно, чтобы число μ было собственным значением оператора A . При этом высота корневого вектора не превосходит кратности m числа μ как корня минимального полинома. Существуют корневые векторы высоты m .

Доказательство. Если μ является собственным значением, то для него существуют ненулевые корневые векторы, например, собственные векторы. Обратно, если для числа μ существует ненулевой корневой вектор X высоты k , то $Z = (A - \mu E)^{k-1} X \neq 0$ и $(A - \mu E)Z = (A - \mu E)^k X = 0$, так что Z есть собственный вектор, соответствующий числу μ , и, следовательно, μ есть собственное значение. Минимальный аннулирующий вектор X полином $(t - \mu)^k$ является делителем минимального полинома оператора A . Поэтому высота k вектора X не превосходит кратности m числа μ как корня минимального полинома.

Остается доказать последнее утверждение теоремы. Пусть $\phi(t) = (t - \mu)^m f(t)$ есть минимальный полином оператора A . Выберем вектор U так, чтобы он не аннулировался оператором $(A - \mu E)^{m-1} f(A)$. Такой вектор U найдется, ибо иначе $\phi(t)$ не был бы минимальным полиномом для A .

Положим $X = f(A)U$. Тогда $(A - \mu E)^{m-1} X = (A - \mu E)^{m-1} f(A)U \neq 0$, но $(A - \mu E)^m X = (A - \mu E)^m f(A)U = \phi(A)U = 0$. Таким образом, X есть корневой вектор высоты m для числа μ .

Б. Свойства оператора, индуцированного на корневом подпространстве. Пусть A оператор в пространстве R , λ его собственное значение кратности m как корня минимального полинома, P корневое подпространство, соответствующее этому собственному

значению. Пусть A_P оператор, индуцированный оператором A на подпространстве P .

Теорема 7.4. Минимальный полином оператора A_P равен $(t - \lambda)^m$, характеристический полином оператора A_P равен $(t - \lambda)^p$, где p — размерность пространства P .

Доказательство. Оператор $(A - \lambda E)^m$ аннулирует все векторы подпространства P , а оператор $(A - \lambda E)^{m-1}$ аннулирует не все векторы P . Следовательно, $(A_P - \lambda E)^m = 0$, а $(A_P - \lambda E)^{m-1} \neq 0$. Отсюда следует, что $(t - \lambda)^m$ есть минимальный полином оператора A_P . Далее, всякое собственное значение оператора является корнем минимального полинома. Следовательно, оператор A_P имеет единственное собственное значение λ и потому характеристический полином оператора A_P равен $(t - \lambda)^p$. Показатель p равен размерности подпространства P , так как степень характеристического полинома любого оператора равна размерности пространства, в котором он определен. Ниже мы установим, что p равно кратности собственного значения как корня характеристического полинома оператора A .

6. Линейная независимость корневых векторов.

Теорема 7.5. Ненулевые корневые векторы, соответствующие попарно различным собственным значениям оператора A , линейно-независимы.

Доказательство. Пусть X_1, \dots, X_s ненулевые корневые векторы оператора A , соответствующие собственным значениям $\lambda_1, \dots, \lambda_s$, причем $\lambda_i \neq \lambda_j$ при $i \neq j$, и пусть k_1, \dots, k_s высоты векторов X_1, \dots, X_s . Обозначим через $f_i(t)$ полином

$$(t - \lambda_1)^{k_1} \dots (t - \lambda_i)^{k_i-1} \dots (t - \lambda_s)^{k_s}.$$

Докажем, что в зависимости

$$c_1X_1 + \dots + c_iX_i + \dots + c_sX_s = 0$$

все коэффициенты могут быть только нулями. Применим к обеим частям равенства оператор $f_i(A)$. Получим

$$c_1f_i(A)X_1 + \dots + c_if_i(A)X_i + \dots + c_sf_i(A)X_s = 0. \quad (2)$$

Ясно, что $f_i(A)X_j = 0$ при $i \neq j$, ибо полином $f_i(t)$ делится на полиномы $(t - \lambda_j)^{k_j}$, $j \neq i$, аннулирующие векторы X_j соответственно.

Далее, $f_i(A)X_i \neq 0$, ибо полином $f_i(t)$ не делится на полином $(t - \lambda_i)^{k_i}$, являющийся минимальным аннулирующим вектор X_i полиномом. Таким образом, равенство (2) превращается в

$$c_if_i(A)X_i = 0,$$

причем $f_i(A)X_i \neq 0$. Следовательно, $c_i = 0$, для всех $i = 1, 2, \dots, s$. Тем самым линейная независимость векторов X_1, \dots, X_s доказана.

7. Разложение пространства в прямую сумму корневых подпространств.

Теорема 7.6. Пространство R есть прямая сумма всех корневых подпространств оператора A .

Доказательство. Векторная сумма R' всех корневых подпространств есть сумма прямая, в силу доказанной выше линейной независимости корневых векторов, соответствующих попарно различным собственным значениям, т. е. принадлежащим к попарно различным корневым подпространствам. Остается только доказать, что R' совпадает со всем пространством R , т. е. что любой вектор X из R может быть разложен в сумму корневых векторов X_i при $i = 1, \dots, s$. Докажем это. Пусть полином $\theta(t)$ есть минимальный аннулирующий полином для вектора X . Разложим его на линейные множители:

$$\theta(t) = (t - \lambda_1)^{k_1} \dots (t - \lambda_s)^{k_s}, \quad \lambda_i \neq \lambda_j.$$

Сомножители $(t - \lambda_1)^{k_1}, \dots, (t - \lambda_s)^{k_s}$ попарно взаимно просты. Следовательно, на основании теоремы 7.1 мы придем к разложению

$$X = X_1 + \dots + X_s,$$

где векторы X_1, \dots, X_s будут аннулироваться соответственно полиномами $(t - \lambda_1)^{k_1}, \dots, (t - \lambda_s)^{k_s}$. Поэтому векторы X_1, \dots, X_s являются корневыми векторами. Тем самым теорема доказана.

Отметим, что если вектор X принадлежит какому-либо инвариантному подпространству, то векторы X_1, \dots, X_s принадлежат тому же подпространству. Это следует из вышеупомянутой теоремы.

Составляющие X_i вектора X по корневым подпространствам будем называть проекциями на эти подпространства.

Из доказанной теоремы вытекает такое следствие. *Размерность корневого подпространства, соответствующего собственному значению λ_i , равна кратности λ_i как корня характеристического полинома оператора A .*

Действительно, характеристический полином $\varphi(t)$ оператора A в силу сказанного ранее в п. 2 есть произведение характеристических полиномов операторов, индуцированных оператором A на корневых подпространствах P_1, \dots, P_s , каждый из которых в свою очередь есть $(\lambda_i - t)^{p_i}$, где p_i — размерность соответствующего корневого подпространства. Итак,

$$\varphi(t) = (\lambda_1 - t)^{p_1} \dots (\lambda_s - t)^{p_s}.$$

откуда следует, что размерности p_1, \dots, p_s являются кратностями собственных значений в характеристическом полиноме оператора A .

8. Канонический базис корневого подпространства. Изучим подробнее строение отдельного корневого подпространства для оператора A . С целью упрощения записи мы будем здесь обозначать

корневое подпространство буквой **P** и соответствующее собственное значение через λ , опустив индексы.

Корневое подпространство P естественно разбивается на „этажи“. Под этажом высоты j мы будем подразумевать совокупность всех векторов высоты j . Этажи не являются подпространствами, так как, в частности, они не содержат нулевого вектора. Однако совокупность векторов, высоты которых не превосходят данного числа j , уже образует подпространство. Действительно, если высоты векторов X_1 и X_2 не превосходят j , то $(A - \lambda E)^j X_1 = (A - \lambda E)^j X_2 = 0$ и, следовательно, $(A - \lambda E)^j (c_1 X_1 + c_2 X_2) = 0$, т. е. высота вектора $c_1 X_1 + c_2 X_2$ не превосходит j . Обозначим указанное подпространство через $P^{(j)}$. Очевидно, что $P^{(j)}$ инвариантно. Далее, $P^{(1)} \subset P^{(2)} \subset \dots \dots \subset P^{(m)} = P$.

Наряду с „горизонтальными“ инвариантными подпространствами $P^{(j)}$ мы рассмотрим инвариантные подпространства совершенно другого рода, так сказать, „вертикальные“. Если X_0 корневой вектор высоты $j \geq 1$, то вектор $X_1 = (A - \lambda E)X_0$ будет иметь высоту $j - 1$. Будем говорить, что вектор X_1 лежит под вектором X_0 . Совокупность векторов X_0, X_1, \dots, X_{i-1} таких, что

$$X_1 = (A - \lambda E) X_0$$

$$\mathbf{X}_2 = (\mathbf{A} - \lambda \mathbf{E}) \mathbf{X}_1$$

卷之三

$$\mathbf{X}_{j+1} = (\mathbf{A} - \lambda \mathbf{E}) \mathbf{X}_{j+2},$$

назовем „башней“. Ясно, что $(A - \lambda E)X_{j-1} = 0$. Высота башни, (т. е. число ее элементов) равна высоте ее „верхнего“ порождающего вектора X_0 . Покажем, что векторы, образующие башню, линейно-независимы. Действительно, пусть

$$c_0\mathbf{x}_0 + c_1\mathbf{x}_1 + \dots + c_{j-1}\mathbf{x}_{j-1} = \mathbf{0}.$$

Применяя к этому равенству последовательно операторы $(A - \lambda E)$, $(A - \lambda E)^2$, ..., $(A - \lambda E)^{j-1}$, получим

$$c_0\mathbf{X}_1 + c_1\mathbf{X}_2 + \dots + c_{j-2}\mathbf{X}_{j-1} = \mathbf{0}$$

$$c_0X_2 + \dots + c_{j-3}X_{j-1} = 0$$

1 2 3 4 5 6 7 8 9 10

$$c_0 X_{j-1} = 0,$$

откуда заключаем, что $c_0 = 0, c_1 = 0, \dots, c_{j-1} = 0$.

Подпространство, натянутое на башню, будет иметь размерность j . Оно инвариантное и циклическое для оператора $A - \lambda E$, а следовательно, инвариантное и циклическое и для оператора A .

Мы установим, что в подпространстве P существует базис, получающийся объединением нескольких башен, не содержащих общих элементов. Такой базис будем называть каноническим базисом корневого подпространства.

Ясно, что каждый выбор канонического базиса определяет разложение подпространства P в прямую сумму инвариантных циклических подпространств. Именно, такими подпространствами будут подпространства, натянутые на базисные векторы, входящие в отдельные башни.

Доказательству существования канонического базиса предпошлем следующую лемму.

Лемма. Если векторы Z_1, \dots, Z_s принадлежат $P^{(j+1)}$ и линейно-независимы относительно $P^{(j)}$, то векторы $(A - \lambda E)Z_1, \dots, (A - \lambda E)Z_s$ принадлежат $P^{(j)}$ и линейно-независимы относительно $P^{(j-1)}$.

Доказательство. Первое утверждение леммы очевидно. Допустим теперь, что

$$c_1(A - \lambda E)Z_1 + \dots + c_s(A - \lambda E)Z_s = V \in P^{(j-1)}.$$

Это значит, что

$$(A - \lambda E)^{j-1}V = (A - \lambda E)^j(c_1Z_1 + \dots + c_sZ_s) = 0.$$

Следовательно,

$$c_1Z_1 + \dots + c_sZ_s \in P^{(j)},$$

что возможно только при $c_1 = \dots = c_s = 0$, в силу линейной независимости векторов Z_1, \dots, Z_s относительно $P^{(j)}$. Тем самым лемма доказана.

Перейдем теперь к доказательству существования канонического базиса.

Пусть $P^{(1)} \subset P^{(2)} \subset \dots \subset P^{(m)} = P$, где, как и раньше, $P^{(j)}$ есть совокупность всех векторов, высоты которых не превосходят j . Выберем произвольный базис X_{11}, \dots, X_{1k_1} подпространства $P^{(m)}$ относительно $P^{(m-1)}$. Число k_1 равно разности размерностей $P^{(m)}$ и $P^{(m-1)}$. Тогда, в силу леммы, векторы $(A - \lambda E)X_{11}, \dots, (A - \lambda E)X_{1k_1}$ принадлежат $P^{(m-1)}$ и будут линейно-независимы относительно $P^{(m-2)}$. Следовательно, они могут быть включены в базис $P^{(m-1)}$ относительно $P^{(m-2)}$. Пусть $(A - \lambda E)X_{11}, \dots, (A - \lambda E)X_{1k_1}; X_{21}, \dots, X_{2k_2}$ базис $P^{(m-1)}$ относительно $P^{(m-2)}$. В силу леммы, векторы

$$(A - \lambda E)^2X_{11}, \dots, (A - \lambda E)^2X_{1k_1}, (A - \lambda E)X_{21}, \dots, (A - \lambda E)X_{2k_2}$$

принадлежат $P^{(m-2)}$ и линейно-независимы относительно $P^{(m-3)}$. Дополним их векторами X_{31}, \dots, X_{3k_3} до базиса $P^{(m-2)}$ относительно $P^{(m-3)}$,

Базис $\mathbf{P}^{(m)}$ отн. $\mathbf{P}^{(m-1)}$	X_{11}	\dots	X_{1k_1}	\dots	X_{1k_1}
Базис $\mathbf{P}^{(m-1)}$ отн. $\mathbf{P}^{(m-2)}$	BX_{11}	\dots	BX_{1k_1}	X_{21}	\dots
\dots	\dots	\dots	\dots	\dots	\dots
Базис $\mathbf{P}^{(2)}$ отн. $\mathbf{P}^{(1)}$	$B^{m-2}X_{11}$	\dots	$B^{m-2}X_{1k_1}$	$B^{m-3}X_{21}$	\dots
Базис $\mathbf{P}^{(1)}$	$B^{m-1}X_{11}$	\dots	$B^{m-1}X_{1k_1}$	$B^{m-2}X_{21}$	\dots

применим к построенному базису оператора $A - \lambda E$, дополним полученную систему векторов до базиса $P^{(m-3)}$ относительно $P^{(m-4)}$ и т. д. На m -м шагу мы придем к базису

$$X_{m1}, \dots, X_{mk_m}$$

подпространства $P^{(1)}$. Объединение всех построенных относительных базисов есть базис P и этот базис, очевидно, будет каноническим.

Построение канонического базиса в наглядной форме описывается схемой, приведенной на левой стороне страницы. В этой схеме $B = A - \lambda E$.

Из построения ясно, что выбор канонического базиса не однозначен. Однако нетрудно убедиться, что любой канонический базис может быть построен посредством описанной конструкции. Поэтому строение любого канонического базиса (число башен данной высоты) будет одинаковым.

9. Канонический базис пространства и каноническая форма Жордана для матрицы оператора. Каноническим базисом пространства R , в котором действует оператор A , называется базис, получающийся в результате объединения канонических базисов всех корневых подпространств, отвечающих оператору. Канонический базис естественно разбивается на башни и соответственно все пространство разбивается в прямую сумму инвариантных циклических подпространств, натянутых на векторы, входящие в отдельные башни. Поэтому матрица оператора в каноническом базисе будет квазидиагональной, состоящей из „ящи-

ков", отвечающих отдельным башням. Выясним вид этих ящиков. Пусть Q одно из инвариантных подпространств, натянутое на башню высоты j , составленную из векторов $X_0, X_1 = (A - \lambda_1 E)X_0, X_2 = (A - \lambda_2 E)X_1, \dots, X_{j-1} = (A - \lambda_{j-1} E)X_{j-2}$. Ясно, что

$$\begin{aligned} \mathbf{A}\mathbf{X}_0 &= \lambda_i \mathbf{X}_0 + \mathbf{X}_1 \\ \mathbf{A}\mathbf{X}_1 &= \lambda_i \mathbf{X}_1 + \mathbf{X}_2 \\ &\dots \\ \mathbf{A}\mathbf{X}_{j-2} &= \lambda_i \mathbf{X}_{j-2} + \mathbf{X}_{j-1} \\ \mathbf{A}\mathbf{X}_{j-1} &= \lambda_i \mathbf{X}_{j-1}. \end{aligned}$$

Таким образом, оператору A на подпространстве Q в выбранном базисе X_0, \dots, X_{r-1} соответствует матрица

$$\begin{bmatrix} \lambda_i & 0 & \dots & 0 & 0 \\ 1 & \lambda_i & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_i & 0 \\ 0 & 0 & \dots & 1 & \lambda_i \end{bmatrix}.$$

Такая матрица называется каноническим ящиком Жордана.

Во всем пространстве оператору A будет соответствовать квазидиагональная матрица, составленная из канонических ящиков Жордана, т. е. матрица вида:

$$\begin{bmatrix} \lambda_1 & 0 & \dots & 0 & 0 \\ 1 & \lambda_1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_1 & 0 \\ 0 & 0 & \dots & 1 & \lambda_1 \end{bmatrix}$$

Число ящиков Жордана равно числу „башен“ и, следовательно, числу первых „этажей“ этих башен, т. е. числу линейно-независимых собственных векторов оператора A . В свою очередь число ящиков Жордана, содержащих одно и то же собственное значение λ_i , равно числу башен, на которые разбивается базис соответствующего λ_i корневого подпространства, т. е. оно равно числу линейно-независимых собственных векторов, принадлежащих собственному значению λ_i . Максимальный порядок ящиков Жордана, содержащих λ_i , равен кратности собственного значения λ_i как корня минимального полинома. Сумма порядков всех канонических ящиков, содержащих λ_i , равна кратности λ_i как корня характеристического полинома.

Пусть оператор A задан посредством матрицы A , соответствующей оператору A в некотором базисе U_1, \dots, U_n . Пусть V_1, \dots, V_n канонический базис для оператора A . В этом базисе оператору A соответствует каноническая матрица Жордана J . Если обозначить буквой C матрицу преобразования координат от базиса U_1, \dots, U_n к базису V_1, \dots, V_n , то $J = C^{-1}AC$, т. е. J получается из A преобразованием подобия. Это преобразование называется приведением матрицы к канонической форме Жордана. Таким образом, знание канонического базиса дает нам как каноническую матрицу J , так и переходную матрицу C . Нетрудно убедиться и в обратном, если каноническая матрица J равна $C^{-1}AC$, то базис V_1, \dots, V_n , связанный с исходным базисом посредством преобразования координат с матрицей C , будет каноническим базисом для оператора A .

Вычисление канонического базиса для оператора, заданного посредством матрицы, довольно сложно. Но часто бывает важно определить лишь каноническую форму для данной матрицы A без вычисления переходной матрицы C , т. е. без вычисления канонического базиса для соответствующего оператора. Это оказывается возможным различными способами. Один из них связан с детальным изучением матрицы $A - tE$.

Обозначим через $D_i(t)$ общий наибольший делитель всех миноров i -го порядка определителя $|A - tE|$. В частности $D_n(t)$ совпадает с характеристическим полиномом. Можно доказать, что все $D_i(t)$, подобно $D_n(t)$, являются общими для класса подобных матриц. Далее, можно доказать, что $D_i(t)$ делится на $D_{i-1}(t)$. Обозначим

$$\frac{D_i(t)}{D_{i-1}(t)} = E_i(t).$$

Очевидно, что $D_n(t) = \prod_{i=1}^n E_i(t)$.

Далее оказывается, что $E_n(t) = \frac{D_n(t)}{D_{n-1}(t)}$ есть минимальный полином матрицы.

Разложим $E_i(t)$ на линейные множители. Тогда

$$E_i(t) = \prod_{j=1}^s (\lambda_j - t)^{m_{ij}}.$$

Здесь s обозначает число различных собственных чисел, $\sum_{i=1}^n m_{ij} = n_j$, $\sum_{j=1}^s \sum_{i=1}^n m_{ij} = n$. Очевидно, что среди показателей m_{ij} лишь немногие отличны от нуля.

Биномы $(\lambda_j - t)^{m_{ij}}$ называются элементарными делителями матрицы A . Знание элементарных делителей позволяет построить каноническую форму. Именно, ящики Жордана строятся, исходя из чисел λ_j , и порядки этих ящиков равны показателям m_{ij} . Число ящиков, содержащих λ_j , равно числу неравных нулю показателей m_{ij} .

В случае, если элементарные делители линейные, т. е. если все отличные от нуля показатели m_{ij} равны единице, ящики Жордана вырождаются в диагональные элементы, а каноническая форма оказывается просто диагональной формой, причем, конечно, одно и то же собственное число будет входить в качестве диагонального элемента столько раз, какова его кратность как корня характеристического полинома.

Верно и обратное, именно: если матрица приводится к диагональному виду, ее элементарные делители линейны. Таким образом, матрицы с различными собственными числами обладают линейными элементарными делителями.

Если все элементарные делители $(\lambda_j - t)^{m_{ij}}$ взаимно просты (что имеет место в том и только в том случае, если $D_{n-1}(t) = 1$), то каждое собственное число входит только в один канонический ящик, причем порядок ящика равен кратности соответствующего собственного числа. В этом и только в этом случае минимальный полином матрицы совпадает с характеристическим.

§ 8. Строение инвариантных подпространств

1. Строение инвариантных подпространств общего вида.

Теорема 8.1. Любое инвариантное подпространство есть прямая сумма пересечений этого подпространства с корневыми.

Доказательство. Пусть T какое-либо инвариантное подпространство. Обозначим через T_i пересечение T с корневым подпространством P_i . Очевидно, что подпространство T_i инвариантно. Покажем, что подпространство разбивается в прямую сумму подпространств $T_1 + \dots + T_s$, где s есть число различных собственных значений оператора. Включение $T_1 + \dots + T_s \subset T$ тривиально, ибо все $T_i \subset T$. Сумма $T_1 + \dots + T_s$ прямая, так как T_i входят в подпространства P_i .

Докажем теперь обратное включение $T \subset T_1 + T_2 + \dots + T_s$. Пусть $X \in T$. Разложим X по корневым подпространствам

$$X = X_1 + X_2 + \dots + X_s.$$

Тогда, как мы видели,¹⁾ $X_i \in T$ и, следовательно, $X_i \in T \cap P_i = T_i$. Тем самым требуемое включение доказано.

Заметим, что некоторые из T_i могут быть и нулевыми подпространствами.

Из доказанной теоремы следует, что всякое инвариантное подпространство есть прямая сумма инвариантных подпространств, содержащихся в корневых.

2. Строение циклического подпространства.

Теорема 8.2. Пусть X_0 произвольный вектор пространства R и Q циклическое подпространство для оператора A , порожденное вектором X_0 . Тогда Q есть прямая сумма циклических подпространств, порожденных проекциями вектора X_0 на корневые подпространства.

Доказательство. По теореме 8.1 Q есть прямая сумма подпространств $Q_i = Q \cap P_i$.

Разложим вектор X_0 по корневым подпространствам

$$X_0 = X_1 + \dots + X_s; \quad X_i \in P_i.$$

Тогда циклические подпространства Q'_i , порожденные векторами X_i , входят соответственно в Q_i , а их прямая сумма Q' входит в Q . Но Q' , очевидно, инвариантное. Следовательно, $Q' = Q$ и все $Q'_i = Q_i$ при $i = 1, 2, \dots, s$.

Пусть высота вектора X_i равна j_i . Тогда векторы

$$X_i, (A - \lambda_i E) X_i, \dots, (A - \lambda_i E)^{j_i - 1} X_i$$

линейно-независимы и порождают инвариантное подпространство, которое совпадает с Q_i вследствие минимальности Q_i . Следовательно, эти векторы образуют базис Q_i и потому размерность Q_i совпадает с высотой j_i вектора X_i . Отсюда непосредственно вытекают следующие теоремы.

Теорема 8.3. Размерность циклического подпространства оператора A , порожденного вектором X_0 , равна сумме высот проекций порождающего вектора X_0 на корневые подпространства.

Теорема 8.4. Минимальный аннулирующий X_0 полином равен $(t - \lambda_1)^{j_1} \dots (t - \lambda_s)^{j_s}$.

Действительно, этот полином, очевидно, аннулирует все проекции X_1, \dots, X_s вектора X_0 . Обратно, если $\theta(A)X_0 = 0$, то $\theta(A)X_1 = \dots = \theta(A)X_s = 0$. Но полином наименьшей степени, аннулирующий X_i , есть $(t - \lambda_i)^{j_i}$, либо $(A - \lambda_i E)^{j_i - 1} X_i \neq 0$.

¹⁾ Замечание к теореме 7.6.

Из теоремы 8.4 следует, что для любого делителя $g(t)$ минимального полинома оператора \mathbf{A} найдется вектор \mathbf{X}_0 , для которого этот делитель будет минимальным аннулирующим полиномом (теорема Лузина — Хлодовского¹⁾), ибо в подпространстве P_i существуют векторы с любой высотой от 0 до m_i включительно.

§ 9. Ортогональность векторов и подпространств

Настоящий параграф, а также и параграф 10 посвящены описанию свойств евклидова и унитарного пространств. Ввиду полного параллелизма теории, мы будем излагать факты и их доказательства в терминах n -мерного унитарного пространства, делая в случае надобности оговорки, касающиеся специфики пространства Эвклида.

1. Ортогональные системы векторов. Два вектора пространства называются ортогональными, если их скалярное произведение равно нулю. Система векторов называется ортогональной, если любые два вектора системы ортогональны друг другу. В последующем, говоря об ортогональной системе, мы будем всегда предполагать, что все векторы этой системы отличны от нуля.

Теорема 9.1. Векторы, образующие ортогональную систему, линейно-независимы.

Доказательство. Пусть $\mathbf{X}_1, \dots, \mathbf{X}_k$ ортогональная система и пусть

$$c_1\mathbf{X}_1 + \dots + c_i\mathbf{X}_i + \dots + c_k\mathbf{X}_k = \mathbf{0}.$$

В силу свойств скалярного произведения имеем

$$\begin{aligned} 0 &= (c_1\mathbf{X}_1 + \dots + c_i\mathbf{X}_i + \dots + c_k\mathbf{X}_k, \mathbf{X}_i) = \\ &= c_1(\mathbf{X}_1, \mathbf{X}_i) + \dots + c_i(\mathbf{X}_i, \mathbf{X}_i) + \dots + c_k(\mathbf{X}_k, \mathbf{X}_i) = c_i |\mathbf{X}_i|^2 \end{aligned}$$

и, так как $|\mathbf{X}_i|^2 > 0$, $c_i = 0$ при любом $i = 1, 2, \dots, k$. Таким образом, единственными значениями для c_1, c_2, \dots, c_k в равенстве $c_1\mathbf{X}_1 + \dots + c_k\mathbf{X}_k = \mathbf{0}$ являются $c_1 = c_2 = \dots = c_k = 0$, т. е. векторы $\mathbf{X}_1, \dots, \mathbf{X}_k$ линейно-независимы. Отсюда вытекает, во-первых, что число векторов, образующих ортогональную систему, не превышает n и, во-вторых, что любая ортогональная система из n векторов образует базис пространства. Такой базис называется ортогональным. Если, кроме того, длины всех векторов ортогонального базиса равны единице, то базис называется ортонормальным.

В арифметическом пространстве, в котором скалярное произведение введено по формуле $(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^n x_i \bar{y}_i$, естественный базис e_1, \dots, e_n является ортонормальным.

¹⁾ Н. Н. Лузин [1], [2], И. Н. Хлодовский [1].

От любой системы линейно-независимых векторов X_1, \dots, X_k можно перейти к ортогональной системе векторов Y_1, \dots, Y_k посредством так называемого процесса ортогонализации. Этот процесс описывается следующей теоремой.

Теорема 9.2. Пусть X_1, \dots, X_k линейно-независимы. Тогда можно построить ортогональную систему векторов Y_1, \dots, Y_k , связанную с исходной соотношениями:

$$\begin{aligned} Y_1 &= X_1 \\ Y_2 &= X_2 + \alpha_{21}X_1 \\ &\dots \\ Y_k &= X_k + \alpha_{k1}X_1 + \dots + \alpha_{kk-1}X_{k-1}. \end{aligned} \tag{1}$$

Для доказательства проведем следующее индуктивное построение. Положим $Y_1 = X_1$. Допустим далее, что векторы Y_1, \dots, Y_{m-1} уже построены и отличны от нуля. Ищем Y_m в виде

$$Y_m = X_m - \gamma_{m1}Y_1 - \dots - \gamma_{m,m-1}Y_{m-1}. \tag{2}$$

Подберем коэффициенты $\gamma_{m1}, \dots, \gamma_{m,m-1}$ так, чтобы $(Y_m, Y_j) = 0$ при $j = 1, \dots, m-1$. Это легко сделать, ибо

$$(Y_m, Y_j) = (X_m, Y_j) - \gamma_{mj}(Y_j, Y_j).$$

Но $(Y_j, Y_j) \neq 0$, так как $Y_j \neq 0$ в силу индукционного предположения и, следовательно, достаточно взять

$$\gamma_{mj} = \frac{(X_m, Y_j)}{(Y_j, Y_j)}. \tag{3}$$

Подставив теперь в равенство (2) вместо Y_1, \dots, Y_{m-1} их выражения через X_1, \dots, X_{m-1} , получим окончательно

$$Y_m = X_m + \alpha_{m1}X_1 + \dots + \alpha_{m,m-1}X_{m-1}.$$

Остается проверить, что $Y_m \neq 0$. Но это очевидно, ибо иначе вектор X_m был бы линейной комбинацией векторов X_1, \dots, X_{m-1} , что противоречит условию теоремы.

Замечание. Линейная независимость векторов X_1, \dots, X_k в процессе доказательства используется лишь для установления того, что каждый построенный вектор отличен от нуля. Поэтому, если процесс ортогонализации применить к системе линейно-зависимых векторов, то по ходу процесса обязательно окажется построен нулевой вектор. Это произойдет в первый раз точно на r -м шагу, если векторы X_1, \dots, X_{r-1} линейно-независимы, а вектор X_r является их линейной комбинацией. Поэтому процесс ортогонализации может

применяться для проверки линейной независимости или установления линейной зависимости данной системы векторов.

От любой ортогональной системы векторов легко перейти далее к ортонормальной системе, деля каждый вектор системы на его длину.

Описанный процесс создает широкий произвол в выборе ортонормального базиса. Действительно, от любого базиса можно перейти к ортонормальному посредством ортогонализации и нормирования.

Скалярное произведение двух векторов очень просто выражается через координаты этих векторов в любом ортонормальном базисе.

Действительно, если U_1, \dots, U_n ортонормальный базис и

$$X = \xi_1 U_1 + \dots + \xi_n U_n, \quad Y = \eta_1 U_1 + \dots + \eta_n U_n,$$

то

$$(X, Y) = (\xi_1 U_1 + \dots + \xi_n U_n, \eta_1 U_1 + \dots + \eta_n U_n) =$$

$$= \sum_{i=1}^n \sum_{j=1}^n (\xi_i U_i, \eta_j U_j) = \sum_{i=1}^n \sum_{j=1}^n \xi_i \bar{\eta}_j (U_i, U_j) = \sum_{i=1}^n \xi_i \bar{\eta}_i. \quad (4)$$

Таким образом, скалярное произведение выражается через координаты векторов в любом ортонормальном базисе по формуле, совпадающей с формулой, выражающей скалярное произведение векторов в арифметическом пространстве через компоненты векторов. Тем самым сопоставление каждому вектору столбца из его координат в ортонормальном базисе дает отображение общего линейного пространства на арифметическое (комплексное для унитарного, вещественное для Евклидова) пространство, изоморфное не только по отношению к действиям сложения и умножения на число, но и по отношению к действию скалярного умножения.

2. Преобразование координат при изменении ортонормального базиса. Пусть U_1, \dots, U_n и U'_1, \dots, U'_n два ортонормальных базиса. Покажем, что матрица преобразования координат при переходе от первого базиса ко второму будет унитарной матрицей, если \mathbb{R} унитарное пространство, и ортогональной матрицей, если \mathbb{R} пространство Эвклида.

Действительно, столбцами этой матрицы

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

будут координаты векторов U'_1, \dots, U'_n в базисе U_1, \dots, U_n . В силу того, что $(U'_i, U'_j) = \delta_{ij}$, где δ_{ij} — символ Кронекера, и в силу ортонормальности базиса U_1, \dots, U_n , имеют место соотношения

$$\sum_{k=1}^n a_{ki} \bar{a}_{kj} = \delta_{ij}$$

(для унитарного пространства) и

$$\sum_{k=1}^n a_{ki} a_{kj} = \delta_{ij}$$

(для евклидова пространства).

Это и означает, что матрица A унитарна (для унитарного пространства) или ортогональна (для евклидова пространства).

3. Ортогонально-дополнительное подпространство. Вектор Z называется ортогональным к подпространству P , если он ортогонален к любому вектору, принадлежащему P . Это обстоятельство записывается в виде $Z \perp P$.

Два пространства называются взаимно ортогональными, если каждый вектор одного ортогонален к любому вектору другого. Взаимно ортогональные подпространства не имеют общих векторов кроме нулевого. Действительно, вектор X , принадлежащий двум взаимно ортогональным подпространствам, удовлетворяет условию $(X, X) = 0$, из которого следует, что X нулевой вектор.

Совокупность всех векторов R , ортогональных ко всем векторам подпространства P , образует, очевидно, подпространство Q , которое называется ортогонально-дополнительным подпространством к P или его ортогональным дополнением. Построить это подпространство можно, например, следующим образом. Пусть подпространство P имеет размерность p . Возьмем базис P и дополним его до базиса всего пространства. К построенному базису применим процесс ортогонализации. Пусть $U_1, \dots, U_p, U_{p+1}, \dots, U_n$ получившийся ортогональный базис. Очевидно, что векторы U_1, \dots, U_p образуют ортогональный базис подпространства P . Покажем, что подпространство, натянутое на векторы U_{p+1}, \dots, U_n , будет искомым ортогонально-дополнительным подпространством Q . Действительно, если $Z \in Q$, то Z ортогонален ко всем U_1, \dots, U_p и, следовательно, в разложении

$$Z = c_1 U_1 + \dots + c_p U_p + c_{p+1} U_{p+1} + \dots + c_n U_n$$

коэффициенты c_1, \dots, c_p равны нулю. Обратно, если

$$Z = c_{p+1} U_{p+1} + \dots + c_n U_n,$$

то Z ортогонален U_1, \dots, U_p и, следовательно, Z ортогонален к любой их линейной комбинации, т. е. к любому вектору, принадлежащему P .

Из приведенного построения следует, что размерность q ортогонального дополнения равна $n - p$. Легко доказать, что ортогональное дополнение к ортогональному дополнению есть исходное подпространство. Действительно, пусть Q есть ортогональное дополнение к P , P_1 ортогональное дополнение к Q . Из определения ясно, что $P \subset P_1$. Но размерности P и P_1 одинаковы. Следовательно, $P_1 = P$.

Далее, из того же построения следует, что все пространство есть прямая сумма подпространства и его ортогонального дополнения. Действительно, объединение базисов этих подпространств есть базис всего пространства. Следовательно, любой вектор X пространства R однозначно представляется в виде

$$X = Y + Z,$$

где $Y \in P$, $Z \perp P$. Вектор Y называется ортогональной проекцией вектора X на подпространство P .

4. Двойственный базис. Пусть U_1, \dots, U_n некоторый базис в пространстве R . Базис V_1, \dots, V_n называется двойственным с U_1, \dots, U_n , если

$$\begin{aligned} (U_i, V_i) &= 1, \\ (U_i, V_j) &= 0 \quad \text{при } i \neq j. \end{aligned} \tag{5}$$

Для каждого базиса существует единственный двойственный базис. Действительно, пусть Q_1 есть подпространство, натянутое на векторы U_2, \dots, U_n , и пусть S_1 ортогональное дополнение к Q_1 . Очевидно, что размерность Q_1 равна $n - 1$, размерность S_1 равна 1. Пусть $Z \neq 0$ любой вектор из S_1 . Тогда $(Z, U_2) = \dots = (Z, U_n) = 0$ и, следовательно, $(Z, U_1) = a_1 \neq 0$. Поэтому вектор $V_1 = \frac{1}{a_1} Z$ удовлетворяет условиям (5). Аналогичным образом строятся и все остальные векторы двойственного базиса. Единственность двойственного базиса следует непосредственно из определения.

Очевидно, что базис будет двойственен самому себе в том и только в том случае, если он ортонормален.

Если базис V_1, \dots, V_n двойственен с базисом U_1, \dots, U_n , то базис U_1, \dots, U_n двойственен с базисом V_1, \dots, V_n , т. е. отношение двойственности базисов симметрично. Поэтому имеет смысл говорить о паре взаимно двойственных базисов (биортогональные базисы).

Введение двойственного базиса дает возможность представлять координаты вектора в виде скалярных произведений. Именно, если координаты вектора X в базисе U_1, \dots, U_n суть числа x_1, \dots, x_n , то

$$x_i = (X, V_i).$$

Действительно, $(X, V_i) = (x_1 U_1 + \dots + x_n U_n, V_i) = x_i$. В свою очередь

$$x'_i = (X, U_i),$$

где x'_1, \dots, x'_n координаты вектора X в базисе V_1, \dots, V_n . В тензорной алгебре координаты x_1, \dots, x_n называются контравариантными относительно базиса U_1, \dots, U_n , а координаты x'_1, \dots, x'_n ковариантными относительно того же базиса. Одновременное употребление тех и других координат представляет большие удобства.

Так, например, скалярное произведение двух векторов выражается через их координаты по отношению к любому (не ортогональному) базису следующей простой формулой

$$(X, Y) = \sum_{i=1}^n x_i \bar{y}_i = \sum_{i=1}^n x'_i \bar{y}_i.$$

Укажем один индуктивный способ построения базиса двойственного к данному, напоминающий процесс ортогонализации. Пусть U_1, \dots, U_n базис, для которого нужно строить двойственный, и пусть $V_1^{(0)}, \dots, V_n^{(0)}$ какой-нибудь другой базис. Допустим, что определители

$$\Delta_1 = (V_1^{(0)}, U_1), \quad \Delta_2 = \begin{vmatrix} (V_1^{(0)}, U_1) & (V_2^{(0)}, U_1) \\ (V_1^{(0)}, U_2) & (V_2^{(0)}, U_2) \end{vmatrix}, \dots$$

$$\dots, \Delta_n = \begin{vmatrix} (V_1^{(0)}, U_1) & \dots & (V_n^{(0)}, U_1) \\ \vdots & \ddots & \vdots \\ (V_1^{(0)}, U_n) & \dots & (V_n^{(0)}, U_n) \end{vmatrix}$$

отличны от нуля.

Построим последовательно системы векторов

$$\{V_1^{(1)}, \dots, V_n^{(1)}\}, \{V_1^{(2)}, \dots, V_n^{(2)}\}, \dots, \{V_1^{(n)}, \dots, V_n^{(n)}\}$$

так, чтобы k -я система удовлетворяла первым k группам условий биортогональности

$$(V_i^{(k)}, U_j) = \delta_{ij} \text{ при } i = 1, 2, \dots, n; j = 1, \dots, k.$$

Возьмем

$$V_1^{(1)} = \frac{1}{(V_1^{(0)}, U_1)} V_1^{(0)}$$

$$V_i^{(1)} = V_i^{(0)} - (V_i^{(0)}, U_1) V_1^{(1)} \text{ при } i > 1.$$

Ясно, что $(V_1^{(1)}, U_1) = 1$ и $(V_i^{(1)}, U_1) = 0$ при $i > 1$.

Заметим, что

$$(V_2^{(1)}, U_2) = (V_2^{(0)}, U_2) - (V_2^{(0)}, U_1) \frac{(V_1^{(0)}, U_2)}{(V_1^{(0)}, U_1)} = \frac{\Delta_2}{\Delta_1} \neq 0.$$

Допустим, что мы уже построили векторы $V_1^{(k-1)}, V_2^{(k-1)}, \dots, V_n^{(k-1)}$, удовлетворяющие поставленным условиям, и убедились в том, что $(V_k^{(k-1)}, U_k) = \frac{\Delta_k}{\Delta_{k-1}} \neq 0$. Положим

$$V_k^{(k)} = \frac{1}{(V_k^{(k-1)}, U_k)} V_k^{(k-1)}$$

$$V_i^{(k)} = V_i^{(k-1)} - (V_i^{(k-1)}, U_k) V_k^{(k)}, \quad i \neq k.$$

Тогда

$$(\mathbf{V}_k^{(k)}, \mathbf{U}_k) = 1, \quad (\mathbf{V}_l^{(k)}, \mathbf{U}_k) = 0 \quad \text{при } l \neq k$$

и

$$\begin{aligned} (\mathbf{V}_i^{(k)}, \mathbf{U}_j) &= (\mathbf{V}_i^{(k-1)}, \mathbf{U}_j) - \frac{(\mathbf{V}_i^{(k-1)}, \mathbf{U}_k)}{(\mathbf{V}_k^{(k-1)}, \mathbf{U}_k)} (\mathbf{V}_k^{(k-1)}, \mathbf{U}_j) = \\ &= (\mathbf{V}_i^{(k-1)}, \mathbf{U}_j) = \delta_{ij} \quad \text{при } j \leq k-1. \end{aligned}$$

Остается убедиться в том, что $(\mathbf{V}_{k+1}^{(k)}, \mathbf{U}_{k+1}) = \frac{\Delta_{k+1}}{\Delta_k}$ и потому $(\mathbf{V}_{k+1}^{(k)}, \mathbf{U}_{k+1}) \neq 0$. С этой целью рассмотрим последовательность матриц

$$\begin{aligned} A^{(k)} &= \begin{bmatrix} (\mathbf{V}_1^{(k)}, \mathbf{U}_1) & \dots & (\mathbf{V}_1^{(k)}, \mathbf{U}_n) \\ \vdots & \ddots & \vdots \\ (\mathbf{V}_n^{(k)}, \mathbf{U}_1) & \dots & (\mathbf{V}_n^{(k)}, \mathbf{U}_n) \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0 & \dots & 0 & (\mathbf{V}_1^{(k)}, \mathbf{U}_{k+1}) & \dots & (\mathbf{V}_1^{(k)}, \mathbf{U}_n) \\ \vdots & \ddots & & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & \dots & 1 & (\mathbf{V}_k^{(k)}, \mathbf{U}_{k+1}) & \dots & (\mathbf{V}_k^{(k)}, \mathbf{U}_n) \\ 0 & 0 & \dots & 0 & (\mathbf{V}_{k+1}^{(k)}, \mathbf{U}_{k+1}) & \dots & (\mathbf{V}_{k+1}^{(k)}, \mathbf{U}_n) \\ \vdots & \ddots & & \ddots & \ddots & \ddots & \ddots \end{bmatrix}. \end{aligned}$$

Матрица $A^{(k)}$ получается из матрицы $A^{(k-1)}$ делением k -й строки на число $(\mathbf{V}_k^{(k-1)}, \mathbf{U}_k) = \frac{\Delta_k}{\Delta_{k-1}}$ и добавлением k -й строки с надлежащим множителем ко всем остальным. Поэтому все главные миноры матрицы $A^{(k)}$, начиная с минора порядка k , равны произведениям соответствующих миноров матрицы $A^{(k-1)}$ на $\frac{\Delta_{k-1}}{\Delta_k}$.

Аналогично, все главные миноры матрицы $A^{(k-1)}$, начиная с минора порядка $k-1$, равны соответствующим минорам матрицы $A^{(k-2)}$, умноженным на $\frac{\Delta_{k-2}}{\Delta_{k-1}}$ и т. д. Применяя это к главному минору порядка $k+1$ матрицы $A^{(k)}$, который, очевидно, равен $(\mathbf{V}_{k+1}^{(k)}, \mathbf{U}_{k+1})$, получим

$$(\mathbf{V}_{k+1}^{(k)}, \mathbf{U}_{k+1}) = \frac{\Delta_{k-1}}{\Delta_k} \cdot \frac{\Delta_{k-2}}{\Delta_{k-1}} \cdots \frac{1}{\Delta_1} \Delta_{k+1} = \frac{\Delta_{k+1}}{\Delta_k}.$$

Вычисление базиса $\mathbf{V}_1, \dots, \mathbf{V}_n$, двойственного к базису $\mathbf{U}_1, \dots, \mathbf{U}_n$, по существу равносильно обращению матрицы A , составленной из координат векторов $\mathbf{U}_1, \dots, \mathbf{U}_n$ по отношению к некоторому ортонормальному базису. Действительно, если матрица B составлена из

координат векторов V_1, \dots, V_n в том же базисе, то

$$B^*A = E,$$

что непосредственно следует из правила умножения матриц и условий ортогональности.

§ 10. Линейные операторы в унитарном пространстве и в евклидовом пространстве

1. Сопряженный оператор. Пусть A оператор, определенный в унитарном пространстве. Оператор A^* называется сопряженным с A , если для любых векторов X, Y выполняется равенство

$$(AX, Y) = (X, A^*Y).$$

Докажем существование и единственность сопряженного оператора. Пусть оператору A в некотором ортонормальном базисе соответствует матрица

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

векторам X и Y в том же базисе соответствуют столбцы из координат $(x_1, x_2, \dots, x_n)'$ и $(y_1, y_2, \dots, y_n)'$. Тогда вектору AX будет соответствовать столбец

$$\begin{bmatrix} a_{11}x_1 + \dots + a_{1n}x_n \\ a_{21}x_1 + \dots + a_{2n}x_n \\ \dots & \dots & \dots \\ a_{n1}x_1 + \dots + a_{nn}x_n \end{bmatrix}.$$

Следовательно,

$$\begin{aligned} (AX, Y) &= a_{11}x_1\bar{y}_1 + \dots + a_{1n}x_n\bar{y}_1 + \\ &\quad + a_{21}x_1\bar{y}_2 + \dots + a_{2n}x_n\bar{y}_2 + \\ &\quad + \dots \dots \dots \dots \dots + \\ &\quad + a_{n1}x_1\bar{y}_n + \dots + a_{nn}x_n\bar{y}_n = \\ &= x_1(a_{11}\bar{y}_1 + a_{21}\bar{y}_2 + \dots + a_{n1}\bar{y}_n) + \\ &\quad + \dots \dots \dots \dots \dots + \\ &\quad + x_n(a_{1n}\bar{y}_1 + a_{2n}\bar{y}_2 + \dots + a_{nn}\bar{y}_n) = \\ &= (X, A^*Y). \end{aligned}$$

где A^* оператор, имеющий в том же базисе матрицу

$$A^* = \begin{bmatrix} \bar{a}_{11} & \bar{a}_{21} & \dots & \bar{a}_{n1} \\ \bar{a}_{12} & \bar{a}_{22} & \dots & \bar{a}_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{a}_{1n} & \bar{a}_{2n} & \dots & \bar{a}_{nn} \end{bmatrix},$$

сопряженную с матрицей A .

Таким образом, в качестве сопряженного оператора может быть взят оператор, имеющий в некотором ортонормальном базисе матрицу, сопряженную с матрицей исходного оператора в том же базисе.

Докажем теперь единственность сопряженного оператора. Пусть A_1^* и A_2^* два оператора, сопряженных с оператором A . Тогда $(AX, Y) = (X, A_1^*Y) = (X, A_2^*Y)$, откуда $(X, (A_1^* - A_2^*)Y) = 0$. Следовательно, вектор $(A_1^* - A_2^*)Y$ ортогонален к любому вектору X пространства R и потому $(A_1^* - A_2^*)Y = 0$ при любом Y . Отсюда заключаем, что $A_1^* - A_2^*$ есть нулевой оператор, т. е. $A_1^* = A_2^*$.

Из определения и единственности сопряженного оператора ясно, что $(A^*)^* = A$.

В евклидовом пространстве сопряженному оператору (по отношению к ортонормальному базису) будет соответствовать матрица, транспонированная с матрицей оператора A (по отношению к тому же базису), ибо операторы в евклидовом пространстве представляются вещественными матрицами.

2. Свойства собственных значений и собственных векторов операторов в унитарном пространстве. Выясним некоторые взаимоотношения между собственными значениями и собственными векторами взаимно сопряженных операторов A и A^* .

Прежде всего отметим, что характеристические полиномы этих операторов имеют комплексно-сопряженные коэффициенты и потому собственные значения оператора A^* будут комплексно сопряжены с собственными значениями оператора A . Это следует из того, что матрица A^* оператора A^* в некотором ортонормальном базисе сопряжена с матрицей A оператора A в том же базисе. В случае, если матрица оператора A вещественна в некотором ортонормальном базисе, то характеристические полиномы и собственные значения операторов A и A^* совпадают.

Взаимосвязи между собственными векторами характеризуются следующей теоремой.

Теорема 10.1. Если X_i собственный вектор оператора A , соответствующий собственному значению λ_i , и Y_j собственный вектор оператора A^* , соответствующий собственному значению $\bar{\lambda}_j$, то $(X_i, Y_j) = 0$ при $\lambda_i \neq \bar{\lambda}_j$.

Доказательство. Подсчитаем двумя способами (AX_i, Y_j) . С одной стороны,

$$(AX_i, Y_j) = (\lambda_i X_i, Y_j) = \lambda_i (X_i, Y_j).$$

С другой стороны,

$$(AX_i, Y_j) = (X_i, A^*Y_j) = (X_i, \bar{\lambda}_j Y_j) = \bar{\lambda}_j (X_i, Y_j).$$

Поэтому

$$(\lambda_i - \bar{\lambda}_j)(X_i, Y_j) = 0$$

и так как $\lambda_i - \bar{\lambda}_j \neq 0$, то

$$(X_i, Y_j) = 0,$$

что и требовалось доказать.

Отсюда следует, что если все собственные значения $\lambda_1, \dots, \lambda_n$ оператора A различны, то для собственных векторов X_1, \dots, X_n оператора A и собственных векторов Y_1, \dots, Y_n оператора A^* имеют место $n^2 - n$ соотношений ортогональности, именно $(X_i, Y_j) = 0$, $i \neq j$. Покажем теперь, что, выбрав каким-либо способом векторы Y_1, \dots, Y_n , мы можем нормировать векторы X_1, \dots, X_n так, что $(X_i, Y_i) = 1$.

Прежде всего покажем, что $(X_i, Y_i) = a_i \neq 0$. Действительно, если бы $(X_i, Y_i) = 0$, то вектор Y_i был бы ортогонален ко всем собственным векторам $X_1, \dots, X_i, \dots, X_n$ оператора A , а следовательно, и ко всем векторам пространства R , что означало бы, что Y_i нулевой вектор. Взяв вместо векторов X_1, \dots, X_n векторы $\frac{1}{a_1} X_1, \dots, \dots, \frac{1}{a_n} X_n$, мы получим требуемое нормирование, ибо

$$\left(\frac{1}{a_i} X_i, Y_i \right) = \frac{1}{a_i} (X_i, Y_i) = 1.$$

Таким образом, системы векторов X_1, \dots, X_n и Y_1, \dots, Y_n после проведения нормировки образуют взаимно двойственные базисы пространства.

3. Две группы соотношений ортогональности для собственных векторов матрицы. Пусть A матрица, собственные значения которой $\lambda_1, \dots, \lambda_n$ различны, и пусть X_1, \dots, X_n соответствующие им собственные векторы (столбцы). Как мы видели, сопряженная матрица A^* имеет собственными значениями числа комплексно-сопряженные с $\lambda_1, \dots, \lambda_n$. Пусть Y_1, \dots, Y_n собственные векторы матрицы A^* , нормированные согласно предыдущему пункту. Составим матрицы

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix} \quad \text{и} \quad Y = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \dots & y_{nn} \end{bmatrix},$$

столбцы которых составлены из компонент собственных векторов матрицы A и матрицы A^* соответственно. Выведенные выше соотношения ортогональности и нормированности в координатной записи имеют вид

$$x_{1i}\bar{y}_{1j} + x_{2i}\bar{y}_{2j} + \dots + x_{ni}\bar{y}_{nj} = \begin{cases} 0 & (i \neq j) \\ 1 & (i = j) \end{cases},$$

что равносильно матричному равенству

$$Y^*X = E,$$

где Y^* матрица, сопряженная с матрицей Y . Заметим, что i -я строка матрицы Y^* состоит из компонент собственного вектора матрицы A' , транспонированной с A , соответствующего собственному значению λ_i . Из равенства $Y^*X = E$ следует, что

$$XY^* = E,$$

что дает вторую группу соотношений ортогональности

$$x_{i1}\bar{y}_{j1} + x_{i2}\bar{y}_{j2} + \dots + x_{in}\bar{y}_{jn} = \begin{cases} 0 & (i \neq j) \\ 1 & (i = j) \end{cases}.$$

4. Свойства ортогональности корневых векторов.

Теорема 10.2. Любой корневой вектор оператора A , соответствующий значению λ , ортогонален к любому корневому вектору оператора A^* , соответствующему собственному значению $\mu \neq \bar{\lambda}$.

Доказательство. Пусть X_0 корневой вектор оператора A высоты m , соответствующий собственному значению λ , и Y_0 корневой вектор сопряженного оператора A^* высоты k , соответствующий собственному значению $\mu \neq \bar{\lambda}$. Построим цепочки корневых векторов:

$$X_1 = (A - \lambda E) X_0$$

$$X_2 = (A - \lambda E) X_1$$

$$\dots \dots \dots$$

$$X_{m-1} = (A - \lambda E) X_{m-2}$$

$$(A - \lambda E) X_{m-1} = 0$$

и

$$Y_1 = (A^* - \mu E) Y_0$$

$$Y_2 = (A^* - \mu E) Y_1$$

$$\dots \dots \dots$$

$$Y_{k-1} = (A^* - \mu E) Y_{k-2}$$

$$(A^* - \mu E) Y_{k-1} = 0.$$

Векторы X_{m-i} и Y_{k-j} будут собственными для операторов A и A^* соответственно, принадлежащими собственным значениям λ и μ . Следовательно, X_{m-i} и Y_{k-j} ортогональны. Доказательство теоремы проведем по индукции. Пусть уже установлено, что $(X_{m-i}, Y_{k-j}) = 0$ при $i+j < l$, $l \geq 3$. Докажем тогда, что $(X_{m-i}, Y_{k-j}) = 0$ и при $i+j = l$. Действительно,

$$\begin{aligned} \lambda(X_{m-i}, Y_{k-j}) &= (\lambda X_{m-i}, Y_{k-j}) = (AX_{m-i} - X_{m-i+1}, Y_{k-j}) = \\ &= (AX_{m-i}, Y_{k-j}) - (X_{m-i+1}, Y_{k-j}) = (X_{m-i}, A^*Y_{k-j}) - \\ &- (X_{m-i+1}, Y_{k-j}) = (X_{m-i}, \mu Y_{k-j}) + (X_{m-i}, Y_{k-j+1}) - \\ &- (X_{m-i+1}, Y_{k-j}) = \bar{\mu}(X_{m-i}, Y_{k-j}), \end{aligned}$$

ибо $(X_{m-i}, Y_{k-j+1}) = 0$ и $(X_{m-i+1}, Y_{k-j}) = 0$ в силу индукционного предположения. Отсюда $(\lambda - \bar{\mu})(X_{m-i}, Y_{k-j}) = 0$ и, следовательно, $(X_{m-i}, Y_{k-j}) = 0$, так как $\lambda \neq \bar{\mu}$.

5. Базис двойственный к каноническому.

Теорема 10.3. *Базис двойственный к каноническому базису оператора A есть канонический для сопряженного оператора A^* с «перевернутыми башнями»; точнее, если X_0, \dots, X_{m-1} «башня», взятая из канонического базиса для оператора A , такая, что*

$$\begin{aligned} AX_0 &= \lambda X_0 + X_1 \\ AX_1 &= \lambda X_1 + X_2 \\ &\dots \\ AX_{m-2} &= \lambda X_{m-2} + X_{m-1} \\ AX_{m-1} &= \lambda X_{m-1}. \end{aligned}$$

то двойственно соответствующие векторы Y_0, \dots, Y_{m-1} удовлетворяют соотношениям

$$\begin{aligned} A^*Y_{m-1} &= \bar{\lambda} Y_{m-1} + Y_{m-2} \\ A^*Y_{m-2} &= \bar{\lambda} Y_{m-2} + Y_{m-3} \\ &\dots \\ A^*Y_0 &= \bar{\lambda} Y_1 + Y_0 \\ A^*Y_0 &= \bar{\lambda} Y_0. \end{aligned}$$

Доказательство. Заметим предварительно, что если оператору A в некотором базисе U_1, \dots, U_n соответствует матрица A , то сопряженному оператору A^* в двойственном базисе V_1, \dots, V_n соответствует матрица A^* , сопряженная с матрицей A .

Действительно, пусть $A = (a_{ij})$ при $i, j = 1, 2, \dots, n$. Тогда $AU_j = \sum_{i=1}^n a_{ij}U_i$, откуда следует, что $a_{ij} = (AU_j, V_i)$. Пусть далее

$B = (b_{ij})$ ($i = 1, \dots, n; j = 1, \dots, n$) матрица, соответствующая оператору A^* в базисе V_1, \dots, V_n . Тогда $b_{ij} = (A^*V_j, U_i)$. Но $(A^*V_j, U_i) = (V_j, AU_i) = (AU_i, V_j) = a_{ji}$. Это и доказывает, что $B = A^*$.

В каноническом базисе оператору A соответствует каноническая матрица Жордана и башне X_0, \dots, X_{m-1} соответствует канонический ящик Жордана

$$\begin{bmatrix} \lambda & & & \\ & 1 & \lambda & \\ & & \ddots & \\ & & & 1 & \lambda \end{bmatrix}.$$

В двойственном базисе оператору A^* соответствует сопряженная матрица. В частности, векторам Y_0, \dots, Y_{m-1} будет соответствовать «ящик»

$$\begin{bmatrix} \bar{\lambda} & 1 & & \\ & \bar{\lambda} & & \\ & & \ddots & \\ & & & 1 & \bar{\lambda} \end{bmatrix}.$$

Это и значит, что

$$\begin{aligned} A^*Y_{m-1} &= \bar{\lambda}Y_{m-1} + Y_{m-2} \\ A^*Y_{m-2} &= \bar{\lambda}Y_{m-2} + Y_{m-3} \\ &\vdots \\ A^*Y_1 &= \bar{\lambda}Y_1 + Y_0 \\ A^*Y_0 &= \bar{\lambda}Y_0, \end{aligned}$$

что и требовалось доказать.

§ 11. Самосопряженный оператор

В настоящем параграфе мы положим в основу не унитарное пространство, а евклидово пространство, так как результаты излагаемой здесь теории, в отличие от общей теории собственных значений, не требуют выхода в комплексное пространство.

Все результаты почти без изменений переносятся и на унитарное пространство. Мы ограничимся лишь формулировками относящихся сюда теорем.

1. Определение. Оператор A называется **самосопряженным**, если для любых векторов X и Y , принадлежащих R , имеет место равенство

$$(AX, Y) = (X, AY),$$

т. е. оператор A совпадает со своим сопряженным. Матрица самосопряженного оператора по отношению к любому ортонормальному базису вещественна (евклидово пространство) и симметрична, так как она должна совпадать со своей транспонированной.

Любая линейная комбинация самосопряженных операторов есть самосопряженный оператор. Далее, если самосопряженный оператор A невырожденный, то обратный для него оператор A^{-1} самосопряжен. Действительно, полагая $A^{-1}X = U$, $A^{-1}Y = V$, имеем $(A^{-1}X, Y) = (U, AV) = (AU, V) = (X, A^{-1}Y)$. Очевидно и обратное, если оператору A в некотором ортонормальном базисе соответствует симметричная матрица, то оператор самосопряженный. Действительно, в этом случае операторам A и A^* отвечает одна и та же матрица, и, следовательно, они совпадают.

Таким образом, если в пространстве выбрать ортонормальный базис, то самосопряженные операторы находятся в естественном одно-однозначном соответствии с симметричными матрицами. Имеется также тесная связь между самосопряженными операторами и квадратичными формами. Именно, скалярное произведение (AX, X) , выраженное через координаты вектора X , есть не что иное как квадратичная форма с матрицей, совпадающей с матрицей A оператора в выбранном ортонормальном базисе. Действительно,

$$(AX, X) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \quad (a_{ij} = a_{ji}).$$

2. Одно свойство инвариантных подпространств самосопряженного оператора.

Теорема II.1. Если P инвариантное подпространство самосопряженного оператора A , то ортогонально-дополнительное подпространство Q есть также инвариантное подпространство.

Доказательство. Пусть $X \in Q$. Это значит, что $(X, Y) = 0$ при любом $Y \in P$. Но тогда при любом $Y \in P$ и $(AX, Y) = (X, AY) = 0$ (ибо $AY \in P$ в силу инвариантности P). Отсюда $AX \in Q$, что и требовалось доказать.

3. Построение системы взаимно ортогональных собственных векторов самосопряженного оператора. Собственные векторы и собственные значения самосопряженного оператора обладают рядом экстремальных свойств, из которых непосредственно вытекают такие важные свойства, как вещественность всех собственных значений и существование ортогонального базиса, составленного из собственных

векторов. В основе этих экстремальных свойств лежит следующая теорема.

Теорема 11.2. Существует максимум отношения $\frac{(AX, X)}{(X, X)}$ при X , пробегающем все пространство (кроме нулевого вектора). Любой вектор, при котором этот максимум достигается, есть собственный вектор оператора A , а величина максимума есть соответствующее собственное значение.

Доказательство. Так как

$$\frac{(AX, X)}{(X, X)} = \frac{(AX, X)}{|X|^2} = \left(A \frac{X}{|X|}, \frac{X}{|X|} \right),$$

то исследуемый максимум равен максимуму (AX, X) при X , изменяющемся на «единичной сфере», т. е. так, что длина X равна единице. Существование максимума (AX, X) непосредственно следует из теоремы Вейерштрасса о достижении точной верхней границы для непрерывной функции на ограниченном замкнутом множестве. Пусть $\lambda_1 = \max_{X \in \mathbb{R}} \frac{(AX, X)}{(X, X)} = (AX_1, X_1)$, где X_1 вектор длины единицы, реализующий максимум. Тогда для любого вектора $X \in \mathbb{R}$

$$(AX, X) \leq \lambda_1 (X, X).$$

Положим $Y = AX_1 - \lambda_1 X_1$ и докажем, что $Y = 0$. Прежде всего покажем, что Y ортогонален к X_1 . Действительно, $(Y, X_1) = (AX_1 - \lambda_1 X_1, X_1) = (AX_1, X_1) - \lambda_1 (X_1, X_1) = 0$ в силу определения λ_1 .

Пусть $X = X_1 + \varepsilon Y$, где ε положительное вещественное число. Имеем

$$(A(X_1 + \varepsilon Y), X_1 + \varepsilon Y) \leq \lambda_1 (X_1 + \varepsilon Y, X_1 + \varepsilon Y),$$

откуда

$$(AX_1, X_1) + \varepsilon (AX_1, Y) + \varepsilon (AY, X_1) + \varepsilon^2 (AY, Y) \leq \lambda_1 (X_1, X_1) + \\ + 2\lambda_1 \varepsilon (X_1, Y) + \lambda_1 \varepsilon^2 (Y, Y).$$

Принимая во внимание, что $(AX_1, X_1) = \lambda_1$; $(AX_1, Y) = (AY, X_1)$; $(X_1, X_1) = 1$ и $(X_1, Y) = 0$, получим

$$2\varepsilon (AX_1, Y) + \varepsilon^2 (AY, Y) \leq \varepsilon^2 \lambda_1 (Y, Y).$$

Поделив обе части на ε и устремив ε к нулю, получим

$$(AX_1, Y) \leq 0.$$

Далее

$$(Y, Y) = (AX_1 - \lambda_1 X_1, Y) = (AX_1, Y) - \lambda_1 (X_1, Y) = (AX_1, Y) \leq 0.$$

Отсюда $Y = 0$, т. е. $AX_1 = \lambda_1 X_1$.

Замечание. Отношение $\frac{(AX, X)}{(X, X)}$ часто называют отношением Релея.

Теорема 11.3. Пусть A самосопряженный оператор. Тогда существует система попарно ортогональных собственных векторов A , образующих базис пространства. Все собственные значения A вещественны.

Доказательство. Пусть X_1 нормированный собственный вектор оператора A , дающий максимум (AX, X) на единичной сфере. Обозначим через P_1 одномерное подпространство, натянутое на X_1 , и через Q_1 его ортогональное дополнение. Очевидно, что P_1 , а следовательно, и Q_1 будут инвариантными подпространствами. Размерность Q_1 равна $n - 1$. Оператор A , рассматриваемый на Q_1 , будет, очевидно, самосопряженным и, следовательно, для него найдется нормированный собственный вектор X_2 , реализующий $\max_{\substack{X \in Q_1; \|X\|=1}} (AX, X)$. Легко видеть, что соответствующее собственное значение λ_2 будет не больше λ_1 . По построению $(X_1, X_2) = 0$.

Пусть P_2 есть подпространство, натянутое на X_1 и X_2 , и Q_2 его ортогональное дополнение (размерности $n - 2$). Подпространство P_2 , а следовательно, и подпространство Q_2 инвариантны. Рассмотрим оператор A на Q_2 . Для этого самосопряженного оператора найдется собственный вектор X_3 , реализующий $\max_{\substack{X \in Q_2; \|X\|=1}} (AX, X)$, и соответствующее собственное значение λ_3 будет не больше λ_2 . Очевидно, что $(X_1, X_3) = 0$ и $(X_2, X_3) = 0$. Продолжая указанный процесс придем к системе попарно ортогональных собственных векторов X_1, \dots, X_n . Соответствующие им собственные значения будут удовлетворять неравенствам $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Покажем, что в этом ряду каждое собственное значение повторяется столько раз, какова его кратность как корня характеристического полинома. Действительно, оператору A в базисе X_1, \dots, X_n соответствует диагональная матрица

$$\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix},$$

и, следовательно, характеристический полином A есть

$$(\lambda_1 - t)(\lambda_2 - t) \dots (\lambda_n - t).$$

Из описанной конструкции следует, что все собственные значения $\lambda_1, \dots, \lambda_n$ вещественные, а также что для каждого кратного собственного значения существует столько попарно ортогональных (и, следовательно, линейно-независимых) собственных векторов, какова кратность этого собственного значения как корня характеристического полинома.

В алгебраической форме теорема 11.3 может быть сформулирована в следующем виде: для любой симметричной матрицы A существует ортогональная матрица P , такая, что $P'AP$ есть диагональная. Действительно, матрицу A можно считать матрицей некоторого самосопряженного оператора \mathbf{A} в ортонормальном базисе. В базисе, составленном из нормированных собственных векторов, матрица Λ оператора \mathbf{A} будет диагональной. Следовательно, $\Lambda = P^{-1}AP$, где P матрица, составленная из компонент нормированных собственных векторов. Матрица P ортогональна, так что $P^{-1} = P'$ и $\Lambda = P'AP$.

Диагональные элементы матрицы Λ являются собственными значениями матрицы A .

Непосредственным следствием проведенной экстремальной конструкции является следующая теорема.

Теорема 11.4. Пусть $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ собственные значения самосопряженного оператора \mathbf{A} , $\mathbf{X}_1, \dots, \mathbf{X}_n$ им принадлежащие попарно ортогональные собственные векторы. Тогда

$$\begin{aligned}\lambda_1 &= \max_{\mathbf{x} \neq 0} \frac{(\mathbf{AX}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})} \\ \lambda_2 &= \max_{\mathbf{x} \perp \mathbf{x}_1} \frac{(\mathbf{AX}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})} \\ &\vdots \\ \lambda_k &= \max_{\substack{\mathbf{x} \perp \mathbf{x}_1 \\ \vdots \\ \mathbf{x} \perp \mathbf{x}_{k-1}}} \frac{(\mathbf{AX}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})}.\end{aligned}$$

Очевидно, что проведенное построение собственных значений и принадлежащих им собственных векторов можно видоизменить, вычисляя вместо последовательных максимумов отношения $\frac{(\mathbf{AX}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})}$ последовательные минимумы этого отношения. При таком построении собственные значения определяются в порядке их возрастания. Поэтому верна следующая теорема.

Теорема 11.5. Пусть $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ собственные числа самосопряженного оператора \mathbf{A} , $\mathbf{X}_1, \dots, \mathbf{X}_n$ соответствующие им попарно ортогональные собственные векторы. Тогда

$$\begin{aligned}\lambda_n &= \min_{\mathbf{x} \neq 0} \frac{(\mathbf{AX}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})} \\ \lambda_{n-1} &= \min_{\mathbf{x} \perp \mathbf{x}_n} \frac{(\mathbf{AX}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})} \\ &\vdots \\ \lambda_k &= \min_{\substack{\mathbf{x} \perp \mathbf{x}_n \\ \vdots \\ \mathbf{x} \perp \mathbf{x}_{k+1}}} \frac{(\mathbf{AX}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})}.\end{aligned}$$

4. Инвариантные подпространства самосопряженного оператора. Пусть $\lambda_1 > \lambda_2 > \dots > \lambda_s$ собственные значения самосопряженного оператора A с кратностями n_1, n_2, \dots, n_s . Как было показано, в пространстве может быть построен базис из собственных векторов оператора A , причем число базисных собственных векторов, соответствующих собственному значению λ_i , равно его кратности n_i . Обозначим через P_i подпространство, натянутое на базисные собственные векторы, соответствующие значению λ_i . Подпространства P_1, \dots, P_s в силу построения взаимно ортогональны, и все пространство R есть прямая сумма подпространств P_1, \dots, P_s . Очевидно, что каждый вектор подпространства P_i есть собственный вектор оператора A , соответствующий собственному значению λ_i . Обратно, каждый собственный вектор, принадлежащий собственному значению λ_i , принадлежит P_i , т. е. P_i есть совокупность всех собственных векторов, принадлежащих λ_i . Действительно, пусть $AX = \lambda_i X$. Разложим вектор X по подпространствам P_i :

$$X = X_1 + \dots + X_i + \dots + X_s, \quad X_j \in P_j.$$

Тогда

$$AX = \lambda_i X = \lambda_i X_1 + \dots + \lambda_i X_i + \dots + \lambda_i X_s.$$

С другой стороны,

$$AX = AX_1 + \dots + AX_i + \dots + AX_s =$$

$$= \lambda_1 X_1 + \dots + \lambda_i X_i + \dots + \lambda_s X_s.$$

В силу однозначности разложения любого вектора по подпространствам P_i , имеем

$$\lambda_i X_j = \lambda_j X_i$$

при всех j , и, следовательно, $X_j = 0$ при $j \neq i$, т. е. $X = X_i \in P_i$. Отсюда вытекает следующее важное свойство собственных векторов самосопряженного оператора A .

Собственные векторы самосопряженного оператора A , принадлежащие различным собственным значениям, ортогональны. Действительно, они принадлежат к взаимно ортогональным подпространствам.

Это свойство собственных векторов легко доказать и непосредственно без использования экстремальной конструкции, ибо если X_i и X_j собственные векторы, принадлежащие λ_i и λ_j и $\lambda_i \neq \lambda_j$, то

$$(\lambda_i - \lambda_j)(X_i, X_j) = (\lambda_i X_i, X_j) - (X_i, \lambda_j X_j) =$$

$$= (AX_i, X_j) - (X_i, AX_j) = (AX_i, X_j) - (AX_i, X_j) = 0,$$

откуда $(X_i, X_j) = 0$.

З а м е ч а н и е. Очевидно, что подпространства P_i являются корневыми подпространствами для самосопряженного оператора A . Таким образом, все корневые векторы для самосопряженного оператора являются просто собственными векторами. Поэтому минимальный

полином для самосопряженного оператора имеет только простые корни.

Из разложения пространства R в прямую сумму подпространств P_1, \dots, P_s и из экстремальных свойств собственных значений самосопряженного оператора (теорем 11.4 и 11.5) непосредственно следует справедливость следующих теорем.

Теорема 11.6. *Если $\lambda_1 > \lambda_2 > \dots > \lambda_s$ попарно различные собственные числа оператора A и P_1, \dots, P_s соответствующие им подпространства собственных векторов, то*

$$\begin{aligned}\lambda_1 &= \max_{\mathbf{x} \neq 0} \frac{(AX, X)}{(X, X)} \\ \lambda_2 &= \max_{\mathbf{x} \perp P_1} \frac{(AX, X)}{(X, X)} \\ &\dots \\ \lambda_k &= \max_{\mathbf{x} \perp P_1 + \dots + P_{k-1}} \frac{(AX, X)}{(X, X)}.\end{aligned}$$

Теорема 11.7. *Если $\lambda_1 > \dots > \lambda_s$ попарно различные собственные числа оператора A и P_1, \dots, P_s соответствующие им подпространства из собственных векторов, то*

$$\begin{aligned}\lambda_s &= \min_{\mathbf{x} \neq 0} \frac{(AX, X)}{(X, X)} \\ \lambda_{s-1} &= \min_{\mathbf{x} \perp P_s} \frac{(AX, X)}{(X, X)} \\ &\dots \\ \lambda_k &= \min_{\mathbf{x} \perp P_s + \dots + P_{k+1}} \frac{(AX, X)}{(X, X)}.\end{aligned}$$

Строение инвариантных подпространств для самосопряженного оператора более просто по сравнению с общим случаем. Действительно, пусть T какое-либо инвариантное подпространство. Оператор \tilde{A} , индуцированный оператором A на T , будет, очевидно, самосопряженным и, следовательно, в подпространстве T существует базис, состоящий из собственных векторов оператора \tilde{A} . Но каждый собственный вектор оператора \tilde{A} есть в то же время собственный вектор для оператора A . Таким образом, любое инвариантное подпространство самосопряженного оператора натянуто на некоторую систему собственных векторов. Далее, подобно всему пространству подпространство разлагается в прямую сумму подпространств T_i , состоящих из собственных векторов, принадлежащих собственному значению λ_i . Ясно, что $T_i = T \cap P_i$.

Если $X \in T$ и $X = X_1 + \dots + X_s$ есть разложение X по подпространствам P_i , то все $X_i \in T_i$. Действительно, $T = T \cap P_1 + \dots + T \cap P_s$ и, следовательно, имеет место так же разложение $X = X'_1 + \dots + X'_s$, где $X'_i \in T \cap P_i$. В силу однозначности разложения вектора X по подпространствам P_i все проекции X'_i совпадают с X_i . Но X'_i принадлежат T .

Строение циклических подпространств описывается теоремой.

Теорема 11.8. Пусть X_0 произвольный вектор пространства R . Тогда циклическое подпространство для самосопряженного оператора A , порожденное вектором X_0 , есть подпространство, натянутое на ненулевые проекции вектора X_0 на подпространства P_1, \dots, P_s .

Эта теорема является непосредственным следствием более общей теоремы 8.2. Мы, однако, приведем ее независимое доказательство. Пусть Q циклическое подпространство, порожденное вектором X_0 , и пусть

$$X_0 = X_{k_1} + \dots + X_{k_j}, \quad X_{k_i} \in P_{k_i},$$

разложение X_0 по подпространствам P_1, \dots, P_s , причем в этом разложении удержаны лишь отличные от нуля проекции. В силу сказанного выше все X_{k_1}, \dots, X_{k_j} принадлежат Q и, следовательно, натянутое на них подпространство Q' содержится в Q . Но Q' инвариантно, ибо оно натянуто на собственные векторы X_{k_1}, \dots, X_{k_j} , и Q' содержит X_0 . Следовательно, $Q \subset Q'$ и потому $Q = Q'$.

5. Положительно-определенные операторы. Среди всех самосопряженных операторов особо важную роль играют положительно-определенные операторы. Самосопряженный оператор A называется положительно-определенным если для любого вектора X , отличного от нуля, $(AX, X) > 0$.

Из определения следует, что ядро положительно-определенного оператора состоит только из нулевого вектора, так что положительно-определенный оператор не вырожден и, следовательно, для него существует обратный оператор A^{-1} .

Матрица положительно-определенного оператора в ортонормальном базисе положительно определена, и обратно, оператор с положительно-определенной матрицей в ортонормальном базисе положительно определен. Действительно, (AX, X) есть квадратичная форма от координат вектора в ортонормальном базисе, матрица которой совпадает с матрицей, сопоставляемой оператору в том же базисе.

Отметим ряд интересных свойств положительно-определенных операторов.

Теорема 11.9. Если A и B положительно-определенные операторы, то оператор $c_1A + c_2B$ положительно определен при $c_1 > 0, c_2 > 0$. Иначе, положительно-определенные операторы образуют выпуклый конус в пространстве операторов.

Доказательство. Имеем $((c_1A + c_2B)X, X) = c_1(AX, X) + c_2(BX, X) > 0$ при $X \neq 0$.

Теорема II.10. Если оператор A положительно определен, то оператор A^{-1} тоже положительно определен.

Действительно,

$$(A^{-1}X, X) = (A^{-1}X, AA^{-1}X) = (AY, Y) > 0.$$

Здесь $Y = A^{-1}X$.

Теорема II.11. Все собственные значения положительно-определенного оператора положительны.

Действительно, если λ собственное значение положительно-определенного оператора A и X соответствующий этому собственному значению собственный вектор, то $AX = \lambda X$ и

$$\lambda = \frac{(AX, X)}{(X, X)} > 0.$$

Теорема II.12. Если A положительно-определенный оператор и C любой невырожденный оператор, то C^*AC положительно определен.

Действительно, $(C^*ACX, X) = (ACX, CX) > 0$, ибо при $X \neq 0$ вектор $CX \neq 0$.

В частности, отсюда следует, что для любого невырожденного оператора C оператор C^*C положительно определен.

Теорема II.13. Для любого положительно-определенного оператора A существует положительно-определенный „квадратный корень“ $A^{1/2}$, т. е. такой положительно-определенный оператор B , что $B^2 = A$.

Действительно, пусть $\lambda_1, \dots, \lambda_n$ собственные значения оператора A и X_1, \dots, X_n попарно ортогональные собственные векторы, им принадлежащие. Тогда $AX_i = \lambda_i X_i$ ($i = 1, 2, \dots, n$). Определим оператор B , положив $BX_i = \sqrt{\lambda_i}X_i$ и линейно распространив оператор на все пространство. Тогда $B^2X_i = \lambda_i X_i = AX_i$, т. е. операторы B^2 и A совпадают на базисе X_1, \dots, X_n . Следовательно, в силу линейности они совпадают и во всем пространстве. Итак, $B^2 = A$. Очевидно, что оператор B положительно-определенный.

Теорема II.14. Если A — самосопряженный оператор и B — положительно-определенный оператор, то все собственные значения оператора BA (а следовательно, и $B^{-1}A$) вещественны.

Действительно, $BA = B^{1/2}(B^{1/2}AB^{1/2})B^{-1/2}$. Поэтому собственные значения оператора BA равняются собственным значениям оператора $B^{1/2}AB^{1/2}$, который, очевидно, самосопряжен.

Теорема II.15. Если A и B положительно-определеные операторы, то все собственные значения оператора BA положительны.

Действительно, в этом случае оператор $B^{\frac{1}{2}}AB^{\frac{1}{2}} = (B^{\frac{1}{2}})^*AB^{\frac{1}{2}}$ положительно определен в силу теоремы 11.12.

Теорема 11.16. Если B положительно-определенный оператор, A самосопряженный оператор и все собственные значения оператора BA положительны, то A положительно определен.

Действительно, если все собственные значения оператора BA положительны, то положительны и все собственные значения оператора $B^{-\frac{1}{2}}(BA)B^{\frac{1}{2}} = B^{\frac{1}{2}}AB^{\frac{1}{2}}$, т. е. самосопряженный оператор $B^{\frac{1}{2}}AB^{\frac{1}{2}}$ положительно определен. В силу теоремы 11.12 будет положительно-определенным и оператор A .

Теорема 11.17. Если A и B перестановочные положительно-определенные операторы, то AB положительно-определенный оператор.

Действительно, в этом случае AB самосопряженный и его положительно-определенность следует из положительности его собственных значений.

Очевидно, что все свойства, описанные в теоремах 11.9--11.17 остаются верными и для положительно-определенных матриц.

Каждый положительно-определенный оператор определяет некоторую метрику пространства, удовлетворяющую всем аксиомам обычной евклидовой метрики. Именно, A -скалярным произведением $(X, Y)_A$ двух векторов X и Y называется скалярное произведение (AX, Y) . При таком определении все четыре аксиомы евклидовой метрики

- 1) $(X, X)_A > 0, X \neq 0$
- 2) $(X, Y)_A = (Y, X)_A$
- 3) $(aX, Y)_A = a(X, Y)_A$
- 4) $(X_1 + X_2, Y)_A = (X_1, Y)_A + (X_2, Y)_A$

выполнены.

Действительно, 3-я и 4-я аксиомы выполнены в силу линейности оператора A , 2-я в силу самосопряженности и 1-я в силу положительной определенности.

В этой метрике роль длины вектора X играет A -длина, т. е. величина $\sqrt{(AX, X)}$. A -ортогональными или сопряженными (относительно A) векторами называются векторы X и Y , для которых $(AX, Y) = 0$. Все теоремы, доказанные в § 9, очевидно, переносятся на пространство с A -метрикой. В частности, справедливы теоремы.

Теорема 11.18. $(AX, Y)^2 \leq (AX, X) \cdot (AY, Y)$.

Теорема 11.19. Попарно A -ортогональные векторы линейно-независимы.

Теорема 11.20. Пусть X_1, \dots, X_k заданная система линейно-независимых векторов. Тогда можно построить A -ортогональную

систему векторов, связанную с исходной соотношениями:

$$\mathbf{S}_1 = \mathbf{X}_1$$

$$\mathbf{S}_2 = \mathbf{X}_2 + \alpha_{21} \mathbf{X}_1$$

$$\vdots \quad \vdots \quad \vdots$$

$$\mathbf{S}_k = \mathbf{X}_k + \alpha_{k1} \mathbf{X}_1 + \dots + \alpha_{kk-1} \mathbf{X}_{k-1}.$$

Ввиду того, что при построении некоторых численных методов будет применяться процесс А-ортогонализации и численная реализация этих методов будет проводиться по формулам процесса, мы приведем доказательство теоремы, хотя оно почти дословно повторяет доказательство теоремы 9.2.

Доказательство. Проведем построение по индукции. Пусть векторы $\mathbf{S}_1, \dots, \mathbf{S}_{m-1}$ уже построены и отличны от нуля. Вектор \mathbf{S}_m ищем в виде

$$\mathbf{S}_m = \mathbf{X}_m - \gamma_{m1} \mathbf{S}_1 - \dots - \gamma_{m m-1} \mathbf{S}_{m-1}. \quad (1)$$

Коэффициенты γ_{mj} определяем из условия А-ортогональности векторов $\mathbf{S}_1, \dots, \mathbf{S}_{m-1}$ к вектору \mathbf{S}_m . В силу А-ортогональности системы векторов $\mathbf{S}_1, \dots, \mathbf{S}_{m-1}$ и в силу того, что $\mathbf{S}_j \neq 0$, $j = 1, \dots, m-1$, имеем

$$\gamma_{mj} = \frac{(\mathbf{X}_m, \mathbf{AS}_j)}{(\mathbf{S}_j, \mathbf{AS}_j)} = \frac{(\mathbf{S}_j, \mathbf{AX}_m)}{(\mathbf{S}_j, \mathbf{AS}_j)}. \quad (2)$$

Подставив в равенство (1) вместо $\mathbf{S}_1, \dots, \mathbf{S}_{m-1}$ их выражения через $\mathbf{X}_1, \dots, \mathbf{X}_{m-1}$, получим, что

$$\mathbf{S}_m = \mathbf{X}_m + \alpha_{m1} \mathbf{X}_1 + \dots + \alpha_{m m-1} \mathbf{X}_{m-1}.$$

Отсюда следует, что вектор \mathbf{S}_m не равен нулю, ибо иначе векторы $\mathbf{X}_1, \dots, \mathbf{X}_m$ были бы линейно-зависимы, что противоречит условию теоремы. База для индукции дается тривиальным случаем $m = 1$.

Итак, система А-ортогональных (сопряженных) векторов определяется рекуррентными соотношениями

$$\mathbf{S}_1 = \mathbf{X}_1, \dots, \mathbf{S}_m = \mathbf{X}_m - \gamma_{m1} \mathbf{S}_1 - \dots - \gamma_{m m-1} \mathbf{S}_{m-1},$$

в которых коэффициенты определяются по формулам (2).

При практических вычислениях более удобно пользоваться несколько иными формулами для коэффициентов. Именно, из равенства

$$\mathbf{X}_j = \mathbf{S}_j + \gamma_{j1} \mathbf{S}_1 + \dots + \gamma_{jj-1} \mathbf{S}_{j-1}$$

следует, что

$$\mathbf{AX}_j = \mathbf{AS}_j + \gamma_{j1} \mathbf{AS}_1 + \dots + \gamma_{jj-1} \mathbf{AS}_{j-1}.$$

Отсюда

$$(\mathbf{S}_m, \mathbf{AX}_j) = 0 \quad \text{при } j < m,$$

$$(\mathbf{S}_j, \mathbf{AX}_j) = (\mathbf{S}_j, \mathbf{AS}_j).$$

Таким образом,

$$\gamma_{mj} = \frac{(S_j, AX_m)}{(S_j, AX_j)}.$$

6. Самосопряженный оператор в унитарном пространстве.

Для унитарного пространства определение самосопряженного оператора совпадает с определением, данным выше для пространства Эвклида. Именно, оператор A называется самосопряженным, если он совпадает со своим сопряженным, т. е. если

$$(AX, Y) = (X, AY)$$

при любых X и Y .

Матрица самосопряженного оператора по отношению к любому ортогональному базису является эрмитовой, ибо она совпадает со своей сопряженной. Для любого вектора X скалярное произведение (AX, X) принимает вещественное значение. Действительно, $(AX, X) = (X, AX) = (AX, X)$. Это обстоятельство делает теорию самосопряженных операторов в унитарном пространстве формально совпадающей с подобной теорией в эвклидовом пространстве. Именно, с некоторыми изменениями проводится конструкция для последовательного построения полной системы попарно ортогональных собственных векторов при помощи экстремальных соображений. Из этой конструкции следует, что все собственные значения самосопряженного оператора вещественны и матрица самосопряженного оператора может быть приведена к диагональной форме. В алгебраической формулировке это обозначает, что любая эрмитова матрица может быть приведена к диагональному виду преобразованием подобия посредством унитарной матрицы.

Далее сохраняются все экстремальные свойства собственных значений, так же как и строение инвариантных подпространств. Скалярное произведение (AX, X) , выраженное через координаты X в ортонормальном базисе, есть форма Эрмита от этих координат с матрицей, равной матрице оператора A в том же базисе.

Самосопряженный оператор называется положительно-определенным, если для любого вектора X , отличного от нуля, $(AX, X) > 0$. Все собственные значения положительно-определенного оператора положительны.

7. Нормальные операторы в унитарном пространстве. Оператор A называется нормальным, если он перестановчен со своим сопряженным оператором, т. е. если $AA^* = A^*A$.

Очевидно, что матрица нормального оператора в любом ортонормальном базисе есть нормальная матрица.

Самосопряженные операторы входят в класс нормальных операторов. В этот же класс входят унитарные операторы, т. е. операторы, удовлетворяющие требованию $A^* = A^{-1}$.

Теория собственных значений и собственных векторов для нормальных операторов очень похожа на соответствующую теорию для самосопряженных операторов. Именно, верна следующая

Теорема 11.21. В пространстве существует ортонормальный базис, состоящий из собственных векторов данного нормального оператора.

§ 12. Квадратичные формы

1. Преобразование квадратичной формы к каноническому виду.

Теорема 12.1. Всякая квадратичная форма с вещественной матрицей коэффициентов может быть приведена к следующему каноническому виду

$$F(x_1, x_2, \dots, x_n) = \alpha_1 y_1^2 + \alpha_2 y_2^2 + \dots + \alpha_n y_n^2.$$

где y_1, y_2, \dots, y_n переменные, связанные с исходными переменными x_1, x_2, \dots, x_n неособенным линейным преобразованием.

Доказательство. Для $n=1$ теорема тривиально выполняется. Допустим, что для форм от $n-1$ переменной теорема доказана. Докажем, что при этом предположении теорема верна и для форм от n переменных. Пусть

Допустим сначала, что $a_{11} \neq 0$. Тогда

$$\begin{aligned} F(x_1, x_2, \dots, x_n) &= a_{11} \left(x_1 + \frac{a_{12}}{a_{11}} x_2 + \dots + \frac{a_{1n}}{a_{11}} x_n \right)^2 - \\ &\quad - a_{11} \left(\frac{a_{12}}{a_{11}} x_2 + \dots + \frac{a_{1n}}{a_{11}} x_n \right)^2 + \varphi(x_2, \dots, x_n) = \\ &= a_{11} \left(x_1 + \frac{a_{12}}{a_{11}} x_2 + \dots + \frac{a_{1n}}{a_{11}} x_n \right)^2 + F_1(x_2, \dots, x_n). \end{aligned}$$

В силу индукционного предположения

$$F_1(x_2, \dots, x_n) = \alpha_2 y_2^2 + \dots + \alpha_n y_n^2,$$

где

$$\begin{bmatrix} y_2 \\ \vdots \\ y_n \end{bmatrix} = B_1 \begin{bmatrix} x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad B_1 \text{ — неособенная матрица.}$$

Положим

$$y_1 = x_1 + \frac{a_{12}}{a_{11}} x_2 + \dots + \frac{a_{1n}}{a_{11}} x_n.$$

Тогда

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & \frac{a_{12}}{a_{11}} & \cdots & \frac{a_{1n}}{a_{11}} \\ 0 & & & \\ \vdots & & & \\ 0 & & & B_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = B \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Ясно, что B — неособенная матрица. В новых переменных квадратичная форма примет искомый вид

$$F(x_1, x_2, \dots, x_n) = \alpha_1 y_1^2 + \alpha_2 y_2^2 + \dots + x_n y_n^2 \text{ при } \alpha_1 = a_{11}.$$

Если же $a_{11} = 0$, то всегда можно сделать предварительно неособенное преобразование переменных, при котором коэффициент при квадрате новой первой переменной окажется отличным от нуля (если только квадратичная форма не равна нулю тождественно). Действительно, если $a_{11} = 0$, но $a_{kk} \neq 0$ при некотором k , то достаточно лишь изменить порядок нумерации переменных. Если же все коэффициенты a_{11}, \dots, a_{nn} равны нулю, но $a_{ij} \neq 0$, то достаточно положить

$$\begin{aligned} x_1 &= y_1 \\ &\vdots \\ x_i &= y_i + y_j \\ &\vdots \\ x_j &= y_j \\ &\vdots \\ x_n &= y_n. \end{aligned}$$

Это преобразование, очевидно, неособенное. Преобразованная квадратичная форма будет иметь ненулевой коэффициент $2a_{ij}$ при y_j^2 .

Замечание. Настоящее доказательство дает также описание вычислительного процесса для приведения квадратичной формы к каноническому виду. Очень часто бывает, что на всех шагах этого процесса имеет место первый случай, и тогда отпадает необходимость вспомогательных преобразований. В этом случае матрица окончательного преобразования B будет треугольной.

2. Положительно-определенная квадратичная форма. Напоминаем, что квадратичная форма $F(x_1, \dots, x_n)$ называется положительно-определенной, если все ее значения положительны, кроме значения при $x_1 = x_2 = \dots = x_n = 0$.

Теорема 12.2. Для того чтобы квадратичная форма была положительно-определенной, необходимо и достаточно, чтобы все коэффициенты после приведения ее к каноническому виду были положительны.

Доказательство. Пусть

$$F(x_1, x_2, \dots, x_n) = \alpha_1 y_1^2 + \alpha_2 y_2^2 + \dots + \alpha_n y_n^2.$$

Если $\alpha_1 > 0, \alpha_2 > 0, \dots, \alpha_n > 0$, то $F(x_1, x_2, \dots, x_n)$, очевидно, положительно определена. Если же $\alpha_i \leq 0$ при некотором i , то, положив $y_1 = 0, \dots, y_i = 1, \dots, y_n = 0$ и найдя соответствующие значения x_1^0, \dots, x_n^0 переменных x_1, x_2, \dots, x_n , получим $F(x_1^0, \dots, x_n^0) = \alpha_i \leq 0$. Теорема доказана.

Установленный критерий положительной определенности очень удобен для практической проверки.

Существует другой критерий положительной определенности квадратичной формы, применение которого не требует приведения формы к каноническому виду.

Теорема 12.3. (Критерий Сильвестера). Квадратичная форма $\sum a_{ij}x_i x_j$ положительно определена в том и только в том случае, если все определители

$$\Delta_1 = a_{11}, \quad \Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \dots, \quad \Delta_n = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}$$

положительны.

Доказательство. Пусть $F(x_1, \dots, x_n) = \sum_{i,j=1}^n a_{ij}x_i x_j$ положительно определена. Тогда $A = B' \Lambda B$, где B матрица, преобразующая форму к каноническому виду, $\Lambda = [\alpha_1, \dots, \alpha_n]$ диагональная матрица, составленная из коэффициентов канонического представления. В силу доказанного выше критерия положительной определенности все $\alpha_i > 0$. Поэтому $\Delta_n = |A| = |B'| \cdot |\Lambda| \cdot |B| = \alpha_1 \alpha_2 \dots \alpha_n |B|^2 > 0$. Вместе с тем, если форма $F(x_1, \dots, x_n)$ положительно определена, то положительно-определенными будут и все формы

$$F_k(x_1, \dots, x_k) = F(x_1, \dots, x_k, 0, \dots, 0) = \sum_{i,j=1}^k a_{ij}x_i x_j.$$

Поэтому все определители Δ_k ($k = 1, 2, \dots, n$) положительны.

Допустим теперь, что $\Delta_1 > 0, \dots, \Delta_n > 0$. Тогда матрица A может быть разложена в произведение

$$A = C \Lambda B,$$

где $\Lambda = [\Delta_1, \frac{\Delta_2}{\Delta_1}, \dots, \frac{\Delta_n}{\Delta_{n-1}}]$. C и B треугольные матрицы с единичными диагональными элементами. Так как матрица A симметрична,

то $C = B'$. Если теперь в квадратичной форме сделать преобразование переменных с матрицей B^{-1} , мы приедем к квадратичной форме с матрицей Λ , т. е. к канонической форме с положительными коэффициентами. Тем самым форма $F(x_1, \dots, x_n)$ положительно определена.

3. Ортогональное преобразование квадратичной формы к каноническому виду.

Теорема 12.4. Любая квадратичная форма с вещественной матрицей коэффициентов может быть приведена к каноническому виду при помощи преобразования переменных с ортогональной матрицей.

Доказательство. Пусть A —матрица квадратичной формы. Тогда, согласно алгебраической формулировке теоремы 11.3, существует ортогональная матрица P такая, что $P'AP = \Lambda$ есть диагональная матрица. Сделав в квадратичной форме преобразование переменных с матрицей P , мы приедем к квадратичной форме в новых переменных с матрицей Λ , т. е. к квадратичной форме $\lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2$. Теорема доказана.

Отметим, что коэффициенты $\lambda_1, \dots, \lambda_n$, полученные при ортогональном преобразовании квадратичной формы к каноническому виду, являются собственными значениями матрицы формы.

4. Закон инерции квадратичных форм. Ясно, что одна и та же квадратичная форма может быть приведена к каноническому виду бесконечным множеством способов. Например, можно сделать сначала какое-либо неособенное преобразование переменных, а затем „спокойтесь“ и применить хотя бы способ, описанный в п. 1. При этом, очевидно, что и сами коэффициенты a_1, a_2, \dots, a_n могут получаться различными. Однако верна следующая важная

Теорема 12.5. Число положительных, отрицательных и нулевых коэффициентов в каноническом представлении квадратичной формы не зависит от способа приведения.

Доказательство. Пусть

$$\begin{aligned} F(x_1, \dots, x_n) &= a_1 y_1^2 + \dots + a_p y_p^2 - a_{p+1} y_{p+1}^2 - \dots - a_{p+q} y_{p+q}^2 = \\ &= \beta_1 z_1^2 + \dots + \beta_r z_r^2 - \beta_{r+1} z_{r+1}^2 - \dots - \beta_{r+s} z_{r+s}^2 \end{aligned}$$

два канонических представления данной квадратичной формы. В этой записи мы предполагаем, что все $a_i > 0$ и $\beta_i > 0$. Нулевые коэффициенты мы опустили. Тем самым переменные y_{p+q+1}, \dots, y_n и z_{r+s+1}, \dots, z_n фактически не входят в преобразованные квадратичные формы. Интерпретируем квадратичную форму как функционал от вектора X , координатами которого в некотором выбранном базисе являются числа x_1, \dots, x_n . Тогда переход от переменных x_1, \dots, x_n к переменным y_1, \dots, y_n так же как и переход к переменным z_1, \dots, z_n , может интерпретироваться как преобразование координат. Обозначим через V_1, \dots, V_n тот базис, в котором координатами

вектора \mathbf{X} являются y_1, \dots, y_n и через W_1, \dots, W_n тот базис, в котором координатами являются z_1, \dots, z_n . Нам надо доказать, что $p = r, q = s$. Допустим, что $p < r$. Рассмотрим подпространство Q , натянутое на векторы $V_{p+1}, \dots, V_{p+q}, \dots, V_n$, и подпространство P , натянутое на векторы W_1, \dots, W_r . Сумма размерностей этих подпространств равна $n - p + r > n$.

По теореме о размерности суммы и пересечения мы заключаем, что размерность $Q \cap P$ больше нуля, ибо размерность векторной суммы Q и P не может быть больше чем n . Следовательно, в $Q \cap P$ существует хотя бы один вектор $X_0 \neq 0$. Пусть y_1^0, \dots, y_n^0 и z_1^0, \dots, z_n^0 его координаты соответственно в базисах V_1, \dots, V_n и W_1, \dots, W_n . Так как $X_0 \in Q$, то $y_1^0 = \dots = y_p^0 = 0$, так как $X_0 \in P$, то $z_{r+1}^0 = \dots = z_n^0 = 0$. Следовательно,

$$F(X_0) = -\alpha_{p+1}y_{p+1}^{02} - \dots - \alpha_{p+q}y_{p+q}^{02},$$

$$F(X_0) = \beta_1z_1^{02} + \dots + \beta_rz_r^{02}.$$

Из первого равенства заключаем, что $F(X_0) \leq 0$, из второго равенства следует, что $F(X_0) > 0$, ибо хотя бы одна из координат z_1^0, \dots, z_r^0 отлична от нуля, так как иначе вектор X_0 был бы нулевым. Полученное противоречие показывает, что предположение, что $p < r$, неверно. Совершенно так же предположение $p > r$ приводит нас к противоречию. Следовательно, $p = r$, и теорема в части, касающейся числа положительных коэффициентов, доказана.

Для доказательства того, что $q = s$, достаточно рассмотреть вместо формы F форму $-F$.

5. Одновременное преобразование двух квадратичных форм к каноническому виду.

Теорема 12.6. Пусть $F(x_1, \dots, x_n)$ и $\Phi(x_1, \dots, x_n)$ две квадратичные формы, причем $\Phi(x_1, \dots, x_n)$ положительно определена. Тогда обе формы можно привести к каноническому виду одним и тем же преобразованием переменных.

Доказательство. Приведем каким-либо преобразованием форму $\Phi(x_1, \dots, x_n)$ к каноническому виду

$$\Phi(x_1, \dots, x_n) = \alpha_1y_1^2 + \dots + \alpha_ny_n^2.$$

То же преобразование сделаем в форме $F(x_1, \dots, x_n)$. Полученную форму обозначим через $F_1(y_1, \dots, y_n)$. Далее, в обоих квадратичных формах положим $y_1 = \frac{z_1}{\sqrt{\alpha_1}}, \dots, y_n = \frac{z_n}{\sqrt{\alpha_n}}$. Коэффициенты этого преобразования вещественны, ибо $\alpha_1 > 0, \dots, \alpha_n > 0$. После этого преобразования получим, что

$$\Phi(x_1, \dots, x_n) = z_1^2 + \dots + z_n^2$$

$$F(x_1, \dots, x_n) = F_2(z_1, \dots, z_n).$$

Теперь приведем форму $F_2(z_1, \dots, z_n)$ к каноническому виду ортогональным преобразованием

$$(z_1, \dots, z_n)' = P(t_1, \dots, t_n)'.$$

После преобразования получим

$$F_2(z_1, \dots, z_n) = \lambda_1 t_1^2 + \dots + \lambda_n t_n^2.$$

При этом же преобразовании форма $z_1^2 + \dots + z_n^2$ преобразуется в $t_1^2 + \dots + t_n^2$. Действительно, матрицей формы $z_1^2 + \dots + z_n^2$ является единичная матрица, матрица преобразованной формы тоже равна единичной, так как $P'EP = E$, в силу ортогональности матрицы P .

Заметим, что числа $\lambda_1, \dots, \lambda_n$ можно определить, не выполняя указанных преобразований. Обозначим матрицу формы $F(x_1, \dots, x_n)$ через A , матрицу формы $\Phi(x_1, \dots, x_n)$ через B и диагональную матрицу $[\lambda_1, \dots, \lambda_n]$ через Λ .

Пусть C матрица линейного преобразования, осуществляющего одновременное приведение обоих форм к каноническому виду указанным выше способом. Тогда $C'AC = \Lambda$, $C'BC = E$ и, следовательно, $\Lambda = tE = C'AC = tC'BC = C'(A - tB)C$. Отсюда $|\Lambda - tE| = |C'| |A - tB| \cdot |C|$ или $(\lambda_1 - t) \dots (\lambda_n - t) = |C|^2 \cdot |A - tB|$.

Таким образом, числа $\lambda_1, \dots, \lambda_n$ оказываются корнями многочлена $|A - tB|$. Уравнение $|A - tB|$ называется обобщенным вековым уравнением.

Проведенное доказательство по существу равносильно следующему рассуждению.

Пусть A матрица формы $F(x_1, \dots, x_n)$ и B матрица формы $\Phi(x_1, \dots, x_n)$. Матрица B по условию положительно определена. Пусть Q ортогональная матрица, преобразующая квадратичную форму с матрицей $B^{-1/2}AB^{-1/2}$ к каноническому виду. Тогда $Q'B^{-1/2}AB^{-1/2}Q = \Lambda$, где Λ диагональна. Обозначим $B^{-1/2}Q = C$. Тогда

$$C'AC = Q'B^{-1/2}AB^{-1/2}Q = \Lambda$$

$$C'BC = Q'B^{-1/2}BB^{-1/2}Q = Q'Q = E.$$

Таким образом, преобразование переменных с матрицей C приводит квадратичную форму F к каноническому виду с коэффициентами $\lambda_1, \dots, \lambda_n$, а квадратичную форму Φ к каноническому виду с коэффициентами 1, ..., 1.

6. Формы Эрмита. Алгебраическое выражение

$$\begin{aligned} F(z_1, \dots, z_n) &= a_{11}z_1\bar{z}_1 + a_{12}z_1\bar{z}_2 + \dots + a_{1n}z_1\bar{z}_n + \\ &+ a_{21}z_2\bar{z}_1 + a_{22}z_2\bar{z}_2 + \dots + a_{2n}z_2\bar{z}_n + \\ &+ \dots \dots \dots \dots \dots \dots + \\ &+ a_{n1}z_n\bar{z}_1 + a_{n2}z_n\bar{z}_2 + \dots + a_{nn}z_n\bar{z}_n. \end{aligned}$$

где z_1, \dots, z_n комплексные переменные, а коэффициенты удовлетворяют условию $a_{ij} = \bar{a}_{ji}$, называется формой Эрмита. Матрица коэффициентов формы Эрмита по самому определению является эрмитовой матрицей. Все значения формы Эрмита вещественны, ибо

$$\overline{F(z_1, \dots, z_n)} = \sum_{i,j} a_{ij} z_i \bar{z}_j = \sum_{i,j} a_{ij} \bar{z}_i z_j = F(z_1, \dots, z_n).$$

Для формы Эрмита справедливы теоремы, аналогичные теоремам для квадратичных форм. Именно,

Теорема 12.7. Любая форма Эрмита может быть приведена неособенным преобразованием переменных к каноническому виду

$$\alpha_1 z_1 \bar{z}_1 + \alpha_2 z_2 \bar{z}_2 + \dots + z_n z_n \bar{z}_n.$$

Форма Эрмита называется положительно-определенной, если все ее значения положительны, кроме значения при

$$z_1 = z_2 = \dots = z_n = 0.$$

Матрица коэффициентов положительно-определенной формы Эрмита называется положительно-определенной эрмитовой матрицей.

Теорема 12.8. Для того чтобы форма Эрмита была положительно-определенной, необходимо и достаточно, чтобы все коэффициенты в ее каноническом разложении были положительны.

Теорема 12.9. Форма Эрмита может быть приведена к каноническому виду при помощи унитарного преобразования переменных.

При таком преобразовании коэффициенты в канонической форме будут собственными значениями матрицы формы Эрмита.

Теорема 12.10. (Закон инерции). Число положительных, отрицательных и нулевых коэффициентов в каноническом представлении формы Эрмита не зависит от способа ее приведения.

Теорема 12.11. Две формы Эрмита, из которых одна положительно-определенная, могут быть приведены к каноническому виду одним и тем же преобразованием переменных.

Доказательства всех этих теорем почти дословно совпадают с доказательствами аналогичных теорем для квадратичных форм.

§ 13. Понятие предела в линейной алгебре

Понятие предела для линейно-алгебраических объектов нам будет нужно преимущественно для описания итерационных методов. Ввиду того, что численные задачи формулируются в терминах матриц, мы определим понятие предела для столбцов, которые мы будем отождествлять с векторами арифметического пространства в их естественном представлении, и для квадратных матриц. Во всем дальнейшем мы будем употреблять термин вектор преимущественно в этом смысле.

1. Предел векторов и матриц. Пусть дана последовательность векторов $X^{(1)}, \dots, X^{(k)}, \dots$ с компонентами $x_1^{(1)}, \dots, x_n^{(1)}; \dots; x_1^{(k)}, \dots, x_n^{(k)}; \dots$. Если у каждой компоненты существует предел $\lim_{k \rightarrow \infty} x_i^{(k)} = x_i$, то вектор X с компонентами x_1, \dots, x_n называется пределом последовательности $X^{(1)}, \dots, X^{(k)}, \dots$, а сама последовательность называется сходящейся к вектору X . Это записывается в виде $X^{(k)} \rightarrow X$ или $\lim_{k \rightarrow \infty} X^{(k)} = X$.

Таким же образом, если имеется последовательность квадратных матриц $A^{(1)}, \dots, A^{(k)}, \dots$ с элементами $a_{ij}^{(1)}, \dots, a_{ij}^{(k)}, \dots$, то пределом последовательности $A^{(1)}, \dots, A^{(k)}, \dots$ называется матрица с элементами $a_{ij} = \lim_{k \rightarrow \infty} a_{ij}^{(k)}$, если все эти пределы существуют.

В соответствии с этим определением предела, бесконечный ряд векторов $X^{(1)} + X^{(2)} + \dots + X^{(k)} + \dots$ называется сходящимся, если существует $\lim_{k \rightarrow \infty} (X^{(1)} + X^{(2)} + \dots + X^{(k)})$, и этот предел называется суммой данного ряда. Очевидно, что для сходимости ряда векторов необходимо и достаточно, чтобы сходились все ряды из одноименных компонент; суммы этих рядов являются компонентами суммы ряда векторов.

Аналогичным образом определяется понятие сходимости ряда из матриц.

Понятие предела без труда распространяется на векторы и операторы в любом линейном пространстве посредством перехода к столбцам из координат (для векторов) и матриц (для операторов) по отношению к некоторому базису. Легко доказать, что выбор базиса не влияет на факт сходимости и на результат предельного перехода.

В некоторых вопросах полезным оказывается понятие сходимости последовательности векторов по направлению.

Последовательность векторов $X^{(1)}, \dots, X^{(k)}, \dots$ называется сходящейся по направлению, если последовательность нормированных векторов $\frac{1}{|X^{(1)}|} X^{(1)}, \dots, \frac{1}{|X^{(k)}|} X^{(k)}, \dots$ сходится в обычном смысле. Ясно, что для сходимости по направлению достаточно, чтобы последовательность $b_k X^{(k)}$ сходилась в обычном смысле к ненулевому вектору. Здесь b_k какие-либо положительные числа. Действительно, если $b_k X^{(k)} \rightarrow X \neq 0$, то $b_k |X^{(k)}| \rightarrow |X|$ и, следовательно,

$$\frac{1}{|X^{(k)}|} X^{(k)} = \frac{1}{b_k |X^{(k)}|} b_k X^{(k)} \rightarrow \frac{1}{|X|} X.$$

Легко видеть, что основные действия (сложение, вычитание, умножение) над векторами и матрицами непрерывны. В частности, если $A^{(k)} \rightarrow A$, то $A^{(k)} X \rightarrow AX$ при любом векторе X . Обратно, если

$A^{(k)}X \rightarrow AX$ при любом векторе X , то $A^{(k)} \rightarrow A$. Действительно, положив $X = e_i = (0, \dots, 1, \dots, 0)'$ при $i = 1, 2, \dots, n$, получим, что каждый столбец матрицы $A^{(k)}$ сходится к соответствующему столбцу матрицы A , и, следовательно, все элементы матрицы $A^{(k)}$ сходятся к соответствующим элементам матрицы A .

Далее, если последовательность квадратных матриц имеет пределом неособенную матрицу A , то при достаточно большом k для $A^{(k)}$ существует обратная и $\lim_{k \rightarrow \infty} (A^{(k)})^{-1} = A^{-1}$.

Действительно, если $A^{(k)} \rightarrow A$, то, очевидно, союзные с матрицами $A^{(k)}$ матрицы $B^{(k)}$ сходятся к матрице B , союзной с A , так как их элементами являются полиномы соответственно от элементов $A^{(k)}$ и элементов A .

По той же причине $|A^{(k)}| \rightarrow |A| \neq 0$, и, следовательно, начиная с некоторого места, $|A^{(k)}| \neq 0$.

Наконец, $(A^{(k)})^{-1} \rightarrow A^{-1}$, ибо

$$(A^{(k)})^{-1} = \frac{1}{|A^{(k)}|} \cdot B_k \rightarrow \frac{1}{|A|} B = A^{-1}.$$

Отметим еще следующую теорему.

Теорема 13.1. Если последовательность матриц $A^{(k)}$ имеет пределом неособенную матрицу A и векторы $F^{(k)}$ сходятся к F , то решения систем

$$A^{(k)}X^{(k)} = F^{(k)}$$

имеют предел, являющийся решением системы $AX = F$.

Действительно, $X^{(k)} = (A^{(k)})^{-1}F^{(k)} \rightarrow A^{-1}F$.

Для матриц, элементы которых являются дифференцируемыми функциями от некоторого параметра t , естественным образом определяется дифференцирование по этому параметру. Именно,

$$\frac{d}{dt} A(t) = \lim_{h \rightarrow 0} \frac{1}{h} (A(t+h) - A(t)).$$

Если

$$A(t) = \begin{bmatrix} a_{11}(t) & \dots & a_{1n}(t) \\ \vdots & \ddots & \vdots \\ a_{n1}(t) & \dots & a_{nn}(t) \end{bmatrix},$$

то, очевидно,

$$A'(t) = \begin{bmatrix} a'_{11}(t) & \dots & a'_{1n}(t) \\ \vdots & \ddots & \vdots \\ a'_{n1}(t) & \dots & a'_{nn}(t) \end{bmatrix}.$$

Легко установить следующие правила дифференцирования

$$\begin{aligned}\frac{d}{dt}(A_1 + A_2) &= \frac{dA_1}{dt} + \frac{dA_2}{dt} \\ \frac{d}{dt}(cA) &= c \frac{dA}{dt} \\ \frac{d}{dt}(A_1 A_2) &= \frac{dA_1}{dt} A_2 + A_1 \frac{dA_2}{dt}.\end{aligned}$$

2. Нормы векторов. В прикладных вопросах бывает важно судить не только о самом факте сходимости последовательностей или рядов, но и о быстроте этой сходимости. С этой целью очень полезно введение так называемой нормы векторов и нормы матриц. Норму можно вводить различными способами, и в различных случаях та или другая норма оказывается более удобной.

Вообще, нормой вектора X называется сопоставляемое этому вектору неотрицательное число $\|X\|$, удовлетворяющее следующим требованиям:

- 1) $\|X\| > 0$ при $X \neq 0$ и $\|0\| = 0$;
- 2) $\|cX\| = |c| \cdot \|X\|$ при любом числовом множителе c ;
- 3) $\|X + Y\| \leq \|X\| + \|Y\|$ („неравенство треугольника“).

Из требований 2) и 3) легко выводится, что

$$\|X - Y\| \geq |\|X\| - \|Y\||.$$

Именно, $\|X\| = \|X - Y + Y\| \leq \|X - Y\| + \|Y\|$, и потому

$$\|X - Y\| \geq \|X\| - \|Y\|.$$

Далее,

$$\|X - Y\| = \|Y - X\| \geq \|Y\| - \|X\|.$$

Следовательно

$$\|X - Y\| \geq |\|X\| - \|Y\||.$$

Каждая норма определяет „единичную сферу“ — множество векторов, норма которых не превосходит 1. Единичная сфера есть центрально симметричное выпуклое тело, т. е. такое множество, которое вместе с каждым вектором X содержит вектор $-X$ (центральная симметрия) и вместе с любыми векторами X_1 и X_2 содержит вектор $tX_1 + (1-t)X_2$, $0 \leq t \leq 1$, опирающийся на отрезок, соединяющий концы векторов X_1 и X_2 (выпуклость). Обратно, любое центрально симметричное выпуклое тело V в вещественном пространстве (в комплексном центральная симметрия должна быть заменена более сильным требованием — вместе с вектором X тело содержит вектор αX , $|\alpha| = 1$) порождает норму $\|X\|_V$, которая определяется как $\inf t$, где $t > 0$, $\frac{1}{t}X \in V$.

Проверка аксиом при таком общем определении нормы не представляет труда.

В дальнейшем мы воспользуемся следующими тремя нормами вектора

$$X = (x_1, x_2, \dots, x_n)',$$

определенными как для вещественного, так и для комплексного арифметического пространства.

1. Первая норма (кубическая).

$$\|X\|_1 = \max_i |x_i|.$$

Множество векторов вещественного пространства с нормой, не превосходящей единицы, заполняет единичный куб

$$-1 \leq x_1 \leq 1, \dots, -1 \leq x_n \leq 1.$$

2. Вторая норма (октаэдрическая).

$$\|X\|_2 = |x_1| + |x_2| + \dots + |x_n|.$$

Множество вещественных векторов, для которых $\|x\|_2 \leq 1$ заполняет n -мерный аналог октаэдра.

3. Третья норма (сферическая).

$$\|X\|_3 = |X| = \sqrt{|x_1|^2 + |x_2|^2 + \dots + |x_n|^2}.$$

Эта норма есть не что иное, как длина вектора. Совокупность векторов, для которых $|X| \leq 1$, заполняет шар единичного радиуса.

Для этих трех норм выполняются все требования 1—3.

Для кубической и октаэдрической это очевидно. Для сферической выполнение требований 1, 2 очевидно; требование 3 выполняется на основании неравенства Коши—Буняковского. Действительно,

$$\begin{aligned} |X+Y|^2 &= (X+Y, X+Y) = (X, X) + (X, Y) + (Y, X) + \\ &\quad + (Y, Y) \leq |X|^2 + 2|(X, Y)| + |Y|^2 \leq |X|^2 + \\ &\quad + 2|X| \cdot |Y| + |Y|^2 = (|X| + |Y|)^2. \end{aligned}$$

Введенные выше нормы связаны следующими неравенствами:

$$\|X\|_1 \leq \|X\|_2 \leq n\|X\|_1 \tag{1}$$

$$\|X\|_1 \leq |X| \leq \sqrt{n}\|X\|_1 \tag{2}$$

$$\frac{1}{\sqrt{n}}\|X\|_n \leq |X| \leq \|X\|_n. \tag{3}$$

Неравенства (1) и (2) очевидны, так же как правое из неравенств (3).

Левое из неравенств (3) легко выводится из неравенства Коши — Буняковского. Именно,

$$\begin{aligned} (|x_1| + |x_2| + \dots + |x_n|)^2 &= \\ &= (|x_1| \cdot 1 + |x_2| \cdot 1 + \dots + |x_n| \cdot 1)^2 \leqslant \\ &\leqslant (|x_1|^2 + |x_2|^2 + \dots + |x_n|^2)(1^2 + 1^2 + \dots + 1^2) = \\ &= n(|x_1|^2 + \dots + |x_n|^2), \end{aligned}$$

откуда посредством извлечения квадратного корня и деления на \sqrt{n} получаем левое неравенство (3).

Легко установить, что необходимым и достаточным условием сходимости последовательности векторов $X^{(k)}$ к вектору X является $\|X^{(k)} - X\| \rightarrow 0$ для каждой из трех введенных норм.

Для первой нормы это очевидно. Для остальных норм это следует из неравенств (1) и (2).

При этом, если $X^{(k)} \rightarrow X$, то $\|X^{(k)}\| \rightarrow \|X\|$. Действительно, $\|X^{(k)}\| - \|X\| \leq \|X - X^{(k)}\| \rightarrow 0$.

3. Нормы матриц. Нормой квадратной матрицы A называется неотрицательное число $\|A\|$, удовлетворяющее условиям

- 1) $\|A\| > 0$, если $A \neq 0$ и $\|0\| = 0$;
- 2) $\|cA\| = |c| \cdot \|A\|$;
- 3) $\|A + B\| \leq \|A\| + \|B\|$;
- 4) $\|AB\| \leq \|A\| \cdot \|B\|$.

Так же как и для нормы векторов, условие $\|A^{(k)} - A\| \rightarrow 0$ является необходимым и достаточным условием того, что $A^{(k)} \rightarrow A$, и так же как для нормы векторов, из $A^{(k)} \rightarrow A$ следует, что $\|A^{(k)}\| \rightarrow \|A\|$.

Часто употребляются следующие две нормы матриц

$$M(A) = n \max_{i,j} |a_{ij}|$$

и

$$N(A) = \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{\text{Sp} A^* A}.$$

То, что эти обе нормы удовлетворяют первым трем условиям, очевидно, так как они являются ($M(A)$ с точностью до множителя n) кубической и сферической нормами матрицы, рассматриваемой как вектор n^2 -мерного арифметического пространства. Остается проверить

выполнение 4-го условия. Пусть $A = (a_{ij})$, $B = (b_{ij})$. Тогда $AB = \left(\sum_{s=1}^n a_{is} b_{sj} \right)$. Имеем

$$\begin{aligned} M(AB) &= n \max \left| \sum_{s=1}^n a_{is} b_{sj} \right| \leq n \max \sum_{s=1}^n |a_{is}| \cdot |b_{sj}| \leq \\ &\leq n \sum_{s=1}^n \frac{1}{n} M(A) \cdot \frac{1}{n} M(B) = M(A) \cdot M(B). \end{aligned}$$

Далее,

$$N^2(AB) = \sum_{i,j} \left| \sum_{s=1}^n a_{is} b_{sj} \right|^2.$$

По неравенству Коши — Буняковского

$$\left| \sum_{s=1}^n a_{is} b_{sj} \right|^2 \leq \sum_{s=1}^n |a_{is}|^2 \sum_{t=1}^n |b_{tj}|^2,$$

откуда

$$N^2(AB) \leq \sum_{i,j,s,t} |a_{is}|^2 |b_{tj}|^2 = \sum_{i,s} |a_{is}|^2 \cdot \sum_{j,t} |b_{tj}|^2 = N^2(A) N^2(B).$$

Итак,

$$N(AB) \leq N(A) N(B).$$

Очевидно, что для любой матрицы A имеет место $N(A) \leq M(A)$.

Норма матриц может быть введена бесконечным множеством способов. Так как в большинстве задач, связанных с оценками, в рассуждении одновременно участвуют как матрицы, так и векторы, целесообразно вводить норму матриц так, чтобы она была разумным образом связана с нормой векторов, введенной в данном рассуждении. Мы будем говорить, что норма матриц согласована с данной нормой векторов, если для любой матрицы A и любого вектора X выполняется неравенство

$$\|AX\| \leq \|A\| \cdot \|X\|.$$

Так, введенная выше норма $M(A)$ согласована с кубической, октаэдрической и сферической нормами вектора. $N(A)$ согласована со сферической нормой вектора.

Действительно, пусть $X = (x_1, \dots, x_n)'$. Тогда

$$\begin{aligned} \|AX\|_1 &= \max_i \left| \sum_j a_{ij} x_j \right| \leq \max_i \sum_j |a_{ij}| \cdot |x_j| \leq \\ &\leq \sum_j \frac{M(A)}{n} \|X\|_1 = M(A) \cdot \|X\|_1. \end{aligned}$$

Далее,

$$\begin{aligned}\|AX\|_n &= \sum_i \left| \sum_j a_{ij} x_j \right| \leq \sum_{i,j} |a_{ij}| \cdot |x_j| \leq \sum_{i,j} \frac{M(A)}{n} \cdot |x_j| = \\ &= M(A) \sum_j |x_j| = M(A) \cdot \|X\|_n.\end{aligned}$$

Итак, норма $M(A)$ согласована с кубической и октаэдрической нормами вектора.

Наконец,

$$\begin{aligned}\|AX\|^2 &= \sum_i \left| \sum_j a_{ij} x_j \right|^2 \leq \sum_i \sum_j |a_{ij}|^2 \sum_s |x_s|^2 = \\ &= N^2(A) \|X\|^2 \leq M^2(A) \|X\|^2.\end{aligned}$$

Из этих неравенств следует, что обе нормы $M(A)$ и $N(A)$ согласованы со сферической нормой вектора.

Укажем конструкцию, которая дает возможность построить наименьшую норму матрицы, согласованную с данной нормой векторов. Именно, примем за норму матрицы A максимум норм векторов AX в предположении, что вектор X пробегает множество всех векторов, норм которых равна единице:

$$\|A\| = \max_{\|X\|=1} \|AX\|.$$

Вследствие непрерывности нормы, для каждой матрицы A этот максимум достигается, т. е. найдется такой вектор X_0 , что $\|X_0\|=1$ и $\|AX_0\|=\|A\|$.

Докажем, что построенная таким образом норма удовлетворяет всем поставленным требованиям 1) — 4) и условию согласованности.

Начнем с проверки первого требования. Пусть $A \neq 0$. Тогда найдется вектор X , $\|X\|=1$, такой, что $AX \neq 0$, а следовательно, и $\|AX\| \neq 0$. Поэтому $\|A\| = \max_{\|X\|=1} \|AX\| > 0$. Если же $A = 0$, то $\|A\| = \max_{\|X\|=1} \|0X\| = 0$.

Второе требование. В силу определения, $\|cA\| = \max_{\|X\|=1} \|cAX\| = |c| \max_{\|X\|=1} \|AX\|$ и, следовательно, $\|cA\| = |c| \cdot \|A\|$.

Далее проверим условие согласованности. Пусть $Y \neq 0$ любой вектор; тогда $X = \frac{1}{\|Y\|} \cdot Y$ удовлетворяет условию $\|X\|=1$. Следовательно, $\|AY\| = \|A(\|Y\|X)\| = \|Y\| \cdot \|AX\| \leq \|Y\| \cdot \|A\|$.

Третье требование. Для матрицы $A+B$ найдем такой вектор X_0 , что $\|A+B\| = \|(A+B)X_0\|$ и $\|X_0\|=1$. Тогда $\|A+B\| = \|(A+B)X_0\| = \|AX_0+BX_0\| \leq \|AX_0\| + \|BX_0\| \leq \|A\| \cdot \|X_0\| + \|B\| \cdot \|X_0\| = \|A\| + \|B\|$.

Наконец, четвертое требование. Для матрицы AB найдем такой вектор X_0 , что $\|X_0\|=1$ и $\|ABX_0\|=\|AB\|$. Тогда $\|AB\|=\|ABX_0\|\leqslant\|A(BX_0)\|\leqslant\|A\|\cdot\|BX_0\|\leqslant\|A\|\cdot\|B\|\cdot\|X_0\|=\|A\|\cdot\|B\|$.

Мы проверили выполнение всех четырех требований и условия согласованности. Построенную таким образом норму матриц мы будем называть подчиненной данной норме векторов. Докажем, что подчиненная норма не больше всякой нормы, согласованной с той же нормой. Действительно, пусть $L(A)$ норма согласования с некоторой нормой вектора, $\|A\|$ — норма, подчиненная той же норме вектора. Тогда найдется вектор X_0 с нормой, равной единице, такой, что

$$\|A\|=\|AX_0\|.$$

Но

$$\|AX_0\|\leqslant L(A)\|X_0\|=L(A),$$

откуда следует, что

$$\|A\|\leqslant L(A).$$

Очевидно, что для всякой нормы матриц, подчиненной какой-либо норме векторов, $\|E\|=1$. Отсюда следует, что норма $M(A)$ и $N(A)$ не подчинены ни одной из норм векторов, ибо $M(E)=N^2(E)=n$.

Построим теперь нормы матриц, подчиненные всем введенным выше нормам для векторов.

$$1. \|X\|_1 = \max_i |x_i|.$$

Подчиненная этой норме векторов норма матриц есть

$$\|A\|_1 = \max_i \sum_{k=1}^n |a_{ik}|.$$

Действительно, пусть $\|X\|_1=1$. Тогда

$$\|AX\|_1 = \max_i \left| \sum_{k=1}^n a_{ik}x_k \right| \leqslant \max_i \sum_{k=1}^n |a_{ik}| \cdot |x_k| \leqslant \max_i \sum_{k=1}^n |a_{ik}|.$$

Следовательно,

$$\max_{\|X\|=1} \|AX\| \leqslant \max_i \sum_{k=1}^n |a_{ik}|.$$

Докажем теперь, что $\max_{\|X\|=1} \|AX\|$ в действительности равен $\max_i \sum_{k=1}^n |a_{ik}|$.

Для этого построим такой вектор X_0 , что $\|X_0\|_1=1$ и $\|AX_0\|=\max_i \sum_{k=1}^n |a_{ik}|$. Именно, пусть $\sum_{k=1}^n |a_{ik}|$ достигает наибольшего значения при $i=j$; тогда в качестве компонент $x_k^{(0)}$ вектора X_0 возьмем $x_k^{(0)} = \frac{|a_{jk}|}{a_{jk}}$, если $a_{jk} \neq 0$, и $x_k^{(0)} = 1$, если $a_{jk} = 0$. Очевидно, что

$\|X_0\| = 1$. Далее, $\left| \sum_{k=1}^n a_{ik} x_k^{(0)} \right| \leq \sum_{k=1}^n |a_{ik}| \leq \sum_{k=1}^n |a_{jk}|$ при $i \neq j$ и $\left| \sum_{k=1}^n a_{jk} x_k^{(0)} \right| = \sum_{k=1}^n |a_{jk}|$. Следовательно,

$$\max_i \left| \sum_{k=1}^n a_{ik} x_k^{(0)} \right| = \sum_{k=1}^n |a_{jk}| = \max_i \sum_{k=1}^n |a_{ik}|.$$

Таким образом, $\|AX_0\|_1 = \max_i \sum_{k=1}^n |a_{ik}|$, что и требовалось доказать.

Очевидно, что норма $\|A\|_1$ может быть представлена в виде

$$\|A\|_1 = \max_i \|A^* e_i\|_\infty.$$

II. $\|X\|_\infty = \sum_{i=1}^n |x_i|$.

Подчиненная этой норме векторов норма матриц есть

$$\|A\|_\infty = \max_k \sum_{i=1}^n |a_{ik}|.$$

Действительно, пусть $\|X\|_\infty = 1$. Тогда

$$\begin{aligned} \|AX\| &= \sum_{i=1}^n \left| \sum_{k=1}^n a_{ik} x_k \right| \leq \sum_{i=1}^n \sum_{k=1}^n |a_{ik}| |x_k| \leq \sum_{k=1}^n |x_k| \left(\sum_{i=1}^n |a_{ik}| \right) \leq \\ &\leq \left(\max_k \sum_{i=1}^n |a_{ik}| \right) \sum_{k=1}^n |x_k| = \max_k \sum_{i=1}^n |a_{ik}|. \end{aligned}$$

Теперь возьмем вектор X_0 следующим образом: пусть $\sum_{i=1}^n |a_{ik}|$ достигает наибольшего значения для столбца с номером j . Положим $x_k^{(0)} = 0$ при $k \neq j$ и $x_j^{(0)} = 1$. Очевидно, что вектор, построенный таким образом, имеет норму, равную единице. Далее,

$$\|AX_0\| = \sum_{i=1}^n \left| \sum_{k=1}^n a_{ik} x_k^{(0)} \right| = \sum_{i=1}^n |a_{ij}| = \max_k \sum_{i=1}^n |a_{ik}|.$$

Таким образом,

$$\max_{\|X\|_\infty = 1} \|AX\| = \max_k \sum_{i=1}^n |a_{ik}|,$$

что и требовалось доказать.

Очевидно, что норма $\|A\|_\infty$ может быть представлена в виде

$$\|A\|_\infty = \max_i \|A e_i\|_\infty.$$

$$\text{III. } |X|^2 = \sum_{k=1}^n |x_k|^2 = (X, X).$$

Подчиненная этой норме векторов норма матриц есть

$$\|A\| = \sqrt{\lambda_1},$$

где λ_1 есть наибольшее собственное число матрицы A^*A . Действительно,

$$\|A\| = \max_{|X|=1} |AX|.$$

Но

$$|AX|^2 = (AX, AX) = (X, A^*AX).$$

Матрица A^*A эрмитова. Пусть λ_1 ее наибольшее собственное значение. Тогда при $|X|=1$, $\max(X, A^*AX) = \lambda_1$. Следовательно,

$$\|A\| = \sqrt{\lambda_1}.$$

Иногда эту норму называют верхней гранью матрицы. Соответственно наименьшее собственное значение матрицы A^*A называют нижней гранью матрицы A . Очевидно, что нижняя грань матрицы A есть число обратное к верхней грани обратной матрицы.

Сравнение различных норм матриц приводит к следующим неравенствам¹⁾

$$\frac{1}{n} M(A) \leq \|A\|_1 \leq M(A) \quad (1)$$

$$\frac{1}{n} M(A) \leq \|A\|_\infty \leq M(A) \quad (2)$$

$$\frac{1}{n} M(A) \leq \|A\| \leq M(A) \quad (3)$$

$$\frac{1}{n} M(A) \leq N(A) \leq M(A) \quad (4)$$

$$\frac{1}{\sqrt{n}} N(A) \leq \|A\| \leq N(A) \quad (5)$$

$$\frac{1}{\sqrt{n}} N(A) \leq \|A\|_1 \leq \sqrt{n} N(A) \quad (6)$$

$$\frac{1}{\sqrt{n}} N(A) \leq \|A\|_\infty \leq \sqrt{n} N(A) \quad (7)$$

$$\frac{1}{\sqrt{n}} \|A\| \leq \|A\|_1 \leq \sqrt{n} \|A\| \quad (8)$$

$$\frac{1}{\sqrt{n}} \|A\| \leq \|A\|_\infty \leq \sqrt{n} \|A\| \quad (9)$$

$$\frac{1}{n} \|A\|_1 \leq \|A\|_\infty \leq n \|A\|_1. \quad (10)$$

¹⁾ Некоторые из этих неравенств имеются в работе Тюринга [1].

Правые неравенства (1), (2), (3), (4), (5) уже отмечались. Все они точные, ибо они превращаются в равенство для матрицы A , у которой $a_{ij} = 1$ ($i, j = 1, \dots, n$).

Докажем левые половины неравенств (1), (2), (3), (4), (5). Имеем

$$\|A\|_1 = \max_i \sum_j |a_{ij}| \geq \max_{i,j} |a_{ij}| = \frac{1}{n} M(A)$$

$$\|A\|_B = \max_j \sum_i |a_{ij}| \geq \max_{i,j} |a_{ij}| = \frac{1}{n} M(A)$$

$$\begin{aligned} \|A\| &= \max_{\|X\|=1} |AX| \geq \max_i |Ae_i| = \\ &= \max_i \sqrt{|a_{i1}|^2 + \dots + |a_{in}|^2} \geq \max_{i,j} |a_{ij}| = \frac{1}{n} M(A) \end{aligned}$$

$$\|A\|^2 = \max \lambda_i \geq \frac{1}{n} (\lambda_1 + \dots + \lambda_n) = \frac{1}{n} \operatorname{Sp} A^* A = \frac{1}{n} N^2(A).$$

Здесь $\lambda_1, \dots, \lambda_n$ собственные значения матрицы $A^* A$. Наконец,

$$N(A) = \sqrt{\sum_{i,j} |a_{ij}|^2} \geq \max_{i,j} |a_{ij}| = \frac{1}{n} M(A).$$

Эти неравенства также точные. Первые четыре из этих неравенств обращаются в равенство для $A = E$, последнее для

$$A = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

Докажем теперь неравенство (8). Имеем

$$\|A\| = \max_{\|X\|=1} |A^* X| \geq \max_i |A^* e_i| \geq \max_i \frac{1}{\sqrt{n}} \|A^* e_i\|_1 = \frac{1}{\sqrt{n}} \|A\|_1.$$

Далее,

$$\begin{aligned} \|A\| &= \max_{\|X\|=1} |AX| \leq \max_{\|X\|=1} \sqrt{n} \|AX\|_1 \leq \\ &\leq \max_{\|X\|=1} \sqrt{n} \|A\|_1 \|X\|_1 \leq \sqrt{n} \|A\|_1. \end{aligned}$$

откуда $\frac{1}{\sqrt{n}} \|A\| \leq \|A\|_1$.

Заменой A на A^* устанавливается неравенство (9). Из неравенств (8) и (9) непосредственно вытекает неравенство (10).

Из (8) и (9) и правого неравенства (5) следуют правые неравенства (6) и (7).

Докажем левое неравенство (6). Имеем

$$\|A\|_1 = \max_i \|A^* e_i\|_1 \geq \max_i |A^* e_i| \geq \\ \geq \frac{1}{\sqrt{n}} \sqrt{|A^* e_1|^2 + \dots + |A^* e_n|^2} = \frac{1}{\sqrt{n}} N(A).$$

Замена A на A^* превращает неравенство (6) в неравенство (7).

Неравенства (6) — (10) точные. Правое неравенство (6), левое (7), правое (8), левое (9) и левое (10) достигается для

$$A = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

левое (6), правое (7), левое (8), правое (9) и правое (10) для

$$A = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \end{bmatrix}.$$

4. Сходимость геометрической прогрессии. Докажем теперь несколько теорем, связанных с понятием предела.

Теорема 13.2. Для того чтобы $A^m \rightarrow 0$, необходимо и достаточно, чтобы все собственные значения матрицы A были по модулю меньше единицы.

Доказательство. Ясно, что при любой фиксированной неособенной матрице C матрицы A^m и $(C^{-1}AC)^m = C^{-1}A^mC$ стремятся или не стремятся к нулю одновременно. Поэтому достаточно доказать справедливость теоремы для канонической матрицы Жордана. Далее, для того чтобы последовательность квазидиагональных матриц одинакового строения стремилась к нулю, необходимо и достаточно, чтобы сходилась к нулю последовательность отдельных ящиков. Тем самым нам нужно установить условия сходимости к нулю лишь для матриц J^m , где J канонический ящик Жордана. Пусть

$$J = \begin{bmatrix} \lambda & & & & \\ 1 & \lambda & & & \\ 1 & 1 & \lambda & & \\ \vdots & \vdots & \ddots & \ddots & \\ 1 & 1 & \dots & \dots & \lambda \end{bmatrix}.$$

Легко видеть, что при $m \geq k - 1$

$$J^m = \begin{bmatrix} \lambda^m & & & \\ \left(\frac{m}{1}\right)\lambda^{m-1} & \lambda^m & & \\ \dots & \dots & \dots & \\ \left(\frac{m}{k-1}\right)\lambda^{m-k+1} & \left(\frac{m}{k-2}\right)\lambda^{m-k+2} & \dots & \lambda^m \end{bmatrix},$$

где k порядок ящика, $\binom{m}{j} = \frac{m(m-1) \dots (m-j+1)}{j!}$.

Для сходимости J^m к нулю необходимо, чтобы $|\lambda| < 1$, ибо диагональными элементами J^m являются λ^m .

Но это условие и достаточное, ибо при этом условии

$$\frac{m(m-1) \dots (m-j+1)}{j!} \lambda^{m-j} \rightarrow 0$$

при $m \rightarrow \infty$, для любого $j = 1, \dots, k - 2$.

Данные в теореме 13.2 условия неудобны для проверки, ибо они требуют знания собственных чисел матрицы A . Поэтому установим некоторые более простые достаточные условия для того, чтобы $A^m \rightarrow 0$.

Теорема 13.3. Для того чтобы $A^m \rightarrow 0$, достаточно, чтобы хоть одна из норм A была меньше единицы.

Доказательство. В силу 4-го требования для нормы, имеем:

$$\|A^m\| \leq \|A^{m-1}\| \cdot \|A\| \leq \dots \leq \|A\|^m.$$

Поэтому, если $\|A\| < 1$, то $\|A^m\| \rightarrow 0$, и, в силу сказанного выше, $A^m \rightarrow 0$.

Сопоставляя теоремы 13.2 и 13.3, мы приходим к следующему выводу.

Теорема 13.4. Модуль каждого собственного числа матрицы не превосходит любой из ее норм.

Доказательство. Пусть $\|A\| = a$. Рассмотрим матрицу $B = \frac{1}{a+\varepsilon} A$, где ε любое положительное число. Тогда

$$\|B\| = \frac{a}{a+\varepsilon} < 1.$$

Следовательно, $B^m \rightarrow 0$ при $m \rightarrow \infty$. В силу теоремы 13.2 ее собственные числа по модулю меньше единицы. Но очевидно, что собственные числа матрицы B равны $\frac{1}{a+\varepsilon} \lambda_i$, где λ_i собственные числа матрицы A . Таким образом, $\frac{|\lambda_i|}{a+\varepsilon} < 1$, т. е. $|\lambda_i| < a + \varepsilon$. Так как ε можно взять как угодно малым, $|\lambda_i| \leq a$.

Теорема 13.5. Для того чтобы ряд

$$E + A + A^2 + \dots + A^m + \dots \quad (1)$$

сходился, необходимо и достаточно, чтобы $A^m \rightarrow 0$ при $m \rightarrow \infty$. В этом случае сумма ряда (1) равна $(E - A)^{-1}$.

Доказательство. Необходимость этого условия очевидна. Покажем, что оно и достаточно. Пусть $A^m \rightarrow 0$. В силу теоремы 13.2 все собственные числа матрицы A по модулю меньше единицы. Следовательно, $|E - A| \neq 0$, и потому существует $(E - A)^{-1}$.

Рассмотрим тождество

$$(E + A + \dots + A^k)(E - A) = E - A^{k+1}.$$

Умножив его справа на $(E - A)^{-1}$, получим

$$E + A + \dots + A^k = (E - A)^{-1} - A^{k+1}(E - A)^{-1}.$$

Отсюда следует, что при $k \rightarrow \infty$

$$E + A + \dots + A^k \rightarrow (E - A)^{-1},$$

ибо $A^{k+1} \rightarrow 0$.

Следовательно,

$$E + A + \dots + A^m + \dots = (E - A)^{-1},$$

что и требовалось доказать.

В силу теоремы 13.2 необходимым и достаточным условием сходимости ряда (1) является неравенство $|\lambda_i| < 1$ для всех собственных чисел матрицы A . Достаточным признаком сходимости в силу теоремы 13.3 является неравенство $\|A\| < 1$ хотя бы для одной из норм. При выполнении этого условия легко дать следующую оценку быстроты сходимости ряда (1).

Теорема 13.6. Если $\|A\| < 1$, то

$$\|(E - A)^{-1} - (E + A + \dots + A^k)\| \leq \frac{\|A\|^{k+1}}{1 - \|A\|}.$$

Доказательство. Имеем:

$$(E - A)^{-1} - (E + A + \dots + A^k) = A^{k+1} + A^{k+2} + \dots$$

Отсюда

$$\begin{aligned} \|(E - A)^{-1} - (E + A + \dots + A^k)\| &\leq \\ &\leq \|A\|^{k+1} + \|A\|^{k+2} + \dots = \frac{\|A\|^{k+1}}{1 - \|A\|}. \end{aligned}$$

Теорема доказана.

5. Некоторые оценки собственных значений. В предыдущем пункте были получены некоторые оценки собственных значений.

Именно, было установлено, что все собственные значения по модулю меньше любой нормы матрицы. Особенно удобными являются оценки при помощи 1-й и 2-й норм, так как они просто выражаются через элементы матрицы.

Сейчас мы выведем оценки, дающие более точную информацию о расположении собственных значений матрицы. Эти оценки носят название оценок Гершгорина, так как они впервые появились в его работе [1].

Теорема 13.7. Пусть $A = (a_{ij})$ матрица с любыми комплексными элементами. Все собственные значения этой матрицы находятся в области D , являющейся объединением кругов

$$|z - a_{ii}| \leq R_i \quad (i = 1, \dots, n),$$

где

$$R_i = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|.$$

Доказательство. Пусть λ любое собственное значение матрицы A и

$$X = (x_1, \dots, x_n)'$$

соответствующий ему собственный вектор. Тогда

$$\sum_{j=1}^n a_{ij} x_j = \lambda x_i \quad (i = 1, \dots, n)$$

или, что то же самое,

$$\sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j = (\lambda - a_{ii}) x_i.$$

Пусть индекс i выбран так, что x_i есть наибольшая по модулю компонента вектора X . Тогда

$$\lambda - a_{ii} = \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} \frac{x_j}{x_i},$$

откуда

$$|\lambda - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \cdot \left| \frac{x_j}{x_i} \right| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| = R_i.$$

Итак, для каждого собственного значения λ найдется круг с центром a_{ii} и радиусом R_i , содержащий это собственное значение. Следовательно, все собственные значения находятся в объединении всех таких кругов.

Замечание 1. Область D Гершгорина целиком лежит внутри круга $|z| \leq \|A\|_1$ и касается его границы. Действительно, если $|z - a_{ii}| \leq R_i$, то $|z| \leq |a_{ii}| + R_i = \sum_{j=1}^n |a_{ij}|$, причем существует такое z , при котором неравенство превращается в равенство. Следовательно, $|z| \leq \max_i \sum_{j=1}^n |a_{ij}| = \|A\|_1$, причем снова найдется число z , осуществляющее равенство.

Таким образом, хотя область Гершгорина есть, вообще говоря, часть круга $|z| \leq \|A\|_1$, но оценка для наибольшего по модулю собственного значения получается такой же, как при помощи 1-й нормы.

Замечание 2. Взяв вместо матрицы A транспонированную, получим другую область D' , которая будет содержаться внутри круга $|z| \leq \|A\|_H$.

Замечание 3. Область Гершгорина может распадаться на несколько связных частей. При этом каждая связная часть содержит внутри себя столько собственных значений, сколько кругов ее составляют.

Для доказательства введем в рассмотрение матрицу

$$A_u = \begin{bmatrix} a_{11} & ua_{12} & \dots & ua_{1n} \\ ua_{21} & a_{22} & \dots & ua_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ua_{n1} & ua_{n2} & \dots & a_{nn} \end{bmatrix},$$

где u вещественный параметр. Ясно, что $A_1 = A$.

Область Гершгорина для матрицы A_u есть, очевидно, объединение кругов $|z - a_{ii}| \leq uR_i$, где R_i радиусы соответствующих кругов для матрицы A . Таким образом, область D_u Гершгорина для A_u при $0 \leq u \leq 1$ содержится в области Гершгорина D для матрицы A . Поэтому при непрерывном изменении u от 0 до 1 собственные значения матриц A_u непрерывно изменяются, не выходя из области D . Следовательно, внутри каждой связной компоненты области D находится одинаковое число собственных значений для любой матрицы A_u , $0 \leq u \leq 1$.

Но собственными значениями матрицы A_0 являются диагональные элементы a_{ii} , являющиеся центрами кругов, образующих D . Следовательно, число собственных значений матрицы A_0 (а следовательно, и матрицы A), лежащих внутри некоторой связной компоненты области D точно равно числу кругов, объединением которых получается эта связная компонента.

За последние годы опубликовано много работ, посвященных дальнейшему уточнению области, содержащей все собственные значения матрицы.

§ 14. Градиент функционала

1. Определение. Пусть $F(\mathbf{X})$ функционал, вообще говоря, не линейный, определенный в вещественном евклидовом пространстве \mathbb{R}^n и принимающий вещественные значения. Пусть вектор \mathbf{X} задан координатами в некотором ортонормированном базисе. Тогда функционал $F(\mathbf{X})$ есть функция $F(x_1, \dots, x_n)$ от переменных координат вектора \mathbf{X} . Мы будем предполагать, что функционал $F(\mathbf{X})$ дифференцируем, для чего достаточно потребовать, чтобы функция $F(x_1, \dots, x_n)$ имела непрерывные частные производные по всем аргументам.

Пусть \mathbf{Y} произвольный вектор единичной длины с координатами y_1, \dots, y_n .

Производной от функционала F в точке \mathbf{X} по направлению \mathbf{Y} называется выражение

$$\frac{\partial F(\mathbf{X})}{\partial \mathbf{Y}} = \lim_{t \rightarrow 0} \frac{F(\mathbf{X} + t\mathbf{Y}) - F(\mathbf{X})}{t} = \frac{d}{dt} F(\mathbf{X} + t\mathbf{Y}) \Big|_{t=0}.$$

Производная $\frac{\partial F(\mathbf{X})}{\partial \mathbf{Y}}$ характеризует скорость изменения функционала F при изменении „аргумента“ в направлении вектора \mathbf{Y} .

Имеем далее

$$F(\mathbf{X} + t\mathbf{Y}) = F(x_1 + ty_1, \dots, x_n + ty_n)$$

и потому

$$\begin{aligned} \frac{\partial F(\mathbf{X})}{\partial \mathbf{Y}} &= \frac{d}{dt} F(x_1 + ty_1, \dots, x_n + ty_n) \Big|_{t=0} = \\ &= \frac{\partial F}{\partial x_1} y_1 + \dots + \frac{\partial F}{\partial x_n} y_n = (\mathbf{Z}, \mathbf{Y}), \end{aligned}$$

где \mathbf{Z} — вектор с координатами $\frac{\partial F}{\partial x_1}, \dots, \frac{\partial F}{\partial x_n}$. Вектор \mathbf{Z} называется градиентом функционала $F(\mathbf{X})$. Из последнего равенства вытекает, что

$$\frac{\partial F(\mathbf{X})}{\partial \mathbf{Y}} = |\mathbf{Z}| \cos(\mathbf{Z}, \mathbf{Y}),$$

ибо $|\mathbf{Y}| = 1$. Отсюда следует, что

$$-|\mathbf{Z}| \leq \frac{\partial F(\mathbf{X})}{\partial \mathbf{Y}} \leq |\mathbf{Z}|,$$

причем $\frac{\partial F(\mathbf{X})}{\partial \mathbf{Y}} = |\mathbf{Z}|$, если направление \mathbf{Y} совпадает с направлением градиента, и $\frac{\partial F(\mathbf{X})}{\partial \mathbf{Y}} = -|\mathbf{Z}|$, если направление \mathbf{Y} противоположно

направлению градиента. Поэтому направление градиента есть направление наибольшей скорости роста функционала F в данной точке, а направление, противоположное градиенту, есть направление наибольшей скорости убывания.

2. Градиенты некоторых функционалов. Пусть A самосопряженный оператор. Определим градиент функционала

$$h(X) = (AX, X).$$

Имеем

$$\frac{h(X + tY) - h(X)}{t} = \frac{(A(X + tY), X + tY) - (AX, X)}{t} = \\ = 2(AX, Y) + t(AY, Y),$$

откуда

$$\frac{\partial h(X)}{\partial Y} = 2(AX, Y),$$

и, следовательно, градиент (AX, X) равен $2AX$. В частности, градиент (X, X) равен $2X$.

Так как в дальнейшем нам важно лишь направление градиента, мы отбрасываем положительный множитель 2 и будем подразумевать под градиентом функционала (AX, X) вектор AX .

Совершенно аналогично находим, что градиент функционала $f(X) = (AX, X) - 2(F, X) + c$ равен (с точностью до множителя 2) вектору $AX - F$.

Вычислим, наконец, градиент функционала $\mu(X) = \frac{(AX, X)}{(X, X)}$.

Имеем

$$\frac{\partial \mu(X)}{\partial Y} = \frac{(X, X) \frac{\partial}{\partial Y} (AX, X) - (AX, X) \frac{\partial}{\partial Y} (X, X)}{(X, X)^2} = \\ = \frac{(X, X) 2(AX, Y) - (AX, X) \cdot 2(X, Y)}{(X, X)^2} = \frac{2}{(X, X)} (AX - \mu(X)X, Y).$$

Следовательно, с точностью до положительного множителя $\frac{2}{(X, X)}$, градиент функционала $\mu(X)$ равен вектору $\xi = AX - \mu(X)X$. Отметим два свойства этого градиента:

- 1) $(X, \xi) = 0$;
- 2) $(\xi, \xi) = (\xi, AX)$.

Действительно,

$$(X, \xi) = (X, AX - \mu(X)X) = (X, AX) - \mu(X)(X, X) = 0$$

в силу определения $\mu(X)$.

Далее,

$$(\xi, \xi) = (\xi, AX - \mu(X)X) = (\xi, AX) - \mu(X)(X, \xi) = (\xi, AX)$$

ибо $(X, \xi) = 0$.

Первое из этих свойств имеет простой геометрический смысл, именно, градиент ξ есть проекция вектора $\eta = AX$ (градиента (AX, X)) на подпространство, ортогональное к вектору X , ибо

$$\eta = \mu(X)X + \xi,$$

причем $\mu(X)X$ пропорционально X и ξ ортогонально к X .

Это обстоятельство легко можно было бы обосновать посредством чисто геометрического рассмотрения, без использования проведенных выше вычислений, но мы на этом не будем останавливаться.

ГЛАВА II

ТОЧНЫЕ МЕТОДЫ РЕШЕНИЯ СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ

Настоящая глава посвящена трем задачам, тесно связанным друг с другом: задаче решения линейной неоднородной алгебраической системы уравнений, задаче обращения матрицы и так называемой задаче исключения.

Все эти задачи теоретически решаются достаточно просто. Однако при большом порядке матрицы, связанной с задачей, фактическое решение этих задач требует большого числа вычислительных операций.

В настоящее время имеется большое количество методов численного решения систем линейных уравнений, и работа над их усовершенствованием интенсивно ведется в наши дни.

Численные методы, решающие указанные задачи, делятся на две группы: точные и итерационные методы. Под точными методами мы подразумеваем методы, которые дают решение задачи при помощи конечного числа элементарных арифметических операций. При этом, если исходные данные, определяющие задачу, заданы точно (например, если они целые или рациональные числа, представленные в виде обыкновенных дробей) и вычисления выполняются точно (например, по правилам действий над обыкновенными дробями), то решение также получается точное. В точных методах число необходимых для решения задачи вычислительных операций зависит только от вида вычислительной схемы и от порядка матрицы, определяющей данную задачу.

Исторически первым методом этой группы является метод, основанный на идее исключения неизвестных. В применении к решению линейной неоднородной системы алгорифм этого метода, связанный с именем Гаусса, состоит из цепочки последовательных исключений, посредством которых данная система превращается в систему с треугольной матрицей, решение которой не составляет труда.

В настоящее время разработано много различных вычислительных схем метода Гаусса в применении ко всем трем упомянутым задачам.

Среди этих схем следует отметить так называемые компактные схемы, требующие минимальной записи или запоминания промежуточных результатов. В компактных схемах используется операция накопления, т. е. вычисления сумм вида $\sum a_k b_k$. Применение этой

операции с двойной точностью, т. е. с сохранением всех значащих цифр отдельных слагаемых с округлением лишь после окончания суммирования, значительно снижает ошибки округления.

Близким к методу Гаусса является метод, основанный на идее окаймления. В основе этого метода лежит представление данной матрицы в виде последовательно вложенных друг в друга матриц, начиная с матрицы первого порядка.

Различные модификации метода исключения, так же как и метод окаймления, по существу связаны с разложением матрицы в произведение двух треугольных матриц, теоретически описанным в § 1, п. 12.

К точным методам обращения матриц относится и метод, основанный на идее постепенного изменения обращаемой матрицы с внесением последовательно поправок к обратной (метод пополнения).

В последние годы получили большое распространение точные методы, основанные на идее построения вспомогательной системы векторов, ортогональных в той или иной метрике. Как будет показано ниже, метод Гаусса может быть включен и в эту группу методов.

Итерационные методы доставляют средство для приближенного решения системы линейных уравнений. Решение системы при помощи итерационных методов получается как предел последовательных приближений, вычисляемых некоторым единообразным процессом. При применении итерационных методов существенным является не только сходимость построенных последовательных приближений, но и быстрота сходимости. В этом отношении каждый итерационный метод не является универсальным; давая быструю сходимость для одних матриц, он может сходиться медленно или даже совсем не сходиться для других матриц. Поэтому при применении итерационных методов важную роль играет предварительная подготовка системы, т. е. замена данной системы ей эквивалентной, устроенной так, чтобы для нее выбранный процесс сходился по возможности быстро, а также различные приемы ускорения.

Настоящая глава книги посвящена изложению простейших методов, входящих в группу точных методов. Другие методы (точные и итерационные) будут излагаться далее на протяжении почти всей книги.

§ 15. Обусловленность матриц

При численном решении систем линейных уравнений даже точными методами возникает несколько источников неточности решения. Одним из них является необходимость округления чисел в процессе вычисления. При этом может случиться, что по ходу вычисления придется столкнуться с явлением исчезновения значащих цифр в результате вычитания двух величин, близких друг другу. Исчезновение значащих цифр может быть причиной настолько значительного снижения точности результата, что из-за этого иногда бывает необходимо изменить схему вычисления или переделать работу с большим числом значащих цифр в промежуточных выкладках,

Второй источник появляется в условиях, когда система линейных уравнений возникает в процессе решения практической задачи, так что элементы матрицы коэффициентов, так же как и свободные члены, известны лишь приближенно, с некоторой степенью точности. Неточность самих исходных данных порождает ошибки в решении, так как изменение коэффициентов системы в пределах заданной точности влечет за собой изменение решения. Займемся более детально исследованием этой причины неточности в решении системы.

Теоретическое решение системы $AX = F$ дается формулой $X = A^{-1}F$, где A^{-1} матрица, обратная к A . Как известно, A^{-1} существует в том и только в том случае, если $|A| \neq 0$. Однако, если элементы матрицы A заданы приближенно, возможно, что даже сам вопрос о том, имеет ли матрица A отличный от нуля определитель или нет, лишен смысла. Именно, может случиться, что при точном вычислении определителя, исходя из приближенных значений элементов матрицы, принятых за точные, определитель оказывается отличным от нуля, но изменение элементов в пределах точности их задания может привести к матрице с нулевым определителем. Ясно, что система с матрицей, обладающей указанным свойством, не может быть решена со сколько-нибудь удовлетворительной точностью. Система практически оказывается несовместной.

Мы будем называть обратную матрицу *устойчивой*, если малым изменениям в элементах матрицы будут соответствовать малые изменения в элементах обратной матрицы.

Очевидно, что для устойчивости обратной матрицы во всяком случае необходимо, чтобы определитель матрицы был бы не слишком мал. Степень малости определителя, препятствующую получению устойчивой обратной матрицы, трудно определить раз навсегда, ибо само понятие „не слишком мал“ вряд ли может быть определено точно. Прежде чем переходить к уточнению этого понятия, рассмотрим следующий пример. Пусть

$$W = \begin{bmatrix} 5 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix}.$$

Нетрудно вычислить, что $|W| = 1$. При этом

$$W^{-1} = \begin{bmatrix} 68 & -41 & -17 & 10 \\ -41 & 25 & 10 & -6 \\ -17 & 10 & 5 & -3 \\ 10 & -6 & -3 & 2 \end{bmatrix}.$$

Если считать, что элементы матрицы W заданы абсолютно точно, то все обстоит благополучно. Однако определитель матрицы W является довольно малым. В самом деле, известная оценка Адамара для значения определителя

$$\Delta \leq \sqrt{\prod_{i=1}^n \sum_{j=1}^n |a_{ij}|^2}$$

дает

$$|W| \leq \sqrt{2534437350} \approx 50000,$$

что означает (в силу достижимости оценки Адамара), что при таких же суммах квадратов модулей элементов строк модуль определителя может доходить до 50 000.

Это дает основание предположить, что матрица W^{-1} будет малоустойчивой, что легко подтверждается следующим рассуждением. Проследим, как изменяется определитель матрицы W при малых изменениях ее первого элемента. Пусть

$$W(\varepsilon) = \begin{bmatrix} 5+\varepsilon & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix}.$$

Тогда $|W(\varepsilon)| = 1 + 68\varepsilon$, откуда следует, что при $\varepsilon = -\frac{1}{68} \approx -0.015$ матрица $W(\varepsilon)$ будет особенной. Таким образом, если считать элементы матрицы W известными с точностью до 0.02, то практически матрица W должна рассматриваться как особенная.

Покажем, как будут меняться все элементы обратной матрицы при незначительном изменении первого элемента матрицы W . Пусть

$$W_1 = \begin{bmatrix} 5.0002 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix}.$$

Тогда

$$|W_1| = 1.0136,$$

$$W_1^{-1} = \begin{bmatrix} 67.088 & -40.450 & -16.772 & 9.866 \\ -40.450 & 24.664 & 9.862 & -5.916 \\ -16.772 & 9.862 & 4.943 & -2.966 \\ 9.866 & -5.916 & -2.966 & 1.980 \end{bmatrix}$$

и, следовательно,

$$W^{-1} - W_1^{-1} = \begin{bmatrix} 0.912 & -0.550 & -0.228 & 0.134 \\ -0.550 & 0.335 & 0.138 & -0.084 \\ -0.228 & 0.138 & 0.057 & -0.033 \\ 0.134 & -0.084 & -0.033 & 0.020 \end{bmatrix}.$$

Еще большее расхождение мы получим, если рассмотрим матрицу

$$W_2 = \begin{bmatrix} 4.99 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix},$$

для которой $|W_2| = 0.320$, и

$$W_2^{-1} = \begin{bmatrix} 204.82 & -128.12 & -53.12 & 31.25 \\ -128.12 & 77.53 & 31.78 & -18.81 \\ -53.12 & 31.78 & 14.03 & -8.31 \\ 31.25 & -18.81 & -8.31 & 5.12 \end{bmatrix}.$$

Интересно отметить, что при этом само определение элементов матриц W^{-1} , W_1^{-1} и W_2^{-1} не встречает трудностей, связанных, например, с исчезновением значащих цифр.

Выясним теперь в общем виде, как могут изменяться элементы обратной матрицы при изменениях элементов самой матрицы. Пусть $A^{-1} = (a_{ij})$.

Из равенства

$$AA^{-1} = E$$

имеем

$$\frac{\partial A}{\partial a_{ij}} A^{-1} + A \frac{\partial A^{-1}}{\partial a_{ij}} = 0,$$

откуда

$$\frac{\partial (A^{-1})}{\partial a_{ij}} = -A^{-1} \frac{\partial A}{\partial a_{ij}} A^{-1}.$$

Но

$$\frac{\partial A}{\partial a_{ij}} = e_{ij},$$

где e_{ij} матрица, элемент i -й строки и j -го столбца которой равен единице, все остальные элементы равны нулю. Следовательно,

$$\begin{aligned} \frac{\partial(A^{-1})}{\partial a_{ij}} &= -A^{-1}e_{ij}A^{-1} = -A^{-1}e_{ii}e_{jj}A^{-1} = \\ &= -\left[\begin{array}{cccc} \alpha_{1i} & 0 & \dots & 0 \\ \alpha_{2i} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{ni} & 0 & \dots & 0 \end{array} \right] \left[\begin{array}{cccc} \alpha_{j1} & \alpha_{j2} & \dots & \alpha_{jn} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{array} \right] = \\ &= -\left[\begin{array}{cccc} \alpha_{1i}\alpha_{j1} & \alpha_{1i}\alpha_{j2} & \dots & \alpha_{1i}\alpha_{jn} \\ \alpha_{2i}\alpha_{j1} & \alpha_{2i}\alpha_{j2} & \dots & \alpha_{2i}\alpha_{jn} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{ni}\alpha_{j1} & \alpha_{ni}\alpha_{j2} & \dots & \alpha_{ni}\alpha_{jn} \end{array} \right], \end{aligned}$$

и потому

$$\frac{\partial \alpha_{kl}}{\partial a_{ij}} = -\alpha_{ki}\alpha_{lj},$$

откуда

$$d\alpha_{kl} = -\sum_{i,j} \alpha_{ki}\alpha_{lj} da_{ij}.$$

Таким образом, изменение каждого элемента обратной матрицы, вызванное изменением элемента a_{ij} , равно этому изменению, умноженному на произведение некоторых двух элементов обратной матрицы. Если элементы обратной матрицы достаточно велики (что при малом определителе бывает всякий раз), то незначительная ошибка в элементах исходной матрицы влечет за собой значительные изменения в элементах обратной. Конечно, в некоторых случаях ошибки от изменения различных элементов матрицы могут компенсировать друг друга.

Будем называть матрицу плохо обусловленной, если соответствующая ей обратная матрица будет неустойчивой.

Плохо обусловленная матрица может оказаться практически особенной, если ее элементы заданы приближенно.

Естественно, что система линейных уравнений с плохо обусловленной матрицей мало устойчива, т. е. ее решение сильно изменяется при малых изменениях как коэффициентов, так и свободных членов. Проследим, как значительна эта неустойчивость. Пусть

$$AX = F$$

данная система. Тогда $X = A^{-1}F$ и, следовательно,

$$\frac{\partial X}{\partial a_{ij}} = \frac{\partial A^{-1}}{\partial a_{ij}} F = -A^{-1}e_{ij}A^{-1}F = -A^{-1}e_{ij}X = -\left[\begin{array}{c} \alpha_{1i}x_j \\ \alpha_{2i}x_j \\ \vdots \\ \alpha_{ni}x_j \end{array} \right].$$

откуда

$$\frac{\partial x_k}{\partial a_{ij}} = -\alpha_{ki}x_j.$$

Аналогично из формулы

$$\frac{\partial X}{\partial f_i} = A^{-1} \frac{\partial F}{\partial f_i} = A^{-1} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha_{1i} \\ \vdots \\ \alpha_{ii} \\ \vdots \\ \alpha_{ni} \end{bmatrix}.$$

следует, что $\frac{\partial x_k}{\partial f_i} = \alpha_{ki}$.

Таким образом, если элементы обратной матрицы велики, то малое изменение как коэффициентов системы, так и свободных членов влечет за собой значительное изменение решения. Так, например, для рассмотренной выше матрицы W „близкие“ свободные члены

$$F_1 = (23, 32, 33, 31)'$$

$$F_2 = (23.001, 31.999, 32.999, 31.001)'$$

$$F_3 = (23.01, 31.99, 32.99, 31.01)'$$

$$F_4 = (23.1, 31.9, 32.9, 31.1)'$$

будут приводить к далеко не „близким“ решениям

$$X_1 = (1, 1, 1, 1)'$$

$$X_2 = (1.136, 0.918, 0.965, 1.021)'$$

$$X_3 = (2.36, 0.18, 0.65, 1.21)'$$

$$X_4 = (14.6, -7.2, -2.5, 3.1)'.$$

Явление плохой обусловленности матрицы было известно давно, по-видимому, еще Гауссу. Однако изучение и количественные характеристики этого явления появились недавно.

Мы видели, что „малость“ определителя является причиной плохой обусловленности матрицы. Однако легко видеть, что одна величина определителя не может вполне характеризовать обусловленность матрицы. Так, например, очевидно, что матрицы, отличающиеся лишь постоянным множителем, надлежит считать одинаково обусловленными. Соответствующие же им определители будут отличаться на n -ю степень множителя. Отсюда следует, что величину определителя нужно сравнивать хотя бы с n -й степенью наибольшего

элемента матрицы или n -й степенью какой-либо нормы матрицы. Однако и этого недостаточно. Так, из матриц

$$\begin{bmatrix} 20 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.05 \end{bmatrix} \text{ и } \begin{bmatrix} 20 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.25 \end{bmatrix}$$

первая обусловлена хуже, чем вторая, хотя у них одинаковы как определители, так и наибольшие элементы.

Действительно, соответствующие обратные матрицы будут

$$\begin{bmatrix} 0.05 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 20 \end{bmatrix} \text{ и } \begin{bmatrix} 0.05 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 4 \end{bmatrix},$$

и потому одинаковые изменения в элементах данных матриц приведут к разным изменениям в элементах обратных матриц, которые в первой матрице будут более значительными.

Для характеристики матрицы с точки зрения ее обусловленности несколькими авторами предложены различные количественные характеристики, называемые числами обусловленности. Это два числа Тюринга¹⁾

$$N\text{-число} = \frac{1}{n} N(A) N(A^{-1}) = \nu(A)$$

$$M\text{-число} = \frac{1}{n} M(A) M(A^{-1}) = \mu(A),$$

где $N(A) = \sqrt{\operatorname{Sp} A' A}$ и $M(A) = n \cdot \max_{ij} |a_{ij}|$ — рассмотренные выше нормы матриц, число Тодда²⁾

$$P\text{-число} = \frac{\max |\lambda_i|}{\min |\lambda_i|} = \rho(A),$$

где λ_i — собственные значения матрицы A , и

$$H\text{-число} = \|A\| \|A^{-1}\|,$$

где $\|A\|$ — третья норма матрицы. Легко видеть, что

$$H\text{-число} = \sqrt{\frac{\rho_1}{\rho_n}} = \eta(A),$$

где ρ_1 и ρ_n наибольшее и наименьшее собственные значения матрицы $A' A$, A' — транспонированная с A матрица. Ясно, что

$$\eta(A) = \sqrt{\rho(A' A)}.$$

¹⁾ Тюринг [1].

²⁾ Тодд [1].

Для симметричных матриц P -число обусловленности совпадает с H -числом.

Числа обусловленности не дают, конечно, исчерпывающей характеристики обусловленности матрицы.

Отметим неравенства, связывающие эти числа:

$$\nu(A) \leq \rho(A) \leq n^2 \nu(A)$$

$$\nu(A) \leq \eta(A) \leq n \nu(A)$$

$$\rho(A) \leq \eta(A).$$

Легко подсчитать, что для ортогональных матриц $\nu(A) = \eta(A) = \rho(A) = 1$. Покажем, что все числа обусловленности не меньше единицы. Для $\rho(A)$ и $\eta(A)$ это очевидно. Для $\nu(A)$ это следует из тождества

$$\begin{aligned} \nu(A) &= \frac{1}{n} \sqrt{(\nu_1 + \dots + \nu_n) \left(\frac{1}{\mu_1} + \dots + \frac{1}{\mu_n} \right)} = \\ &= \frac{1}{n} \sqrt{n^2 + \sum_{i>j} \left(\frac{\sqrt{\mu_i}}{\sqrt{\mu_j}} - \frac{\sqrt{\mu_j}}{\sqrt{\mu_i}} \right)^2}. \end{aligned}$$

где μ_1, \dots, μ_n — собственные значения матрицы $A'A$.

Числа обусловленности $\nu(A)$ и $\eta(A)$ имеют следующий вероятностный смысл. Рассмотрим систему линейных уравнений $AX = F$, где вектор F задан точно, а элементы матрицы A являются независимыми случайными величинами со средними значениями a_{ij} и одинаковой дисперсией σ^2 , которая предполагается очень малой по сравнению с величинами коэффициентов. Тогда N -число обусловленности показывает во сколько раз отношение среднего квадратичного ошибок неизвестных к среднему квадратичному самих неизвестных превосходит отношение среднего квадратичного ошибок коэффициентов системы к среднему квадратичному самих коэффициентов. H -число дает отношение наибольшей полуоси к наименьшей полуоси для эллипса для рассеяния вектора, компонентами которого являются ошибки неизвестных¹⁾.

Легко подсчитать, что для матрицы W , рассмотренной выше,

$$P\text{-число} = H\text{-число} \approx \frac{30.28868}{0.01015005} \approx 2984$$

$$N\text{-число} = \frac{1}{4} \sqrt{933} \sqrt{9708} \approx 752$$

$$M\text{-число} = \frac{1}{4} 40 \times 272 = 2720.$$

¹⁾ Тьюриг [1], Д. К. Фаддеев [4].

Числа обусловленности матрицы W еще не очень велики. В практических расчетах иногда приходится встречаться с системами, матрицы которых имеют числа обусловленности свыше 20 000.

На вопрос о том, какой характер неопределенности влечет за собой плохая обусловленность системы, в известной степени отвечают следующие рассуждения.

Пусть A невырожденная матрица, которую для простоты мы будем считать вещественной, пусть U_1, U_2, \dots, U_n ортонормальная система собственных векторов для матрицы $A'A$ и пусть $\mu_1, \mu_2, \dots, \mu_n$ соответствующие им собственные значения. Тогда векторы $V_i = \frac{1}{\sqrt{\mu_i}} AU_i$ образуют ортонормальную систему собственных векторов для матрицы AA' . Действительно,

$$AA'V_i = \frac{1}{\sqrt{\mu_i}} AA'AU_i = \sqrt{\mu_i} AU_i = \mu_i V_i;$$

далее

$$\begin{aligned} (V_i, V_j) &= \frac{1}{\sqrt{\mu_i \mu_j}} (AU_i, AU_j) = \\ &= \frac{1}{\sqrt{\mu_i \mu_j}} (A'AU_i, U_j) = \frac{\mu_i}{\sqrt{\mu_i \mu_j}} (U_i, U_j) = \delta_{ij}. \end{aligned}$$

Пусть теперь дана система

$$AX = F$$

с матрицей A .

Разложим вектор F по векторам V_1, V_2, \dots, V_n

$$F = \sum_{i=1}^n c_i V_i.$$

Ясно, что $c_i = (F, V_i)$.

Далее ищем решение в виде

$$X = \sum_{i=1}^n d_i U_i.$$

Подстановка в систему дает

$$\sum_{i=1}^n d_i \sqrt{\mu_i} V_i = \sum_{i=1}^n c_i V_i,$$

откуда следует, что

$$d_i = \frac{c_i}{\sqrt{\mu_i}}. \quad (1)$$

Можно убедиться в том, что при малых изменениях матрицы A как собственные значения μ_i матрицы $A'A$, так и системы векторов $\{U_i\}$ и $\{V_i\}$ изменяются мало, независимо от обусловленности матрицы A . Однако при плохой обусловленности матрицы A среди μ_i

имеются очень малые числа, и даже малые их изменения могут оказаться относительно большими. Коэффициенты d_i формулы (1), отвечающие малым μ_i , окажутся плохо определенными, так что решение „разбалтывается“ в направлениях векторов U_i , соответствующих малым μ_i , что мы уже видели раньше на основе вероятностных расуждений.

Если коэффициенты системы известны точно (что бывает, например, при решении методом сеток краевых задач для уравнений эллиптического типа с постоянными коэффициентами), то источником неопределенности может служить только неточность в коэффициентах c_i , соответствующих малым μ_i , которая, согласно формуле (1), увеличивается во много раз при переходе к коэффициентам d_i . Однако и это явление может не иметь места, если по условию задачи вектор свободных членов ортогонален или почти ортогонален к векторам V_i , соответствующим малым μ_i , и потому решение остается устойчивым при небольших допустимых (без нарушения упомянутой почти-ортогональности) изменениях свободных членов.

§ 16. Метод Гаусса

В этом параграфе мы изложим наиболее простой и естественный метод для решения системы уравнений, основанный на последовательном исключении неизвестных. Как уже говорилось выше, он связан с именем Гаусса. Метод имеет много различных вычислительных схем. Мы начнем изложение с рассмотрения так называемой схемы единственного деления.

Итак, пусть дана система уравнений

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= f_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= f_2 \\ \vdots &\quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= f_n. \end{aligned} \tag{1}$$

Допустим, что коэффициент $a_{11} \neq 0$. Разделим коэффициенты первого из уравнений (1) (включая и свободный член) на коэффициент a_{11} , который мы будем называть ведущим элементом (1-го шага), и обозначим

$$\begin{aligned} b_{1j} &= \frac{a_{1j}}{a_{11}} \quad (j > 1) \\ g_1 &= \frac{f_1}{a_{11}}. \end{aligned} \tag{2}$$

Мы получим новое уравнение

$$x_1 + b_{12}x_2 + \dots + b_{1n}x_n = g_1. \tag{3}$$

Исключим теперь неизвестное x_1 из всех уравнений (1), начиная со второго, посредством вычитания из этих уравнений уравнения (3),

умноженного соответственно на числа $a_{21}, a_{31}, \dots, a_{n1}$. Преобразованные уравнения будут

$$\begin{aligned} a_{22 \cdot 1}x_2 + \dots + a_{2n \cdot 1}x_n &= f_{2 \cdot 1} \\ \dots &\dots \dots \dots \dots \dots \dots \\ a_{n2 \cdot 1}x_2 + \dots + a_{nn \cdot 1}x_n &= f_{n \cdot 1}, \end{aligned} \quad (4)$$

где

$$\begin{aligned} a_{ij \cdot 1} &= a_{ij} - a_{i1}b_{1j} \quad (i, j \geq 2) \\ f_{i \cdot 1} &= f_i - a_{i1}g_1. \end{aligned} \quad (5)$$

Разделим далее коэффициенты первого из преобразованных уравнений на ведущий элемент 2-го шага $a_{22 \cdot 1}$, который будем считать отличным от нуля. Мы получим уравнение

$$x_2 + b_{23}x_3 + \dots + b_{2n}x_n = g_2,$$

где

$$b_{2j} = \frac{a_{2j \cdot 1}}{a_{22 \cdot 1}}, \quad g_2 = \frac{f_{2 \cdot 1}}{a_{22 \cdot 1}}.$$

Исключая неизвестное x_2 из уравнений (4), начиная со второго, мы придем к уравнениям

$$\begin{aligned} a_{33 \cdot 2}x_3 + \dots + a_{3n \cdot 2}x_n &= f_{3 \cdot 2} \\ \dots &\dots \dots \dots \dots \dots \dots \\ a_{n3 \cdot 2}x_3 + \dots + a_{nn \cdot 2}x_n &= f_{n \cdot 2}, \end{aligned}$$

где

$$\begin{aligned} a_{ij \cdot 2} &= a_{ij \cdot 1} - a_{i2 \cdot 1}b_{2j} \quad (i, j \geq 3). \\ f_{i \cdot 2} &= f_{i \cdot 1} - a_{i2 \cdot 1}g_2. \end{aligned}$$

Продолжаем процесс по той же схеме. На m -м шагу мы получим уравнения

$$\begin{aligned} x_m + b_{mm+1}x_{m+1} + \dots + b_{mn}x_n &= g_m \\ a_{m+1 \cdot m+1}x_{m+1} + \dots + a_{m+1 \cdot n}x_n &= f_{m+1 \cdot m} \\ \dots &\dots \dots \dots \dots \dots \dots \\ a_{nm+1 \cdot m}x_{m+1} + \dots + a_{nn \cdot m}x_n &= f_{n \cdot m}, \end{aligned} \quad (6)$$

где

$$\begin{aligned} b_{mj} &= \frac{a_{mj \cdot m-1}}{a_{mm \cdot m-1}} \quad (j \geq m+1), \\ g_m &= \frac{f_{m \cdot m-1}}{a_{mm \cdot m-1}} \end{aligned} \quad (7)$$

$$\begin{aligned} a_{ij \cdot m} &= a_{ij \cdot m-1} - a_{im \cdot m-1}b_{mj} \\ f_{i \cdot m} &= f_{i \cdot m-1} - a_{im \cdot m-1}g_m. \end{aligned}$$

Объединив все первые уравнения каждого шага, мы получим систему

$$\begin{aligned}
 x_1 + b_{12}x_2 + b_{13}x_3 + \dots + b_{1n}x_n &= g_1 \\
 x_2 + b_{23}x_3 + \dots + b_{2n}x_n &= g_2 \\
 \vdots &\quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\
 x_n &= g_n
 \end{aligned} \tag{8}$$

с треугольной матрицей, эквивалентную исходной системе. Отметим, что описанный процесс возможен только при условии неравенства нулю всех ведущих элементов a_{mm-m-1} , так как по ходу процесса на них приходится производить деление.

Из системы (8) значения для неизвестных находим последовательно от x_n к x_1 по очевидным формулам

$$x_m = g_m - b_{m+1}x_{m+1} - \dots - b_{mn}x_n.$$

Итак, для решения данной системы по схеме единственного деления мы сначала строим вспомогательную треугольную систему, а затем решаем ее. Процесс нахождения коэффициентов треугольной системы мы будем называть прямым ходом, а процесс получения ее решения обратным.

В табл. II.1 приведена схема единственного деления для системы четырех уравнений и численный пример.

Таблица II.1

Схема единственного деления

a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	1	0.17	-0.25	0.54	0.3	1.76
a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	a_{26}	0.47	1	0.67	-0.32	0.5	2.32
a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	a_{36}	-0.11	0.35	1	-0.74	0.7	1.20
a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	a_{46}	0.55	0.43	0.36	1	0.9	3.24
1	b_{12}	b_{13}	b_{14}	b_{15}	b_{16}	1	0.17	-0.25	0.54	0.3	1.76
	a_{22-1}	a_{23-1}	a_{24-1}	a_{25-1}	a_{26-1}		0.9201	0.7875	-0.5738	0.3590	1.4928
	a_{32-1}	a_{33-1}	a_{34-1}	a_{35-1}	a_{36-1}		0.3687	0.9725	-0.6806	0.7330	1.3936
	a_{42-1}	a_{43-1}	a_{44-1}	a_{45-1}	a_{46-1}		0.3365	0.4975	0.7030	0.7350	2.2720
	1	b_{23}	b_{24}	b_{25}	b_{26}		1	0.85589	-0.62363	0.39017	1.62243
		a_{33-2}	a_{34-2}	a_{35-2}	a_{36-2}			0.65693	-0.45067	0.58914	0.79540
		a_{43-2}	a_{44-2}	a_{45-2}	a_{46-2}			0.20949	0.91285	0.60371	1.72605
		1	b_{34}	b_{35}	b_{36}			1	-0.68602	0.89681	1.21079
			a_{44-3}	a_{45-3}	a_{46-3}				1.05656	0.41584	1.47240
			1	x_4	\bar{x}_4				1	0.39358	1.39358
			1	x_3	\bar{x}_3				1	1.16681	2.16682
			1	x_2	\bar{x}_2					-0.36304	0.63695
1				x_1	\bar{x}_1	1				0.44089	1.44089

Поясним табл. II.1. В схеме мы обозначили свободные члены исходных и преобразованных уравнений теми же буквами, что и коэффициенты уравнений, но со вторым индексом б. Это позволяет объединить формулы, по которым преобразуются коэффициенты и свободные члены данной системы. Шестой столбец служит для контроля. Этот контроль основан на следующем обстоятельстве. Если в данной системе сделать замену $\bar{x}_i = x_i + 1$, то для определения \bar{x}_i мы получим систему с прежними коэффициентами и свободными членами, равными суммам элементов строк матрицы коэффициентов (включая свободные члены). Поэтому, если составить сумму элементов каждой строки исходной матрицы (включая свободные члены) и провести над ними все те операции, что и над остальными элементами, то, при отсутствии вычислительных ошибок, должны получаться числа, равные суммам элементов вновь построенных строк. В конце процесса, после окончания обратного хода, должны получиться числа \bar{x}_i , равные $x_i + 1$. Иногда более целесообразным является использование этой идеи для контроля каждой построенной строки. С этой целью

Таблица II.1a

Схема единственного деления в симметричном случае

a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	1.00	0.42	0.54	0.66	0.3	2.92
a_{22}	a_{23}	a_{24}	a_{25}	a_{26}		1.00	0.32	0.44	0.5	0.6	2.68
a_{33}		a_{34}	a_{35}	a_{36}			1.00	0.22	0.7	2.78	
		a_{44}	a_{45}	a_{46}				1.00	0.9	3.22	
1	b_{12}	b_{13}	b_{14}	b_{15}	b_{16}	1	0.42	0.54	0.66	0.3	2.92
	$a_{22} \cdot 1$	$a_{23} \cdot 1$	$a_{24} \cdot 1$	$a_{25} \cdot 1$	$a_{26} \cdot 1$		0.82360	0.09320	0.16280	0.37400	1.45360
		$a_{33} \cdot 1$	$a_{34} \cdot 1$	$a_{35} \cdot 1$	$a_{36} \cdot 1$			0.70840	-0.13640	0.53800	1.20320
			$a_{44} \cdot 1$	$a_{45} \cdot 1$	$a_{46} \cdot 1$				0.56440	0.70200	1.29280
	1	b_{23}	b_{24}	b_{25}	b_{26}		1	0.11316	0.19767	0.45410	1.76493
		$a_{33} \cdot 2$	$a_{34} \cdot 2$	$a_{35} \cdot 2$	$a_{36} \cdot 2$			0.69785	-0.15482	0.49568	1.03871
			$a_{44} \cdot 2$	$a_{45} \cdot 2$	$a_{46} \cdot 2$				0.53222	0.62807	1.00347
		1	b_{34}	b_{35}	b_{36}			1	-0.22185	0.71030	1.48844
			$a_{41} \cdot 3$	$a_{45} \cdot 3$	$a_{46} \cdot 3$				0.49787	0.73804	1.23591
			1	x_4	\bar{x}_4				1	1.48240	2.48240
				x_3	\bar{x}_3					1.03917	2.03916
				x_2	\bar{x}_2					0.04348	1.04348
				x_1	\bar{x}_1	1				-1.25780	-0.25779

в контрольный столбец записываются суммы элементов строящихся строк и полученные числа лишь сравниваются с результатами контрольных действий.

Число умножений и делений, нужных для нахождения решения системы n уравнений по схеме единственного деления, равно $\frac{n}{3}(n^2 + 3n - 1)$.

В случае, если матрица коэффициентов системы симметрична, т. е. $a_{ij} = a_{ji}$, мы имеем, очевидно, $a_{ij \cdot k} = a_{ji \cdot k}$. Поэтому можно опустить запись элементов, расположенных ниже диагонали. Схема единственного деления, приспособленная для симметричного случая, показана в табл. II.1a.

Первый элемент строки, опущенный при записи коэффициентов (и нужный нам для вычислений элементов вспомогательных матриц), мы легко находим как верхний элемент столбца, включающего диагональный элемент данной строки. Контрольный столбец по-прежнему содержит суммы всех элементов каждой строки, включая и опущенные при записи.

Число вычислительных операций в схеме единственного деления значительно сокращается, если матрица коэффициентов системы почти треугольна, в частности, если она трехдиагональна.

В случае, если нужно решить несколько систем с данной матрицей, естественно искать решения одновременно, выписывая свободные члены в соседних столбцах. Контрольная сумма образуется как сумма элементов строк расширенной матрицы. Схема решения нескольких уравнений дана в табл. II.2.

Схема единственного деления очень проста и удобна. Однако она не является универсальной, в том смысле, что для ее применимости нужно, чтобы все ведущие элементы были отличны от нуля. Это обстоятельство, однако, не может быть предсказано без вычислений, которые в той или иной форме эквивалентны самому применению схемы. Близость ведущих элементов к нулю может быть причиной значительной потери точности.

Поэтому схему единственного деления целесообразно несколько видоизменить, не предписывая а priori порядка исключаемых неизвестных.

Наилучшим вариантом является схема единственного деления по главным элементам. В этой схеме в качестве исключаемой на m -м шагу неизвестной выбирается та, коэффициент при которой на предыдущем шагу был наибольшим по модулю. При вычислении по схеме главных элементов исчезновение значащих цифр может происходить, только если система плохо обусловлена, так что происходящая при этом потеря точности неизбежна по существу дела.

Отметим, что в разобранном примере главные элементы совпадают с коэффициентами $a_{mm \cdot m-1}$, так что обе схемы совпадают.

Таблица II. 2

Схема единственного деления. Несколько систем уравнений

1.00	0.42	0.54	0.66	0.25	0.3	0.15	3.32
0.42	1.00	0.32	0.44	0.45	0.5	0.30	3.43
0.54	0.32	1.00	0.22	0.65	0.7	0.45	3.88
0.66	0.44	0.22	1.00	0.85	0.9	0.60	4.67
<hr/>							
1	0.42	0.54	0.66	0.25	0.3	0.15	3.32
	0.82360	0.09320	0.16280	0.34500	0.37400	0.23700	2.03560
	0.09320	0.70840	-0.13640	0.51500	0.53800	0.36900	2.08720
	0.16280	-0.13640	0.56440	0.68500	0.70200	0.50100	2.47880
<hr/>							
	1	0.11316	0.19767	0.41889	0.45410	0.28776	2.47159
		0.69785	-0.15482	0.47596	0.49568	0.34218	1.85685
		-0.15482	0.53222	0.61680	0.62807	0.45415	2.07643
<hr/>							
		1	-0.22185	0.68201	0.71030	0.49033	2.66082
			0.49787	0.72239	0.73804	0.51006	2.48838
<hr/>							
			1	1.45096	1.48249	1.06466	4.99805
				1.00394	1.03917	0.72652	3.76964
				0.01847	0.04348	-0.00490	1.05705
1	1			-1.25752	-1.25780	-0.94294	-2.45828

Последовательные исключения неизвестных, преобразующие данную систему в систему с треугольной матрицей, можно проводить и по другим вычислительным схемам.

В схеме деления и вычитания на каждом шагу делятся все уравнения на коэффициент при исключаемой неизвестной и затем само исключение производится вычитанием одного уравнения из всех остальных.

В схеме умножения и вычитания на первом шагу неизвестное x_1 исключается из i -го уравнения посредством умножения этого уравнения на a_{11} и вычитанием 1-го уравнения, умноженного на a_{11} . На последующих шагах применяется тот же прием, так что коэффициенты вспомогательных уравнений $\tilde{a}_{ij \cdot m}$ на m -м шагу вычисляются по формулам

$$\tilde{a}_{ij \cdot m} = \tilde{a}_{mm \cdot m-1} \tilde{a}_{ij \cdot m-1} - \tilde{a}_{im \cdot m-1} \tilde{a}_{mj \cdot m-1}$$

$$\tilde{f}_{im} = \tilde{a}_{mm \cdot m-1} f_{i \cdot m-1} - \tilde{a}_{im \cdot m-1} \tilde{f}_{m \cdot m-1}$$

Существуют и другие вычислительные схемы. В частности, каждую из описанных схем можно применять как с заранее предписаным порядком исключения неизвестных, так и выбирая порядок исключения по ходу процесса, например, по главным элементам.

Вычислительные схемы метода Гаусса основаны на приведении системы к системе с правой треугольной матрицей посредством линейного комбинирования уравнений, что равносильно умножению слева матрицы системы (и одновременно столбца свободных членов) на некоторые вспомогательные матрицы. Заслуживает внимания модификация метода исключения, при котором в качестве вспомогательных матриц, управляющих линейным комбинированием уравнений, берутся элементарные матрицы вращений. При этом вычисления располагаются следующим образом. Берем

$$c_{21} = \frac{a_{11}}{\sqrt{a_{11}^2 + a_{21}^2}}; \quad s_{21} = -\frac{a_{21}}{\sqrt{a_{11}^2 + a_{21}^2}}$$

(если $a_{11} = a_{21} = 0$, берем $c_{21} = 1$, $s_{21} = 0$) и первые два уравнения системы заменяем уравнениями

$$\begin{aligned} c_{21}y_1 - s_{21}y_2 &= c_{21}f_1 - s_{21}f_2 \\ s_{21}y_1 + c_{21}y_2 &= s_{21}f_1 + c_{21}f_2. \end{aligned}$$

Здесь через y_1 и y_2 обозначены левые части первых двух уравнений исходной системы. После преобразования во втором уравнении коэффициент при неизвестном x_1 будет равен нулю.

Далее аналогичным образом обрабатывается преобразованное первое уравнение с третьим исходным уравнением, полученное новое первое с исходным четвертым и т. д. После $n-1$ -го шага процесса мы придем к системе вида

$$\begin{aligned} a_{11}^{(1)}x_1 + a_{12}^{(1)}x_2 + \dots + a_{1n}^{(1)}x_n &= f_1^{(1)} \\ a_{22}^{(1)}x_2 + \dots + a_{2n}^{(1)}x_n &= f_2^{(1)} \\ \vdots &\quad \vdots \\ a_{n2}^{(1)}x_2 + \dots + a_{nn}^{(1)}x_n &= f_n^{(1)}. \end{aligned}$$

Теперь тот же процесс применим к системе с отброшенным первым уравнением. После $\frac{n(n-1)}{2}$ шагов мы придем к системе с треугольной матрицей, которая решается обычным обратным ходом.

Эта схема исключения требует приблизительно в 4 раза больше вычислительных операций, чем схема единственного деления, но отличается большой стабильностью и мало чувствительна к „провалам“ промежуточных определителей системы.

В табл. II. 3 дается решение системы, ранее решенной в табл. II. 1 методом единственного деления. В последнем столбце таблицы помещены числа c_{ij} и s_{ij} , в ненумерованных строках — вспомогательные величины, нужные для их вычисления.

Результативная система с треугольной матрицей располагается в строках 1'', 2'', 3'', 4'', исходная — в строках 1, 2, 3, 4. Строки 1', 2', 3', 4' и 1'', 2'', 3'', 4'' заняты промежуточными линейными комбинациями исходных уравнений. Их удобно получать в порядке, совпадающем с порядком следования строк в таблице. В последней строке таблицы расположено решение системы. Шестой столбец во всех нумерованных строках контрольный.

Таблица II. 3

Решение линейной системы методом вращений

1	1.00	0.17	-0.25	0.54	0.3	1.76	0.90503
2	0.47	1.00	0.67	-0.32	0.5	2.32	-0.42536
	1.2209	1.10494					
1'	1.10495	0.57922	0.05873	0.35260	0.48419	2.57969	0.99508
3	-0.11	0.35	1.00	-0.74	0.7	1.20	0.09906
	1.23301	1.11041					
1''	1.11041	0.54176	-0.04062	0.42417	0.41247	2.44813	0.89610
4	0.55	0.43	0.36	1.00	0.9	3.24	-0.44385
	1.53551	1.23916					
1'''	1.23916	0.67627	0.12339	0.82395	0.76908	3.63184	
2'		0.83272	0.71271	-0.51930	0.32491	1.35104	0.89900
3'		0.40566	1.00090	-0.70143	0.74452	1.44964	-0.43795
		0.85798	0.92627				
2''		0.92627	1.07907	-0.77404	0.61816	1.84945	0.98799
4'		0.14489	0.34063	0.70783	0.62342	1.81676	-0.15455
		0.87897	0.93753				
2'''		0.93754	1.11875	-0.65536	0.70708	2.10890	
3''			0.58768	-0.40316	0.52703	0.71154	0.96072
4''			0.16978	0.81895	0.52040	1.50913	-0.27755
			0.37419	0.61171			
3'''			0.61172	-0.16002	0.65077	1.10245	
4'''				0.89868	0.35368	1.25236	
	0.44089	-0.36302	1.16679	0.39356			

Вместо элементарных матриц вращения для приведения системы к равносильной системе с треугольной матрицей можно использовать и другие ортогональные матрицы. Удобными оказываются матрицы отражений от $n - 1$ -мерной плоскости¹⁾. Пусть W вектор единичной длины, ортогональный к заданной $n - 1$ -мерной плоскости Q . Тогда матрица отражения относительно Q есть $U = E - 2WW'$ (здесь WW' есть произведение столбца на строку, т. е. матрица ранга 1). Действительно, если вектор Z ортогонален к W , то

$$UZ = Z - 2WW'Z = Z,$$

ибо $W'Z = (W, Z) = 0$, а $UW = W - 2WW'W = -W$, ибо $W'W = -(W, W) = 1$.

Вектор W можно всегда подобрать так, чтобы матрица U переводила любой заданный вектор S в вектор заданного направления, например в вектор, направленный по одному из координатных векторов e . Для этого достаточно взять

$$W = \frac{1}{\rho} (S - \alpha e),$$

где $\alpha = \pm |S|$, $\rho = |S - \alpha e| = \sqrt{2\alpha^2 - 2\alpha(S, e)}$.

Действительно, при таком выборе W имеем

$$US = S - 2(W, S)W = S - \rho W = S - S + \alpha e = \alpha e,$$

так как $2(W, S) = \frac{1}{\rho} (2\alpha^2 - 2\alpha(S, e)) = \rho$.

Выбор знака для α , вообще говоря, безразличен. Им можно распорядиться так, чтобы число ρ было большим из двух возможных, для чего следует взять $\text{sign } \alpha = -\text{sign}(S, e)$.

Для любого вектора Y имеем

$$UY = Y - 2(W, Y)W. \quad (9)$$

Для преобразования системы $AX = F$ к треугольной системе поступаем так. На первом шагу вектор W строится по первому столбцу $S = A_1$ матрицы A и по вектору $e = e_1$. После умножения обеих частей системы на матрицу U (что равносильно применению формулы (9) к столбцам матрицы A и к вектору F) приходим к системе с матрицей вида $\begin{bmatrix} \alpha & v \\ 0 & B \end{bmatrix}$, где B квадратная матрица $n - 1$ -го порядка. Далее процесс повторяется для системы с матрицей B и т. д.

После $n - 1$ -го шага мы придем к треугольной системе. Легко видеть, что матрица этой системы, с точностью до знаков строк, совпадает с матрицей, полученной методом вращений.

¹⁾ Хаусхольдер [13].

Приведем результат первого шага для системы табл. II. 1. Имеем

$$\alpha = -1.23915 \quad \rho = 2.35569$$

$$W = (0.95053, 0.19952, -0.04670, 0.23348)'$$

$$\begin{bmatrix} \alpha & v \\ 0 & B \end{bmatrix} = \begin{bmatrix} -1.23917 & -0.67628 & -0.12339 & -0.82397 \\ -0.00001 & 0.82236 & 0.69658 & -0.60630 \\ 0.00001 & 0.39158 & 0.99378 & -0.67299 \\ -0.00001 & 0.22213 & 0.39110 & 0.66497 \end{bmatrix}$$

$$UF = (-0.76908, 0.27560, 0.75252, 0.63740)'.$$

Метод единственного деления может быть следующим образом обобщен на системы, матрицы которых разбиты на клетки с квадратными диагональными клетками. Разбив соответственно решение и столбец свободных членов на векторы, размерность которых равна порядкам диагональных клеток, запишем систему в виде

$$A_{11}X_1 + A_{12}X_2 + \dots + A_{1n}X_n = F_1$$

$$A_{21}X_1 + A_{22}X_2 + \dots + A_{2n}X_n = F_2$$

$$A_{n1}X_1 + A_{n2}X_2 + \dots + A_{nn}X_n = F_n.$$

Пусть $|A_{11}| \neq 0$. Найдем матрицу A_{11}^{-1} и умножим на нее слева первое уравнение системы. Получим матричное уравнение

$$X_1 + B_{12}X_2 + \dots + B_{1n}X_n = G_1,$$

где

$$B_{12} = A_{11}^{-1} A_{12}, \dots, B_{1n} = A_{11}^{-1} A_{1n}, G_1 = A_{11}^{-1} F_1.$$

Умножим это уравнение слева на A_{21}, \dots, A_{n1} и вычтем, соответственно, из 2-го, ..., n -го уравнений исходной системы. Получим систему

$$A_{22 \cdot 1}X_2 + \dots + A_{2n \cdot 1}X_n = F_{2 \cdot 1}$$

$$A_{n2+1}X_2 + \cdots + A_{nn+1}X_n = F_{n+1}$$

где

$$A_{ij-1} = A_{ij} - A_{ii}B_{ij} \quad (i, j \geq 2)$$

Поступая аналогично дальше, придем в конце концов к системе с квазитреугольной матрицей с единичными диагональными клетками

$$X_1 + B_{12}X_2 + \dots + B_{1n}X_n = G_1$$

$$A_1 X_1 + \dots + A_{2n} X_n = G_1$$

$$X_n \equiv G_{n+1}$$

Неизвестные векторы X_1, \dots, X_n находятся последовательно от X_n к X_1 по формулам

$$X_i = G_i - B_{ii+1}X_{i+1} - \dots - B_{in}X_n.$$

Разбиение матрицы на клетки, вообще говоря, не изменяет числа элементарных операций, нужных для решения системы. Однако можно получить значительный выигрыш в объеме работы, если существует разбиение, при котором возникают упрощения при обращении ведущих клеток.

§ 17. Вычисление определителей

Метод Гаусса, развитый в предыдущем параграфе для решения линейной системы, может быть применен и для вычисления определителей. Мы остановимся отдельно на описании соответствующей схемы единственного деления, так как вычисление определителей часто встречается в приложениях. Пусть

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

и пусть $a_{11} \neq 0$. Вынесем элемент a_{11} из первой строки. Тогда, применяя обозначения § 16, получим:

$$\Delta = a_{11} \begin{vmatrix} 1 & b_{12} & \dots & b_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Затем из каждой строки отнимем первую строку, умноженную соответственно на первый элемент этой строки. Мы получим, очевидно,

$$\Delta = a_{11} \begin{vmatrix} 1 & b_{12} & \dots & b_{1n} \\ 0 & a_{22 \cdot 1} & \dots & a_{2n \cdot 1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2 \cdot 1} & \dots & a_{nn \cdot 1} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22 \cdot 1} & \dots & a_{2n \cdot 1} \\ \vdots & \ddots & \vdots \\ a_{n2 \cdot 1} & \dots & a_{nn \cdot 1} \end{vmatrix}.$$

Далее, с оставшимся определителем $n - 1$ -го порядка поступаем совершенно таким же образом, если только $a_{22 \cdot 1} \neq 0$.

Продолжая процесс, мы получим, что искомый определитель равен произведению ведущих элементов:

$$\Delta = a_{11}a_{22 \cdot 1} \dots a_{nn \cdot n-1}.$$

Если на каком-либо шагу окажется, что $a_{ii \cdot i-1} = 0$ или $a_{ii \cdot i-1}$ близко к нулю (что влечет за собой уменьшение точности вычислений), можно предварительно изменить порядок строк и столбцов определителя так, чтобы в левом верхнем углу оказался неисчезающий элемент.

Наилучший в смысле надежности результат получится, если на каждом шагу процесса переводить в левый верхний угол наибольший по абсолютной величине элемент матрицы, для которой вычисляется определитель.

Число умножений и делений, нужных для вычисления определителя n -го порядка, равно $\frac{n-1}{3}(n^2 + n + 3)$.

В табл. II. 4 дана схема вычисления определителя и численный пример.

Таблица II. 4.

**Вычисление определителя по схеме единственного деления
(с исключением по строкам)**

a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	1.00	0.17	-0.25	0.54	1.46
a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	0.47	1.00	0.67	-0.32	3.82
a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	-0.11	0.35	1.00	-0.74	0.50
a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	0.55	0.43	0.36	1.00	2.34
1	b_{12}	b_{13}	b_{14}	b_{15}	1	0.17	-0.25	0.54	1.46
$a_{22} \cdot 1$	$a_{23} \cdot 1$	$a_{24} \cdot 1$	$a_{25} \cdot 1$		0.9201	0.7875	-0.5738	1.1338	
$a_{32} \cdot 1$	$a_{33} \cdot 1$	$a_{34} \cdot 1$	$a_{35} \cdot 1$		0.3687	0.9725	-0.6806	0.6606	
$a_{42} \cdot 1$	$a_{43} \cdot 1$	$a_{44} \cdot 1$	$a_{45} \cdot 1$		0.3365	0.4975	0.7030	1.5370	
	1	b_{23}	b_{24}	b_{25}		1	0.85589	-0.62363	1.23226
		$a_{33} \cdot 2$	$a_{34} \cdot 2$	$a_{35} \cdot 2$			0.65693	-0.45067	0.20627
		$a_{43} \cdot 2$	$a_{44} \cdot 2$	$a_{45} \cdot 2$			0.20949	0.91285	1.12234
		1	b_{34}	b_{35}			1	-0.68602	0.31399
			$a_{44} \cdot 3$	$a_{45} \cdot 3$				1.05656	1.05656
				$\Delta =$	1.00 × 0.9201 ×	0.65693 ×	1.05656 =	0.63863	

Схема единственного деления для вычисления определителей может применяться не только с исключением по строкам, но и по столбцам (табл. II.5). Это, очевидно, равносильно вычислению определителя транспонированной матрицы по строкам.

Из рассмотрения процесса вычисления определителя мы видим, что он, за исключением последнего умножения, совпадает с прямым ходом процесса Гаусса, примененным для системы с матрицей, для которой вычисляется определитель.

Известные формулы Крамера (§ 1) показывают, что решение линейной системы можно находить в виде $x_i = \frac{\Delta_i}{\Delta}$; $i = 1, \dots, n$, где Δ обозначает определитель из коэффициентов системы, Δ_i — определитель из коэффициентов, в котором элементы i -го столбца заменены свободными членами. Таким образом, для решения системы приходится вычислять $n+1$ определитель n -го порядка,

Таблица II. 5

**Вычисление определителя по схеме единственного деления
(с исключением по столбцам)**

1.00	0.17	- 0.25	0.54	1.00
0.47	1.00	0.67	- 0.32	0.47
- 0.11	0.35	1.00	- 0.74	- 0.11
0.55	0.43	0.36	1.00	0.55
1.91	1.95	1.78	0.48	1.91
	0.9201	0.7875	- 0.5738	1
	0.3687	0.9725	- 0.6806	0.40072
	0.3365	0.4975	0.7030	0.36572
	1.6253	2.2575	- 0.5514	1.76644
		0.65693	- 0.45067	1
		0.20949	0.91285	0.31889
		0.86643	0.46218	1.31891
			1.05653	
			1.05656	
1	×	0.9201	×	0.65693 × 1.05656 = 0.63863

Сравнивая процесс Гаусса для решения системы с процессом вычисления определителя, мы видим, что объем вычислений для решения системы лишь немногим превышает вычисление одного определителя. Поэтому пользоваться формулами Крамера для численного решения системы нецелесообразно. По сути дела, в способе Гаусса производятся вычисления всех определителей Δ и Δ_i одновременно, причем деление на $\Delta = a_{11}a_{22}\dots a_{nn} \cdot n - 1$ производится постепенно, по одному множителю на каждом шаге.

Получение нулей при вычислении определителя можно осуществить также посредством линейного комбинирования строк матрицами вращений. Это требует значительно большего числа операций, чем

схема единственного деления, но дает значительно более стабильный процесс вычислений.

Для вычисления определителя можно так же разбить его матрицу на клетки с квадратными диагональными клетками и затем „получить нули“ подобно тому, как это делается в соответственно видоизмененной схеме единственного деления для решения систем линейных уравнений.

В конце концов искомый определитель окажется произведением определителей „ведущих“ клеток

$$\Delta = |A_{11}| \cdot |A_{22,1}| \cdots |A_{nn,n-1}|.$$

§ 18. Компактные схемы для решения неоднородной линейной системы

В § 16 мы видели, что решение системы линейных уравнений по схеме единственного деления свелось к определению коэффициентов $a_{ij,k}$ преобразованных уравнений (включая свободные члены) и коэффициентов b_{ij} уравнений окончательной треугольной системы. При этом, для получения решения данной системы нам нужны только коэффициенты b_{ij} , а числа $a_{ij,k}$ играют вспомогательную роль и нужны лишь затем, чтобы определить числа b_{ij} .

Покажем¹⁾, что числа b_{ij} можно получать процессом накопления, позволяющим избежать вычисления и записи всех коэффициентов $a_{ij,k}$.

Выделим элементы 1-го столбца каждой вспомогательной матрицы $a_{ij,j-1}$, $i \geq j$, обозначив их через c_{ij} , $i \geq j$.

Анализируя процесс вычисления коэффициентов вспомогательных матриц, мы видим, что

$$\begin{aligned} a_{ij,k} &= a_{ij,k-1} - a_{ik,k-1} b_{kj} = a_{ij,k-1} - c_{ik} b_{kj} = \\ &= a_{ij,k-2} - c_{ik-1} b_{k-1,j} - c_{ik} b_{kj} = \dots = \\ &= a_{ij} - c_{ii} b_{1j} - c_{i2} b_{2j} - \dots - c_{ik} b_{kj} = a_{ij} - \sum_{l=1}^k c_{il} b_{lj}. \end{aligned} \quad (1)$$

Таким образом, любой элемент $a_{ij,k}$ выражается посредством накопления через выделенные нами элементы c_{ij} и числа b_{ij} .

В частности, для самих элементов c_{ij} , $i \geq j$, и b_{ij} , $i < j$, имеют место рекуррентные формулы

$$c_{ij} = a_{ij,j-1} = a_{ij} - \sum_{l=1}^{j-1} c_{il} b_{lj} \quad (i \geq j) \quad (2)$$

$$b_{ij} = \frac{a_{ij,i-1}}{a_{ii,i-1}} = \frac{a_{ij} - \sum_{l=1}^{i-1} c_{il} b_{lj}}{c_{ii}} \quad (i < j).$$

¹⁾ Дуайр [1].

По этим же формулам определяются, очевидно, и свободные члены преобразуемых уравнений. Схема проведения прямого хода по формулам (2) называется компактной. Обратный ход остается таким же, как и в развернутой схеме единственного деления.

Вычисления по компактной схеме удобно располагать так, как это указано в табл. II. 6.

Таблица II. 6

Компактная схема метода единственного деления

a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	1	0.17	-0.25	0.54	0.3	1.76
a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	a_{26}	0.47	1	0.67	-0.32	0.5	2.32
a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	a_{36}	-0.11	0.35	1	-0.74	0.7	1.20
a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	a_{46}	0.55	0.43	0.36	1	0.9	3.24
$c_{11} \mid 1$	b_{12}	b_{13}	b_{14}	b_{15}	b_{16}	1	0.17	-0.25	0.54	0.3	1.76
c_{21}	$c_{22} \mid 1$	b_{23}	b_{24}	b_{25}	b_{26}	0.47	0.9201	-0.85589	-0.62363	0.39017	1.62243
c_{31}	c_{32}	$c_{33} \mid 1$	b_{34}	b_{35}	b_{36}	-0.11	0.3687	0.65693	-0.68602	0.89681	1.21080
c_{41}	c_{42}	c_{43}	$c_{44} \mid 1$	b_{45}	b_{46}	0.55	0.3365	0.20949	1.05637	0.39357	1.39357
			1	x_4	\bar{x}_4				1	0.39357	1.39357
				x_3	\bar{x}_3					1.16681	2.16682
				x_2	\bar{x}_2		1			-0.36305	0.63694
1				x_1	\bar{x}_1	1				0.44069	1.44090

Здесь вычисление элементов c и b мы производим последовательно по углам, начиная с вычисления элементов c :

c_{11}	b_{12}	b_{13}	b_{14}	1-й шаг
c_{21}	c_{22}	b_{23}	b_{24}	2-й шаг
c_{31}	c_{32}	c_{33}	b_{34}	3-й шаг
c_{41}	c_{42}	c_{43}		

При этом любой элемент получается как разность между соответствующим элементом a и суммой попарных произведений недиагональных элементов c , находящихся в данной строке (слева), и элементов b , находящихся в данном столбце (выше). Конечно, при вычислении элементов b нужно еще произвести деление на соответствующий диагональный элемент c .

Так,

$$\begin{aligned} c_{43} &= a_{43} - c_{41}b_{13} - c_{42}b_{23} = \\ &= 0.36 + 0.55 \cdot 0.25 - 0.3365 \cdot 0.85589 = 0.20949 \\ b_{34} &= \frac{a_{34} - c_{31}b_{14} - c_{32}b_{24}}{c_{33}} = \\ &= \frac{-0.74 + 0.11 \cdot 0.54 + 0.3687 \cdot 0.62363}{0.65693} = -0.68602. \end{aligned}$$

Скажем несколько слов о контроле, применяющемся при вычислении по компактной схеме. Так же, как и раньше, составляем столбец из контрольных сумм и над ним проделываем те же операции, что и над столбцом свободных членов.

При этом каждое число преобразованного столбца должно совпадать с суммой элементов соответствующих строк матрицы B , расширенной посредством присоединения преобразованного столбца свободных членов. Действительно, матрица B есть матрица коэффициентов системы, получаемой из данной после окончания прямого хода по схеме единственного деления.

Вычисление по компактной схеме требует предварительной фиксации ведущих элементов, так что схеме главных элементов невозможно придать компактную форму.

Компактная схема формально почти без изменений может быть распространена на решение систем линейных уравнений, матрицы которых разбиты на клетки с квадратными диагональными клетками.

Обозначив через C_{ij} матрицу $A_{ij}, j \geq i$, мы, так же как в числовом случае, будем иметь

$$\begin{aligned} C_{ij} &= A_{ij} - \sum_{l=1}^{j-1} C_{il}B_{lj} \quad (i \geq j) \\ B_{ij} &= C_{ii}^{-1} \left(A_{ij} - \sum_{l=1}^{i-1} C_{il}B_{lj} \right) \quad (i < j). \end{aligned}$$

Эти формулы позволяют последовательно вычислять матрицы C_{ij} и B_{ij} в том же порядке, как и в числовом случае.

§ 19. Связь метода Гаусса с разложением матрицы на множители

В § 1 было показано, что систему n линейных уравнений с n неизвестными можно записать в матричной форме

$$AX = F, \quad (1)$$

где A данная неособенная матрица, X и F столбцы, состоящие из значений неизвестных и свободных членов, которые мы будем рассматривать как векторы арифметического пространства.

Метод Гаусса, проведенный с фиксированным порядком ведущих элементов, состоит в том, что данная система заменяется равносильной треугольной системой посредством линейного комбинирования уравнений, что сводится к линейному комбинированию строк A . При этом нам приходится (пользуясь схемой единственного деления) кроме делений на ведущие элементы добавлять к элементам строк числа, пропорциональные элементам предшествующих строк, т. е. совершать над матрицей A элементарные преобразования типа $a)$ и $b')$ пункта 11 § 1.

Результат нескольких преобразований этого вида, как было там показано, равносителен умножению матрицы A слева на левую треугольную матрицу

$$\Gamma = \begin{bmatrix} \gamma_{11} & 0 & \dots & 0 \\ \gamma_{21} & \gamma_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} \end{bmatrix}. \quad (2)$$

В результате этих преобразований мы переходим к системе с правой треугольной матрицей

$$B = \begin{bmatrix} 1 & b_{12} & \dots & b_{1n} \\ 0 & 1 & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}. \quad (3)$$

Таким образом, $\Gamma A = B$, т. е. $A = \Gamma^{-1}B$ и, следовательно, матрица A разлагается в произведение двух треугольных матриц.

Компактная схема осуществляет это разложение. Действительно, $\Gamma^{-1} = C$, где элементы матрицы C определены по формулам (2) § 8, ибо из этих формул следует, что

$$a_{ij} = c_{ij} + \sum_{l=1}^{j-1} c_{il}b_{lj} = \sum_{l=1}^j c_{il}b_{lj} \quad (i \geq j)$$

(из формул для c_{ij}) и

$$a_{ij} = \sum_{l=1}^{i-1} c_{il}b_{lj} + c_{ii}b_{ij} = \sum_{l=1}^i c_{il}b_{lj} \quad (i < j)$$

(из формул для b_{ij}).

Последние формулы означают, что

$$A = CB.$$

Так как диагональные элементы матрицы B равны единице, такое разложение единственно.

Компактная схема, примененная к клеточной матрице, очевидно, осуществляет ее разложение в произведение двух квазитреугольных матриц.

Компактная запись для схем метода Гаусса, отличных от схемы единственного деления, также связана с разложением матрицы в произведение двух треугольных, но с другим выбором диагональных элементов.

Заметим, что компактные схемы закрепляют порядок исключения, и потому они применимы только в том случае, как это следуем из теоремы 1.1, когда все определители a_{ii} ,

не равны нулю.

Покажем, что в случае, когда матрица A симметрична,

$$b_{ik} = \frac{c_{ki}}{c_{ii}}. \quad (4)$$

Действительно, $A = CB$, $A' = B'C'$ и, так как $A = A'$,

$$CB = B'C' = B' \left[\begin{array}{cccc|ccccc} c_{11} & 0 & \dots & 0 & 1 & \frac{c_{21}}{c_{11}} & \dots & \frac{c_{n1}}{c_{11}} \\ 0 & c_{22} & \dots & 0 & 0 & 1 & \dots & \frac{c_{n2}}{c_{22}} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & c_{nn} & 0 & 0 & \dots & 1 \end{array} \right].$$

Отсюда, в силу единственности разложения матрицы A в произведение двух треугольных матриц, следует, что

$$B = \left[\begin{array}{cccc|ccccc} 1 & \frac{c_{21}}{c_{11}} & \dots & \frac{c_{n1}}{c_{11}} \\ 0 & 1 & \dots & \frac{c_{n2}}{c_{22}} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{array} \right].$$

Таким образом, элементы матрицы B находятся посредством деления элементов матрицы C на диагональные элементы. Тем не менее компактная схема требует n^2 записей и в случае, если матрица A симметрична (табл. II. 7), так как в рекуррентные соотношения (2) § 18 входят как числа b_{ik} , так и числа c_{ki} . Можно дать рекуррентные соотношения, в которых элементы b_{ik} выражаются друг через друга и через диагональные элементы c_{ii} . Однако так построенные формулы оказываются более сложными.

Таблица II.7

**Компактная схема метода единственного деления
в симметричном случае**

a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	1.00	0.42	0.54	0.66	0.3	2.92	
a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	a_{26}	0.42	1.00	0.32	0.44	0.5	2.68	
a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	a_{36}	0.54	0.32	1.00	0.22	0.7	2.78	
a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	a_{46}	0.66	0.44	0.22	1.00	0.9	3.22	
c_{11}	$1/b_{12}$	b_{13}	b_{14}	b_{15}	b_{16}	1.00	1/1	0.42	0.54	0.66	0.3	2.92
c_{21}	c_{22}	$1/b_{23}$	b_{24}	b_{25}	b_{26}	0.42	0.82360	1/1	0.11316	0.19767	0.45410	1.76493
c_{31}	c_{32}	c_{33}	$1/b_{34}$	b_{35}	b_{36}	0.54	0.09320	0.69785	1/-0.22185	0.71030	1.48844	
c_{41}	c_{42}	c_{43}	c_{44}	$1/b_{45}$	b_{46}	0.66	0.16280	-0.15482	0.49787	1/1	1.48240	2.48239
			1	x_4	\bar{x}_4					1	1.48240	2.48239
			1	x_5	\bar{x}_5						1.03917	2.03916
1				x_6	\bar{x}_6						0.04348	1.04348
				x_7	\bar{x}_7	1					-1.25780	-0.25779

§ 20. Метод квадратных корней

В этом параграфе мы покажем, что в случае, когда матрица системы симметрична, нахождение решения можно еще упростить,¹⁾ так как в этом случае матрицу можно разложить в произведение двух транспонированных друг другу треугольных матриц.

Итак, пусть

$$A = S' S, \quad (1)$$

где

$$S = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ 0 & s_{22} & \dots & s_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & s_{nn} \end{bmatrix}. \quad (2)$$

Определим элементы матрицы S . Имеем, в силу правила умножения матриц:

$$a_{ij} = s_{1i}s_{1j} + s_{2i}s_{2j} + \dots + s_{ii}s_{ij} \quad i < j$$

$$a_{ii} = s_{1i}^2 + s_{2i}^2 + \dots + s_{ii}^2 \quad i = j.$$

1) Т. Банахевич [4], [5].

Отсюда получаем формулы для определения s_{ij} :

$$s_{ii} = \sqrt{a_{ii}}, \quad s_{ij} = \frac{a_{ij}}{s_{ii}},$$

$$s_{ii} = \sqrt{a_{ii} - \sum_{l=1}^{i-1} s_{il}^2} \quad (i > 1); \quad s_{ij} = \frac{a_{ij} - \sum_{l=1}^{i-1} s_{il} s_{lj}}{s_{ii}} \quad (j > i) \quad (3)$$

$$s_{ij} = 0 \quad (i > j).$$

Далее, решение системы сводится к решению двух треугольных систем. Действительно, равенство

$$AX = F$$

равносильно двум равенствам

$$S'K = F \text{ и } SX = K.$$

Элементы вектора K определяются по рекуррентным формулам, аналогичным формулам для s_{ij} . Именно,

$$k_1 = \frac{f_1}{s_{11}}, \quad k_i = \frac{f_i - \sum_{l=1}^{i-1} s_{il} k_l}{s_{ii}} \quad (i > 1). \quad (4)$$

Окончательное решение находится по формулам

$$x_n = \frac{k_n}{s_{nn}}, \quad x_i = \frac{k_i - \sum_{l=i+1}^n s_{il} x_l}{s_{ii}} \quad (i < n). \quad (5)$$

В схеме применяется обычный контроль, причем при составлении контрольных элементов мы учитываем все элементы матрицы. Аналогично компактной схеме контрольным равенством является

$$\bar{k}_i = \sum_{k=1}^n s_{ik} + k_i.$$

В методе квадратных корней приходится записывать только $\frac{n(n+1)}{2}$ элементов матрицы S и $2n$ компонент векторов K и X .

В табл. II. 8 приведено решение системы по методу квадратных корней.

Таблица II. 8

Метод квадратных корней

a_{11}	a_{12}	a_{13}	a_{14}	f_1	\bar{f}_1	1.00	0.42	0.54	0.66	0.3	2.92
a_{22}	a_{23}	a_{24}	f_2	\bar{f}_2		1.00	0.32	0.44	0.5	2.68	
a_{33}	a_{34}	f_3	\bar{f}_3			1.00	0.22	0.3	0.7	2.78	
	a_{44}	f_4	\bar{f}_4				1.00	0.9	3.22		
s_{11}	s_{12}	s_{13}	s_{14}	k_1	\bar{k}_1	1.00	0.42	0.54	0.66	0.3	2.92
	s_{21}	s_{23}	s_{24}	k_2	\bar{k}_2		0.90752	0.10270	0.17039	0.41211	1.60173
		s_{34}	s_{31}	k_3	\bar{k}_3			0.83537	-0.18533	0.50336	1.24340
			s_{41}	k_4	\bar{k}_4				0.70560	1.04397	1.75157
x_1	x_2	x_3	x_4			-1.25778	0.04349	1.03017	1.48238		
\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_4			-0.23779	1.04349	2.00817	2.48238		

Вычисление элементов s_{ij} (так же как и элементов k_i и \bar{k}_i) производится последовательно по строкам. Любой диагональный элемент вычисляется как корень квадратный из разности между соответствующим элементом a и суммой квадратов всех вычисленных элементов s , находящихся в том же столбце. Недиагональный элемент s_{ij} получается вычитанием из элемента a_{ij} суммы произведений соответствующих элементов s , взятых из столбцов с номерами i и j . Полученная разность делится на диагональный элемент строки.

Так,

$$s_{34} = \frac{a_{34} - s_{13}s_{14} - s_{23}s_{24}}{s_{33}} = \frac{0.22 - 0.54 \cdot 0.66 - 0.10270 \cdot 0.17039}{0.83537} = -0.18533.$$

Обратный ход происходит по формулам (б).

По существу прямой ход метода квадратных корней равносителен приведению квадратичной формы с матрицей A к сумме квадратов посредством преобразования переменных с треугольной матрицей.

Если A положительно-определенная матрица, то метод квадратных корней протекает без всяких осложнений. Если же A не положительно-определенная, то возможно вырождение процесса за счет того, что некоторый коэффициент s_{ii} может оказаться равным нулю или близким к нулю. Может оказаться также, что подкоренные выражения для некоторых s_{ii} окажутся отрицательными. Однако это не внесет существенных затруднений.

Действительно, в этом случае в строку, для которой $s_{ii}^2 < 0$, будут входить чисто мнимые числа, действия с которыми ничуть не сложнее, чем с вещественными числами.

Поясним сказанное на примере.

2	-1 -2	1 3 1	4 5 6	6 5 11
1.41421	-0.70711 1.58114 <i>t</i>	0.70711 -2.21359 <i>t</i> 2.32379	2.82843 -4.42720 <i>t</i> 5.93857	4.24265 -5.05965 <i>t</i> 8.26235
1.11111	0.77776	2.55555		
2.11111	1.77776	3.55555		

В настоящее время метод квадратных корней широко используется для решения симметричных систем и может быть рекомендован читателю, как один из наиболее эффективных методов.

§ 21. Обращение матрицы

Как уже говорилось во введении, задача решения линейной неоднородной системы и задача обращения матрицы тесно связаны друг с другом.

Действительно, если для матрицы A известна ее обратная матрица A^{-1} , то, умножая равенство

$$AX = F \quad (1)$$

на A^{-1} слева, мы получим

$$X = A^{-1}F. \quad (2)$$

Обратно, определение элементов обратной матрицы можно свести к решению n систем вида

$$\sum_{k=1}^n a_{ik} a_{kj} = \delta_{ij} \quad \begin{matrix} j = 1, \dots, n \\ i = 1, \dots, n \end{matrix} \quad (3)$$

где δ_{ij} — символ Кронекера.

Последнее вытекает из определения обратной матрицы

$$AA^{-1} = E$$

и правила умножения матриц.

Применяющийся в строительной механике прием определения решения системы при помощи так называемых чисел влияния¹⁾ есть не что иное, как решение системы посредством построения обратной матрицы. Сами числа влияния суть элементы обратной матрицы.

Численное решение n систем уравнений, дающих элементы обратной матрицы, можно осуществлять, например, по методу единственного деления для нескольких систем с общей матрицей коэффициентов (см. табл. II. 9).

Таблица II. 9

Обращение матрицы. Схема единственного деления

1.00	0.42	0.54	0.66	1	0	0	0	3.62
0.42	1.00	0.32	0.44	0	1	0	0	3.18
0.54	0.32	1.00	0.22	0	0	1	0	3.08
0.66	0.44	0.22	1.00	0	0	0	1	3.32
1	0.42	0.54	0.66	1	0	0	0	3.62
0.82360	0.09320	0.16280	-0.42	1	0	0	0	1.65960
0.09320	0.70840	-0.13640	-0.54	0	1	0	0	1.12520
0.16280	-0.13640	0.56440	-0.66	0	0	1	0	0.93080
	1	0.11316	0.19767	-0.50996	1.21418	0	0	2.01506
		0.69785	-0.15482	-0.49247	-0.11316	1	0	0.93740
		-0.15482	0.53222	-0.57698	-0.19767	0	1	0.60275
		1	-0.22185	-0.70570	-0.16216	1.43297	0	1.34327
			0.49787	-0.68624	-0.22278	0.22185	1	0.81071
			1	-1.37835	-0.44746	0.44560	2.00856	1.62836
				-1.01149	-0.26143	1.53183	0.44560	1.70452
				-0.12304	1.33221	-0.26142	-0.44746	1.50030
1				2.50759	-0.12303	-1.01149	-1.37834	0.99472

В результате мы получим матрицу, состоящую из строк обратной матрицы, расположенных в противоположном порядке. Для контроля вычисления и оценки точности результата целесообразно произвести умножение A на A^{-1} .

¹⁾ А. А. Уманский [1].

Теорема о разложении матрицы в произведение двух треугольных матриц дает возможность построить компактную схему для вычисления элементов обратной матрицы.¹⁾ Эта схема требует всего $2n^2$ записей, причем из них n^2 записей будут давать элементы обратной матрицы.

Пусть

$$A = CB, \quad (4)$$

причем элементы треугольных матриц C и B определяются по формулам (2) § 18, которые мы здесь перепишем:

$$c_{ij} = a_{ij} - \sum_{l=1}^{j-1} c_{il} b_{lj} \quad (l \geq j),$$

$$b_{ij} = \frac{a_{ij} - \sum_{l=1}^{i-1} c_{il} b_{lj}}{c_{ii}} \quad (l < j; b_{ii} = 1).$$

Обозначим элементы обратной матрицы $A^{-1} = D$ через d_{ij} .

Имеем, очевидно, что

$$D = B^{-1}C^{-1}. \quad (5)$$

Покажем, что элементы d_{ij} можно определить, не обращая матриц B и C .

Умножая равенство (5) на C справа, получим

$$DC = B^{-1}. \quad (6)$$

Матрица B^{-1} , очевидно, также треугольная, с единицами по главной диагонали. Поэтому мы знаем ее $\frac{n(n+1)}{2}$ элементов (из них $\frac{n(n-1)}{2}$ будут нулями и остальные n единицами).

Аналогично, умножая равенство (5) на B слева, получим

$$BD = C^{-1}. \quad (7)$$

Так как матрица C^{-1} треугольная, то $\frac{n(n-1)}{2}$ ее элементов будут нулями.

Легко видеть, что система, полученная объединением упомянутых выше $\frac{n(n+1)}{2}$ равенств системы $DC = B^{-1}$ и $\frac{n(n-1)}{2}$ равенств системы $BD = C^{-1}$, является рекуррентной системой, дающей возможность определить n^2 элементов обратной матрицы.

¹⁾ Уо и Дайр [1].

Мы выпишем ее для $n = 4$.

$i = 1$	2	3	4
$c_{11}d_{11} + c_{21}d_{12} + c_{31}d_{13} + c_{41}d_{14} =$	1	0	0
$c_{22}d_{12} + c_{32}d_{13} + c_{42}d_{14} =$	1	0	0
$c_{33}d_{13} + c_{43}d_{14} =$	1	0	
$c_{44}d_{14} =$	1		
$j =$	2	3	4
$d_{1j} + b_{12}d_{2j} + b_{13}d_{3j} + b_{14}d_{4j} =$	0	0	0
$d_{2j} + b_{23}d_{3j} + b_{24}d_{4j} =$	0	0	
$d_{3j} + b_{34}d_{4j} =$	0		

Из уравнений первой группы, при $i = 4$, определяются последовательно d_{44} , d_{43} , d_{42} , d_{41} . Затем, из уравнений второй группы, при $j = 4$, определяются d_{34} , d_{24} , d_{14} . Далее процесс идет аналогично. И мы пользуемся по очереди формулами первой и второй группы. Именно, из уравнений первой группы, при $i = 3$, определяются d_{31} , d_{32} и d_{33} и из уравнений второй группы, при $j = 3$, d_{23} и d_{13} ; из уравнений первой группы, при $i = 2$, d_{21} и d_{22} , и из уравнений второй группы, при $j = 2$, d_{12} . Наконец, из уравнений первой группы, при $i = 1$, определяем d_{11} . Обращение матрицы по компактной схеме показано в табл. II. 10.

Таблица II. 10

Компактная схема обращения матрицы

1.00	0.42	0.54	0.66				
0.42	1.00	0.32	0.44				
0.54	0.32	1.00	0.22				
0.66	0.44	0.22	1.00				
1.00	1.042	0.54	0.66	2.50759	-0.12303	-1.01149	-1.37834
0.42	0.82360	1.011316	0.19767	-0.12303	1.33221	-0.26143	-0.44745
0.54	0.09320	0.69785	1.022185	-1.01149	-0.26143	1.53184	0.44560
0.66	0.16280	-0.15482	0.49787	1.037834	-0.44745	0.44560	2.00855

Компактная схема обращения матрицы может быть распространена на клеточные матрицы с квадратными диагональными клетками. Разложив матрицу A в произведение CB двух квазитреугольных матриц, мы ищем обратную матрицу $D = A^{-1}$ тоже в клеточном виде. Тогда, аналогично числовому случаю, клетки D_{ij} обратной

матрицы находятся последовательно из соотношений (мы их приводим для $n = 4$):

$$\begin{array}{l} i=1 \quad 2 \quad 3 \quad 4 \\ D_{11}C_{11} + D_{12}C_{21} + D_{13}C_{31} + D_{14}C_{41} = E \quad 0 \quad 0 \quad 0 \\ D_{12}C_{22} + D_{13}C_{32} + D_{14}C_{42} = \quad E \quad 0 \quad 0 \\ D_{13}C_{33} + D_{14}C_{43} = \quad E \quad 0 \\ D_{14}C_{44} = \quad E \\ j= \quad 2 \quad 3 \quad 4 \\ D_{1j} + B_{12}D_{2j} + B_{13}D_{3j} + B_{14}D_{4j} = \quad 0 \quad 0 \quad 0 \\ D_{2j} + B_{23}D_{3j} + B_{24}D_{4j} = \quad 0 \quad 0 \\ D_{3j} + B_{34}D_{4j} = \quad 0 \end{array}$$

Порядок определения матриц D_{ij} такой же, как и в случае числовых матриц.

§ 22. Задача исключения

Эта задача в простейшем случае состоит в вычислении значения линейной формы

$$c_1x_1 + c_2x_2 + \dots + c_nx_n, \quad (1)$$

где c_1, c_2, \dots, c_n данные числа, а x_1, x_2, \dots, x_n решение системы

$$\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = f_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = f_2 \\ \dots \dots \dots \dots \dots \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = f_n \end{array} \quad (2)$$

определитель которой отличен от нуля.

Естественный ход решения этой задачи заключается в определении чисел x_1, x_2, \dots, x_n в явном виде и в подстановке их в выражение (1). Однако этого можно избежать следующим образом.

Запишем матрицу коэффициентов системы (2) и припишем справа от нее столбец свободных членов, а под ней строку, состоящую из коэффициентов вычисляемой линейной формы, взятых с обратными знаками. В правом нижнем углу поставим число 0. Мы придем к схеме:

$$\begin{array}{c|c} a_{11} & a_{12} \dots a_{1n} \\ a_{21} & a_{22} \dots a_{2n} \\ \dots & \dots \dots \dots \\ a_{n1} & a_{n2} \dots a_{nn} \\ \hline -c_1 & -c_2 \dots -c_n \end{array} \left| \begin{array}{c} f_1 \\ f_2 \\ \dots \\ f_n \\ 0 \end{array} \right. \quad (3)$$

или, в сокращенной записи,

$$\begin{array}{c|c} A & F \\ \hline \hline -C & 0 \end{array}. \quad (3')$$

Пусть $\gamma_1, \gamma_2, \dots, \gamma_n$ обозначают решение системы уравнений:

$$\begin{aligned} a_{11}\gamma_1 + a_{21}\gamma_2 + \dots + a_{n1}\gamma_n &= c_1 \\ a_{12}\gamma_1 + a_{22}\gamma_2 + \dots + a_{n2}\gamma_n &= c_2 \\ \dots &\dots \dots \dots \dots \dots \\ a_{1n}\gamma_1 + a_{2n}\gamma_2 + \dots + a_{nn}\gamma_n &= c_n. \end{aligned} \quad (4)$$

Тогда

$$\begin{aligned} f_1\gamma_1 + f_2\gamma_2 + \dots + f_n\gamma_n &= \\ &= (a_{11}x_1 + \dots + a_{1n}x_n)\gamma_1 + \\ &+ (a_{21}x_1 + \dots + a_{2n}x_n)\gamma_2 + \\ &+ \dots \dots \dots \dots + \\ &+ (a_{n1}x_1 + \dots + a_{nn}x_n)\gamma_n = \\ &= (a_{11}\gamma_1 + \dots + a_{n1}\gamma_n)x_1 + \\ &+ (a_{12}\gamma_1 + \dots + a_{n2}\gamma_n)x_2 + \\ &+ \dots \dots \dots \dots + \\ &+ (a_{1n}\gamma_1 + \dots + a_{nn}\gamma_n)x_n = \\ &= c_1x_1 + c_2x_2 + \dots + c_nx_n. \end{aligned}$$

Таким образом, вычисление формы $c_1x_1 + c_2x_2 + \dots + c_nx_n$ можно заменить вычислением формы $f_1\gamma_1 + f_2\gamma_2 + \dots + f_n\gamma_n$. С другой стороны, очевидно, что если мы умножим первую строку схемы (3) на γ_1 , вторую на γ_2, \dots, n -ю на γ_n и добавим к последней строке, то мы получим строку, элементы которой, расположенные за двойной чертой, равны нулю; элемент в правом нижнем углу, очевидно, равен искомому числу. Обратно, если каким-либо способом мы подберем линейную комбинацию n строк так, что добавление ее к последней строке дает нулевую строку (до черты), то коэффициенты этой комбинации будут решениями системы (4), и, следовательно, элемент в правом нижнем углу будет равен искомому числу. Это следует из единственности решения системы (4). Таким образом, нет необходимости находить числа $\gamma_1, \gamma_2, \dots, \gamma_n$. Нужно только аннулировать последнюю строку за счет добавления подходящей линейной комбинации первых n строк. Это можно сделать обычным прямым ходом процесса Гаусса, примененным к схеме (3).

Очевидно, что такой же прием может быть применен для вычисления линейной неоднородной формы $c_1x_1 + \dots + c_nx_n + d$. Разница будет только в том, что в правом нижнем углу нужно записать вместо 0 свободный член d , так что исходная схема имеет вид:

$$\begin{array}{c|c} a_{11} & a_{12} \dots a_{1n} \\ \hline \vdots & \vdots \\ a_{n1} & a_{n2} \dots a_{nn} \\ \hline -c_1 - c_2 \dots -c_n & d \end{array} \quad (3'')$$

Этим способом можно найти и решение системы (2) без применения обратного хода. Действительно, выражения x_1, x_2, \dots, x_n являются частными случаями линейной формы (1) с коэффициентами $(1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, 0, \dots, 1)$. Одновременное их определение может быть осуществлено по схеме исключения с одновременной записью в нижнем левом углу строк $(-1, 0, \dots, 0), (0, -1, \dots, 0), \dots, (0, 0, \dots, -1)$, которые вместе составляют матрицу $-E$. В правом нижнем углу вместо числа 0 мы помещаем нулевой столбец.

Таким образом, исходная схема для решения системы имеет вид:

$$\begin{array}{c|c} A & F \\ \hline -E & 0 \end{array} \quad (5)$$

Получив нулевую матрицу в левом нижнем углу схемы за счет добавления подходящих линейных комбинаций первых n строк, мы получим в правом нижнем углу столбец, составленный из значений неизвестных.

Обращение матрицы равносильно, как мы видели, решению n систем частного вида, свободные члены которых образуют единичную матрицу. Решение их в совокупности можно осуществить при помощи схемы

$$\begin{array}{c|c} A & E \\ \hline -E & 0 \end{array} \quad (6)$$

где E по-прежнему обозначает единичную матрицу, а в правом нижнем углу расположена нулевая матрица n -го порядка. После аннулирования всех строк в левом нижнем углу за счет добавления подходящих линейных комбинаций первых n строк, мы получим в правом нижнем углу матрицу A^{-1} .

Схема (6) может быть видоизменена следующим образом. Очевидно, что аннулирование матрицы $-E$, находящейся в левом нижнем углу схемы (6), требует такого же линейного комбинирования первых n строк схемы, как при получении единичной матрицы E .

в левом верхнем углу вместо матрицы A . Поэтому вычисления можно проводить так. Составим матрицу $[A, E]$. Посредством деления первой строки на a_{11} сделаем первый элемент первой строки равным единице. Затем получим нуль в первом столбце за счет добавления ко всем строкам первой, умноженной на подходящие множители. Далее получим единицу на месте второго элемента второй строки, посредством деления, и нуль на месте всех остальных элементов второго столбца, за счет добавления второй строки к остальным (в частности к первой). Через n шагов мы получим на месте матрицы A единичную матрицу, а тогда на месте E окажется матрица A^{-1} .

По ходу процесса можно как угодно переставлять строки, что позволяет избегать деления на числа, близкие к нулю, если матрица не очень плохо обусловлена.

Описанному процессу можно дать другое толкование. Вспомним, что линейное комбинирование строк равносильно умножению слева на некоторую матрицу. Обозначим через B_1, B_2, \dots, B_n последовательные матрицы, умножение на которые соответствует окончанию 1-го, 2-го, ..., n -го шагов процесса. Ясно, что матрицы, получающиеся по ходу процесса, будут $[A, E], [B_1 A, B_1], [B_2 A, B_2], \dots, [B_n A, B_n]$.

У матрицы $B_1 A$ первый столбец совпадает с первым столбцом единичной матрицы, у матрицы $B_2 A$ уже первые два столбца совпадают со столбцами единичной матрицы и т. д.

Обозначим через $V_i^{(k)}$ вектор, компоненты которого равны элементам i -й строки матрицы B_k , через A_j j -й столбец матрицы A . Из сказанного выше ясно, что

$$(A_j, V_i^{(k)}) = \delta_{ij} \text{ при } j \leq k.$$

Сам процесс построения векторов $V_i^{(k)}$ в точности совпадает с описанным в § 9 процессом построения двойственного базиса. Двойственный базис здесь строится к базису, состоящему из столбцов матрицы A , исходя из системы координатных векторов.

Выпишем формулы, соответствующие описанному процессу.

$$V_1^{(0)} = (1, 0, \dots, 0)', V_2^{(0)} = (0, 1, \dots, 0)', \dots, V_n^{(0)} = (0, 0, \dots, 1)'$$

$$V_k^{(k)} = \frac{1}{(A_k, V_k^{(k-1)})} V_k^{(k-1)}$$

$$V_i^{(k)} = V_i^{(k-1)} - (A_k, V_i^{(k-1)}) V_k^{(k)}.$$

Вместо единичных векторов можно брать любую другую систему линейно-независимых векторов.

Однако при неудачном выборе начальной системы векторов может случиться, что одно из скалярных произведений $(A_k, V_k^{(k-1)})$ окажется равным нулю или станет близким к нулю. В частности, для

координатной системы векторов это будет иметь место, если один из главных миноров матрицы A равен нулю или близок к нулю.

Следует отметить, что если взять в качестве системы векторов $V_1^{(0)}, V_2^{(0)}, \dots, V_n^{(0)}$ столбцы A_k матрицы A , то, как легко видеть $(A_k, V_k^{(k-1)}) \neq 0$. Это следует из того (§ 12), что матрица

$$\begin{bmatrix} (A_1, A_1) & (A_1, A_2) & \dots & (A_1, A_n) \\ (A_2, A_1) & (A_2, A_2) & \dots & (A_2, A_n) \\ \vdots & \vdots & \ddots & \vdots \\ (A_n, A_1) & (A_n, A_2) & \dots & (A_n, A_n) \end{bmatrix} = A'A$$

положительно определена, и, следовательно, все ее главные миноры положительны.

Отметим также, что процесс проходит хорошо, если за векторы $V_1^{(0)}, \dots, V_n^{(0)}$ взять строки матрицы, близкой к обратной.

В работе Хестинса [4] рекомендуется для обращения плохо обусловленных матриц пользоваться описанным процессом, проводя его сначала исходя из столбцов матрицы A , а затем, с целью уточнения (возможно несколько раз), исходя из полученного приближения к обратной матрице.

Результату исключения можно придать матричную форму. Это дает возможность получить некоторые обобщения.

Именно, решение системы x_1, \dots, x_n в матричном виде можно записать как столбец

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A^{-1}F,$$

а значение линейной формы $c_1x_1 + \dots + c_nx_n$ — как число $CA^{-1}F$.

Такое представление указывает путь для вычисления более сложного выражения $CA^{-1}B$, где C и B некоторые прямоугольные матрицы, из которых C состоит из p столбцов, а B из p строк, причем число столбцов B и число строк C безразлично. Действительно, элемент i -й строки и k -го столбца матрицы $CA^{-1}B$ есть число $c_iA^{-1}b_k$, где c_i есть i -я строка матрицы C , а b_k k -й столбец матрицы B .

Поэтому вычисление элементов матрицы $CA^{-1}B$ можно производить методом исключения, примененным к схеме

$$\begin{array}{c|c} A & B \\ \hline \hline -C & 0 \end{array} \quad (7)$$

После аннулирования элементов, находящихся в левом нижнем углу, за счет добавления линейной комбинации первых n строк мы получим в правом нижнем углу матрицу $CA^{-1}B$.

Очевидно, что при одновременном изменении знаков всех элементов матриц B и C результат не меняется, так что схема (7) равносильна схеме

$$\begin{array}{c|c} A & B \\ \hline \hline C & 0 \end{array} \quad (7')$$

Вычисление обратной матрицы для матрицы A можно осуществлять при помощи схемы

$$\begin{array}{c|c} BA & B \\ \hline \hline -E & 0 \end{array}, \quad (8)$$

в которой в качестве матрицы B может быть взята произвольная невырожденная матрица. Применение схемы (8) по существу равносильно применению описанного выше процесса биортогонализации исходя из системы векторов, компоненты которых образуют строки матрицы B . В частности, процесс при $B = A'$ равносителен процессу биортогонализации столбцов матрицы A .

Для вычисления произведения $CA^{-1}B$ можно кроме схемы (7) пользоваться транспонированной (с точностью до знаков B и C) схемой

$$\begin{array}{c|c} A' & C' \\ \hline \hline -B' & 0 \end{array}. \quad (9)$$

При вычислении по этой схеме мы получим в правом нижнем углу матрицу $(CA^{-1}B)'$, так как $(CA^{-1}B)' = B'(A')^{-1}C'$.

В частности, для вычисления значения линейной формы $c_1x_1 + c_2x_2 + \dots + c_nx_n$ можно также пользоваться схемой

$$\begin{array}{c|c} A' & C' \\ \hline \hline -F' & 0 \end{array}, \quad (9')$$

где A' — матрица, транспонированная с матрицей коэффициентов системы, C' — столбец, составленный из коэффициентов вычисляемой линейной формы, F' — строка из свободных членов. Справедливость этого построения легко обосновывается непосредственно, без ссылки на приведенные результаты. Действительно, если мы умножим первые n строк схемы (9') на x_1, x_2, \dots, x_n и добавим к последней,

мы получим нуль в левом нижнем углу и $c_1x_1 + c_2x_2 + \dots + c_nx_n$ в правом нижнем углу.

Аналогично, для решения системы $AX = F$ мы можем пользоваться схемой

$$\begin{array}{c|c} A' & E \\ \hline \hline -F' & 0 \end{array}. \quad (10)$$

Схемы (7) и (9) могут быть применены для вычисления $A^{-1}B$. Для этого достаточно положить $C = E$ в этих схемах.

Для обращения матрицы наряду со схемой (6) можно пользоваться схемой

$$\begin{array}{c|c} A' & E \\ \hline \hline -E & 0 \end{array}. \quad (11)$$

При этом после окончания процесса мы получим в правом нижнем углу матрицу, транспонированную с A^{-1} .

Таким образом, для каждой из разобранных задач мы имеем две схемы, определяемые или матрицей A или ее транспонированной. Отметим, что целесообразно пользоваться той схемой, которая в нижнем левом углу содержит наименьшее число строк. Так, в задаче решения линейной системы без обратного хода целесообразно пользоваться схемой с транспонированной матрицей, в задаче вычисления произведения $A^{-1}B$ — схемой с транспонированной матрицей, в случае, если число столбцов в B меньше n , и схемой с данной матрицей, если число столбцов в B больше n .

В табл. II. 11 дается обращение матрицы методом биортогонализации столбцов. В графах I, II, III, IV записываются компоненты столбцов матрицы A и компоненты последовательно вычисляемых векторов $V_i^{(k)}$. В графе V, для контроля, записываются $(A_k, V_i^{(k)})$, в графе VI числа $c_{ik} = (A_k, V_i^{(k-1)})$, в графе VII числа $c_i = \frac{1}{c_{ii}}$. В последних четырех строках граф I—IV получается A^{-1} .

Контрольное умножение обратной матрицы на матрицу A показало, что максимальный элемент $E = A^{-1}A$ по модулю не больше 0.00005.

В табл. II. 12 мы приводим решение системы, проведенное по схеме (10).

Как обычно, последний столбец является контрольным. Его элементами являются строчные суммы.

Отметим, что при решении системы по схеме (10) число операций немного превосходит число операций по способу Гаусса. Однако однообразие процесса, именно — отсутствие обратного хода, нередко делает этот способ более удобным.

Таблица II. 13

Обращение матрицы методом биортогонализации столбцов

I	II	III	IV	V	VI	VII
1.00	0.47	-0.11	0.55			
0.17	1.00	0.35	0.43			
-0.25	0.67	1.00	0.36			
0.54	-0.32	-0.74	1.00			
0.65125	0.30609	-0.07164	0.35819	0.999997	1.5355	0.65125
-0.37575	0.74350	0.41003	0.12984	0.000004	0.8380	
-0.34958	0.62320	1.01095	0.30523	-0.000004	0.1529	
-0.12493	-0.63252	-0.66687	0.63429	0.000001	1.0210	
0.88455	-0.15555	-0.32623	0.27757	-0.000002	0.54575	
-0.42749	0.84589	0.46650	0.14772	1.000011	0.87896	1.13771
0.09879	-0.26401	0.52166	0.15029	-0.000010	1.04885	
-0.38759	-0.11279	-0.38024	0.72505	0.000007	-0.61442	
1.03020	-0.54478	0.44286	0.49914	-0.000002	-0.55166	
-0.74255	1.68785	-1.19713	-0.33157	0.000002	1.19330	
0.26402	-0.70557	1.39414	0.40165	0.999997	0.37418	2.67251
-0.36175	-0.18186	-0.24377	0.76437	0.000006	-0.09798	
1.43425	-0.34165	0.71514	-0.35461	0.000009	0.90206	
-0.91580	1.60075	-1.31388	0.03450	-0.000001	-0.38678	
0.14682	-0.76449	1.31516	0.64930	0.000001	-0.26166	
-0.44792	-0.22518	-0.30184	0.94645	0.999992	0.80762	1.23821

В табл. II. 13 мы приводим вычисление произведения $A^{-1}B$. Здесь в качестве A взята матрица

$$\begin{bmatrix} 1.00 & 0.42 & 0.54 & 0.66 \\ 0.42 & 1.00 & 0.32 & 0.44 \\ 0.54 & 0.32 & 1.00 & 0.22 \\ 0.66 & 0.44 & 0.22 & 1.00 \end{bmatrix}$$

а в качестве B — матрица

$$\begin{bmatrix} 0.25 & 0.30 & 0.15 & 0.20 \\ 0.45 & 0.50 & 0.30 & 0.40 \\ 0.65 & 0.70 & 0.45 & 0.60 \\ 0.85 & 0.90 & 0.60 & 0.80 \end{bmatrix}.$$

Вычисление $A^{-1}B$ произведено по схеме (9) при $C' = E$. В результате вычислений

$$A^{-1}B = \begin{bmatrix} -1.25753 & -1.25780 & -0.94295 & -1.25728 \\ 0.01847 & 0.04348 & -0.00491 & -0.00655 \\ 1.00394 & 1.03916 & 0.72652 & 0.96871 \\ 1.45096 & 1.48240 & 1.06466 & 1.41957 \end{bmatrix}.$$

Таблица II.12

Решение системы линейных уравнений методом исключения

1.00	0.42	0.54	0.66	1	0	0	0	3.62
0.42	1.00	0.32	0.44	0	1	0	0	3.18
0.54	0.32	1.00	0.22	0	0	1	0	3.08
0.66	0.44	0.22	1.00	0	0	0	1	3.32
-0.3	-0.5	-0.7	-0.9	0	0	0	0	-2.40
<hr/>								
1	0.42	0.54	0.66	1	0	0	0	3.62
	0.82360	0.09320	0.16280	-0.42	1	0	0	1.65960
	0.09320	0.70840	-0.13640	-0.54	0	1	0	1.12520
	0.16280	-0.13640	0.56440	-0.66	0	0	1	0.93050
	-0.37400	-0.53800	-0.70200	0.30	0	0	0	-1.31400
<hr/>								
	1	0.11316	0.19767	-0.50996	1.21418	0	0	2.01506
		0.69785	-0.15482	-0.49247	-0.11316	1	0	0.93740
		-0.15482	0.53222	-0.57698	-0.19767	0	1	0.60275
		-0.49568	-0.62807	0.10928	0.45410	0	0	-0.56037
<hr/>								
		1	-0.22185	-0.70570	-0.16216	1.43297	0	1.34327
			0.49737	-0.68624	-0.22278	0.22185	1	0.81071
			-0.73804	-0.24052	0.37372	0.71029	0	0.10546
<hr/>								
			1	-1.37835	-0.44746	0.44560	2.00856	1.62836
				-1.25780	0.04348	1.03916	1.48240	1.30725

Таблица II. 13

Вычисление произведения $A^{-1}B$

1.00	0.42	0.54	0.66	1	0	0	0	3.62
0.42	1.00	0.32	0.44	0	1	0	0	3.18
0.54	0.32	1.00	0.22	0	0	1	0	3.08
0.66	0.44	0.22	1.00	0	0	0	1	3.32
-0.25	-0.45	-0.65	-0.85	0	0	0	0	-2.20
-0.30	-0.50	-0.70	-0.90	0	0	0	0	-2.40
-0.15	-0.30	-0.45	-0.60	0	0	0	0	-1.50
-0.20	-0.40	-0.60	-0.80	0	0	0	0	-2.00
1	0.42	0.54	0.66	1	0	0	0	3.62
0.82360	0.09320	0.16280	-0.42	1	0	0	0	1.65960
0.09320	0.70840	-0.13640	-0.54	0	1	0	0	1.12520
0.16280	-0.13640	0.56440	-0.66	0	0	1	0	0.93080
-0.34500	-0.51500	-0.68500	0.25	0	0	0	0	-1.29500
-0.37400	-0.53800	-0.70200	0.30	0	0	0	0	-1.31400
-0.23700	-0.36900	-0.50100	0.15	0	0	0	0	-0.95700
-0.31600	-0.49200	-0.66800	0.20	0	0	0	0	-1.27600
1	0.11316	0.19767	-0.50996	1.21418	0	0	0	2.01506
	0.69785	-0.15482	-0.49247	-0.11316	1	0	0	0.93740
	-0.15482	0.53222	-0.57698	-0.19767	0	1	0	0.60275
	-0.47596	-0.61680	0.07406	0.41889	0	0	0	-0.59980
	-0.49568	-0.62807	0.10928	0.45410	0	0	0	-0.56037
	-0.34218	-0.45415	0.02914	0.28776	0	0	0	-0.47943
	-0.45624	-0.60554	0.03885	0.38368	0	0	0	-0.63924
	1	-0.22185	-0.70570	-0.16216	1.43297	0	0	1.34327
		0.49787	-0.68624	-0.22278	0.22185	1	0	0.81071
		-0.72239	-0.26182	0.34171	0.68204	0	0	0.03954
		-0.73804	-0.24052	0.37372	0.71029	0	0	0.10546
		-0.53006	-0.21234	0.23227	0.49033	0	0	-0.01979
		-0.70676	-0.28312	0.30970	0.65878	0	0	-0.02639
	1	-1.37835	-0.44746	0.44560	2.00856	0	0	1.62836
		-1.25753	0.01847	1.00394	1.45096	0	0	1.21585
		-1.25780	0.04348	1.03916	1.48240	0	0	1.30725
		-0.94295	-0.00491	0.72652	1.06466	0	0	0.84334
		-1.25728	-0.00655	0.96871	1.41957	0	0	1.12447

§ 23. Исправление элементов обратной матрицы

Обращение матрицы по всем приведенным выше схемам не дает уверенности в точности полученных результатов из-за неизбежных округлений, влияние которых на конечный результат трудно оценить. Для контроля точности матрицы D_0 , полученной из данной матрицы A каким-либо процессом обращения, следует составить произведение AD_0 . Отклонение этого произведения от единичной матрицы указывает степень неточности полученных результатов.

Пусть это контрольное вычисление покажет нам, что приближение D_0 к A^{-1} таково, что $\|R_0\| \leq k < 1$, где

$$R_0 = E - AD_0. \quad (1)$$

В качестве нормы матриц удобно брать первую или вторую норму, введенные в § 13, как наиболее легко вычисляемые.

При этом условии элементы обратной матрицы A^{-1} могут быть вычислены со сколь угодно большой точностью при помощи следующего итерационного процесса.¹⁾

Образуем последовательность матриц

$$\begin{aligned} D_1 &= D_0(E + R_0), \quad R_1 = E - AD_1 \\ D_2 &= D_1(E + R_1), \quad R_2 = E - AD_2 \\ &\dots \dots \dots \dots \\ D_m &= D_{m-1}(E + R_{m-1}), \quad R_m = E - AD_m. \end{aligned} \quad (2)$$

Проверим, что матрица $R_m = E - AD_m$ равна $R_0^{2^m}$. Действительно,

$$\begin{aligned} R_m &= E - AD_m = E - AD_{m-1}(E + R_{m-1}) = \\ &= E - (E - R_{m-1})(E + R_{m-1}) = R_{m-1}^2 = R_{m-2}^4 = \dots = R_0^{2^m}. \end{aligned} \quad (3)$$

Отсюда следует, что

$$D_m = A^{-1}(E - R_0^{2^m}). \quad (4)$$

Последняя формула показывает, что D_m стремится к A^{-1} , причем сходимость процесса очень быстрая. Дадим оценку погрешности, принимая во внимание, что $A^{-1} = D_0(AD_0)^{-1} = D_0(E - R_0)^{-1}$:

$$\begin{aligned} \|D_m - A^{-1}\| &= \|A^{-1}R_0^{2^m}\| = \|D_0(E - R_0)^{-1}R_0^{2^m}\| \leq \\ &\leq \|D_0\| \|(E - R_0)^{-1}\| \|R_0^{2^m}\| \leq \|D_0\| \frac{k^{2^m}}{1-k}. \end{aligned} \quad (5)$$

Из этой оценки видно, что как только начальное приближение выбрано так, что $\|E - AD_0\| \leq k < 1$, число верных десятичных знаков возрастает в геометрической прогрессии.

¹⁾ Хотеллинг [1]. Отметим, что Хотеллинг пользовался нормой $N(A)$.

Последовательные приближения следует вычислять, раскрыв скобки в формулах (2), именно:

$$D_m = D_{m-1}(E + R_{m-1}) = D_{m-1} + D_{m-1}(E - AD_{m-1}). \quad (6)$$

Второе слагаемое будет играть при этом роль малой поправки к первому.

Иногда, исходя из матрицы R_0 , целесообразно образовать матрицы $R_1 = R_0^2$, $R_2 = R_0^4 = (R_1)^2$ посредством последовательного возвышения в квадрат и затем воспользоваться формулой (6).

Замечание. В случае, если A симметричная матрица и в качестве начального приближения D_0 взята симметричная матрица, то и все последующие приближения будут матрицами симметричными, хотя матрица R_0 может оказаться и не симметричной матрицей. Действительно, из формулы (6) следует, что $D_m = 2D_{m-1} - D_{m-1}AD_{m-1}$, откуда, допустив, что $A' = A$, $D'_{m-1} = D_{m-1}$, получим

$$\begin{aligned} D'_m &= 2D'_{m-1} - (D_{m-1}AD_{m-1})' = 2D'_{m-1} - D'_{m-1}A'D'_{m-1} = \\ &= 2D_{m-1} - D_{m-1}AD_{m-1} = D_m. \end{aligned}$$

В качестве примера применим описанный процесс для уточнения элементов матрицы A^{-1} для

$$A = \begin{bmatrix} 1.00 & 0.42 & 0.54 & 0.66 \\ 0.42 & 1.00 & 0.32 & 0.44 \\ 0.54 & 0.32 & 1.00 & 0.22 \\ 0.66 & 0.44 & 0.22 & 1.00 \end{bmatrix}. \quad (7)$$

За начальное приближение возьмем результат обращения матрицы A по методу единственного деления (удерживая 4 знака после запятой) из табл. II. 9.

$$D_0 = \begin{bmatrix} 2.5076 & -0.1230 & -1.0115 & -1.3783 \\ -0.1230 & 1.3322 & -0.2614 & -0.4475 \\ -1.0115 & -0.2614 & 1.5318 & 0.4456 \\ -1.3783 & -0.4475 & 0.4456 & 2.0086 \end{bmatrix}.$$

Контрольное вычисление дает для R_0 следующее значение:

$$R_0 = 10^{-6} \begin{bmatrix} -52 & -18 & 20 & -50 \\ -60 & 8 & -10 & 10 \\ -18 & -34 & 26 & -10 \\ -66 & 20 & 10 & -54 \end{bmatrix}.$$

Отсюда мы видим, что $\|R_0\|_1 \leq 0.000150$, $\|R_0\|_H \leq 0.000196$. Для оценки погрешности берем $\|R_0\|_1$. В силу формулы (5) имеем, принимая во внимание, что $\|D_0\|_1 < 5$:

$$\|D_1 - A^{-1}\|_1 < 5 \frac{(0.00015)^2}{1 - 0.00015} < 0.0000001.$$

Таким образом, D_1 дает для A^{-1} значение верное с точностью, по крайней мере, до единицы седьмого знака для каждого ее элемента.

Вычисляя, получим:

$$D_1 = \begin{bmatrix} 2.50758616 & -0.12303930 & -1.01148870 & -1.37834207 \\ -0.12303930 & 1.33221281 & -0.26142705 & -0.44745375 \\ -1.01148870 & -0.26142705 & 1.53182667 & 0.44560858 \\ -1.37834207 & -0.44745375 & 0.44560858 & 2.00855152 \end{bmatrix} \quad (8)$$

Контрольное вычисление $E - AD_1$ дает:

$$E - AD_1 = 10^{-8} \begin{bmatrix} 2 & 0 & 0 & 1 \\ 1 & 0 & -1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}.$$

Уточненное значение матрицы A^{-1} позволяет получить уточненное решение неоднократно рассматриваемой системы

$$\begin{aligned} x_1 + 0.42x_2 + 0.54x_3 + 0.66x_4 &= 0.3 \\ 0.42x_1 + x_2 + 0.32x_3 + 0.44x_4 &= 0.5 \\ 0.54x_1 + 0.32x_2 + x_3 + 0.22x_4 &= 0.7 \\ 0.66x_1 + 0.44x_2 + 0.22x_3 + x_4 &= 0.9. \end{aligned} \quad (9)$$

Именно,

$$x_1 = 1.2577938, \quad x_2 = 0.0434873, \quad x_3 = 1.0391663, \quad x_4 = 1.4823929. \quad (10)$$

§ 24. Обращение матрицы при помощи разбиения на клетки

Иногда бывает целесообразно при обращении матрицы предварительно разбить ее на клетки. Мы рассмотрим формулы для обращения матрицы порядка n , разбитой на четыре клетки по схеме

$$S = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right],$$

где A и D квадратные матрицы порядков p и q ; $p + q = n$.

Будем искать обратную матрицу также в виде клеточной матрицы

$$S^{-1} = \left[\begin{array}{c|c} K & L \\ \hline M & N \end{array} \right],$$

где K и N снова квадратные матрицы порядков p и q .

Согласно правилу умножения клеточных матриц, должны иметь место следующие матричные равенства

$$AK + BM = E$$

$$AL + BN = 0$$

$$CK + DM = 0$$

$$CL + DN = E.$$

Умножив третье равенство слева на BD^{-1} и вычитая из первого, получим

$$(A - BD^{-1}C)K = E,$$

откуда

$$K = (A - BD^{-1}C)^{-1}.$$

Из третьего равенства, далее, находим

$$M = -D^{-1}CK.$$

Подобным же образом из второго и четвертого уравнений находим

$$N = (D - CA^{-1}B)^{-1}$$

и

$$L = -A^{-1}BN.$$

Конечно, эти формулы выведены в предположении, что все указанные обращения матриц выполнимы.

Таким образом, обращение матрицы порядка n сводится к обращению четырех матриц, из которых две имеют порядок p и две порядок q , и к некоторым матричным умножениям.

Выведенные формулы можно изменить так, что для вычисления матриц K , L , M и N нужно обратить лишь две матрицы порядков p и q .

Именно, легко проверить, что

$$N = (D - CA^{-1}B)^{-1}; \quad M = -NCA^{-1}$$

$$L = A^{-1}BN; \quad K = A^{-1} - A^{-1}BM$$

и аналогично

$$K = (A - BD^{-1}C)^{-1}; \quad L = KBD^{-1}$$

$$M = -D^{-1}CK; \quad N = D^{-1} - D^{-1}CL.$$

Последние формулы показывают, что метод разбиения удобно применять в том случае, когда какая-либо диагональная клетка легко обращается.

В качестве примера найдем обращение матрицы

$$\left[\begin{array}{cc|cc} 1.00 & 0.42 & 0.54 & 0.66 \\ 0.42 & 1.00 & 0.32 & 0.44 \\ \hline 0.54 & 0.32 & 1.00 & 0.22 \\ 0.66 & 0.44 & 0.22 & 1.00 \end{array} \right].$$

Вычисление проводим следующим образом.

1) Вычисляем матрицу A^{-1}

$$A^{-1} = \begin{bmatrix} 1.21418 & -0.50996 \\ -0.50996 & 1.21418 \end{bmatrix}$$

и образуем произведения

$$A^{-1}B = \begin{bmatrix} 0.49247 & 0.57698 \\ 0.11316 & 0.19767 \end{bmatrix}; \quad CA^{-1} = \begin{bmatrix} 0.49247 & 0.11316 \\ 0.57698 & 0.19767 \end{bmatrix};$$

$$CA^{-1}B = \begin{bmatrix} 0.30215 & 0.37482 \\ 0.37482 & 0.46778 \end{bmatrix}.$$

Вычисляя последнюю матрицу дважды, как $C(A^{-1}B)$ и $(CA^{-1})B$, мы получаем контроль предыдущих вычислений.

2) Образуем матрицу

$$D - CA^{-1}B = \begin{bmatrix} 0.69785 & -0.15482 \\ -0.15482 & 0.53222 \end{bmatrix}$$

и находим ее обратную

$$N = (D - CA^{-1}B)^{-1} = \begin{bmatrix} 1.53183 & 0.44560 \\ 0.44560 & 2.00855 \end{bmatrix}.$$

3) Вычисляем матрицы

$$M = -NCA^{-1} = \begin{bmatrix} -1.01148 & -0.26142 \\ -1.37834 & -0.44745 \end{bmatrix}$$

$$L = -A^{-1}BN = \begin{bmatrix} -1.01148 & -1.37834 \\ -0.26142 & -0.44745 \end{bmatrix}$$

$$K = A^{-1} - A^{-1}BM = \begin{bmatrix} 2.50758 & -0.12305 \\ -0.12305 & 1.33221 \end{bmatrix}.$$

Таким образом, искомая обратная матрица будет:

$$\begin{bmatrix} 2.50758 & -0.12305 & -1.01148 & -1.37834 \\ -0.12305 & 1.33221 & -0.26142 & -0.44745 \\ -1.01148 & -0.26142 & 1.53183 & 0.44560 \\ -1.37834 & -0.44745 & 0.44560 & 2.00855 \end{bmatrix}.$$

Изложенный метод по существу совпадает с описанной выше компактной схемой обращения клеточной матрицы для случая разбиения матрицы на четыре клетки.

§ 25. Метод окаймления

В этом параграфе мы рассмотрим вычислительные схемы для обращения матрицы и решения линейной системы, основанные на идеи окаймления.

Данную матрицу A мы будем рассматривать как результат окаймления матрицы $n = 1$ -го порядка, для которой мы будем считать известной обратную матрицу. Итак, пусть

$$A = A_n = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1, n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2, n-1} & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-1, 1} & a_{n-1, 2} & \cdots & a_{n-1, n-1} & a_{n-1, n} \\ \hline a_{n1} & a_{n2} & \cdots & a_{n, n-1} & a_{nn} \end{array} \right] = \left[\begin{array}{cc|c} A_{n-1} & u_n \\ v_n & a_{nn} \end{array} \right].$$

Здесь A_{n-1} обозначает упомянутую матрицу $n = 1$ -го порядка,

$$v_n = (a_{n1}, \dots, a_{n, n-1}), \quad u_n = (a_{1n}, \dots, a_{n-1, n})'.$$

Матрицу A^{-1} ищем также в виде окаймленной матрицы

$$D_n = A_n^{-1} = \left[\begin{array}{cc} P_{n-1} & r_n \\ q_n & \frac{1}{a_n} \end{array} \right],$$

где P_{n-1} матрица, q_n строка, r_n столбец и $\frac{1}{a_n}$ число, которые нам нужно определить.

По правилу умножения окаймленных матриц имеем:

$$\begin{aligned} AA^{-1} &= \begin{bmatrix} A_{n-1} & u_n \\ v_n & a_{nn} \end{bmatrix} \begin{bmatrix} P_{n-1} & r_n \\ q_n & \frac{1}{a_n} \end{bmatrix} = \\ &= \begin{bmatrix} A_{n-1}P_{n-1} + u_nq_n, & A_{n-1}r_n + \frac{u_n}{a_n} \\ v_nP_{n-1} + a_{nn}q_n, & v_nr_n + \frac{a_{nn}}{a_n} \end{bmatrix} = \begin{bmatrix} E & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

Отсюда

$$A_{n-1}P_{n-1} + u_nq_n = E \quad (1)$$

$$v_nP_{n-1} + a_{nn}q_n = 0 \quad (2)$$

$$A_{n-1}r_n + \frac{u_n}{a_n} = 0 \quad (3)$$

$$v_nr_n + \frac{a_{nn}}{a_n} = 1. \quad (4)$$

Из равенства (3) имеем

$$r_n = -\frac{A_{n-1}^{-1}u_n}{a_n}.$$

Подставляя значение r_n в (4), получим

$$\alpha_n = a_{nn} - v_nA_{n-1}^{-1}u_n. \quad (5)$$

Далее, из (1) имеем

$$P_{n-1} = A_{n-1}^{-1} - A_{n-1}^{-1}u_nq_n \quad (6)$$

и потому, на основании (2) и (5):

$$\begin{aligned} v_nA_{n-1}^{-1} - v_nA_{n-1}^{-1}u_nq_n + a_{nn}q_n &= \\ &= v_nA_{n-1}^{-1} - (a_{nn} - \alpha_n)q_n + a_{nn}q_n = v_nA_{n-1}^{-1} + \alpha_nq_n = 0. \end{aligned}$$

Отсюда

$$q_n = -\frac{v_nA_{n-1}^{-1}}{\alpha_n}.$$

Наконец,

$$P_{n-1} = A_{n-1}^{-1} + \frac{A_{n-1}^{-1}u_nv_nA_{n-1}^{-1}}{\alpha_n}.$$

Таким образом, окончательно:

$$A^{-1} = \begin{bmatrix} A_{n-1}^{-1} + \frac{A_{n-1}^{-1}u_nv_nA_{n-1}^{-1}}{\alpha_n}, & -\frac{A_{n-1}^{-1}u_n}{\alpha_n} \\ -\frac{v_nA_{n-1}^{-1}}{\alpha_n}, & \frac{1}{\alpha_n} \end{bmatrix}, \quad (7)$$

где $\alpha_n = a_{nn} - v_nA_{n-1}^{-1}u_n$.

Очевидно, что построенная формула является частным случаем формул обращения матрицы методом разбиения на клетки при $p = n - 1$ и $q = 1$.

Выведенная формула кладется в основу метода обращения матрицы посредством последовательного окаймления. Именно, последовательно строятся обратные матрицы для матриц

$$(A_{11}), \quad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \dots,$$

из которых каждая следующая получается из предыдущей посредством окаймления. Каждый шаг этого процесса осуществляется на основании формулы (7). Именно, если A_{n-1}^{-1} уже известна, для нахождения A_n^{-1} нужно произвести следующие действия:

1) Вычислить столбец — $A_{n-1}^{-1}u_n$. Элементы столбца $\beta_{1n}, \dots, \beta_{n-1,n}$ находятся при помощи накоплений.

2) Вычислить строку — $v_n A_{n-1}^{-1}$ с элементами $\gamma_{n1}, \dots, \gamma_{n,n-1}$.

3) Вычислить число

$$\alpha_n = a_{nn} + \sum_{i=1}^{n-1} a_{ni} \beta_{in} = a_{nn} + \sum_{i=1}^{n-1} a_{in} \gamma_{in}.$$

(Двойное вычисление числа α_n является хорошим контролем предшествующих вычислений).

4) Наконец, найти элементы d_{ik} обратной матрицы по формулам:

$$d_{ik} = d'_{ik} + \frac{\beta_{in} \gamma_{nk}}{\alpha_n} \quad i, k \leq n-1$$

$$d_{in} = \frac{\beta_{in}}{\alpha_n}; \quad d_{nk} = \frac{\gamma_{nk}}{\alpha_n} \quad i, k \leq n-1; \quad d_{nn} = \frac{1}{\alpha_n}.$$

Здесь d'_{ik} — элементы матрицы A_{n-1}^{-1} .

В случае симметричной матрицы A схема сокращается, очевидно, ровно вдвое.

В табл. II. 14 дано обращение матрицы (7) § 23 при помощи последовательного окаймления.

Каждый шаг процесса оформлен по следующей схеме

A_{n-1}^{-1}	u_n	$-A_{n-1}^{-1}u_n$
v_n	a_{nn}	
$-v_n A_{n-1}^{-1}$		α_n

Метод окаймления может быть полезен и в применении к решению системы линейных уравнений. Особенно целесообразно использовать этот метод в случае, когда нужно решать систему, для которой уже ранее решена усеченная система, получающаяся из данной

Таблица II. 14

Обращение матрицы методом окаймления

1				
1	0.42	-0.42		
0.42	1			
-0.42		0.82360		
1.21418	-0.50996	0.54	-0.49247	
-0.50996	1.21418	0.32	-0.11316	
0.54	0.32	1		
-0.49247	-0.11316		0.69785	
1.56172	-0.43010	-0.70570	0.66	-0.68624
-0.43010	1.23253	-0.16216	0.44	-0.22277
-0.70570	-0.16216	1.43297	0.22	0.22186
0.66	0.44	0.22	1	
-0.68624	-0.22277	0.22186		0.49787
2.50759	-0.12304	-1.01149	-1.37835	
-0.12304	1.33221	-0.26143	-0.44745	
-1.01149	-0.26143	1.53183	0.44562	
-1.37835	-0.44745	0.44562	2.00856	

вычеркиванием одного уравнения и одного неизвестного. Такая ситуация часто встречается в приложениях. Например, при решении задач математической физики по методу Б. Г. Галеркина или Ритца может оказаться, что решение, полученное в результате использования $n - 1$ -й координатной функции, недостаточно точно; если для построения более точного решения достаточно добавление еще одной координатной функции, то новая система для определения коэффициентов получается из предшествующей системы окаймлением.

Метод окаймления может быть применен к решению системы линейных уравнений следующим образом.

Пусть система имеет вид

$$A_n X_n = F_n.$$

Обозначим

$$A_n = \begin{bmatrix} A_{n-1} & u_n \\ v_n & a_{nn} \end{bmatrix}; \quad F_n = \begin{bmatrix} F_{n-1} \\ f_n \end{bmatrix}.$$

Тогда

$$\begin{aligned} X_n &= \begin{bmatrix} A_{n-1}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} F_{n-1} \\ f_n \end{bmatrix} + \\ &\quad + \frac{1}{a_n} \begin{bmatrix} A_{n-1}^{-1} u_n v_n A_{n-1}^{-1} - A_{n-1}^{-1} u_n \\ -v_n A_{n-1}^{-1} & 1 \end{bmatrix} \begin{bmatrix} F_{n-1} \\ f_n \end{bmatrix} = \\ &= \begin{bmatrix} A_{n-1}^{-1} F_{n-1} \\ 0 \end{bmatrix} + \frac{1}{a_n} \begin{bmatrix} A_{n-1}^{-1} u_n v_n A_{n-1}^{-1} F_{n-1} - A_{n-1}^{-1} u_n f_n \\ -v_n A_{n-1}^{-1} F_{n-1} + f_n \end{bmatrix}. \end{aligned}$$

Но $A_{n-1}^{-1} F_{n-1}$ есть решение усеченной системы, т. е. системы

$$\begin{aligned} a_{11}x_1 + \dots + a_{1, n-1}x_{n-1} &= f_1 \\ \dots \dots \dots \dots \dots \dots & \\ a_{n-1, 1}x_1 + \dots + a_{n-1, n-1}x_{n-1} &= f_{n-1}, \end{aligned}$$

которое мы обозначим X_{n-1} .

Аналогично

$$-A_{n-1}^{-1} u_n = -A_{n-1}^{-1} \begin{bmatrix} a_{1n} \\ \dots \\ a_{n-1, n} \end{bmatrix} = Q_{n-1}$$

есть решение системы с той же матрицей коэффициентов, но с другими свободными членами.

Зная X_{n-1} и Q_{n-1} , мы легко вычислим X_n . Действительно,

$$\begin{aligned} X_n &= \begin{bmatrix} X_{n-1} \\ 0 \end{bmatrix} + \frac{1}{a_{nn} + v_n Q_{n-1}} \begin{bmatrix} -Q_{n-1} v_n X_{n-1} + Q_{n-1} f_n \\ -v_n X_{n-1} + f_n \end{bmatrix} = \\ &= \begin{bmatrix} X_{n-1} \\ 0 \end{bmatrix} + \frac{f_n - v_n X_{n-1}}{a_{nn} + v_n Q_{n-1}} \begin{bmatrix} Q_{n-1} \\ 1 \end{bmatrix}. \end{aligned}$$

Таким образом, для нахождения X_n нам надо кроме X_{n-1} вычислить еще Q_{n-1} .

Если усеченная система была решена по методу Гаусса, то для нахождения Q_{n-1} в схему нужно добавить еще один столбец. Вычисление прямого хода выполняется лишь для этого столбца, с использованием уже известной левой половины схемы. Затем Q_{n-1} определяется обычным обратным ходом.

§ 26. Эскалаторный метод

При обращении матрицы методом окаймления существенную роль при переходе от обращения матрицы A_{k-1} к обращению окаймленной матрицы A_k играет вычисление выражений $-A_{k-1}^{-1}u_k$ и $-v_k A_{k-1}^{-1}$.

При этом компоненты вектора $-A_{k-1}^{-1}u_k$ представляют собой не что иное, как решение системы уравнений:

$$\begin{aligned} a_{11}z_1 + a_{12}z_2 + \dots + a_{1, k-1}z_{k-1} + a_{1k} &= 0 \\ \dots &\dots \\ a_{k-1, 1}z_1 + a_{k-1, 2}z_2 + \dots + a_{k-1, k-1}z_{k-1} + a_{k-1, k} &= 0. \end{aligned}$$

Аналогично, компоненты $-v_k A_{k-1}^{-1}$ образуют решение транспонированной системы.

Последовательное решение таких систем при $k = 2, \dots, n, n+1$ лежит в основе так называемого эскалаторного метода решения систем линейных уравнений¹⁾. Тем самым эскалаторный метод тесно связывается с методом окаймления.

Существенным достоинством метода является наличие надежного контроля, при помощи которого можно регулировать точность вычислений, добиваясь выполнения контрольных равенств за счет привлечения большего числа значащих цифр. Следует, однако, заметить, что эскалаторный метод не является в этом отношении единственным среди точных методов решения систем. Надежным контролем обладает также рассмотренный выше прием биортогонализации столбцов и целый ряд других методов, которые будут рассмотрены ниже.

Мы изложим эскалаторный метод лишь в применении к системам с симметричной матрицей, придав вычислительной схеме компактную форму. При этом обнаружится связь эскалаторного метода с методом Гаусса.

Итак, пусть дана система уравнений:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + a_{1, n+1} &= 0 \\ \dots &\dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n + a_{n, n+1} &= 0, \end{aligned} \tag{1}$$

матрица коэффициентов которой симметрична.

¹⁾ Моррис [4].

Обозначим через $z_{1k}, \dots, z_{k-1, k}$ решение системы:

$$a_{11}z_1 + \dots + a_{1, k-1}z_{k-1} + a_{1k} = 0 \quad (2)$$

$$a_{k-1, 1}z_1 + \dots + a_{k-1, k-1}z_{k-1} + a_{k-1, k} = 0.$$

Числа $z_{i, n+1} = x_i$ очевидно образуют решение системы (1). Поэтому, если мы установим способ последовательного вычисления чисел z_{ik} для $i < k \leq n+1$, то тем самым дадим способ построения искомого решения.

Допустим сначала, что уже вычислены все числа z_{ik} для $i < k \leq n$. Построим с их помощью матрицу

$$Z = \begin{bmatrix} 1 & z_{12} & z_{13} & \dots & z_{1n} \\ 0 & 1 & z_{23} & \dots & z_{2n} \\ 0 & 0 & 1 & \dots & z_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}. \quad (3)$$

Легко видеть, что матрица AZ имеет нули выше главной диагонали. Действительно,

$$C_1 = AZ =$$

$$= \begin{bmatrix} a_{11}, & a_{11}z_{12} + a_{12}, & \dots, & a_{11}z_{1n} + a_{12}z_{2n} + \dots + a_{1n} \\ a_{21}, & a_{21}z_{12} + a_{22}, & \dots, & a_{21}z_{1n} + a_{22}z_{2n} + \dots + a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1}, & a_{n1}z_{12} + a_{n2}, & \dots, & a_{n1}z_{1n} + a_{n2}z_{2n} + \dots + a_{nn} \end{bmatrix} \quad (4)$$

и, в силу определения чисел z_{ij} , все элементы, лежащие выше главной диагонали, равны нулю. Ненулевые элементы матрицы C_1 вычисляются по формулам

$$c_{ij} = a_{i1}z_{1j} + a_{i2}z_{2j} + \dots + a_{i, j-1}z_{j-1, j} + a_{ij}, \quad i \geq j. \quad (5)$$

Отсюда вытекает связь этого метода с методом Гаусса, рассматриваемого в свете разложения матрицы на множители.

Действительно, из (4) имеем

$$A = C_1 Z^{-1}. \quad (6)$$

Но матрица Z^{-1} — треугольная матрица с единичной диагональю и нулевыми элементами ниже диагонали, C_1 — треугольная матрица с нулевыми элементами выше диагонали. Сравнивая это разложение с разложением матрицы, соответствующим схеме единственного деления метода Гаусса ($A = CB$), мы получаем, в силу единственности такого разложения, что

$$Z^{-1} = B, \quad C_1 = C. \quad (7)$$

Представим матрицу C в виде

$$C = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \gamma_{21} & 1 & 0 & \dots & 0 \\ \gamma_{31} & \gamma_{32} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \dots & 1 \end{bmatrix} = \begin{bmatrix} c_{11} & & & & \\ & c_{22} & & & \\ & & c_{33} & & \\ & & & \ddots & \\ & & & & c_{nn} \end{bmatrix} = \Gamma A. \quad (8)$$

Здесь $\gamma_{ij} = \frac{c_{ij}}{c_{jj}}$, $i > j$.

Докажем, что матрица $\Gamma = CA^{-1}$ является матрицей, транспонированной к матрице Z^{-1} . При этом мы будем существенно пользоваться симметричностью матрицы A . Имеем:

$$A = \Gamma A Z^{-1} = A' = (Z^{-1})' \cdot \Lambda \Gamma'.$$

Но $(Z^{-1})'$ — треугольная матрица с нулями выше диагонали, Γ' — треугольная матрица с нулями ниже диагонали и единицами на диагонали. В силу единственности разложения, $\Gamma' = Z^{-1}$.

Это обстоятельство позволяет указать рекуррентные формулы для последовательного определения чисел z_{ij} . Допустим, действительно, что мы уже вычислили элементы первых k столбцов матрицы Z . Тогда по формулам (5) можно определить k столбцов матрицы C , следовательно — k столбцов матрицы Γ , т. е. k строк матрицы Z^{-1} . Для продолжения процесса нам нужно вычислить элементы $(k+1)$ -го столбца матрицы Z . Так как диагональный элемент этого столбца равен единице, а элементы ниже диагонали равны нулю, нам остается дать формулы для вычисления $z_{i,k+1}$, где $i \leq k$. Из равенства $Z\Gamma' = E$ мы получаем, в силу правила умножения матриц, следующие рекуррентные формулы:

$$\begin{aligned} \gamma_{k+1,1} + \gamma_{k+1,2} z_{12} + \dots + \gamma_{k+1,k} z_{1k} + z_{1,k+1} &= 0 \\ \gamma_{k+1,2} + \dots + \gamma_{k+1,k} z_{2k} + z_{2,k+1} &= 0 \\ \dots &\dots \\ \gamma_{k+1,k} + z_{k,k+1} &= 0, \end{aligned} \quad (9)$$

определяющие элементы $(k+1)$ -го столбца матрицы Z .

Упомянутым выше надежным контролем является действительное обращение в нуль нижедиагональных элементов матрицы Γ' .

Приведем компактную схему эскалаторного метода для системы 4 уравнений. (Напомним, что свободные члены для единообразия вычисления мы записываем в левых частях уравнений).

a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	1	z_{12}	z_{13}	z_{14}	x_1
a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	0	1	z_{23}	z_{24}	x_2
a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	0	0	1	z_{34}	x_3
a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	0	0	0	1	x_4
c_{11}	1	γ_{21}	γ_{31}	γ_{41}	γ_{51}				
0	c_{22}	1	γ_{32}	γ_{42}	γ_{52}				
0	0	c_{33}	1	γ_{43}	γ_{53}				
0	0	0	c_{44}	1	γ_{54}				

Схема состоит из трех частей. В первой части записана матрица коэффициентов системы (симметричная по условию). Во второй части мы записываем постепенно элементы столбцов матрицы Z . В нижней части записываются диагональные элементы матрицы C и элементы матрицы Γ' . Отметим, что элементы k -й строки матрицы Γ' мы получаем, умножая столбцы матрицы A на k -й столбец матрицы Z и деля полученные суммы на элемент c_{kk} . Столбец матрицы Z заполняется по формулам (9). В табл. II. 15 дан иллюстративный пример.

Таблица II. 15
Компактная схема эскалаторного метода

1.00	0.42	0.54	0.66	-0.3	1	-0.42	-0.49247	-0.69624	-1.25780
0.42	1.00	0.32	0.44	-0.5	0	1	-0.11316	-0.22278	0.04348
0.54	0.32	1.00	0.22	-0.7	0	0	1	0.22185	1.03917
0.66	0.44	0.22	1.00	-0.9	0	0	0	1	1.48240
1	1	0.42	0.54	0.66	-0.3				
0	0.82360	1	0.11316	0.19767	-0.45410				
0.000003	0.000003	0.69785	1	-0.22185	-0.71030				
-0.000009	-0.000009	-0.000009	0.49787	1	-1.48240				

§ 27. Метод Перселла

С решением $X = (x_1, \dots, x_n)$ системы уравнений

$$\begin{aligned} a_{11}x_1 + \dots + a_{1n}x_n - f_1 &= 0 \\ \dots &\dots \dots \dots \dots \dots \dots \\ a_{n1}x_1 + \dots + a_{nn}x_n - f_n &= 0 \end{aligned} \quad (1)$$

связывается¹⁾ вектор $Z = (x_1, \dots, x_n, 1)' = (X, 1)'$ арифметического $(n+1)$ -мерного пространства R_{n+1} в естественном представлении. Уравнения системы истолковываются как условия ортогональности вектора Z с векторами

$$A_i = (a_{i1}, \dots, a_{in}, -f_i)' (i = 1, 2, \dots, n).$$

Решение разыскивается следующим образом. Шаг за шагом строятся базисы подпространств убывающих размерностей

$$R_{n+1} = R^{(0)} \supset R^{(1)} \supset \dots \supset R^{(n)},$$

где $R^{(k)}$ — подпространство, состоящее из векторов, ортогональных к векторам A_1, \dots, A_k .

Каждый последующий базис

$$V_{k+1}^{(k)}, \dots, V_{n+1}^{(k)}$$

строится из предыдущего

$$V_k^{(k-1)}, \dots, V_{n+1}^{(k-1)}$$

в виде двучленных линейных комбинаций

$$V_i^{(k)} = V_i^{(k-1)} - c_i^{(k)} V_k^{(k-1)} \quad (i = k+1, \dots, n+1). \quad (2)$$

Коэффициенты $c_i^{(k)}$ определяются из условия ортогональности к вектору A_k , что дает

$$c_i^{(k)} = \frac{(A_k, V_i^{(k-1)})}{(A_k, V_k^{(k-1)})}. \quad (3)$$

Для осуществимости процесса нужно, чтобы все скалярные произведения $(A_k, V_k^{(k-1)})$ были отличными от нуля.

В качестве базиса $R^{(0)} = R_{n+1}$ берется естественный базис

$$V_1^{(0)} = (1, 0, \dots, 0)', \dots, V_{n+1}^{(0)} = (0, 0, \dots, 1)'.$$

Из хода процесса ясно, что на всех шагах вектор $V_{n+1}^{(k)}$ будет иметь $(n+1)$ -ю компоненту, равную единице. Единственный базисный вектор $V_{n+1}^{(n)}$ подпространства $R^{(n)}$ ортогонален ко всем векторам A_1, \dots, A_n и имеет последнюю компоненту, равную единице. Таким образом, вектор $V_{n+1}^{(n)}$ дает численное решение системы (1).

Метод Перселя по существу очень близок методам, связанным с треугольной факторизацией и, в частности, если матрица системы симметрична, с эскалаторным методом.

¹⁾ Персель [1].

Таблица II. 16

Решение системы уравнений методом Перселя. Несимметричная матрица коэффициентов

	$V_1^{(0)}$	$V_2^{(0)}$	$V_3^{(0)}$	$V_4^{(0)}$	$V_5^{(0)}$	$V_6^{(1)}$	$V_3^{(1)}$	$V_4^{(1)}$	$V_5^{(1)}$	$V_6^{(2)}$	$V_3^{(2)}$	$V_4^{(2)}$	$V_5^{(2)}$	$V_6^{(3)}$	$V_3^{(3)}$	$V_4^{(3)}$	$V_5^{(3)}$	$V_6^{(4)} = (X, 1)$	
I	1	0	0	0	0	-0.17	0.25	-0.54	0.3	0.385550	-0.64602	0.23367	-0.37470	0.58836	0.44089				
	0	1	0	0	0	1	0	0	0	-0.85589	0.62363	0.39017	0.03647	-0.37740	-0.36304				
	0	0	1	0	0	0	1	0	0	1	0	0	1	0.88681	0.89681	1.16681			
	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0.39357	0.39357		
	0	0	0	0	1	0	0	0	1	0	0	1	0	1	1	1	1		
II						0	0	0	0	-0.000001	-0.000003	-0.000001	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005	
										-0.000005	0.656933	-0.68602	-0.89681	1.06566	1.06566	1.06566	1.06566	1.06566	
III	1	0.17	-0.25	0.54	-0.3	0.9201	0.85589	-0.85589	-0.62363	-0.39017	0.656933	-0.68602	-0.89681	-0.89681	-0.89681	-0.89681	-0.89681	-0.89681	

Таблица II. 17

Решение системы уравнений методом Перселя. Симметричная матрица коэффициентов

	$V_1^{(0)}$	$V_2^{(0)}$	$V_3^{(0)}$	$V_4^{(0)}$	$V_5^{(0)}$	$V_6^{(1)}$	$V_3^{(1)}$	$V_4^{(1)}$	$V_5^{(1)}$	$V_6^{(2)}$	$V_3^{(2)}$	$V_4^{(2)}$	$V_5^{(2)}$	$V_6^{(3)}$	$V_3^{(3)}$	$V_4^{(3)}$	$V_5^{(3)}$	$V_6^{(4)} = (X, 1)$	
I	1	0	0	0	0	-0.42	-0.54	-0.66	0.3	-0.49247	-0.57698	0.10938	-0.68623	-0.24052	-1.25779				
	0	1	0	0	0	1	0	0	0	-0.11316	-0.19767	0.45410	-0.22277	0.37372	0.04349				
	0	0	1	0	0	0	1	0	0	1	0	0	0.22185	0.71029	1.08916				
	0	0	0	1	0	0	0	1	0	0	1	0	1	0	1	1	1		
	0	0	0	0	1	0	1	0	0	0	0	1	0	1	1	1	1		
II						0	0	0	0	0.000003	-0.000011	0.000002	0.000006	-0.000011	-0.000006				
										0.000003	-0.000002	-0.000002	0.000005	-0.000011	-0.000006				
III	1	0.42	0.54	0.66	-0.3	0.83360	0.11316	0.19757	-0.45410	0.69786	-0.22185	-0.71029	0.49788	-1.48240					

Действительно, если составить матрицу

$$Z = \begin{bmatrix} 1 & z_{12} & z_{13} & \dots & z_{1n} \\ 1 & z_{22} & \dots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & & & & 1 \end{bmatrix},$$

столбцы которой состоят из первых n компонент векторов $V_1^{(0)}$, $V_2^{(1)}, \dots, V_n^{(n-1)}$, то эта матрица будет очевидно правой треугольной, а матрица AZ , в силу условий ортогональности будет левой треугольной матрицей. Если матрица A симметрична, то матрица Z совпадает с соответствующей матрицей эскалаторного метода (§ 26, (3)).

В табл. II.16 и II.17 приводится решение систем с несимметричной и симметричной матрицами, рассматривавшимися ранее в табл. II.1 и II.1a.

Схема состоит из трех частей. В первой части записываются последовательно вычисляемые компоненты базисных векторов $V_i^{(k)}$, во второй осуществляется контроль вычислений посредством проверки выполнения условий ортогональности. В первой строке третьей части записываются $(A_k, V_k^{(k-1)})$, во второй строке коэффициенты $c_i^{(k)}$.

§ 28. Метод пополнения для обращения матрицы

Метод пополнения для обращения матрицы основывается на следующей идее. Пусть B — неособенная матрица, обратная для которой известна, u — некоторый столбец, v — некоторая строка, $A = B + uv$. Ясно, что

$$uv = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} (v_1, v_2, \dots, v_n) = \begin{bmatrix} u_1 v_1 & u_1 v_2 & \dots & u_1 v_n \\ u_2 v_1 & u_2 v_2 & \dots & u_2 v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n v_1 & u_n v_2 & \dots & u_n v_n \end{bmatrix}$$

есть матрица ранга один.

Покажем, что матрица, обратная к A , находится по формуле¹⁾

$$A^{-1} = B^{-1} - \frac{1}{\gamma} B^{-1} uv B^{-1}, \quad (1)$$

где $\gamma = 1 + v B^{-1} u$.

Конечно, предполагается, что $\gamma \neq 0$.

¹⁾ Дуайр и Уо [1].

Действительно,

$$(B + uv) \left(B^{-1} - \frac{1}{\gamma} B^{-1}uvB^{-1} \right) = E + uvB^{-1} - \frac{1}{\gamma} uvB^{-1} - \\ - \frac{1}{\gamma} u(vB^{-1}u)vB^{-1} = E + uvB^{-1} - \\ - \frac{1}{\gamma} uvB^{-1} - \frac{1}{\gamma} (\gamma - 1) uvB^{-1} = E.$$

Тем самым справедливость формулы (1) доказана. Установленная связь показывает, что элементы матрицы A^{-1} легко находятся, если элементы матрицы B^{-1} известны.

Полученную формулу можно, в частности, применять в случае, если матрица A получается из матрицы B изменением одной строки, т. е. если

$$A = B + V,$$

где V матрица, все элементы которой равны нулю, кроме элементов изменяемой строки. Пусть эта строка имеет номер k . Тогда

$$V = uv = e_k v,$$

где v — ненулевая строка матрицы V и e_k столбец, k -й элемент которого равен единице, а все остальные равны нулю. Следовательно,

$$A^{-1} = B^{-1} - \frac{1}{1 + v(B^{-1}e_k)} (B^{-1}e_k)(vB^{-1}) = \\ = B^{-1} - \frac{1}{1 + v\beta_k} \beta_k (vB^{-1}) = B^{-1} - \frac{1}{1 + (v', \beta_k)} \beta_k (vB^{-1}),$$

где β_k — k -й столбец матрицы B^{-1} . Обозначив через α_j j -й столбец матрицы A^{-1} , получим

$$\alpha_j = \beta_j - \frac{(v', \beta_j)}{1 + (v', \beta_k)} \beta_k. \quad (2)$$

Метод пополнения для обращения матрицы заключается в следующем. Данная матрица $A = (a_{ij})$ рассматривается как последний член последовательности $A_0 = E$, $A_1, \dots, A_n = A$, причем переход от матрицы A_{k-1} к матрице A_k осуществляется посредством замены k -й строки матрицы A_{k-1} на k -ю строку матрицы A . Таким образом, матрица A^{-1} получается в результате n -кратного применения описанного выше процесса. Выпишем формулы перехода на k -м шагу. Пусть $\alpha_j^{(k)}$ — j -й столбец матрицы A_k^{-1} . Тогда

$$\alpha_j^{(k)} = \alpha_j^{(k-1)} - \frac{(v'_k, \alpha_j^{(k-1)})}{1 + (v'_k, \alpha_k^{(k-1)})} \alpha_k^{(k-1)}. \quad (3)$$

Здесь $v_k = (a_{k1}, \dots, a_{k,k-1}, \dots, a_{kn})$.

В табл. II.18 дано обращение матрицы табл. II. 1а методом пополнения.

Каждый шаг процесса происходит по следующей схеме

	$a_1^{(k-1)}$...	$\left \begin{array}{c} \\ a_k^{(k-1)} \\ \end{array} \right $...	$a_n^{(k-1)}$	v'_k
$1 + p_k$	p_1	...	p_k	...	p_n	

Здесь $p_j = (v'_k, a_j^{(k-1)})$; $w_j = \frac{p_j}{1 + p_k}$.

В последней части таблицы расположена матрица A^{-1} .

Метод пополнения можно проводить по другой вычислительной схеме, несколько более развернутой¹⁾.

Обозначим $c_{ij}^{(k-1)} = (v'_i, a_j^{(k-1)})$.

В этих обозначениях

$$a_j^{(k)} = a_j^{(k-1)} - \frac{c_{kj}^{(k-1)}}{1 + c_{kk}^{(k-1)}} a_k^{(k-1)}. \quad (4)$$

Числа $c_{ij}^{(k)}$ связаны соотношением

$$c_{ij}^{(k)} = c_{ij}^{(k-1)} - \frac{c_{ik}^{(k-1)} c_{kj}^{(k-1)}}{1 + c_{kk}^{(k-1)}}.$$

Действительно,

$$\begin{aligned} c_{ij}^{(k)} &= (v'_i, a_j^{(k)}) = (v'_i, a_j^{(k-1)}) - \frac{(v'_k, a_j^{(k-1)})}{1 + (v'_k, a_k^{(k-1)})} (a_k^{(k-1)}, v'_i) = \\ &= c_{ij}^{(k-1)} - \frac{c_{kj}^{(k-1)} c_{ik}^{(k-1)}}{1 + c_{kk}^{(k-1)}}. \end{aligned}$$

Обозначив далее через $\gamma_j^{(k)} — j\text{-й столбец матрицы } C_k = (c_{ij}^{(k)})$, будем иметь

$$\gamma_j^{(k)} = \gamma_j^{(k-1)} - \frac{c_{kj}^{(k-1)}}{1 + c_{kk}^{(k-1)}} \gamma_k^{(k-1)}. \quad (5)$$

Таким образом, матрица C_k получается из матрицы C_{k-1} совершенно так же, как матрица A_k^{-1} из матрицы A_{k-1}^{-1} .

Для перехода от матрицы A_{k-1}^{-1} к матрице A_k^{-1} (и от C_{k-1} к C_k) нужно кроме матрицы A_{k-1}^{-1} знать лишь элементы k -й строки матрицы C_{k-1} . На следующем шаге понадобятся соответственно элементы $k+1$ -й строки матрицы C_k , для построения которой по формуле (5) в свою очередь нужно знать элементы $k+1$ -й строки

1) А. П. Ершов [1].

Таблица II. 18

Обращение матрицы методом пополнения

	$\begin{vmatrix} 1 \\ 0 \\ 0 \\ 0 \end{vmatrix}$	0	0	0	0
1	0	0.42	0.54	0.66	
	0	0.42	0.54	0.66	
	1	$\begin{vmatrix} -0.42 \\ 1 \\ 0 \\ 0 \end{vmatrix}$	$\begin{vmatrix} -0.54 \\ 0 \\ 1 \\ 0 \end{vmatrix}$	$\begin{vmatrix} -0.66 \\ 0 \\ 0 \\ 1 \end{vmatrix}$	$\begin{vmatrix} 0.42 \\ 0 \\ 0.32 \\ 0.44 \end{vmatrix}$
0.82360	0.42	-0.17640	0.09320	0.16280	
	0.50996	-0.21418	0.11316	0.19767	
	1.21418	-0.50996	$\begin{vmatrix} -0.49247 \\ -0.11316 \end{vmatrix}$	$\begin{vmatrix} -0.57698 \\ -0.19767 \end{vmatrix}$	$\begin{vmatrix} 0.54 \\ 0.32 \end{vmatrix}$
	-0.50996	1.21418			
	0	0	1	0	0
	0	0	0	1	0.22
0.69785	0.49247	0.11316	-0.30215	-0.15482	
	0.70570	0.16216	-0.43297	-0.22185	
	1.56172	-0.43010	-0.70569	$\begin{vmatrix} -0.68623 \\ -0.22277 \end{vmatrix}$	$\begin{vmatrix} 0.66 \\ 0.44 \end{vmatrix}$
	-0.43010	1.23253	-0.16215		
	-0.70570	-0.16216	1.43297	0.22185	0.22
	0	0	0	1	0
0.49788	0.68624	0.22277	-0.22185	-0.50212	
	1.37832	0.44744	-0.44559	-1.00852	
	2.50756	-0.12305	-1.01147	-1.37831	
	-0.12305	1.33231	-0.26141	-0.44744	
	-1.01148	-0.26142	1.53182	0.44559	
	-1.37832	-0.44744	0.44559	2.00852	

Таблица II. 19

Обращение матрицы при помощи „склеенной“ схемы
метода пополнения

	$\begin{vmatrix} 1 \\ 0.42 \\ 0.54 \\ 0.66 \end{vmatrix}$	0	0	0	1
	$\begin{vmatrix} 0 \\ 0.42 \\ 0.54 \\ 0.66 \end{vmatrix}$	0	0.32	0.44	1.18
	$\begin{vmatrix} 0.42 \\ 0.54 \\ 0.66 \end{vmatrix}$	0.32	0	0.22	1.08
	$\begin{vmatrix} 0.54 \\ 0.66 \end{vmatrix}$	0.44	0.22	0	1.32
1	0	0.42	0.54	0.66	
	0	0.42	0.54	0.66	1.62
	1	-0.42	-0.54	-0.66	-0.62
	0	1	0	0	1
	0.54	0.09320	-0.29160	-0.13640	0.2052
0.66	0.16280	-0.13640	-0.43560	0.2508	
0.82360	0.42	-0.17640	0.09320	0.16280	
	0.50996	-0.21418	0.11316	0.19767	0.60661
	1.21418	-0.50996	-0.49247	-0.57698	-0.36523
	-0.50996	1.21418	-0.11316	-0.19767	0.39339
	0	0	1	0	1
0.57698	0.19767	-0.15482	-0.46778	0.15205	
0.69785	0.49247	0.11316	-0.30215	-0.15482	
	0.70570	0.16216	-0.43297	-0.22185	0.21304
	1.56172	-0.43010	-0.70569	-0.68623	-0.26030
	-0.43010	1.23253	-0.16215	-0.22277	0.41751
	-0.70570	-0.16216	1.43297	0.22185	0.78696
0	0	0	1	1	
0.49788	0.68624	0.22278	-0.22185	-0.50213	
	1.37832	0.44746	-0.44559	-1.00854	0.37165
	2.50756	-0.12304	-1.01147	-1.37832	-0.00527
	-0.12305	1.33221	-0.26141	-0.44746	0.50029
	-1.01148	-0.26143	1.53182	0.44559	0.70450
-1.37832	-0.44746	0.44559	2.00854	0.62835	

матрицы C_{k-1} и т. д. Элементы же первых $k-1$ строк матрицы C_{k-1} вообще не понадобятся в дальнейших вычислениях, и вычислять их нет необходимости. С другой стороны, для матрицы A_{k-1}^{-1} достаточно вычислять элементы первых $k-1$ строк, так как последующие строки совпадают со строками единичной матрицы. Поэтому, целесообразно ввести в рассмотрение „склеенные“ из A_{k-1}^{-1} и C_{k-1} матрицы S_{k-1} и \tilde{S}_{k-1} , из которых первая S_{k-1} имеет первые $k-1$ строки совпадающими с первыми $k-1$ строками матрицы A_{k-1}^{-1} и последние $n-k+1$ строк из C_{k-1} , вторая — первые k строк из A_{k-1}^{-1} , последние $n-k$ строк из C_{k-1} ; матрица \tilde{S}_{k-1} получается из S_{k-1} заменой k -й строки на k -ю строку единичной матрицы. Очевидно, что элементами k -й строки матрицы S_{k-1} будут числа $c_{kj}^{(k-1)}$.

В силу формул (4) и (5) матрица S_k получается из матрицы \tilde{S}_{k-1} по формулам

$$\sigma_j^{(k)} = \tilde{\sigma}_j^{(k-1)} - \frac{c_{kj}^{(k-1)}}{1 + c_{kk}^{(k-1)}} \sigma_k^{(k-1)}. \quad (6)$$

Здесь $\sigma_j^{(k)}$ обозначает j -й столбец матрицы S_k , $\tilde{\sigma}_j^{(k-1)}$ обозначает j -й столбец матрицы \tilde{S}_{k-1} . За исходную матрицу S_0 нужно взять, очевидно, $S_0 = A - E$. Матрица $S_n = \tilde{S}_n$ равна A^{-1} .

В табл. II.19 дано обращение матрицы по этому варианту методом пополнения. Поясним заполнение таблицы II.19. В строках 1—4, 7—10, 13—16, 19—22 последовательно записываются матрицы S_0 , \tilde{S}_1 , \tilde{S}_2 и \tilde{S}_3 . Строки, совпадающие со строками единичной матрицы, целесообразно вписывать в схему заранее. В строках 5, 11, 17 и 23 записываются k -е строки матриц S_{k-1} , в строках 6, 12, 18 и 24 соответствующие множители $\frac{c_{kj}^{(k-1)}}{1 + c_{kk}^{(k-1)}}$. В тех же строках, слева от схемы, записываются знаменатели $1 + c_{kk}^{(k-1)}$. Наконец, в строках 25—28 записывается матрица $S_n = A^{-1}$.

Для контроля к каждой матрице \tilde{S}_k пристраивается столбец $\tilde{\sigma}_{n+1}^{(k)}$, составленный из строчных сумм. Суммируются также элементы строк, составленных из множителей $\frac{c_{kj}^{(k-1)}}{1 + c_{kk}^{(k-1)}}$. Контрольные столбцы (за исключением одного элемента, всегда равного единице) связаны соотношением

$$\tilde{\sigma}_{n+1}^{(k)} = \sigma_{n+1}^{(k-1)} - \tilde{\sigma}_k^{(k-1)} \sum \frac{\sigma_{kj}^{(k-1)}}{1 + c_{kk}^{(k-1)}}.$$

ГЛАВА III

ИТЕРАЦИОННЫЕ МЕТОДЫ РЕШЕНИЯ СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ

Перейдем теперь к описанию итерационных методов решения систем линейных уравнений. Эти методы дают решение системы в виде предела последовательности некоторых векторов, построение которых осуществляется посредством единообразного процесса, называемого процессом итераций.

В современной литературе описано большое количество итерационных методов, основанных на различных принципах. Как правило, вычислительные схемы этих методов просты и удобны при использовании вычислительной техники. Однако каждый итерационный процесс имеет свою ограниченную область применимости, так как, во-первых, процесс итераций может оказаться расходящимся для данной системы и, во-вторых, сходимость процесса может быть настолько медленной, что практически оказывается невозможным достичнуть удовлетворительной близости к решению.

Отметим, что хотя задача обращения матрицы эквивалентна решению n частных систем с одной и той же матрицей, итерационные методы редко употребляются для этой цели.

Настоящая глава посвящена описанию общих принципов построения итерационных процессов, детальному рассмотрению простейшего итерационного процесса — метода последовательных приближений в его различных модификациях и координатным релаксационным методам.

§ 29. Принципы построения итерационных процессов

Основные итерационные процессы для решения линейных систем могут быть описаны посредством следующей общей схемы.

Пусть дана система линейных уравнений

$$AX = F \quad (1)$$

с неособенной матрицей A . Строится последовательность векторов $X^{(1)}, X^{(2)}, \dots, X^{(k)}, \dots$ по рекуррентным формулам

$$X^{(k)} = X^{(k-1)} + H^{(k)}(F - AX^{(k-1)}), \quad (2)$$

где $H^{(1)}, H^{(2)}, \dots$ — некоторая последовательность матриц, $X^{(0)}$ — начальное приближение, вообще говоря, произвольное. Различный выбор последовательности матриц $H^{(k)}$ приводит к различным итерационным процессам.

Итерационные процессы, протекающие по формуле (2), обладают тем свойством, что для каждого из них точное решение X^* является неподвижной точкой. Это значит, что если за начальное приближение $X^{(0)}$ взято X^* , то все последующие приближения будут также равны X^* .

Обратно, всякий итерационный процесс, для которого X^* является неподвижной точкой, протекающий по формулам

$$X^{(k)} = C^{(k)} X^{(k-1)} + Z^{(k)}, \quad (3)$$

где $C^{(k)}$ последовательность матриц, $Z^{(k)}$ последовательность векторов, может быть представлен в виде (2). Действительно, для X^* имеем

$$X^* = C^{(k)} X^* + Z^{(k)},$$

откуда

$$\begin{aligned} X^{(k)} &= X^* + C^{(k)} (X^{(k-1)} - X^*) = X^{(k-1)} + (C^{(k)} - E) (X^{(k-1)} - X^*) = \\ &= X^{(k-1)} + (E - C^{(k)}) A^{-1} A (X^* - X^{(k-1)}) = \\ &= X^{(k-1)} + H^{(k)} (F - AX^{(k-1)}) \end{aligned}$$

при

$$H^{(k)} = (E - C^{(k)}) A^{-1}.$$

Нетрудно дать необходимые и достаточные условия для того, чтобы итерационный процесс (2) сходился к решению при любом начальном векторе.

Действительно,

$$\begin{aligned} X^* - X^{(k)} &= X^* - X^{(k-1)} - H^{(k)} (AX^* - AX^{(k-1)}) = \\ &= (E - H^{(k)} A) (X^* - X^{(k-1)}). \end{aligned}$$

Отсюда

$$X^* - X^{(k)} = (E - H^{(k)} A) (E - H^{(k-1)} A) \dots (E - H^{(1)} A) (X^* - X^{(0)}).$$

Для того, чтобы $X^* - X^{(k)} \rightarrow 0$ при любом начальном векторе $X^{(0)}$, необходимо и достаточно (см. § 13, п. 1), чтобы матрица

$$T^{(k)} = (E - H^{(k)} A) (E - H^{(k-1)} A) \dots (E - H^{(1)} A)$$

стремилась к нулю, для чего, в свою очередь, достаточно, чтобы любая норма матрицы $T^{(k)}$ стремилась к нулю. Конечно, выведенное условие дает лишь общую точку зрения для построения условий сходимости конкретных итерационных процессов.

Простейшими среди итерационных процессов являются стационарные итерационные процессы, в которых матрицы $H^{(k)}$ не зависят от номера шага k . В частности, при $H^{(k)} = E$ получается классический процесс последовательных приближений. Любой стационарный процесс с $H \neq E$ можно рассматривать как процесс последовательных приближений, примененный к равносильной системе

$$HAX = HF,$$

так сказать „подготовленной“ к применению метода последовательных приближений. Конечно, осуществлять такую подготовку на самом деле обычно нет необходимости, и такого рода рассмотрение стационарных процессов лишь дает удобное средство для их теоретического исследования.

Близкими к стационарным итерационным процессам являются циклические, в которых матрицы $H^{(k)}$ периодически повторяются через некоторое число p шагов. Ясно, что из каждого циклического процесса можно получить равносильный ему стационарный, принимая за один шаг стационарного процесса результат применения полного цикла из p шагов исходного циклического процесса.

Нестационарные итерационные процессы, в свою очередь, можно подразделить на два типа. Это, во-первых, нестационарные итерационные процессы в буквальном смысле, когда изменение матрицы $H^{(k)}$ осуществляется на каждом шагу. Во-вторых, сюда можно включить стационарные процессы с ускорением сходимости посредством замены, время от времени, стационарной матрицы H , определяющей процесс, на некоторые, специальным образом подобранные, матрицы $H^{(k)}$.

Выбор матрицы H для стационарного процесса и матриц $H^{(k)}$ для нестационарного может осуществляться многими различными способами на основании различных принципов.

Возможно построение матриц $H^{(k)}$ так, чтобы итерационный процесс сходился к решению для возможно более широкого класса систем уравнений. Возможна противоположная точка зрения, в силу которой при построении матриц $H^{(k)}$ максимально используются частные особенности данной системы для получения итерационного процесса, обладающего быстрой сходимостью. Естественно, что для применения итерационного процесса, построенного исходя из последнего принципа, нужно располагать возможно большей информацией о матрице коэффициентов системы, в частности о расположении ее собственных значений.

Важным принципом построения итерационных процессов является принцип релаксации. Под этим понимается принцип выбора матриц $H^{(k)}$ из некоторого заранее очерченного класса матриц так, чтобы

на каждом шагу процесса уменьшалась какая-либо величина, характеризующая точность решения системы.

Среди релаксационных методов наиболее разработаны координатные, в которых матрицы $H^{(k)}$ подобраны так, что на каждом шагу меняются одна или несколько компонент последовательных приближений, и градиентные, в которых матрицы $H^{(k)}$ являются скалярными.

О точности приближенного решения X системы $AX = F$ естественно судить по величине (в том или другом смысле) вектора ошибки $Y = X^* - X$. Однако вектор ошибки не может быть вычислен без знания точного решения системы и может лишь оцениваться. Вектором, характеризующим точность приближенного решения X системы $AX = F$, может служить также вектор невязки (невязка) $r = F - AX$. Ясно, что $r = AY$. Таким образом, релаксация может быть построена на уменьшении любой нормы каждого из этих векторов.

При положительно-определеных матрицах A удобной мерой точности является так называемая функция ошибки

$$f(X) = (AY, Y) = (Y, r) = (A^{-1}r, r).$$

В силу положительной определенности A всегда $f(X) \geq 0$, причем $f(X) = 0$ только при $X = X^*$. Ясно, что

$$\begin{aligned} f(X) &= (X^* - X, F - AX) = (X^*, F) - (X, F) - (X^*, AX) + (X, AX) = \\ &= (AX, X) - 2(X, F) + (X^*, F). \end{aligned}$$

Функция ошибки не может быть вычислена, если не известно точное решение. Однако ее значения лишь постоянным слагаемым отличаются от значений функционала

$$f_0(X) = (AX, X) - 2(X, F),$$

которые могут быть вычислены без знания X^* . Поэтому мы можем судить об убывании функции ошибки сравнивая соответствующие значения функционала $f_0(X)$.

Другим важным принципом построения итерационных процессов является принцип последовательного „подавления компонент“ вектора ошибки в разложении его по собственным векторам матрицы коэффициентов системы.

Релаксационным градиентным методам будет посвящена глава VII. В главе IX будут рассмотрены методы, основанные на идеи подавления компонент.

§ 30. Метод последовательных приближений

Наиболее простым итерационным процессом является процесс последовательных приближений.

Под процессом последовательных приближений понимается следующий итерационный процесс. Система уравнений

$AX = F$ записывается в виде

$$X = BX + G, \quad (1)$$

где $B = E - A$, $G = F$ и последовательные приближения вычисляются по формуле

$$X^{(k)} = BX^{(k-1)} + G, \quad (2)$$

начиная с некоторого исходного приближения $X^{(0)}$, которое может выбираться, вообще говоря, произвольно. Очевидно, что процесс последовательных приближений является частным случаем общего итерационного процесса (2) § 29; именно это будет стационарный процесс, в котором $H = E$. Действительно,

$$X^{(k)} = (E - A)X^{(k-1)} + G = X^{(k-1)} + (F - AX^{(k-1)}).$$

Легко дать формулу для выражения $X^{(k)}$ непосредственно через начальное приближение $X^{(0)}$. Именно

$$X^{(k)} = B^k X^{(0)} + (E + B + \dots + B^{k-1})G. \quad (3)$$

Действительно, при $k = 1$ это верно, а при остальных k формула легко проверяется методом математической индукции.

Отметим, что если процесс последовательных приближений сходится, то он сходится к решению системы. Действительно, если $X^{(k)} \rightarrow X^*$, то предельный переход в равенстве

$$X^{(k)} = BX^{(k-1)} + G$$

дает $X^* = BX^* + G$, т. е. X^* удовлетворяет данной системе.

Теорема 30.1. Для сходимости процесса последовательных приближений при любом начальном векторе $X^{(0)}$ необходимо и достаточно, чтобы все собственные значения матрицы B были бы по модулю меньше единицы.

Доказательство. Пусть $X^{(k)} \rightarrow X^*$. Тогда, как мы видели, X^* есть решение системы и, следовательно, $X^* - X^{(k)} = B(X^* - X^{(k-1)}) = \dots = B^k(X^* - X^{(0)})$, откуда $B^k(X^* - X^{(0)}) \rightarrow 0$. Так как это должно иметь место при любом векторе $X^{(0)}$, необходимо, чтобы $B^k \rightarrow 0$, для чего, в свою очередь, необходимо, чтобы все собственные значения матрицы B были меньше единицы по модулю [теорема 13.2].

Достаточность условия непосредственно вытекает из формулы (3), ибо $B^k \rightarrow 0$ и $E + B + \dots + B^{k-1} \rightarrow (E - B)^{-1} = A^{-1}$, если все собственные значения матрицы B меньше единицы по модулю.

Так как условие теоремы 30.1 трудно проверяется, судить о сходимости процесса последовательных приближений лучше при помощи достаточных признаков, связанных непосредственно с элементами матрицы B . Некоторые достаточные признаки вытекают из теоремы 30.2.

Теорема 30.2. Для того чтобы процесс последовательных приближений сходился, достаточно, чтобы какая-либо норма матрицы B была меньше единицы.

Доказательство. Действительно, если $\|B\| < 1$, то все собственные значения матрицы B меньше единицы и потому на основании теоремы 30.1 процесс последовательных приближений сходится.

Дадим теперь оценки быстроты сходимости процесса последовательных приближений в терминах нормы. При этом выбор нормы векторов совершенно безразличен, но норма матриц должна быть согласована с выбранной нормой векторов.

Теорема 30.3. Если $\|B\| < 1$, то

$$\|X^* - X^{(k)}\| \leq \|B\|^k \|X^{(0)}\| + \frac{\|G\| \|B\|^k}{1 - \|B\|}. \quad (4)$$

Доказательство. Имеем

$$\begin{aligned} \|X^* - X^{(k)}\| &= \|(E - B)^{-1} G - (E + B + \dots + B^{k-1}) G - B^k X^{(0)}\| \leq \\ &\leq \|(E - B)^{-1} G - (E + B + \dots + B^{k-1}) G\| + \|B^k X^{(0)}\| \leq \\ &\leq \|(E - B)^{-1} G - (E + B + \dots + B^{k-1}) G\| + \|B^k\| \|X^{(0)}\| \leq \\ &\leq \|B\|^k \|X^{(0)}\| + \frac{\|G\| \|B\|^k}{1 - \|B\|}. \end{aligned}$$

Часто бывает важно сравнить точность двух последовательных приближений, т. е. сравнить величины $\|X^* - X^{(k)}\|$ и $\|X^* - X^{(k-1)}\|$. Такое сравнение можно проводить на основании следующей теоремы.

Теорема 30.4. $\|X^* - X^{(k)}\| \leq \|B\| \|X^* - X^{(k-1)}\|$.

Доказательство. Действительно, из равенств

$$X^* = BX^* + G, \quad X^{(k)} = BX^{(k-1)} + G$$

следует, что

$$X^* - X^{(k)} = B(X^* - X^{(k-1)}).$$

Отсюда

$$\|X^* - X^{(k)}\| = \|B(X^* - X^{(k-1)})\| \leq \|B\| \|X^* - X^{(k-1)}\|. \quad (5)$$

Введенные нами в § 13 векторные нормы (кубическая, октоэдрическая и сферическая) и согласованные с ними нормы матриц дают следующие легко проверяемые достаточные признаки сходимости процесса последовательных приближений и оценки быстроты его сходимости.

I. Если $\sum_{j=1}^n |b_{ij}| \leq \mu < 1$ при $i = 1, 2, \dots, n$, то процесс последовательных приближений сходится, причем

$$|x_i - x_i^{(k)}| \leq \mu \max_j |x_j - x_j^{(k-1)}|, \quad (6)$$

где $X^* = (x_1, \dots, x_n)'$ и $X^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})'$.

II. Если $\sum_{i=1}^n |b_{ij}| \leq v < 1$ при $j = 1, 2, \dots, n$, то процесс последовательных приближений сходится, причем

$$\sum_{i=1}^n |x_i - x_i^{(k)}| \leq v \sum_{i=1}^n |x_i - x_i^{(k-1)}|. \quad (7)$$

III. Если $\sum_{i,k=1}^n b_{ik}^2 \leq p < 1$, то процесс последовательных приближений сходится и

$$\sqrt{\sum_{i=1}^n (x_i - x_i^{(k)})^2} \leq \sqrt{p} \sqrt{\sum_{i=1}^n (x_i - x_i^{(k-1)})^2}. \quad (8)$$

Укажем еще на один путь построения достаточных признаков сходимости процесса последовательных приближений.

В уравнении

$$X = BX + G \quad (1)$$

с матрицей

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix}$$

введем новые неизвестные $x_i = p_i z_i$, где p_i некоторые положительные числа.

Тогда система (1) превратится в систему

$$p_i z_i = \sum_{j=1}^n b_{ij} p_j z_j + g_i$$

или

$$z_i = \sum_{j=1}^n b_{ij} \cdot \frac{p_j}{p_i} z_j + \frac{1}{p_i} g_i. \quad (9)$$

Очевидно, что компоненты последовательных приближений $X^{(k)}$ для системы (1) и $Z^{(k)}$ для системы (9) тоже связаны соотношениями $x_i^{(k)} = p_i z_i^{(k)}$, если только эти соотношения имеют место для исходных приближений $X^{(0)}$ и $Z^{(0)}$. Поэтому процессы последовательных приближений для систем (1) и (9) сходятся или расходятся одновременно, и, следовательно, всякое достаточное условие сходимости процесса последовательных приближений для системы (9) является вместе с тем достаточным условием сходимости для системы (1).

Таким образом, если можно указать такие положительные числа p_1, \dots, p_n , что выполняется одно из условий

- 1) $\sum_{j=1}^n |b_{ij}| \cdot \frac{p_j}{p_i} < 1 \quad \text{при } i = 1, \dots, n$
- 2) $\sum_{i=1}^n |b_{ij}| \cdot \frac{p_j}{p_i} < 1 \quad \text{при } j = 1, \dots, n$ (10)
- 3) $\sum_{i,j=1}^n \frac{b_{ij}^2 p_j^2}{p_i^2} < 1,$

то процесс последовательных приближений для системы (1) сходится.

Замечание. При практическом вычислении итераций мы можем поступать двумя способами.

- 1) Положим $X^{(0)} = G$. Тогда

$$X^{(k)} = G + BG + \dots + B^k G.$$

Для вычисления $X^{(k)}$ мы вычисляем последовательно векторы $G, BG, \dots, B^k G$ и находим их сумму. Это удобно вследствие единобразия процесса вычисления, а также потому, что каждое последующее слагаемое является лишь поправкой к сумме предыдущих. Недостатком этого способа является возможное накопление ошибок от округления с возрастанием числа слагаемых.

- 2) Вычисление ведется непосредственно по формулам

$$X^{(k)} = BX^{(k-1)} + G.$$

Здесь каждое приближение является как бы исходным и поэтому нет необходимости на первых шагах процесса проводить вычисления с большой точностью; возникающие ошибки впоследствии сглаживаются.

В качестве примера найдем решение системы

$$\begin{aligned} 0.78x_1 - 0.02x_2 - 0.12x_3 - 0.14x_4 &= 0.76 \\ -0.02x_1 + 0.86x_2 - 0.04x_3 + 0.06x_4 &= 0.08 \\ -0.12x_1 - 0.04x_2 + 0.72x_3 - 0.08x_4 &= 1.12 \\ -0.14x_1 + 0.06x_2 - 0.08x_3 + 0.74x_4 &= 0.68. \end{aligned} \quad (11)$$

Решая эту систему по схеме единственного деления, находим

$$x_1 = 1.534965$$

$$x_2 = 0.122010$$

$$x_3 = 1.975156$$

$$x_4 = 1.412955.$$

Для применения процесса последовательных приближений приведем систему к виду $X = BX + G$, положив $B = E - A$. Получим

$$\begin{aligned}x_1 &= 0.22x_1 + 0.02x_2 + 0.12x_3 + 0.14x_4 + 0.76 \\x_2 &= 0.02x_1 + 0.14x_2 + 0.04x_3 - 0.06x_4 + 0.08 \\x_3 &= 0.12x_1 + 0.04x_2 + 0.28x_3 + 0.08x_4 + 1.12 \\x_4 &= 0.14x_1 - 0.06x_2 + 0.08x_3 + 0.26x_4 + 0.68.\end{aligned}\quad (12)$$

Нетрудно видеть, что достаточные условия сходимости процесса последовательных приближений выполнены.

Таблица III. I

Схема вычислений по формуле $X^{(k)} = \sum_{l=0}^k B^l G$

1) $B^l G$, $l = 0, \dots, 14$.

G	0.76	0.08	1.12	0.68	2.64
BG	0.3984	0.0304	0.4624	0.3680	1.2592
	0.195264	0.008640	0.207936	0.186624	0.598464
	0.09421056	0.00223488	0.09692928	0.09197568	0.28535040
	0.04527913	0.00055572	0.04589292	0.04472340	0.13645117
	0.02174095	0.00013570	0.02188361	0.02160525	0.06536551
	0.01043649	0.00003285	0.01047017	0.01040364	0.03134316
	0.00500961	0.00000792	0.00501763	0.00500170	0.01503686
	0.00240463	0.00000190	0.00240654	0.00240272	0.00721580
	0.00115422	0.00000046	0.00115468	0.00115376	0.00346312
	0.00055403	0.00000011	0.00055414	0.00055392	0.00166219
	0.00026593	0.00000003	0.00026596	0.00026591	0.00079783
	0.00012765	0.00000001	0.00012765	0.00012764	0.00038295
	0.00006127		0.00006127	0.00006127	0.00018381
$B^{14}G$	0.00002941		0.00002941	0.00002941	0.00008823

2) $X^{(k)} = \sum_{l=0}^k B^l G$ для $k = 12, 13, 14$.

$X^{(12)}$	1.53484720	0.12200958	1.97503858	1.41283762	
$X^{(13)}$	1.53490847	0.12200958	1.97509985	1.41289889	
$X^{(14)}$	1.53493788	0.12200958	1.97512926	1.41292830	

Для сравнения хода итерационного процесса в разных вариантах проведем его тремя способами. Именно:

1) Вычисление последовательных приближений производим по формуле $X^{(k)} = \sum_{l=0}^k B^l G$ (см. табл. III. 1).

2) Вычисление последовательных приближений производим по формуле $X^{(k)} = BX^{(k-1)} + G$ при $X^{(0)} = G$ (см. табл. III. 2).

3) Снова вычисляем $X^{(k)} = BX^{(k-1)} + G$ при $X^{(0)} = (1, 0, 0, 0)'$ (см. табл. III. 3).

Поясним табл. III. 1. Первая часть таблицы содержит компоненты последовательно вычисляемых векторов $B^l G$. Последний столбец является контрольным. В нем записываются числа $\sum_{j=1}^n c_j x_j^{(k)}$, где $c_j = -\sum_{i=1}^n b_{ij}$ (числа c_j должны быть вычислены заранее и приписаны

Таблица III. 2

Вычисление приближений по формуле

$$X^{(k)} = BX^{(k-1)} + G; X^{(0)} = G$$

$X^{(0)}$	0.76	0.08	1.12	0.68	2.64
$X^{(1)}$	1.1584	0.1104	1.5824	1.0480	3.8992
$X^{(2)}$	1.3537	0.1190	1.7903	1.2346	4.4977
$X^{(3)}$	1.4479	0.1213	1.8873	1.3266	4.7830
$X^{(4)}$	1.4932	0.1218	1.9332	1.3713	4.9195
$X^{(5)}$	1.5149	0.1220	1.9551	1.3929	4.9849
$X^{(6)}$	1.5253	0.1220	1.9655	1.4033	5.0162
$X^{(7)}$	1.5303	0.1220	1.9705	1.4083	5.0312
$X^{(8)}$	1.53273	0.12201	1.97292	1.41072	5.03838
$X^{(9)}$	1.53389	0.12201	1.97408	1.41188	5.04187
$X^{(10)}$	1.53445	0.12201	1.97464	1.41244	5.04354
$X^{(11)}$	1.53472	0.12201	1.97491	1.41271	5.04434
$X^{(12)}$	1.53485	0.12201	1.97504	1.41284	5.04473
$X^{(13)}$	1.534910	0.122010	1.975101	1.412900	5.044920
$X^{(14)}$	1.5349385	0.1220096	1.9751299	1.4129289	5.0450069

в качестве дополнительной строки матрицы B). Но очевидно, что $\sum_{j=1}^n c_j x_j^{(k)} = \sum_{j=1}^n x_j^{(k+1)}$, так что элементы контрольного столбца равны суммам остальных элементов, лежащих в той же строке. Вторая часть таблицы дает приближенное решение системы, которое мы получаем, суммируя соответствующие компоненты вычисленных векторов.

Из сравнения табл. III.1, III.2 и III.3 с результатом, полученным по схеме единственного деления, мы видим, что сходимость процесса во всех трех вариантах приблизительно одинаковая; 14-е приближение дает в данном примере результат, верный с точностью до единицы четвертого знака.

Таблица III.3

Вычисление приближений по формуле

$$X^{(k)} = BX^{(k-1)} + G, X^{(0)} = (1, 0, 0, 0)'$$

$X^{(0)}$	1	0	0	0	1
$X^{(1)}$	0.98	0.10	1.24	0.82	3.14
$X^{(2)}$	1.2412	0.1140	1.6544	1.1236	4.1332
$X^{(12)}$	1.534769	0.122010	1.974961	1.412761	5.04451
$X^{(13)}$	1.534871	0.122010	1.975063	1.412862	5.044805
$X^{(14)}$	1.534920	0.122010	1.975111	1.412910	5.044951

Сходимость процесса последовательных приближений можно сильно улучшить, применяя различные приемы ускорения. Процесс последовательных приближений с применением приемов ускорения сходимости в большинстве случаев укладывается в общую схему итерационных процессов с нарушением стационарности. Целесообразный выбор приемов ускорения требует предварительной информации о расположении собственных значений матрицы. Мы вернемся к этому вопросу в гл. IX.

§ 31. Подготовка системы линейных уравнений к виду, удобному для применения метода последовательных приближений. Метод простой итерации

Условия сходимости метода последовательных приближений требуют, чтобы матрица коэффициентов системы $AX=F$ была, в том или ином смысле, близка к единичной матрице. Если это условие не выполнено или „плохо выполнено“, систему целесообразно пред-

вариально подготовить к применению метода последовательных приближений. Подготовка состоит в переходе от данной системы $AX = F$ к равносильной системе

$$HAX = HF,$$

где H некоторая неособенная матрица, которая выбирается так, чтобы матрица HA была бы близка к единичной, т. е. матрица H была бы близка к A^{-1} .

Применение метода последовательных приближений к подготовленной системе, как было указано, равносильно применению стационарного итерационного процесса

$$X^{(k)} = X^{(k-1)} + H(F - AX^{(k-1)})$$

к исходной системе.

Выбор матрицы H может быть осуществлен с использованием частных особенностей данной системы. Рассмотрим некоторые наиболее употребительные способы подготовки, использующие лишь довольно поверхностные сведения о матрице коэффициентов.

Пусть матрица A положительно определена. Тогда система $AX = F$ всегда может быть подготовлена к виду, в котором метод последовательных приближений будет сходящимся. Действительно, вычислив, например, первую норму μ матрицы A , мы получим, что все собственные значения матрицы A заключены в открытом интервале $(0, \mu)$. Положим

$$H = \frac{2}{\mu} E. \quad (1)$$

Система $AX = F$ преобразуется к виду

$$X = \left(E - \frac{2}{\mu} A\right) X + \frac{2}{\mu} F = BX + G. \quad (2)$$

Собственные значения матрицы $B = E - \frac{2}{\mu} A$ будут заключены в открытом интервале $(-1, 1)$ и, следовательно, метод последовательных приближений будет сходящимся.

В качестве примера возьмем систему (9) § 23. Здесь $\mu = 2.62$. Выполнив вычисления, получим

$$B = \begin{bmatrix} 0.23664122 & -0.32061069 & -0.41221374 & 0.50381679 \\ -0.32061069 & 0.23664122 & -0.24427481 & -0.33587786 \\ -0.41221374 & -0.24427481 & 0.23664122 & -0.16793893 \\ 0.50381679 & -0.33587786 & -0.16793893 & 0.23664122 \end{bmatrix},$$

$$G = \begin{bmatrix} 0.22900763 \\ 0.38167939 \\ 0.53435115 \\ 0.68702290 \end{bmatrix}. \quad (3)$$

В § 23 было найдено, что решение системы $X = (-1.2577938, 0.0434873, 1.0391663, 1.4823929)'$.

В табл. III. 4 приведем несколько последовательных приближений, положив $X^{(0)} = G$.

Таблица III. 4

Вычисление приближений по формуле $X^{(k)} = BX^{(k-1)} + G$.

$X^{(0)}$	0.22900763	0.38167939	0.53435115	0.68702290	1.83206107
$X^{(1)}$	-0.4055708	0.0372939	0.3577880	0.5162869	0.5057980
$X^{(20)}$	-1.2354227	0.0473327	1.0287985	1.4668834	1.3075919
$X^{(25)}$	-1.2505812	0.0442080	1.0348124	1.4760729	1.3045121
$X^{(40)}$	-1.2574335	0.0435403	1.0389819	1.4821204	1.3072091
$X^{(50)}$	-1.2577475	0.0434939	1.0391421	1.4823572	1.3072456
$X^{(74)}$	-1.2577935	0.0434874	1.0391661	1.4823926	1.3072526

Из приведенных результатов видно, что метод последовательных приближений в данном случае сходится довольно медленно.

В прикладных вопросах часто встречаются системы, в которых диагональные элементы матрицы A значительно преобладают над остальными элементами матрицы. В этом случае подготовка системы осуществляется так.

Перепишем систему $AX = F$ в развернутом виде

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= f_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= f_2 \\ \vdots &\quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= f_n. \end{aligned} \tag{4}$$

Поделим каждое уравнение системы (4) на диагональный элемент. Мы получим систему

$$\begin{aligned} x_1 + \frac{a_{12}}{a_{11}}x_2 + \dots + \frac{a_{1n}}{a_{11}}x_n &= \frac{f_1}{a_{11}} \\ \frac{a_{21}}{a_{22}}x_1 + x_2 + \dots + \frac{a_{2n}}{a_{22}}x_n &= \frac{f_2}{a_{22}} \\ \vdots &\quad \vdots \\ \frac{a_{n1}}{a_{nn}}x_1 + \frac{a_{n2}}{a_{nn}}x_2 + \dots + x_n &= \frac{f_n}{a_{nn}} \end{aligned}$$

или в матричной записи

$$X = BX + G, \quad (5)$$

где

$$B = \begin{bmatrix} 0 & -\frac{a_{12}}{a_{11}} & \dots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & \dots & -\frac{a_{2n}}{a_{22}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & \dots & 0 \end{bmatrix}, \quad G = \begin{bmatrix} \frac{f_1}{a_{11}} \\ \frac{f_2}{a_{22}} \\ \vdots \\ \frac{f_n}{a_{nn}} \end{bmatrix}. \quad (6)$$

Для применения процесса итерации нет необходимости на самом деле делать преобразование системы (4) в систему (5). Последовательные приближения можно вычислять по формулам:

$$\begin{aligned} a_{11}x_1^{(k)} &= f_1 - a_{12}x_2^{(k-1)} - \dots - a_{1n}x_n^{(k-1)} \\ \vdots &\quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{nn}x_n^{(k)} &= f_n - a_{n1}x_1^{(k-1)} - \dots - a_{n,n-1}x_{n-1}^{(k-1)}. \end{aligned} \quad (7)$$

Описанная модификация процесса последовательных приближений имеет название метода простой итерации или метода Якоби.

Упомянутое преобразование системы (4) в систему (5), очевидно, равносильно умножению системы (4) слева на матрицу

$$H = \begin{bmatrix} \frac{1}{a_{11}} & 0 \\ & \frac{1}{a_{22}} \\ 0 & & \frac{1}{a_{nn}} \end{bmatrix}.$$

Таким образом, $H = D^{-1}$, где D —диагональная матрица $[a_{11}, \dots, a_{nn}]$. Отсюда следует, что необходимое и достаточное условие сходимости процесса простой итерации состоит в том, что все собственные значения матрицы $B = E - D^{-1}A$ по модулю меньше единицы.

Это условие можно представить в другой форме. Именно,

$$\begin{aligned} |B - tE| &= |E - D^{-1}A - tE| = \\ &= |D^{-1}| |D - A - tD| = (-1)^n |D^{-1}| |A - D + tD|. \end{aligned}$$

Таким образом, для сходимости процесса простой итерации необходимо и достаточно, чтобы все корни уравнения

$$|A - D + tD| = \begin{vmatrix} a_{11}t & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}t & \dots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn}t \end{vmatrix} = 0 \quad (8)$$

по модулю были бы меньше единицы.

В случае, если матрица A коэффициентов системы симметричная (или эрмитова) с положительными диагональными элементами, необходимо и достаточно условию сходимости метода простой итерации можно придать следующую легко проверяемую форму (Ю. М. Гаврилов [2]).

Для того чтобы метод простой итерации для системы $A\bar{X} = F$ с симметричной матрицей A , имеющей положительные диагональные элементы, сходился, необходимо и достаточно, чтобы матрицы A и $A = 2D - A$ (отличающиеся друг от друга знаками недиагональных элементов) были бы положительно-определенными.

Действительно, в этом случае, в силу положительной определенности матрицы D , все собственные значения матрицы $D^{-1}A$ вещественны (теорема 11.14). Поэтому для сходимости процесса необходимо и достаточно, чтобы собственные значения матрицы $D^{-1}A = E - (E - D^{-1}A)$ заключались в интервале $(0, 2)$, т. е. чтобы собственные значения матриц $D^{-1}A$ и $2E - D^{-1}A$ были положительны. Но в силу теоремы 11.16 это равносильно положительной определенности матриц A и $2D - A$.

Введенные в § 30 достаточные условия сходимости процесса последовательных приближений, будучи применены к системе (5), дают следующие достаточные условия сходимости метода простой итерации:

- I. $\sum_{j=1}^n' \left| \frac{a_{ij}}{a_{ii}} \right| < 1 \quad (i = 1, 2, \dots, n)$
- II. $\sum_{i=1}^n' \left| \frac{a_{ij}}{a_{ii}} \right| < 1 \quad (j = 1, 2, \dots, n)$
- III. $\sum_{i,j=1}^n' \left(\frac{a_{ij}}{a_{ii}} \right)^2 < 1.$

Здесь знак штрих у суммы показывает, что при суммировании опускаются значения $i = j$.

Если воспользоваться обобщенными признаками (10) § 30, положив $p_i = \frac{1}{|a_{ii}|}$, то мы получим еще следующие достаточные признаки сходимости простой итерации:

- I. $\sum_{j=1}^n' \left| \frac{a_{ij}}{a_{jj}} \right| < 1 \quad (i = 1, 2, \dots, n)$
- II. $\sum_{i=1}^n' \left| \frac{a_{ij}}{a_{jj}} \right| < 1 \quad (j = 1, 2, \dots, n)$ (10)
- III. $\sum_{i,j=1}^n' \left(\frac{a_{ij}}{a_{jj}} \right)^2 < 1.$

Допустим теперь, что задана система

$$AX = F,$$

в которой преобладание главной диагонали не имеет места.

Подбор вспомогательной матрицы H может быть осуществлен, например, грубым обращением матрицы A по методу Гаусса.

Часто оказывается целесообразным в качестве матрицы H взять матрицу, обратную к матрице

$$\begin{bmatrix} a_{11} & a_{12} & & 0 \\ a_{21} & a_{22} & & \\ & & a_{33} & a_{34} \\ & & a_{43} & a_{44} \\ & 0 & & \ddots \end{bmatrix}.$$

Обращение такой матрицы не представляет труда, ибо сводится к обращению матриц второго порядка. Именно,

$$H = \begin{bmatrix} \frac{a_{22}}{\Delta_1} & -\frac{a_{12}}{\Delta_1} & 0 \\ -\frac{a_{21}}{\Delta_1} & \frac{a_{11}}{\Delta_1} & \\ 0 & & \ddots \end{bmatrix},$$

где Δ_1 определитель $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$.

В случае, если матрица A симметрична и $A = R + S$, где R положительно-определенная матрица, обратная для которой известна,

процесс последовательных приближений, примененный к системе, подготовленной к виду

$$X = -R^{-1}SX + R^{-1}F,$$

имеет следующий простой критерий сходимости. Для сходимости процесса необходимо и достаточно, чтобы матрицы $R+S$ и $R-S$ были положительно-определенными.

Действительно, собственные значения матрицы $R^{-1}S$ вещественны, так как R положительно определена (теорема 11.14). Для сходимости процесса необходимо и достаточно, чтобы собственные значения матрицы $R^{-1}S$ были по модулю меньше единицы, т. е. заключены в интервале $(-1, 1)$. Для этого же, в свою очередь, необходимо и достаточно, чтобы все собственные значения матриц $E+R^{-1}S$ и $E-R^{-1}S$ были бы положительны. Это имеет место в том и только в том случае (теорема 11.16), если матрицы $R(E+R^{-1}S)=R+S$ и $R(E-R^{-1}S)=R-S$ положительно-определенны.

§ 32. Одношаговый циклический процесс

Пусть система линейных уравнений $AX=F$ представлена в виде

$$X = BX + G, \quad (1)$$

где $B = E - A$, $G = F$. Обозначим компоненты искомого вектора решения через x_1, \dots, x_n . Одношаговый циклический процесс (часто называемый также методом Зейделя) напоминает процесс последовательных приближений с той разницей, что при вычислении k -го приближения для i -й компоненты учитываются вычисленные уже ранее k -е приближения для компонент $x_1^{(k)}, \dots, x_{i-1}^{(k)}$.

Именно, вычисление последовательных приближений ведется по формулам

$$x_i^{(k)} = \sum_{j=1}^{i-1} b_{ij} x_j^{(k)} + \sum_{j=i}^n b_{ij} x_j^{(k-1)} + g_i \quad (2)$$

(вместо $x_i^{(k)} = \sum_{j=1}^n b_{ij} x_j^{(k-1)} + g_i$ в методе последовательных приближений).

Одношаговый циклический процесс может быть истолкован двумя способами как разновидность общего итерационного процесса. Именно, за один шаг процесса можно принять переход от вектора $(x_1^{(k)}, \dots, x_{i-1}^{(k)}, x_i^{(k-1)}, \dots, x_n^{(k-1)})'$ к вектору $(x_1^{(k)}, \dots, x_i^{(k)}, x_{i+1}^{(k-1)}, \dots, x_n^{(k-1)})'$ (или от вектора $(x_1^{(k)}, \dots, x_n^{(k)})'$ к вектору $(x_1^{(k+1)}, x_2^{(k)}, \dots, x_n^{(k)})'$), или же за один шаг можно считать результат применения полного цикла, т. е. переход от вектора $(x_1^{(k-1)}, \dots,$

$\dots, x_n^{(k-1)})'$ к вектору $(x_1^{(k)}, \dots, x_n^{(k)})'$. В первом истолковании метод не будет стационарным, но будет циклическим. Именно, как легко видеть, в этом случае в качестве матриц, определяющих процесс, берутся, поочереди, матрицы $e_{11}, e_{22}, \dots, e_{nn}$, где

$$e_{ii} = \begin{bmatrix} & & & & (i) \\ 0 & \dots & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & 0 & \dots & 0 \end{bmatrix} (i). \quad (3)$$

Точнее, $H_{n(k-1)+i} = e_{ii}$.

Во втором истолковании процесс будет стационарным. Исследуем его подробнее.

Уравнение

$$X = BX + G$$

представим в виде

$$X = (M + N)X + G, \quad (4)$$

где

$$M = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ b_{21} & 0 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ b_{n1} & b_{n2} & \dots & b_{n,n-1} & 0 \end{bmatrix}, \quad N = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ 0 & b_{22} & \dots & b_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & b_{nn} \end{bmatrix}. \quad (5)$$

В этих обозначениях формулы

$$x_i^{(k)} = \sum_{j=1}^{i-1} b_{ij} x_j^{(k)} + \sum_{j=i}^n b_{ij} x_j^{(k-1)} + g_i$$

можно представить в матричной форме в виде

$$X^{(k)} = MX^{(k)} + NX^{(k-1)} + G. \quad (6)$$

Отсюда следует, что

$$X^{(k)} = (E - M)^{-1} NX^{(k-1)} + (E - M)^{-1} G.$$

Таким образом, один полный цикл одношагового циклического процесса для системы (4) оказывается равносильным одному шагу процесса последовательных приближений, примененного к системе

$$X = (E - M)^{-1} NX + (E - M)^{-1} G,$$

которая равносильна исходной системе

$$X = (M + N)X + G$$

и может быть получена из нее умножением слева на неособенную матрицу $(E - M)^{-1}$.

Из такого представления процесса следует, что для его сходимости необходимо и достаточно, чтобы все собственные значения матрицы $S = (E - M)^{-1}N$ были по модулю меньше единицы. Эти собственные значения являются корнями полинома $|S - tE|$. Умножив обе части этого уравнения на $|E - M|$ и воспользовавшись теоремой о произведении определителей двух матриц, мы преобразуем уравнение к виду

$$|N - (E - M)t| = 0$$

или, в развернутой форме, к виду

$$\begin{vmatrix} b_{11} - t & b_{12} & \dots & b_{1n} \\ b_{21}t & b_{22} - t & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1}t & b_{n2}t & \dots & b_{nn} - t \end{vmatrix} = 0. \quad (7)$$

Таким образом, для сходимости одношагового циклического процесса необходимо и достаточно, чтобы все корни уравнения (7) были бы по модулю меньше единицы.

Обратимся теперь к рассмотрению одного достаточного условия сходимости одношагового циклического процесса.

Именно, пусть

$$\|B\|_1 = \max_i \sum_{j=1}^n |b_{ij}| \leq \mu < 1.$$

Как мы видели, в этом случае для метода последовательных приближений имеет место оценка

$$\|X^* - X^{(k)}\| \leq \mu \|X^* - X^{(k-1)}\|, \quad (8)$$

где в качестве нормы векторов взята кубическая норма, т. е. $\max_i |x_i|$.

Покажем, что при этом условии одношаговый циклический процесс сходится, и для него имеет место несколько лучшая оценка.

Действительно, если $X^* = (x_1, \dots, x_n)'$, то

$$x_i = \sum_{j=1}^n b_{ij}x_j + g_i \quad (i = 1, \dots, n). \quad (9)$$

Вычитая из (9) равенства (2), получим

$$\begin{aligned} x_i - x_i^{(k)} &= \sum_{j=1}^{i-1} b_{ij}(x_j - x_j^{(k)}) + \sum_{j=i}^n b_{ij}(x_j - x_j^{(k-1)}) \\ |x_i - x_i^{(k)}| &\leq \sum_{j=1}^{i-1} |b_{ij}| |x_j - x_j^{(k)}| + \sum_{j=i}^n |b_{ij}| |x_j - x_j^{(k-1)}|. \end{aligned} \quad (10)$$

Обозначим

$$\sum_{j=1}^{i-1} |b_{ij}| = \beta_i, \quad \sum_{j=i}^n |b_{ij}| = \gamma_i.$$

Тогда из неравенства (10) следует, что

$$|x_i - x_i^{(k)}| \leq \beta_i \|X^* - X^{(k)}\| + \gamma_i \|X^* - X^{(k-1)}\|.$$

Взяв для i то значение i_0 , при котором $|x_i - x_i^{(k)}|$ достигает максимума, получим

$$\|X^* - X^{(k)}\| \leq \frac{\gamma_{i_0}}{1 - \beta_{i_0}} \|X^* - X^{(k-1)}\|, \quad (11)$$

ибо

$$\|X^* - X^{(k)}\| = \max_i |x_i - x_i^{(k)}| = |x_{i_0} - x_{i_0}^{(k)}|.$$

Обозначим

$$\max \frac{\gamma_i}{1 - \beta_i} = \mu'.$$

Тогда

$$\|X^* - X^{(k)}\| \leq \mu' \|X^* - X^{(k-1)}\|. \quad (12)$$

Установим, что

$$\mu' \leq \mu.$$

Действительно,

$$\beta_i + \gamma_i = \sum_{j=1}^n |b_{ij}| \leq \mu$$

и

$$\beta_i + \gamma_i - \frac{\gamma_i}{1 - \beta_i} = \frac{\beta_i(1 - \beta_i - \gamma_i)}{1 - \beta_i} \geq 0.$$

Отсюда

$$\mu \geq \max(\beta_i + \gamma_i) \geq \max \frac{\gamma_i}{1 - \beta_i} = \mu'.$$

Здесь знак равенства возможен, только если $\max_i \sum_{j=1}^n |b_{ij}|$ достигается при $i = 1$ (или если $\beta_i = 0$), и понижение оценки по сравнению с оценкой (8) будет наилучшим, если расположить уравнения в порядке возрастания $\sum_{j=1}^n |b_{ij}|$, принимая за первое то уравнение, в котором эта сумма наименьшая.

Однако одношаговый циклический процесс не всегда оказывается более выгодным, чем метод последовательных приближений. Иногда одношаговый циклический процесс сходится медленнее процесса последовательных приближений. Возможно даже, что одношаговый циклический процесс расходится, хотя метод последовательных приближений сходится. Области сходимости этих двух процессов различны и лишь частично перекрываются.

Приведем примеры, показывающие различие областей сходимости одношагового циклического процесса и процесса последовательных приближений.

Пример 1. Пусть

$$B = \begin{bmatrix} 5 & -5 \\ 1 & 0.1 \end{bmatrix}.$$

Тогда собственные значения матрицы B определяются из уравнения $(0.1-t)(5-t)+5=0$ и потому $\max|\lambda_i|>1$. Процесс последовательных приближений расходится.

Образуем матрицу

$$S = (E - M)^{-1}N = \begin{bmatrix} 5 & -5 \\ 5 & -4.9 \end{bmatrix}.$$

Собственные значения этой матрицы определяются из уравнения $t^2 - 0.1t + 0.5 = 0$. Очевидно, $\max|\lambda_i| < 1$. Одношаговый циклический процесс сходится.

Пример 2. Пусть

$$B = \begin{bmatrix} 2.3 & -5 \\ 1 & -2.3 \end{bmatrix}.$$

Собственные значения матрицы B определяются из уравнения $-(2.3-t)(2.3+t)+5=t^2-0.29=0$; $\max|\lambda_i| < 1$. Процесс последовательных приближений сходится. Нетрудно проверить, что в этом случае одношаговый циклический процесс расходится. Действительно,

$S = (E - M)^{-1}N = \begin{bmatrix} 2.3 & -5 \\ 2.3 & -7.3 \end{bmatrix}$ и собственные значения этой матрицы по модулю больше единицы.

Одношаговый циклический процесс теоретически тождественен с процессом последовательных приближений, примененным к надлежащим образом подготовленной данной системе и это обстоятельство было нами использовано для получения условий сходимости процесса. Однако фактически при проведении процесса вычислительная схема не совпадает с вычислительной схемой эквивалентного процесса последовательных приближений и, в частности, вычисление „подготавливающей матрицы“ $H = (E - M)^{-1}$ не должно быть осуществлено. Это обстоятельство и заставляет выделить одношаговый циклический процесс как самостоятельный итерационный метод.

В качестве примера найдем решение системы (14) § 30, приведенной к виду

$$x_1 = 0.22x_1 + 0.02x_2 + 0.12x_3 + 0.14x_4 + 0.76$$

$$x_2 = 0.02x_1 + 0.14x_2 + 0.04x_3 - 0.06x_4 + 0.08$$

$$x_3 = 0.12x_1 + 0.04x_2 + 0.28x_3 + 0.08x_4 + 1.12$$

$$x_4 = 0.14x_1 - 0.06x_2 + 0.08x_3 + 0.26x_4 + 0.68.$$

Достаточные условия сходимости одношагового циклического процесса, очевидно, выполнены.

В качестве начального приближения берем вектор свободных членов. Последовательные приближения помещены в табл. III.5.

Таблица III.5

Вычисление решения системы одношаговым циклическим процессом

$X^{(0)} = G$	0.76	0.08	1.12	0.68
$X^{(1)}$	1.1584	0.1184	1.6317	1.1424
$X^{(2)}$	1.3730	0.1208	1.8379	1.3090
$X^{(3)}$	1.4683	0.1213	1.9204	1.3723
$X^{(4)}$	1.5080	0.1216	1.9533	1.3969
$X^{(5)}$	1.5242	0.1218	1.9665	1.4066
$X^{(6)}$	1.5307	0.1219	1.9717	1.4104
$X^{(7)}$	1.5333	0.1219	1.9738	1.4118
$X^{(8)}$	1.5343	0.1220	1.9746	1.4125
$X^{(9)}$	1.53469	0.12201	1.97493	1.41278
$X^{(10)}$	1.53485	0.12201	1.97507	1.41289
$X^{(11)}$	1.53492	0.12201	1.97512	1.41293
$X^{(12)}$	1.534947	0.122009	1.975141	1.412945
$X^{(13)}$	1.5349579	0.1220094	1.9751507	1.4129513
$X^{(14)}$	1.5349622	0.1220010	1.9751541	1.4129538

Таблица III.6

Вычисление решения системы одношаговым циклическим процессом

$X^{(0)} = G$	0.22900763	0.38167939	0.53435515	0.68702290
$X^{(1)}$	-0.40557078	0.24074649	0.65379631	0.86327494
$X^{(20)}$	-1.2560487	0.0439969	1.0383844	1.4810254
$X^{(25)}$	-1.2574501	0.0435866	1.0390124	1.4821242
$X^{(40)}$	-1.2577909	0.0434880	1.0391650	1.4823907
$X^{(49)}$	-1.2577935	0.0434873	1.0391661	1.4823927

Сравнивая найденные последовательные приближения с решением системы, найденным по методу единственного деления (см. § 30), мы видим, что четырнадцатое приближение дает решение с точностью

до трех единиц шестого знака. Одношаговый циклический процесс в рассматриваемом примере сходится быстрее, чем процесс последовательных приближений (см. табл. III.1, III.2 и III.3).

То же заключение будет верно и в применении к системе (3) § 31, что видно из сравнения табл. III.4 и табл. III.6, в которой дается решение упомянутой системы циклическим одношаговым процессом.

§ 33. Метод Некрасова.

Так же как в методе последовательных приближений, данную систему $AX=F$ можно подготавливать к виду, удобному для применения одношагового циклического процесса различными способами. Наиболее употребительной является модификация одношагового циклического процесса, параллельная методу простой итерации.

Необходимое и достаточное условие сходимости этой модификации циклического одношагового процесса впервые было найдено П. А. Некрасовым¹⁾, что дает право называть ее методом Некрасова.

Система $AX=F$ записывается в виде

$$a_{ii}x_i = -\sum_{j=1}^{i-1} a_{ij}x_j - \sum_{j=i+1}^n a_{ij}x_j + f_i \quad (1)$$

или, что то же самое, в виде

$$x_i = -\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j - \sum_{j=i+1}^n \frac{a_{ij}}{a_{ii}} x_j + \frac{f_i}{a_{ii}},$$

и последовательные приближения определяются по формулам

$$x_i^{(k)} = -\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k)} - \sum_{j=i+1}^n \frac{a_{ij}}{a_{ii}} x_j^{(k-1)} + \frac{f_i}{a_{ii}}. \quad (2)$$

Необходимые и достаточные условия сходимости этой модификации процесса легко получаются из приведенных выше необходимых и достаточных условий сходимости в общем случае.

Действительно, выбранная подготовка системы $AX=F$ к виду $X=BX+G$ основана на предварительном умножении системы на диагональную матрицу $D^{-1}=[a_{11}, \dots, a_{nn}]^{-1}$. Следовательно, $B=E-D^{-1}A$. Обозначим

$$L = \begin{bmatrix} 0 & 0 & \dots & 0 \\ a_{21} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 0 & a_{12} & \dots & a_{1n} \\ 0 & 0 & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{bmatrix}. \quad (3)$$

¹⁾ П. А. Некрасов [1].

Тогда

$$A = L + D + R$$

так что в прежних обозначениях

$$M = -D^{-1}L, \quad N = -D^{-1}R,$$

$$S = (E - M)^{-1}N = -(E + D^{-1}L)^{-1}D^{-1}R = -(D + L)^{-1}R.$$

Характеристический полином

$$|-(D + L)^{-1}R - tE|$$

матрицы

$$-(D + L)^{-1}R$$

после умножения на

$$|-(D + L)|$$

принимает вид

$$|R + (D + L)t|.$$

Следовательно, для сходимости метода Некрасова необходимо и достаточно, чтобы все корни уравнения (уравнение Некрасова)

$$\begin{vmatrix} a_{11}t & a_{12} & \dots & a_{1n} \\ a_{21}t & a_{22}t & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1}t & a_{n2}t & \dots & a_{nn}t \end{vmatrix} = 0 \quad (4)$$

были бы по модулю меньше единицы.

Большое количество достаточных признаков сходимости метода дано в переписке П. А. Некрасова и Мемке.¹⁾

В частности, достаточными признаками являются признаки I, II, I', II' § 30.

В случае, если матрица A из коэффициентов системы $AX = F$ симметрична или эрмитова, существует еще одно важное условие сходимости метода Некрасова. Именно, если матрица A положительно определена, то метод Некрасова для системы $AX = F$ сходится. Это условие в предположении положительности диагональных элементов матрицы A оказывается также необходимым.²⁾

Для доказательства положим

$$A = R^* + D + R, \quad (5)$$

где D — диагональная матрица, составленная из диагональных элементов матрицы A , R — треугольная матрица, образованная элементами матрицы A , лежащими выше главной диагонали, R^* — ее сопряженная.

¹⁾ Мемке и П. А. Некрасов [1].

²⁾ Рейх [1], Шнейдер [1], Островский [8].

Как мы видели выше, необходимым и достаточным условием сходимости метода Некрасова является требование, чтобы все собственные числа матрицы $S = -(D + R^*)^{-1}R$ были бы по модулю меньше единицы.

Приведем оценку для модулей собственных значений этой матрицы, из которой непосредственно следует, что все они меньше единицы.

Обозначим через λ_0 наименьшее собственное значение матрицы A . Так как матрица A положительно-определенная, $\lambda_0 > 0$. Далее, обозначим $\Lambda_0 = \|R\| = \|R^*\|$. Пусть U собственный вектор длины единица для матрицы $(D + R^*)^{-1}R$, принадлежащий некоторому собственному значению λ . Тогда

$$(D + R^*)^{-1}RU = \lambda U$$

или

$$RU = (D + R^*)\lambda U = \lambda DU + \lambda R^*U.$$

Далее,

$$(RU, U) = \lambda(DU, U) + \lambda(R^*U, U) = \lambda(AU, U) - \lambda(RU, U).$$

Обозначим

$$(AU, U) = p$$

$$(DU, U) = d$$

$$(RU, U) = a + ib$$

$$(R^*U, U) = (U, RU) = a - ib.$$

Тогда

$$\lambda = \frac{a + ib}{p - a - ib},$$

откуда

$$|\lambda|^2 = \frac{a^2 + b^2}{(p - a)^2 + b^2}.$$

Но

$$p = (AU, U) = (DU, U) + (RU, U) + (R^*U, U) = d + 2a,$$

и потому

$$(p - a)^2 = p^2 - 2ap + a^2 = p(p - 2a) + a^2 = pd + a^2.$$

Таким образом,

$$|\lambda|^2 \leq \frac{a^2 + b^2}{pd + a^2 + b^2} \leq \frac{a^2 + b^2}{\lambda_0 d_0 + a^2 + b^2},$$

где d_0 есть наименьшее собственное значение матрицы D , ибо $d \geq d_0$, $p \geq \lambda_0$. Очевидно, что $d_0 = \min_i a_{ii}$. Далее,

$$a^2 + b^2 = |a + ib|^2 = |(RU, U)|^2 \leq |RU|^2 |U|^2 \leq \|R\|^2 = \Lambda_0^2.$$

Следовательно,

$$|\lambda| \leq \frac{\Lambda_0}{\sqrt{\lambda_0 d_0 + \Lambda_0^2}}. \quad (6)$$

Отметим, что $d_0 \geq \lambda_0$, ибо

$$d_0 = \min_i a_{ii} = \min_i (Ae_i, e_i) \geq \min_{\|X\|=1} (AX, X) = \lambda_0.$$

Поэтому справедлива оценка¹⁾

$$|\lambda| \leq \frac{\Lambda_0}{\sqrt{\lambda_0^2 + \Lambda_0^2}}, \quad (7)$$

несколько более грубая, чем оценка (6).

Из полученных оценок следует, что все собственные значения матрицы $(D + R^*)^{-1}R$ по модулю меньше единицы. Тем самым сходимость процесса Некрасова доказана.

Докажем теперь необходимость высказанных условий.

Пусть A симметричная или эрмитова матрица с положительными диагональными элементами,

$$X = (x_1^{(k)}, \dots, x_i^{(k)}, x_{i+1}^{(k-1)}, \dots, x_n^{(k-1)})$$

и

$$X' = (x_1^{(k)}, \dots, x_{i+1}^{(k)}, x_{i+2}^{(k-1)}, \dots, x_n^{(k-1)})$$

два соседние последовательные приближения в k -м цикле. Тогда

$$X' = X + e_{ii} D^{-1}(F - AX).$$

Пусть X^* решение системы, $Y = X^* - X$, $Y' = X^* - X'$ соответствующие векторы ошибки. Тогда

$$Y' = Y - e_{ii} D^{-1} AY = Y - \frac{1}{a_{ii}} r_i e_i,$$

где r_i — i -я компонента вектора $r = F - AX = AY$, e_i — вектор, у которого i -я компонента равна единице, а остальные нули.

Вычислим значение функционала $f(X') = (AY', Y')$ (совпадающего с функцией ошибки, если A положительно определена). Имеем

$$f(X') = (AY', Y') = (AY, Y) - 2 \frac{r_i}{a_{ii}} (AY, e_i) + \frac{r_i^2}{a_{ii}^2} (Ae_i, e_i).$$

Но

$$(AY, e_i) = r_i, \quad (Ae_i, e_i) = a_{ii},$$

и потому

$$f(X') = (AY', Y') = (AY, Y) - \frac{r_i^2}{a_{ii}} \leq (AY, Y). \quad (8)$$

Если матрица A не положительно определена и $|A| \neq 0$, то можно найти начальный вектор $X^{(0)}$ так, что $(AY^{(0)}, Y^{(0)}) < 0$. Тогда в силу (8) на протяжении всего процесса будет $(AY^{(k)}, Y^{(k)}) \leq (AY^{(0)}, Y^{(0)}) < 0$.

¹⁾ Островский [8].

и, следовательно, предельное соотношение $Y^{(k)} \rightarrow 0$ невозможно. Поэтому процесс будет расходящимся.

Установленный критерий показывает, что если матрица системы симметричная с положительными диагональными элементами, то область сходимости метода Некрасова шире области сходимости метода простой итерации.

Действительно, для сходимости метода Некрасова необходима и достаточна, в этом случае, положительная определенность матрицы A , для сходимости же метода простой итерации необходимым и достаточным условием является положительная определенность матриц A и $2D - A$.

В табл. III.7 приводится решение системы (9) § 23 по методу Некрасова.

Таблица III.7

Решение системы по методу Некрасова

$X^{(0)}$	0	0	0	0
$X^{(1)}$	0.3	0.374	0.41832	0.4454096
$X^{(20)}$	-1.2577875	0.0434903	1.0391641	1.4823879
$X^{(22)}$	-1.2577922	0.0434881	1.0391657	1.4823916
$X^{(23)}$	-1.2577935	0.0434874	1.0391662	1.4823927

§ 34. Методы полной релаксации

Начиная с этого параграфа и до § 38 включительно, мы будем рассматривать преимущественно системы с положительно-определенными матрицами, оговаривая каждый раз случаи, когда это требование не выполняется.

Пусть X^* — точное решение системы $AX = F$ с положительно-определенной матрицей A , X — некоторый вектор, $f(X)$ — функция ошибки. Ставится задача, как изменить i -ю компоненту вектора X , чтобы для измененного вектора X' значение функции ошибки было бы наименьшим. Пусть

$$X' = X + \alpha e_i.$$

Тогда

$$Y' = Y - \alpha e_i,$$

и

$$\begin{aligned} f(X') &= (AY', Y') = (A(Y - \alpha e_i), Y - \alpha e_i) = \\ &= (AY, Y) - 2\alpha(AY, e_i) + \alpha^2(Ae_i, e_i) = (AY, Y) + \alpha^2 a_{ii} - 2\alpha r_i = \\ &= f(X) + \frac{1}{a_{ii}}(a_{ii}\alpha - r_i)^2 - \frac{r_i^2}{a_{ii}}, \end{aligned} \quad (1)$$

где r_i — i -я компонента вектора невязки для приближения X .

Ясно, что $f(X')$ будет иметь минимальное значение при

$$\alpha = \frac{r_i}{a_{ii}},$$

и это минимальное значение равно

$$f(X) - \frac{r_i^2}{a_{ii}}.$$

Вычислим теперь i -ю компоненту невязки для приближения X' . Она равна

$$(F - AX', e_i) = (F - AX - \alpha A e_i, e_i) = \\ = (F - AX, e_i) - \alpha (A e_i, e_i) = r_i - \alpha a_{ii} = 0.$$

т. е. приближение X' удовлетворяет i -му уравнению системы $AX = F$. Другими словами, i -я компонента приближения X' может быть вычислена из i -го уравнения системы $AX = F$, в которое вместо остальных неизвестных подставлены компоненты вектора X . Именно так проходит один шаг в каком-либо цикле процесса Некрасова. Тем самым процессу Некрасова может быть дано следующее истолкование: на каждом шагу минимизируется функция ошибки за счет изменения одной компоненты предыдущего приближения, номера же этих компонент циклически чередуются от 1 до n .

Если, используя отдельные шаги процесса Некрасова, отказаться от цикличности в выборе изменяемых неизвестных, то мы придем к более общей группе методов, называемых методами полной релаксации.

При такой постановке имеется широкий произвол в выборе последовательности номеров изменяемых компонент (ведущих индексов). Так, например, решая систему десяти уравнений с десятью неизвестными, занумерованными числами от 0 до 9, можно в качестве «управления» процессом взять хотя бы десятичную запись числа $e = 2.718\ 281\ 828\ 459\ 045\ 235\ 36\dots$, т. е. менять на первом шагу вторую компоненту, на втором—седьмую, на третьем— первую и т. д. Ясно, что для фактического проведения релаксационного процесса должен быть выбран какой-либо разумный принцип управления процессом, т. е. принцип выбора последовательности номеров изменяемых компонент. О некоторых принципах управления процессом релаксации будет сказано ниже.

Конечно, не всякий процесс полной релаксации сходится к решению. Так, например, если выбранная последовательность номеров изменяемых компонент совсем не содержит хотя бы одного номера, то все поправки $X^{(k+1)} - X^{(k)}$ будут содержаться в $(n-1)$ -мерном подпространстве, и если $X^* - X^{(0)}$ не содержится в этом подпространстве, процесс не может сходиться к X^* .

Достаточное условие для сходимости процесса к решению будет дано в § 37.

§ 35. Неполная релаксация

Вместо полной минимизации функции ошибки на каждом отдельном шагу процесса можно заботиться лишь об уменьшении функции ошибки. Процессы, построенные исходя из этого принципа, называются процессами неполной релаксации.

Выясним, как изменить одну компоненту приближения с тем, чтобы функция ошибки уменьшалась.

Пусть $X' = X + \alpha e_i$. Тогда

$$f(X') - f(X) = \frac{1}{a_{ii}} (a_{ii}\alpha - r_i)^2 - \frac{r_i^2}{a_{ii}}.$$

Для того, чтобы значение $f(X') - f(X)$ было отрицательным, необходимо и достаточно, чтобы

$$(a_{ii}\alpha - r_i)^2 < r_i^2,$$

откуда

$$|a_{ii}\alpha - r_i| < |r_i|$$

и, следовательно,

$$\alpha = q \frac{r_i}{a_{ii}}, \quad (1)$$

при $0 < q < 2$.

При $q = 1$ будет иметь место полная минимизация функции ошибки или, как говорят, полная релаксация. Неполная релаксация называется нижней, если $0 < q < 1$ и верхней — если $1 < q < 2$.

При неполной релаксации функция ошибки изменяется по формуле

$$f(X') = f(X) - \frac{q(2-q)}{a_{ii}} r_i^2. \quad (2)$$

На каждом отдельном шагу процесса метод полной релаксации является наивыгоднейшим, так как он обеспечивает максимальное уменьшение функции ошибки за один шаг. Однако при проведении большего числа шагов может оказаться, что неполная релаксация дает лучший результат.

Число q при неполной релаксации может изменяться от шага к шагу. Если процесс неполной релаксации берется циклическим при постоянном или циклически меняющемся q , то процесс можно рассматривать как частный случай общего одношагового циклического процесса, примененного к системе, подготовленной к виду

$$X = (E - QD^{-1}A)X + QD^{-1}F,$$

где $D = [a_{11}, \dots, a_{nn}]$; $Q = [q_1, \dots, q_n]$ (q_1, \dots, q_n — значения множителей релаксации).

Действительно, формулы для вычисления компонент результата k -го цикла будут

$$x_i^{(k)} = x_i^{(k-1)} + \\ + q_i \frac{f_i - (a_{i1}x_1^{(k)} + \dots + a_{i(i-1)}x_{i-1}^{(k)} + a_{ii}x_i^{(k-1)} + \dots + a_{in}x_n^{(k-1)})}{a_{ii}}. \quad (3)$$

Отметим, что последние формулы определяют итерационный процесс и для систем с не положительно-определенными матрицами. Однако в этом случае конечно нельзя уже говорить о релаксационных свойствах процесса.

Легко найти необходимое и достаточное условие сходимости процесса. Действительно, в обозначениях § 32 и § 33 будем иметь

$$M = -QD^{-1}L, \quad N = E - QD^{-1}(D + R)$$

и

$$S = (E - M)^{-1}N = (E + QD^{-1}L)^{-1}(E - Q - QD^{-1}R) = \\ = (D + QL)^{-1}(D - DQ - QR),$$

так что необходимым и достаточным условием сходимости процесса (3) будет требование, чтобы все собственные значения матрицы $(D + QL)^{-1}(D - DQ - QR)$ были по модулю меньше единицы. Собственные значения этой матрицы, очевидно, являются корнями уравнения

$$\left| \begin{array}{cccc} (t + q_1 - 1)a_{11} & q_1 a_{12} & \dots & q_1 a_{1n} \\ t q_2 a_{21} & (t + q_2 - 1)a_{22} & \dots & q_2 a_{2n} \\ \dots & \dots & \dots & \dots \\ t q_n a_{n1} & t q_n a_{n2} & \dots & (t + q_n - 1)a_{nn} \end{array} \right| = 0 \quad (4)$$

Быстрота же сходимости метода будет зависеть от величины наибольшего модуля корней этого уравнения.

Полученный критерий сходимости процесса (3) представляет интерес главным образом в случае, если матрицы системы не положительно определены, так как в случае положительно-определенной матрицы циклический процесс неполной релаксации всегда сходится, как это будет показано в § 37. Однако и для этого случая критерий представляет интерес, так как он дает средство для исследования быстроты сходимости процесса.

Рассмотрим подробнее процесс (3) при $q_1 = q_2 = \dots = q_n = q$.

Покажем, что за счет малого отклонения q от единицы почти всегда можно добиться уменьшения наибольшего модуля корней уравнения (4) и, следовательно, получить процесс с более быстрой сходимостью, чем процесс Некрасова.

Пусть $q = 1 - \varepsilon$, где ε малое вещественное число. Тогда уравнение (4) равносильно уравнению

$$\begin{vmatrix} \frac{t-\varepsilon}{1-\varepsilon} a_{11} & a_{12} & \cdots & a_{1n} \\ ta_{21} & \frac{t-\varepsilon}{1-\varepsilon} a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ta_{n1} & ta_{n2} & \cdots & \frac{t-\varepsilon}{1-\varepsilon} a_{nn} \end{vmatrix} = 0. \quad (5)$$

Положим $t = \lambda_0 + k\varepsilon$, где λ_0 наибольший по модулю корень уравнения Некрасова $|tL + tD + R| = 0$. Тогда, с точностью до ε^2 , имеем $\frac{t-\varepsilon}{1-\varepsilon} = \lambda_0 + (k-1+\lambda_0)\varepsilon$, и уравнение (5), с точностью до малых 2-го порядка, переходит в уравнение

$$|T_0 + \varepsilon(k-1+\lambda_0)D + k\varepsilon L| = 0. \quad (6)$$

Здесь $T_0 = \lambda_0 L + \lambda_0 D + R$. Легко видеть, что, с точностью до малых 2-го порядка, справедлива следующая формула

$$|A + \varepsilon B| = |A| + \varepsilon \operatorname{Sp} B\tilde{A},$$

где \tilde{A} — матрица, союзная с матрицей A . Следовательно, уравнение (6) примет вид

$$|T_0| + \varepsilon [(k-1+\lambda_0) \operatorname{Sp} D\tilde{T}_0 + k \operatorname{Sp} L\tilde{T}_0] = O(\varepsilon^2).$$

Принимая во внимание, что $|T_0| = 0$, получим, с точностью до малых 1-го порядка, что

$$k = \frac{(1-\lambda_0) \operatorname{Sp} D\tilde{T}_0}{\operatorname{Sp} D\tilde{T}_0 + \operatorname{Sp} L\tilde{T}_0}. \quad (7)$$

Итак, число $\lambda' = \lambda_0 + k\varepsilon$, где k определяется по формуле (7), является приближенным значением наибольшего по модулю корня уравнения (5).

Сравним модули λ_0 и λ' . Ясно, что

$$|\lambda'|^2 = |\lambda_0|^2 + 2\varepsilon \operatorname{Re}(k\bar{\lambda}_0) + \varepsilon^2 |k|^2.$$

Поэтому, при достаточно малом ε , можно добиться того, чтобы $|\lambda'|^2$ был меньше $|\lambda_0|^2$, если только $\operatorname{Re}(k\bar{\lambda}_0) \neq 0$. При этом, если $\operatorname{Re}(k\bar{\lambda}_0) < 0$, следует взять $\varepsilon > 0$, т. е. прибегнуть к нижней релаксации, если $\operatorname{Re}(k\bar{\lambda}_0) > 0$, то нужно взять $\varepsilon < 0$, т. е. перейти к верхней релаксации.

Формула (7) равносильна следующей формуле

$$k = \frac{(1-\lambda_0)(DX_0, Y_0)}{((D+L)X_0, Y_0)}, \quad (8)$$

где X_0 ненулевой вектор, определенный из уравнения $T_0 X_0 = 0$, а Y_0 ненулевой вектор, определенный из уравнения $T_0^* Y_0 = 0$. Векторы X_0 и Y_0 определяются с точностью до скалярных множителей однозначно, если предположить, что λ_0 простой корень уравнения $|tL + tD + R| = 0$. Но, как можно показать, только в этом случае имеет смысл и формула (7).

Вопрос о выборе множителя q , приводящего к процессу с наибольшей быстротой сходимости, в общем случае не решен.

Для матриц 2-го порядка исчерпывающее исследование проведено А. М. Островским [7]. Пусть

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Ненулевым корнем уравнения

$$\begin{vmatrix} a_{11}t & a_{12} \\ a_{21}t & a_{22}t \end{vmatrix} = 0$$

является, очевидно, $\lambda_0 = u = \frac{a_{12}a_{21}}{a_{11}a_{22}}$, так что для сходимости метода Некрасова необходимо и достаточно, чтобы $|u| < 1$.

Для положительно-определенной матрицы A имеет место неравенство

$$0 < u < 1.$$

Применяя формулу (7), получим

$$k = 2(1 - u), \quad k\bar{\lambda}_0 = 2u(1 - u).$$

Поэтому при $0 < u < 1$ (в частности, для положительно-определенных матриц) верхняя релаксация (по крайней мере при малых ϵ) дает более быструю сходимость, чем полная, а при $-1 < u < 0$ более быстрая сходимость будет при нижней релаксации.

В той же работе А. М. Островского дано оптимальное значение для множителя q . Именно,

$$q_0 = \frac{2}{1 + \sqrt{1 - u}},$$

причем соответствующее значение для модуля λ' есть

$$|\lambda'| = \frac{|u|}{1 + \sqrt{1 - u}}.$$

Отметим, что изменение быстроты сходимости при неполной релаксации может быть довольно значительным. Так, для

$$A = \begin{bmatrix} 1 & 0,6 \\ 0,6 & 1 \end{bmatrix}$$

будем иметь при полной релаксации $\lambda_0 = \frac{9}{25} = 0.36$. Оптимальная же неполная релаксация будет при $q = \frac{10}{9}$, причем соответствующее значение λ' равно $\frac{1}{9} = 0.111\dots$

Для не положительно-определеных матриц Островским показано в той же работе, что при $u > 1$ процесс (3) расходится при всех значениях q из интервала $(0, 2)$. Если же $u < -1$, то процесс (3) будет сходящимся при $0 < q < \frac{2}{1 + \sqrt{|u|}}$, так что область сходимости процесса (3) при $q < 1$ будет шире, чем область сходимости процесса Некрасова.

Для положительно-определенных матриц третьего порядка можно показать посредством довольно громоздких выкладок¹⁾, что

$$\operatorname{Re}(k\lambda_0) > 0,$$

так что верхняя релаксация (по крайней мере при малых ϵ) приводит к более быстрой сходимости, чем полная.

Таблица III. 8

Решение системы методом Некрасова с нижней релаксацией. $q = 0.8$

$X^{(0)}$	0	0	0	0
$X^{(1)}$	0.24	0.31936	0.3745638	0.4149420
$X^{(20)}$	-1.2560832	0.0440084	1.0384154	1.4810565
$X^{(25)}$	-1.2575067	0.0435748	1.0390402	1.4821686
$X^{(40)}$	-1.2577924	0.0434877	1.0391657	1.4823918
$X^{(45)}$	-1.2577935	0.0434874	1.0391661	1.4823927

Таблица III. 9

Решение системы методом Некрасова с верхней релаксацией. $q = 1.1$

$X^{(0)}$	0	0	0	0
$X^{(1)}$	0.33	0.39754	0.4340459	0.4529715
$X^{(16)}$	-1.2577939	0.0434866	1.0391661	1.4823932
$X^{(18)}$	-1.2577935	0.0434873	1.0391661	1.4823928

¹⁾ Д. К. Фаддеев [3].

Таблица III. 10

Решение системы методом Некрасова с верхней релаксацией. $q = 1.2$

$X^{(0)}$	0	0	0	0
$X^{(1)}$	0.36	0.41856	0.4459930	0.4561382
$X^{(15)}$	-1.2577988	0.0434852	1.0391678	1.4823963
$X^{(16)}$	-1.2577939	0.0434869	1.0391662	1.4823930
$X^{(19)}$	-1.2577936	0.0434873	1.0391661	1.4823928

В табл. III. 8, III. 9 и III. 10 приводятся результаты вычисления решения системы (9) § 23 по методу Некрасова с неполной релаксацией при $q = 0.8$, $q = 1.1$ и $q = 1.2$.

Сравнение этих таблиц с табл. III. 7 показывает, что в данном примере верхняя релаксация дает лучший результат, чем полная. Отметим, что дальнейшее увеличение множителя релаксации приводит уже к худшему результату. Так, при $q = 1.3$ имеем

$$X^{(19)} = (-1.2577970, 0.0434871, 1.0391679, 1.4823915)',$$

а результат, аналогичный $X^{(18)}$ табл. III. 9, получается лишь при $k = 26$.

§ 36. Исследование итерационных методов для систем с квазитрехдиагональными матрицами

В настоящем параграфе будет исследоваться быстрота сходимости метода простой итерации (процесса Якоби) и циклического релаксационного метода с постоянным множителем релаксации для систем с положительно-определенными квазитрехдиагональными матрицами вида

$$A = \begin{bmatrix} D_1 & W_1 & & & \\ W'_1 & D_2 & W_2 & & \\ & W'_2 & D_3 & W_3 & \\ & & \ddots & \ddots & \\ & & & \ddots & D_{m-1} & W_{m-1} \\ & & & & W'_{m-1} & D_m \end{bmatrix}, \quad (1)$$

где D_1, \dots, D_m — диагональные матрицы (быть может разных порядков), W_1, W_2, \dots, W_{m-1} — некоторые прямоугольные матрицы. Очевидно, что все диагональные элементы матрицы A положительны.

Результаты, здесь излагаемые, принадлежат Янгу [2].

Всякая матрица A указанного вида обладает „свойством A^* “ (Янг), состоящим в том, что номера строк и столбцов могут быть разбиты на два непересекающихся множества P и Q так, что если $a_{ij} \neq 0$ и $i \in P$ и $j \in Q$ или $i \in Q$ и $j \in P$. Именно, такими множествами будут совокупности номеров строк, соответствующих нечетным клеткам D_1, D_3, \dots с одной стороны и четным D_2, D_4, \dots с другой.

Верно и обратное, что любая симметричная матрица, обладающая „свойством A^* “, может быть приведена к квазитрехдиагональному виду за счет одновременного изменения нумерации строк и столбцов; достаточно, например, первые номера отдать множеству P , последующие — множеству Q .

После такой перенумерации матрица примет вид

$$\begin{bmatrix} \tilde{D}_1 & W \\ W' & \tilde{D}_2 \end{bmatrix}, \quad (2)$$

где \tilde{D}_1 и \tilde{D}_2 — диагональные матрицы, W — некоторая прямоугольная матрица. В частности, общая квазитрехдиагональная матрица может быть преобразована к виду (2) при

$$\tilde{D}_1 = \begin{bmatrix} D_1 & & & \\ & D_3 & & \\ & & \ddots & \\ & & & D_{2k-1} \end{bmatrix}, \quad \tilde{D}_2 = \begin{bmatrix} D_2 & & & \\ & D_4 & & \\ & & \ddots & \\ & & & D_{2k} \end{bmatrix}$$

$$W = \begin{bmatrix} W_1 & & & \\ W'_2 & W_3 & & \\ & \ddots & \ddots & \\ & & W'_{2k-2} & W_{2k-1} \end{bmatrix} \quad (\text{если } m = 2k)$$

и

$$\tilde{D}_1 = \begin{bmatrix} D_1 & & & \\ & D_3 & & \\ & & \ddots & \\ & & & D_{2k+1} \end{bmatrix}, \quad \tilde{D}_2 = \begin{bmatrix} -D_2 & & & \\ & D_4 & & \\ & & \ddots & \\ & & & D_{2k} \end{bmatrix}$$

$$W = \begin{bmatrix} W_1 \\ W'_2 & W_3 \\ & \ddots & \ddots \\ & & W'_{2k-2} & W_{2k-1} \\ & & & W'_{2k} \end{bmatrix} \quad (\text{если } m = 2k + 1)$$

Системы с положительно-определенными матрицами, обладающими „свойством A “, возникают, например, при решении некоторых уравнений в частных производных эллиптического типа разностными методами. Установление нумерации, в которой матрица становится квазитрехдиагональной, связано с тем или другим естественным выбором нумерации узлов.

Прежде всего заметим, что при исследовании сходимости метода простой итерации и релаксационных циклических методов с постоянным множителем релаксации для матриц с положительными диагональными элементами мы можем считать, без нарушения общности, что все диагональные элементы равны единице.

Действительно, если D диагональная матрица, составленная из диагональных элементов матрицы A , и $\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$, то \tilde{A} имеет единичные диагональные элементы. Быстрота сходимости метода простой итерации обуславливается наибольшим модулем собственных значений матрицы $B = E - D^{-1}A$. Быстрота сходимости релаксационного циклического метода с постоянным множителем релаксации q определяется наибольшим модулем собственных значений матрицы

$$S_q = (D + qL)^{-1}(D - qD - qR),$$

где D , L и R диагональная, поддиагональная и наддиагональная части матрицы A . Для матрицы $\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ диагональная, поддиагональная и наддиагональная части будут, соответственно, E , $D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$, $D^{-\frac{1}{2}} R D^{-\frac{1}{2}}$. Поэтому

$$\begin{aligned} \tilde{B} &= E - E^{-1}\tilde{A} = E - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = \\ &= D^{\frac{1}{2}} (E - D^{-1}A) D^{-\frac{1}{2}} = D^{\frac{1}{2}} B D^{-\frac{1}{2}}, \end{aligned}$$

$$\begin{aligned} \tilde{S}_q &= \left(E + qD^{-\frac{1}{2}} L D^{-\frac{1}{2}} \right)^{-1} \left(E - qE - qD^{-\frac{1}{2}} R D^{-\frac{1}{2}} \right) = \\ &= D^{\frac{1}{2}} (D + qL)^{-1} D^{\frac{1}{2}} D^{-\frac{1}{2}} (D - qD - qR) D^{-\frac{1}{2}} = D^{\frac{1}{2}} S_q D^{-\frac{1}{2}}, \end{aligned}$$

Таким образом, матрицы \tilde{B} и \tilde{S}_q , построенные исходя из матрицы \tilde{A} , подобны матрицам B и S_q и, следовательно, их собственные значения соответственно совпадают.

Это замечание позволяет при исследовании указанных методов для положительно-определенных квазитрехдиагональных матриц ограничиться рассмотрением матриц вида

$$A = \begin{bmatrix} E_1 & W_1 & & & \\ W'_1 & E_2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & W_{m-1} & \\ & & & W'_{m-1} & E_m \end{bmatrix},$$

где E_1, E_2, \dots, E_m — единичные матрицы.

Если все наддиагональные клетки квазитрехдиагональной матрицы умножить на некоторое число α , а все поддиагональные на обратное число α^{-1} , то определитель матрицы не изменится.

Действительно, ясно что

$$\begin{aligned} & \begin{bmatrix} D_1 & \alpha W_1 & & & \\ \alpha^{-1} W'_1 & D_2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \alpha W_{m-1} & \\ & & & \alpha^{-1} W'_{m-1} & D_m \end{bmatrix} = \\ & = \begin{bmatrix} E_1 & & & & \\ & \alpha E_2 & & & \\ & & \ddots & & \\ & & & \alpha^{m-1} E_m & \end{bmatrix}^{-1} \begin{bmatrix} D_1 & W_1 & & & \\ W'_1 & D_2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & W_{m-1} & \\ & & & W'_{m-1} & D_m \end{bmatrix} \times \\ & \times \begin{bmatrix} E_1 & & & & \\ & \alpha E_2 & & & \\ & & \ddots & & \\ & & & \alpha^{m-1} E_m & \end{bmatrix}, \end{aligned}$$

откуда непосредственно следует справедливость сказанного.

Это свойство квазитрехдиагональной матрицы позволяет связать характеристический полином матрицы S_q с характеристическим поли-

номом матрицы B . (Мы считаем диагональные элементы матрицы A единичными). Действительно,

$$|tE - S_q| = |t(E + qL) - (E - qE - qR)| =$$

$$= \begin{vmatrix} (t+q-1)E_1 & qW_1 \\ tqW'_1 & (t+q-1)E_2 \\ & \ddots \\ & & qW_{m-1} \\ & tqW'_{m-1} & (t+q-1)E_m \end{vmatrix} =$$

$$= \begin{vmatrix} (t+q-1)E_1 & q\sqrt{t}W_1 \\ q\sqrt{t}W'_1 & (t+q-1)E_2 \\ & \ddots \\ & & q\sqrt{t}W_{m-1} \\ & q\sqrt{t}W'_{m-1} & (t+q-1)E_m \end{vmatrix} =$$

$$= q^n t^{\frac{n}{2}} \begin{vmatrix} \frac{t+q-1}{q\sqrt{t}} E_1 & W_1 \\ W'_1 & \frac{t+q-1}{q\sqrt{t}} E_2 \\ & \ddots \\ & & W_{m-1} \\ & W'_{m-1} & \frac{t+q-1}{q\sqrt{t}} E_m \end{vmatrix} =$$

$$= q^n t^{\frac{n}{2}} F\left(\frac{t+q-1}{q\sqrt{t}}\right), \quad (3)$$

где $F(t)$ есть характеристический полином матрицы

$$B = E - A = \begin{bmatrix} 0 & -W_1 & & & \\ -W'_1 & 0 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & -W_{m-1} \\ & & & -W'_{m-1} & 0 \end{bmatrix}.$$

Полином $F(t)$ обладает свойством $F(-t) = (-1)^n F(t)$. Действительно,

$$F(-t) = \begin{vmatrix} -tE_1 & W_1 \\ W'_1 & -tE_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ W_{m-1} & \\ W'_{m-1} & -tE_m \end{vmatrix} =$$

$$= (-1)^n \begin{vmatrix} tE_1 & -W_1 \\ -W'_1 & tE_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ -W_{m-1} & \\ -W'_{m-1} & tE_m \end{vmatrix} = (-1)^n F(t).$$

Следовательно,

$$F(t) = t^{n-2k} (t^2 - c_1)(t^2 - c_2) \dots (t^2 - c_k). \quad (4)$$

Здесь k — число пар ненулевых корней полинома $F(t)$. Так как матрица B симметрична, то все корни полинома $F(t)$ вещественны. Поэтому $c_1 > 0; c_2 > 0; \dots; c_k > 0$. Положим $c_i = \mu_i^2, i = 1, 2, \dots, k; \mu_i > 0$. Таким образом, собственными значениями матрицы B являются 0 (кратности $n - 2k$) и $\pm \mu_1, \pm \mu_2, \dots, \pm \mu_k$. Нулевой корень может отсутствовать при четном n .

Так как матрица $A = E - B$ положительно определена, все числа $1 \pm \mu_i > 0$, откуда $\mu_i < 1$. Следовательно, в наших условиях метод простой итерации сходится. Быстрота сходимости определяется наибольшим из чисел μ_i , которое мы обозначим через μ .

Характеристическим полиномом матрицы S_q в силу (3) и (4) является

$$F_q(t) = q^n t^{\frac{n}{2}} \frac{(t+q-1)^{n-2k}}{q^{n-2k} t^{\frac{n}{2}-k}} \prod_{i=1}^k \left[\frac{(t+q-1)^2}{tq^2} - \mu_i^2 \right] =$$

$$= (t+q-1)^{n-2k} \prod_{i=1}^k [(t+q-1)^2 - q^2 \mu_i^2 t].$$

В частности, при $q = 1$ (метод Некрасова)

$$F_1(t) = t^{n-k} \prod_{i=1}^k (t - \mu_i^2).$$

Его корнями являются 0 (кратности $n - k$) и числа $\mu_1^2, \mu_2^2, \dots, \mu_k^2$. Отсюда мы заключаем, что метод Некрасова сходится вдвое быстрее, чем метод простой итерации.

Выясним теперь вопрос о наиболее целесообразном выборе множителя релаксации q .

Нулевым собственным значениям матрицы B соответствуют собственные значения матрицы S_q , равные $1 - q$. Собственным значениям $\pm \mu_i$ соответствуют два собственных значения матрицы S_q , определяемые из квадратного уравнения

$$(t + q - 1)^2 - q^2 \mu_i^2 t = 0.$$

Корни этого уравнения суть

$$\left(\frac{\mu_i q \pm \sqrt{\mu_i^2 q^2 - 4q + 4}}{2} \right)^2.$$

При $0 < q \leq \frac{2}{1 + \sqrt{1 - \mu_i^2}} = q_i$ эти корни будут вещественными и положительными, причем большим из них является

$$\left(\frac{\mu_i q + \sqrt{\mu_i^2 q^2 - 4q + 4}}{2} \right)^2.$$

При $q_i < q < 2$ корни становятся комплексными и их модули равны $q - 1$.

На плоскости q, t кривая третьего порядка $(t + q - 1)^2 - q^2 \mu_i^2 t = 0$ имеет двойную точку при $q = 0, t = 1$ и выпуклую петлю в полосе $0 \leq q \leq q_i$, касающуюся прямой $t = 0$ при $q = 1$ и прямой $q = q_i$ при $t = q_i - 1$.

Так как каждая прямая, параллельная оси q , пересекает кривую не более чем в двух точках, то прямая $t = 1$, проходящая через двойную точку, более не пересекает кривую. Поэтому петля кривой целиком расположена ниже прямой $t = 1$, и верхняя часть MN петли опускается при изменении q от 0 до q_i . Таким образом, график модуля большего из двух корней, соответствующих данному μ_i , имеет вид, изображенный сплошной линией на рис. 1.

При возрастании μ_i точка N сдвигается вправо, а участок кривой MN поднимается. Таким образом, наибольшее по модулю собственное значение матрицы S_q соответствует $\tilde{\mu}$.

Наи выгоднейшим значением множителя релаксации, очевидно, является

$$\tilde{q} = \frac{2}{1 + \sqrt{1 - \tilde{\mu}^2}}.$$

Интересно отметить, что при таком выборе множителя релаксации все собственные значения матрицы $S_{\tilde{q}}$ становятся по модулю

равными $\tilde{q} - 1 = \frac{1 - \sqrt{1 - \mu^2}}{1 + \sqrt{1 - \mu^2}}$. Действительно, нулевым корням матрицы B соответствуют корни $\tilde{q} - 1$, корням $\pm\mu_i$, при $\mu_i < \tilde{\mu}$, соответствуют пары сопряженных комплексных корней, равных по модулю $\tilde{q} - 1$, и, наконец, для $\pm\mu_i$, при $\mu_i = \tilde{\mu}$, оба корня совпадают и равны $\tilde{q} - 1$.

При $\tilde{q} < q < 2$ по тем же соображениям все собственные значения матрицы S_q по модулю равны $q - 1$.

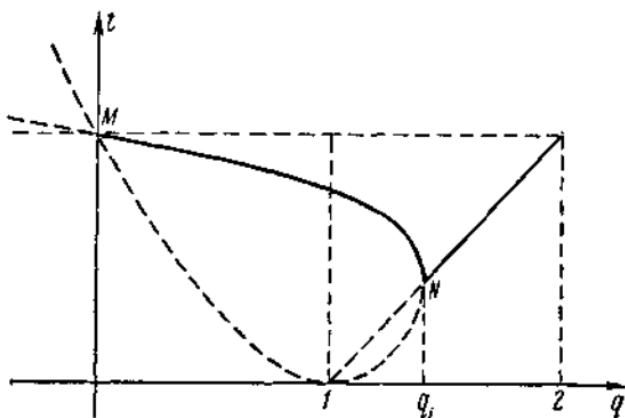


Рис. 1.

График зависимости от q наибольшего модуля собственных значений матрицы S_q имеет в точке $q = \tilde{q}$ резкий минимум с вертикальной касательной слева и касательной справа, образующей угол $\frac{\pi}{4}$ с осью абсцисс. Поэтому при отклонении q от \tilde{q} должно получаться резкое уменьшение быстроты сходимости, особенно при отклонении в сторону уменьшения.

В заключение заметим, что установленная связь между собственными значениями матриц B и S_q сохраняется для любой квазитрехдиагональной матрицы с единичными диагональными блоками, без предположения о симметрии и положительной определенности.

§ 37. Теорема сходимости

Теорема 37.1. Если в процессе неполной (или полной) релаксации для системы с положительно-определенной матрицей выполнены условия:

а) последовательность ведущих индексов i_1, \dots, i_k, \dots имеет интервал повторяемости, т. е. в каждом отрезке длины l i_{k+1}, \dots, i_{k+l} этой последовательности присутствуют хотя бы по одному разу все числа $1, 2, \dots, n$;

b) множители релаксации удовлетворяют условию $\varepsilon < q_k < 1 - \varepsilon$ при $0 < \varepsilon < 1$,

то процесс сходится к решению системы. Более того, существует число θ , $0 < \theta < 1$ такое, что $|X^* - X^{(k)}| \leq \theta^k$.

Доказательство. Пусть процесс решения системы $AX = F$ проходит по формуле

$$X^{(k)} = X^{(k-1)} + q_k \frac{r_{i_k}^{(k-1)}}{a_{i_k i_k}} e_{i_k} = X^{(k-1)} + m_k e_{i_k}, \quad (1)$$

где

$$m_k = \frac{r_{i_k}^{(k-1)}}{a_{i_k i_k}} q_k. \quad (2)$$

Здесь через i_k обозначен номер компоненты, изменяемой на k -м шагу процесса. Пусть $f(X^{(k)})$ есть значение функции ошибки на k -м шагу, $Y^{(k)}$ вектор ошибки на k -м шагу.

Тогда

$$f(X^{(k-1)}) - f(X^{(k)}) = \frac{(r_{i_k}^{(k-1)})^2}{a_{i_k i_k}} q_k (2 - q_k) = \frac{2 - q_k}{q_k} a_{i_k i_k} m_k^2, \quad (3)$$

$$Y^{(k)} = Y^{(k-1)} - m_k e_{i_k}. \quad (4)$$

Положительные числа $\frac{2 - q_k}{q_k} a_{i_k i_k}$ ограничены сверху и снизу, так что существуют такие константы γ_1 и γ_2 , что

$$\gamma_1 m_k^2 < f(X^{(k-1)}) - f(X^{(k)}) < \gamma_2 m_k^2.$$

Так как ряд из положительных членов $\sum_{k=1}^{\infty} f(X^{(k-1)}) - f(X^{(k)})$, очевидно, сходится, то сходящимся будет и ряд $\sum_{k=0}^{\infty} \gamma_1 m_k^2$, а вместе с ним

и ряд $\sum_{k=0}^{\infty} m_k^2$ и, следовательно, $m_k \rightarrow 0$. Так как компонента $r_{i_k}^{(k-1)}$ вектора невязки отличается от m_k ограниченным сверху и снизу множителем $\frac{a_{i_k i_k}}{q_k}$, мы устанавливаем, что $r_{i_k}^{(k-1)} \rightarrow 0$. Таким образом,

компонента вектора невязки с номером, равным номеру компоненты приближения, меняющейся на следующем шагу, стремится к нулю. Для сходимости процесса достаточно показать, что и все остальные компоненты вектора невязки стремятся к нулю при $k \rightarrow \infty$. Пусть t любое натуральное число, $1 \leq t \leq n$ и пусть $k > t$. Обозначим через k_t номер последнего шага, предшествующего k -му, при

котором менялась i -я компонента приближения. В силу условия о существовании промежутка повторяемости длины l имеет место неравенство $0 \leq k - k_i < l$. В силу доказанного $r_i^{(k_i-1)} \rightarrow 0$. Имеем далее

$$|r_i^{(k)}| \leq |r_i^{(k_i-1)}| + |r_i^{(k)} - r_i^{(k_i-1)}| \leq c |m_{k_i}| + |r_i^{(k)} - r_i^{(k_i-1)}|.$$

Оценим последнее слагаемое. Для этого предварительно оценим $|r^{(k)} - r^{(m)}|$, где $m < k$. Имеем

$$|Y^{(k)} - Y^{(m)}| = |m_k e_{ik}| = m_k.$$

Следовательно,

$$|Y^{(k)} - Y^{(m)}| \leq \sum_{v=m+1}^k |m_v|,$$

и потому

$$|r^{(k)} - r^{(m)}| = |A(Y^{(k)} - Y^{(m)})| \leq \|A\| \sum_{v=m+1}^k |m_v|.$$

Отсюда

$$|r_i^{(k)} - r_i^{(k_i-1)}| \leq |r^{(k)} - r^{(k_i-1)}| \leq \|A\| \sum_{v=k_i}^k |m_v| \leq \|A\| \sum_{v=k-l}^k |m_v|,$$

ибо $k_i \geq k - l$.

Следовательно,

$$|r_i^{(k)}| \leq c |m_{k_i}| + \|A\| \sum_{v=k-l}^k |m_v| \leq c_1 \sum_{v=k-l}^k |m_v|.$$

Таким образом, $r_i^{(k)} \rightarrow 0$ при $k \rightarrow \infty$ и тем самым сходимость процесса доказана.

Оценим быстроту сходимости. Из последнего неравенства следует, что

$$|r_i^{(k)}|^2 \leq c_1^2 (l+1) \sum_{v=k-l}^k |m_v|^2.$$

Следовательно,

$$|r^{(k)}|^2 = \sum_{i=1}^n |r_i^{(k)}|^2 \leq c_1^2 n (l+1) \sum_{v=k-l}^k |m_v|^2$$

и

$$f(X^{(k)}) = (A^{-1}r^{(k)}, r^{(k)}) \leq \|A^{-1}\| |r^{(k)}|^2 \leq c_2 M_k,$$

где

$$c_2 = c_1^2 n (l+1) \|A^{-1}\|, \quad M_k = \sum_{v=k-l}^k |m_v|^2.$$

С другой стороны,

$$f(X^{(k)}) = \sum_{v=k}^{\infty} (f(X^{(v)}) - f(X^{(v+1)})) \geq \gamma_1 \sum_{v=k}^{\infty} |m_v|^2 \geq \frac{\gamma_1}{t+1} \sum_{v=k}^{\infty} M_v.$$

Таким образом,

$$\sum_{v=k}^{\infty} M_v \leq c_3 M_k \quad (5)$$

при всех k и при некотором $c_3 > 1$. Обозначим

$$\sum_{v=k}^{\infty} M_v = S_k.$$

Тогда неравенство (5) представится в виде

$$S_k \leq c_3 (S_k - S_{k+1}),$$

откуда

$$S_{k+1} \leq \frac{c_3 - 1}{c_3} S_k = \theta_1 S_k \quad (\text{где } 0 < \theta_1 < 1)$$

и, следовательно, $S_k \leq c_4 \theta_1^k$. Далее,

$$f(X^{(k)}) = \sum_{v=k+1}^{\infty} (f(X^{(v-1)}) - f(X^{(v)})) \leq \gamma_2 \sum_{v=k+1}^{\infty} m_v^2 < \gamma_2 S_k < c_4 \gamma_2 \theta_1^k.$$

Итак,

$$f(X^{(k)}) = O(\theta_1^k).$$

Вместе с этой оценкой справедливы и оценки

$$|r^{(k)}| = O(\theta_1^{k/2}); \quad |Y^{(k)}| = O(\theta_1^{k/2}) = O(\theta^k),$$

где $\theta = \theta_1^{1/2}$.

Сходимость процесса может быть доказана и при более слабых предположениях относительно множителей релаксации (А. М. Островский [8]).

Из доказанной теоремы следует, в частности, сходимость циклического релаксационного процесса с постоянным множителем релаксации q , удовлетворяющим неравенству $0 < q < 2$.

Сделаем еще одно замечание, касающееся этого случая. Полином

$$F_q(t) = \begin{vmatrix} (t+q-1)a_{11} & qa_{12} & \cdots & qa_{1n} \\ tqa_{21} & (t+q-1)a_{22} & \cdots & qa_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ tqa_{n1} & tqa_{n2} & \cdots & (t+q-1)a_{nn} \end{vmatrix}.$$

максимум модулей корней которого определяет быстроту сходимости процесса, имеет старший коэффициент $a_{11}a_{22}\dots a_{nn}$ и свободный член $(q-1)^n a_{11}a_{22}\dots a_{nn}$. Поэтому произведение его корней равно $(-1)^n(q-1)^n$. Отсюда следует, что максимум его корней не меньше $|q-1|$ и может равняться этому числу, только если все его корни равны по модулю $|q-1|$.

Если $0 < q < 2$, то корни полинома $F_q(t)$ строго меньше единицы по модулю, так как релаксационный процесс сходится. В силу непрерывной зависимости корней от коэффициентов полинома, на границах интервала для q , т. е. при $q=0$ и $q=2$, корни $F_q(t)$ не превосходят единицы. Но, при $q=0$ и $q=2$, $1=|q-1|$ и, следовательно, в силу сделанного выше замечания, все корни $F_q(t)$ равны по модулю единице.

Это обстоятельство для $q=0$ легко проверяется непосредственно, ибо $F_0(t)=a_{11}\dots a_{nn}(t-1)^n$. Для $q=2$ доказанное обстоятельство не тривиально. Для квазитреугольных матриц оно было установлено в § 36 прямым подсчетом.

§ 38. Управление релаксацией

Вместо того, чтобы при проведении процесса координатной релаксации задаваться последовательностью ведущих индексов a priori (как это делается, например, в циклическом процессе или в процессе, где управление задается десятичной записью числа e), можно выбирать индекс i_k на каждом шагу, исходя из результатов предыдущего шага. Так, например, самое быстрое убывание функции ошибки при переходе от $X^{(k-1)}$ к $X^{(k)}$ получается, если выбирать индекс i_k так, чтобы

число $\frac{(r_{i_k}^{(k-1)})^2}{a_{i_k} i_k}$ было бы наибольшим среди всех чисел $\frac{(r_i^{(k-1)})^2}{a_{ii}}$, где

$i=1, \dots, n$. Иными словами, индекс i_k выбирается так, чтобы

число $\frac{|r_{i_k}^{(k-1)}|}{\sqrt{a_{i_k} i_k}}$ было бы наибольшим. Процесс релаксации с таким управлением носит название процесса Зейделя.

С именем Гаусса связывается процесс, при котором ведущий индекс выбирается из условия $\max \frac{|r_i^{(k-1)}|}{a_{ii}}$. Наконец, правило управления Сауссела определяется вычислением $\max |r_i^{(k-1)}|$.

Теорема 38.1. Если ведущий индекс i_k релаксационного координатного процесса для системы с положительно-определенной матрицей A выбирается так, что

$$|r_{i_k}^{(k-1)}| \geq \gamma |r_i^{(k-1)}| \quad (0 < \gamma \leq 1, i=1, \dots, n), \quad (1)$$

и если $\varepsilon < q_k < 2 - \varepsilon$, то процесс неполной релаксации сходится к решению системы X^* и существует число θ , $0 < \theta < 1$ такое, что $|X^* - X^{(k)}| < \theta^k$.

Из этой теоремы следует сходимость процессов неполной релаксации Гаусса, Зейделя и Сауссела, так как эти процессы удовлетворяют условию теоремы при $\gamma = \frac{\min a_{ii}}{\max a_{ii}}$ для процесса Гаусса, при $\gamma = \sqrt{\frac{\min a_{ii}}{\max a_{ii}}}$ для процесса Зейделя и при $\gamma = 1$ для процесса Сауссела. Действительно, для метода Сауссела это очевидно. В методе Гаусса имеем

$$\frac{|r_{i_k}^{(k-1)}|}{a_{i_k i_k}} \geq \frac{r_i^{(k-1)}}{a_{ii}} \quad \text{при } i = 1, \dots, n,$$

и потому

$$|r_{i_k}^{(k-1)}| \geq \frac{a_{i_k i_k}}{a_{ii}} |r_i^{(k-1)}| \geq \frac{\min a_{ii}}{\max a_{ii}} |r_i^{(k-1)}|.$$

Аналогичные неравенства имеют место и для метода Зейделя.

Доказательство теоремы. Введем те же обозначения, что и при доказательстве теоремы предыдущего параграфа. Повторяя доказательство упомянутой теоремы, получим, что ряд $\sum m_k^2$ сходится, $m_k \rightarrow 0$ и $r_{i_k}^{(k-1)} \rightarrow 0$. Далее, из неравенства

$$|r_i^{(k-1)}| \leq \frac{1}{\gamma} |r_{i_k}^{(k-1)}| \quad (2)$$

заключаем, что $|r_i^{(k-1)}| \rightarrow 0$ при всех $i = 1, \dots, n$. Тем самым сходимость процесса доказана.

Для оценки быстроты сходимости используем вытекающее из (2) неравенство

$$|r^{(k-1)}| \leq \frac{n}{\gamma} |r_{i_k}^{(k-1)}| \leq c_1 m_k.$$

Далее,

$$f(X^{(k-1)}) = (A^{-1} r^{(k-1)}, r^{(k-1)}) \leq \|A^{-1}\| c_1^2 m_k^2.$$

С другой стороны,

$$f(X^{(k-1)}) - f(X^{(k)}) \geq \gamma_1 m_k^2,$$

и, следовательно,

$$f(X^{(k-1)}) \geq \gamma_1 \sum_{i=k}^{\infty} m_i^2.$$

Положив

$$s_k = \sum_{i=k}^{\infty} m_i^2,$$

получим

$$\gamma_1 s_k \leq \|A^{-1}\| c_1 m_k^2,$$

откуда

$$s_k \leq c_2 (s_k - s_{k+1}) \quad (c_2 > 1)$$

и

$$s_{k+1} \leq \frac{c_2 - 1}{c_2} s_k = \theta_1 s_k \quad (0 < \theta_1 < 1).$$

Таким образом,

$$s_k < c_3 \theta_1^k$$

и, следовательно,

$$f(X^{(k)}) \leq \gamma_2 \sum_{i=k}^{\infty} m_i^2 = \gamma_2 s_k \leq c_4 \theta_1^k. \quad (3)$$

Справедливы и оценки

$$|r^{(k)}| = O(\theta^k), \quad |Y^{(k)}| = O(\theta^k), \quad \theta = \theta_1^{1/2}.$$

При пользовании релаксационными методами с управлением Сауселла, Гаусса, Зейделя на каждом шагу нужно вычислять все компоненты вектора невязки, но затем при определении ведущего индекса для следующего шага учитывается только одна. Поэтому представляют интерес такие вычислительные схемы, в которых использовались бы все эти компоненты. Для построения таких схем следует воспользоваться соотношением между двумя соседними векторами невязки.

Именно, из соотношения

$$X^{(k)} = X^{(k-1)} + q_k \frac{r^{(k-1)}_{i_k}}{a_{i_k i_k}} e_{i_k}$$

следует

$$r^{(k)} = r^{(k-1)} - q_k \frac{r^{(k-1)}_{i_k}}{a_{i_k i_k}} A e_{i_k} = r^{(k-1)} - q_k \frac{r^{(k-1)}_{i_k}}{a_{i_k i_k}} A_{i_k},$$

где через A_{i_k} обозначен столбец с номером i_k матрицы A .

Переходя к компонентам, получим

$$r_i^{(k)} = r_i^{(k-1)} - r_{i_k}^{(k-1)} \frac{a_{ii_k}}{a_{i_k i_k}} q_k.$$

Положим

$$\gamma_i^{(k)} = \frac{r_i^{(k)}}{a_{ii}}, \quad \sigma_i^{(k)} = \frac{r_i^{(k)}}{\sqrt{a_{ii}}}. \quad (4)$$

Тогда

$$\begin{aligned} r_i^{(k)} &= \gamma_i^{(k-1)} - \gamma_{i_k}^{(k-1)} \frac{a_{ii_k}}{a_{i_k i_k}} q_k \\ \sigma_i^{(k)} &= \sigma_i^{(k-1)} - \sigma_{i_k}^{(k-1)} \frac{a_{ii_k}}{\sqrt{a_{ii}} \sqrt{a_{i_k i_k}}} q_k. \end{aligned}$$

Положив

$$\frac{a_{ij}}{a_{jj}} = b_{ij}; \quad \frac{a_{ij}}{a_{ii}} = c_{ij}; \quad \frac{a_{ij}}{\sqrt{a_{ii} a_{jj}}} = h_{ij}, \quad (5)$$

получим

$$\begin{aligned} r_i^{(k)} &= r_i^{(k-1)} - q_k r_{i_k}^{(k-1)} b_{ii_k} \\ \gamma_i^{(k)} &= \gamma_i^{(k-1)} - q_k \gamma_{i_k}^{(k-1)} c_{ii_k} \\ \sigma_i^{(k)} &= \sigma_i^{(k-1)} - q_k \sigma_{i_k}^{(k-1)} h_{ii_k}, \end{aligned} \quad (6)$$

или в векторной форме

$$\begin{aligned} r^{(k)} &= r^{(k-1)} - q_k r_{i_k}^{(k-1)} B_{i_k} \\ \gamma^{(k)} &= \gamma^{(k-1)} - q_k \gamma_{i_k}^{(k-1)} C_{i_k} \\ \sigma^{(k)} &= \sigma^{(k-1)} - q_k \sigma_{i_k}^{(k-1)} H_{i_k}. \end{aligned} \quad (6')$$

Здесь B_j , C_j и H_j j -е столбцы соответствующих матриц (b_{ij}) , (c_{ij}) и (h_{ij}) , которые, очевидно, равны соответственно $D^{-1}A$, AD^{-1} , $D^{-1/2}AD^{-1/2}$. Для проведения выбранного процесса соответствующая вспомогательная матрица должна быть составлена заранее. Заранее должны быть выбраны и коэффициенты релаксации q_1 , q_2 , ...

Далее вычисления располагаются так.

При помощи начального приближения составляются n чисел $r_i^{(0)}$ (или $\gamma_i^{(0)}$ или $\sigma_i^{(0)}$). Из них выбирается наибольшее по модулю, подчеркивается и его номер принимается за i_1 . По формулам (6) составляются числа $r_i^{(1)}$ (или $\gamma_i^{(1)}$ или $\sigma_i^{(1)}$), из них выбирается наибольшее по модулю, подчеркивается и его номер принимается за i_2 и т. д. После того, как выполнено достаточное число N шагов, компоненты приближения находятся по формулам:

$$X_i^{(N)} = X_i^{(0)} + \frac{1}{a_{ii}} \sum' q_k r_i^{(k-1)}$$

или

$$X_i^{(N)} = X_i^{(0)} + \sum' q_k \gamma_i^{(k-1)}$$

или

$$X_i^{(N)} = X_i^{(0)} + \frac{1}{\sqrt{a_{ii}}} \sum' q_k \sigma_i^{(k-1)}.$$

Сумма \sum' распространяется на те значения k , для которых $r_i^{(k-1)}$ (или $\gamma_i^{(k-1)}$ или $\sigma_i^{(k-1)}$) были подчеркнуты.

Для уменьшения влияния ошибок округления, процесс следует время от времени прерывать с тем, чтобы, вычислив приближение, найти невязку непосредственно и начать процесс сначала,

Вычислительная схема становится особенно простой, если все $a_{ij} = 1$. В этом случае все три способа управления совпадают и нет надобности в вычислении вспомогательных матриц, ибо каждая из них совпадает с исходной.

Если систему $AX = F$ преобразовать посредством подстановки $X = D^{-\frac{1}{2}}Y$ и умножением на $D^{-\frac{1}{2}}$ слева к виду

$$D^{-\frac{1}{2}}AD^{-\frac{1}{2}}Y = D^{-\frac{1}{2}}F,$$

матрица которой симметрична и имеет равные единице диагональные элементы, то применение релаксационного процесса с управлением равносильно применению метода Зейделя к исходной системе.

Таблица III. 11

Вычисление последовательных приближений релаксационным процессом с управлением при $q = 1$

k	X	0	0	0	0	t_k
1	$F - AX$	0.3	0.5	0.7	0.9	4
2		-0.294	0.104	0.502	0	3
3		-0.56508	-0.05664	0	-0.11044	1
4		0	0.1806936	0.3051432	0.2625128	3
5		-0.16477733	0.08304778	0	0.19538130	4
6		-0.29372899	-0.00291999	-0.04298389	0	1
7		0	0.12044619	0.11562976	0.19386113	4
8		-0.12794835	0.03514729	0.07298031	0	1
	$\sum r_{i_k}^{(k-1)}$	-0.98675734	0	0.80714320	1.28924243	
	X	-0.98675734	0	0.80714320	1.28924243	
9	$F - AX$	
.....		
100		-0.00000007	-0.00000004	0.00000003	0	
	X	-1.2577936	0.0434874	1.0391661	1.4823928	

В табл. III. 11 приведены результаты применения полного релаксационного процесса ($q_k = 1$) с управлением к системе (9) § 23.

Последовательные векторы невязки вычисляются по формуле

$$r^{(k)} = r^{(k-1)} - r_{i_k}^{(k-1)} A_{ik},$$

где через A_j обозначен j -й столбец матрицы A .

Таким образом, решение системы, с точностью до $2 \cdot 10^{-7}$ в каждой компоненте, получено через 100 элементарных шагов, что эквивалентно приблизительно 25 простым итерациям.

Применение процесса с управлением при $q_k = 1.2$ приводит к такому же результату через 52 элементарных шага, при $q_k = 0.8$ через 148 элементарных шагов.

§ 39. Релаксация по длине вектора невязки

Рассмотрим теперь процессы, основанные на минимизации или уменьшении длины вектора невязки за счет изменения, на каждом шагу, одной компоненты предыдущего приближения.

Пусть X некоторое приближение и i номер изменяемой компоненты. Положим

$$X' = X + \delta e_i.$$

Тогда

$$r' = r - \delta A_i,$$

и потому

$$\begin{aligned} (r', r') &= (r - \delta A_i, r - \delta A_i) = (r, r) - 2\delta(r, A_i) + \delta^2(A_i, A_i) = \\ &= (r, r) + \frac{1}{(A_i, A_i)} [(r, A_i) - \delta(A_i, A_i)]^2 - \frac{(r, A_i)^2}{(A_i, A_i)}. \end{aligned}$$

Здесь A_i — i -й столбец матрицы A . Величина (r', r') будет минимальной при

$$\delta = \frac{(r, A_i)}{(A_i, A_i)},$$

так что в случае полной релаксации можно взять

$$X' = X + \frac{(r, A_i)}{(A_i, A_i)} e_i. \quad (1)$$

Если заданная система предварительно приведена к такому виду, что все $(A_i, A_i) = 1$ (что всегда можно сделать посредством замены неизвестных $x_i = \frac{\bar{x}_i}{\sqrt{(A_i, A_i)}}$), мы будем иметь

$$X' = X + (r, A_i) e_i \quad (2)$$

$$(r', r') = (r, r) - (r, A_i)^2. \quad (3)$$

Неполная релаксация в этом случае может проводиться по формулам

$$X' = X + q(r, A_i) e_i \quad (0 < q < 2), \quad (4)$$

причем

$$(r', r') = (r, r) - q(2 - q)(r, A_i)^2. \quad (5)$$

Нетрудно видеть, что метод релаксации по длине вектора невязки равносителен релаксационному методу по функции ошибки для системы $A'AX = A'F$, полученной из данной системы $AX = F$ первой трансформацией Гаусса. Действительно,

$$(r, r) = (AY, AY) = (A'AY, Y),$$

где Y вектор ошибки, так что квадрат длины вектора невязки для системы $AX = F$ есть функция ошибки для системы $A'AX = A'F$.

Если последовательность номеров изменяемых компонент задана a priori (например, если применять циклический процесс), то проведение процесса по формулам (2) или (4) не требует фактического выполнения трансформации Гаусса, т. е. вычисления матрицы $A'A$. Если же осуществлять управление процессом, требуя на каждом шагу максимального убывания длины вектора невязки за счет выбора номера изменяемой компоненты, то, как справедливо отмечает Покорина [1], целесообразно предварительно вычислить матрицу $A'A$ и вести процесс (при $(A_i, A_i) = 1$) по формулам

$$r^{(k)} = r^{(k-1)} + r_{i_k}^{(k-1)} B_{i_k},$$

где i_k номер наибольшей по модулю компоненты вектора $r^{(k-1)}$, B_i есть i -й столбец матрицы $A'A$. Это равносильно проведению релаксационного процесса с управлением для системы $A'AX = A'F$.

§ 40. Групповая релаксация

Пусть $AX = F$ данная система.

Отдельный шаг метода групповой релаксации заключается в следующем. Выделяется группа G индексов и при переходе от предшествующего приближения к следующему изменяются только компоненты с индексами из выбранной группы G . Изменение осуществляется так, что уравнения, индексы которых входят в группу G , удовлетворяются точно. Иначе говоря, изменяемые компоненты суть решения системы уравнений

$$\sum_{j \in G} a_{ij} x'_j = f_i - \sum_{m \notin G} a_{im} x_m, \quad (1)$$

где индекс i пробегает всю группу G .

Легко видеть, что если матрица A положительно-определенная, то один шаг групповой релаксации минимизирует функцию ошибок

в подпространстве, натянутом на единичные векторы с индексами, образующими группу G .

Метод групповой релаксации допускает много модификаций, в зависимости от принципа выбора групп G на каждом шагу. Простейшей модификацией является циклический групповой процесс, в котором все индексы раз навсегда разбиваются на несколько не перекрывающихся групп и по ходу процесса эти группы циклически чередуются. Этот процесс может рассматриваться как процесс последовательных приближений при надлежащей подготовке системы. Для того чтобы показать это, положим, для простоты, что группы состоят из последовательных индексов и чередование осуществляется в порядке их возрастания. Обозначим через D квазидиагональную матрицу, „вырезанную“ из матрицы A в соответствии с разбиением индексов на группы, через L матрицу, состоящую из элементов матрицы A , лежащих налево от элементов матрицы D , через R матрицу, состоящую из элементов A , лежащих направо от элементов матрицы D . Тогда

$$A = L + D + R. \quad (2)$$

Пусть $X^{(k)}$ приближение, полученное после завершения k циклов процесса. Тогда $X^{(k+1)}$ и $X^{(k)}$, очевидно, связаны соотношением

$$LX^{(k+1)} + DX^{(k+1)} + RX^{(k)} = F,$$

и потому

$$X^{(k+1)} = (L + D)^{-1}F - (L + D)^{-1}RX^{(k)}. \quad (3)$$

Таким образом, изучаемый процесс равносителен процессу последовательных приближений, примененному к системе, подготовленной к виду

$$X = -(L + D)^{-1}RX + (L + D)^{-1}F.$$

Отсюда вытекает необходимое и достаточное условие сходимости процесса, которое заключается в том, чтобы собственные значения матрицы $(L + D)^{-1}R$ были по модулю меньше единицы. Если матрица A положительно-определенная эрмитова матрица, то процесс сходится всегда. Действительно, в этом случае собственные значения матрицы $(L + D)^{-1}R = (R^* + D)^{-1}R$ по модулю не превосходят $\frac{\Lambda_0}{\sqrt{\lambda_0 d_0 + \Lambda_0^2}}$.

Здесь через Λ_0 обозначено $\|R\| = \|R^*\|$, через λ_0 — наименьшее собственное значение матрицы A , через d_0 — наименьшее собственное значение матрицы D . Доказательство этого неравенства проводится совершенно аналогично доказательству, проведенному для процесса Некрасова. При этом используется то обстоятельство, что вырезанная матрица D всегда положительно определена. Справедливо неравенство $d_0 \geq \lambda_0$.

При свободной групповой релаксации допускается не только не циклическое чередование заранее выбранных групп, но и изменение

состава группы на каждом шагу. Как мы видели, отдельный шаг с выбранной группой $G^{(k)}$ осуществляется по формулам

$$\sum_{j \in G^{(k)}} a_{ij} x_j^{(k+1)} = f_i - \sum_{m \in G^{(k)}} a_{im} x_m^{(k)} \quad (i \in G^{(k)})$$

$$x_j^{(k+1)} = x_j^{(k)} \text{ при } j \notin G^{(k)}.$$

Эти формулы равносильны формулам

$$\sum_{j \in G^{(k)}} a_{ij} [x_j^{(k+1)} - x_j^{(k)}] = f_i - \sum_{m=1}^n a_{im} x_m^{(k)} = r_i^{(k)} \quad (i \in G^{(k)}) \quad (4)$$

$$x_j^{(k+1)} = x_j^{(k)}, \quad j \notin G^{(k)}.$$

Обозначив

$$x_j^{(k+1)} - x_j^{(k)} = \delta_j^{(k)},$$

получим

$$x_j^{(k+1)} = x_j^{(k)} + \delta_j^{(k)} \quad (j \in G^{(k)})$$

$$x_j^{(k+1)} = x_j^{(k)} \quad (j \notin G^{(k)}),$$

где $\delta_j^{(k)}$ есть решения системы

$$\sum_{j \in G^{(k)}} a_{ij} \delta_j^{(k)} = r_i^{(k)} \quad (i \in G^{(k)}). \quad (5)$$

При неполной релаксации вычисление проводится по формулам

$$x_j^{(k+1)} = x_j^{(k)} + q_k \delta_j^{(k)} \quad (j \in G^{(k)})$$

$$x_j^{(k+1)} = x_j^{(k)} \quad (j \notin G^{(k)}), \quad (6)$$

где $0 < q_k < 2$.

Для положительно-определенной матрицы имеет место следующая теорема сходимости. Если существует число l такое, что любая последовательность групп длины l содержит каждый индекс хотя бы раз, и если $\varepsilon < q_k < 2 - \varepsilon$, то процесс групповой свободной неполной релаксации сходится к решению системы. (Островский [8]).

Доказательство этой теоремы аналогично доказательству соответствующей теоремы для одношагового процесса.

Как правило, сходимость группового процесса оказывается более быстрой по сравнению со сходимостью соответствующего одношагового процесса.

Дополнительная же работа, заключающаяся в решении вспомогательных систем, легко осуществляется при наличии подпрограммы для решения системы фиксированного порядка.

Конечно, применять метод групповой релаксации имеет смысл лишь для систем с большим числом уравнений.

ГЛАВА IV

ПОЛНАЯ ПРОБЛЕМА СОБСТВЕННЫХ ЗНАЧЕНИЙ

Под полной проблемой собственных значений понимается проблема нахождения всех собственных значений матрицы A , так же как и принадлежащих этим собственным значениям собственных векторов (или векторов, образующих канонический базис). Напомним, что собственными значениями матрицы A называются корни ее характеристического полинома, т. е. корни уравнения

$$|A - tE| = \begin{vmatrix} a_{11} - t & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - t & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - t \end{vmatrix} = (-1)^n [t^n - p_1 t^{n-1} - \dots - p_n] = 0.$$

Определение компонент собственного вектора требует решения системы n однородных уравнений с n неизвестными; для вычисления всех собственных векторов матрицы, вообще говоря, требуется решить n систем вида

$$(A - \lambda_i E) X_i = 0,$$

где $X_i = (x_{1i}, \dots, x_{ni})$ — собственный вектор матрицы A , принадлежащий собственному значению λ_i .

Как уже отмечалось в § 1, п. 8 коэффициенты p_i характеристического полинома являются, с точностью до знака, суммами всех миноров определителя матрицы A порядка i , опирающихся на главную диагональ. Непосредственное вычисление коэффициентов p_i является чрезвычайно громоздким и требует огромного числа операций.

Совершенно естественно, поэтому, появление специальных вычислительных приемов, упрощающих численное решение поставленных задач. Большинство методов, дающих решение полной проблемы собственных значений, включает предварительное вычисление коэффициентов характеристического полинома, которое осуществляется теми или иными средствами, минуя вычисление многочисленных миноров. Собственные значения вычисляются затем по какому-либо методу для приближенного вычисления корней полинома. Одним из лучших

способов приближенного вычисления корней полинома $\varphi(t) = a_0 t^n + a_1 t^{n-1} + \dots + a_{n-1} t + a_n$ является способ Ньютона. Именно, если t_0 есть некоторое исходное приближение к корню, то последовательные приближения t_1, t_2, \dots вычисляются по формуле

$$t_i = t_{i-1} - \frac{\varphi(t_{i-1})}{\varphi'(t_{i-1})} \quad (i = 1, 2, \dots).$$

Если t_0 взято достаточно близко к искомому корню λ , то последовательные приближения t_i сходятся к λ с квадратичной сходимостью, т. е. погрешность последующего приближения будет иметь порядок квадрата предыдущей:

$$|\lambda - t_i| \leq c |\lambda - t_{i-1}|^2,$$

где c некоторая константа.

Вычисление значений полинома $\varphi(t)$ и его производной в данной точке t_0 следует проводить по схеме Хорнера:

$$\begin{array}{cccccc} a_0 & a_1 & a_2 & \dots & a_{n-1} & a_n \| t_0 \\ a_0 & b_1 & b_2 & \dots & b_{n-1} & b_n \\ a_0 & c_1 & c_2 & \dots & c_{n-1} & \end{array}$$

которая заполняется по рекуррентным формулам

$$b_i = b_{i-1} t_0 + a_i \quad (i = 1, 2, \dots, n)$$

$$c_i = c_{i-1} t_0 + b_i \quad (i = 1, 2, \dots, n-1).$$

Тогда

$$b_n = \varphi(t_0), \quad c_{n-1} = \varphi'(t_0).$$

Числа же a_0, b_1, \dots, b_{n-1} будут коэффициентами частного $\varphi_1(t)$ от деления полинома $\varphi(t)$ на $t - t_0$. Соответственно числа a_0, c_1, \dots, c_{n-1} будут коэффициентами частного $\varphi_2(t)$ от деления $\varphi_1(t)$ на $t - t_0$.

По большей части собственные векторы матрицы удается определить, используя промежуточные результаты вычислений, проведенных для определения коэффициентов характеристического полинома. Конечно, для определения собственного вектора, принадлежащего тому или другому собственному значению, это собственное значение должно быть уже вычислено. Методы этой группы являются точными, т. е. если их осуществлять для матриц, элементы которых заданы точно (рациональными числами) и вычисления проводить точно (по правилам действий над обыкновенными дробями), то в результате будет получено точное значение коэффициентов характеристического полинома, и компоненты собственных векторов окажутся выражеными точными формулами через собственные значения.

Наряду с точными методами для решения проблемы собственных значений имеются методы итерационные, в которых собственные значения получаются как пределы некоторых числовых последовательностей.

довательностей, так же как и компоненты принадлежащих им собственных векторов. В итерационных методах, как правило, собственные значения вычисляются непосредственно, без предварительного вычисления коэффициентов характеристического полинома. Это существенно упрощает задачу, так как вычисление корней полинома, коэффициенты которого известны, достаточно трудоемко.

Однако итерационные методы более приспособлены к решению частичной проблемы собственных значений. Под частичной проблемой мы подразумеваем задачу нахождения одного или нескольких собственных значений и соответствующих им собственных векторов.

Полная и частичная проблемы собственных значений совершенно различны как по методам их решения, так и по области приложений. Решение полной проблемы для матриц даже не очень высокого порядка неизбежно оказывается весьма громоздким, и возможность решения частичной проблемы, минуя тяжести решения полной, является очень ценной для практики.

Настоящая глава посвящена изложению точных методов для решения полной проблемы собственных значений. Итерационные методы решения полной проблемы будут рассмотрены в гл. VIII, частичная же проблема будет изучаться, начиная со следующей главы.

Отметим, что все предлагаемые ниже методы (как в этой главе, так и в последующих), кроме метода Леверье (1840 г.) и метода Якоби (1846 г.), появились в тридцатых годах нашего столетия или позднее.

При изложении численных методов мы будем, как правило, предполагать элементы матриц вещественными.

§ 41. Устойчивость проблемы собственных значений

При постановке проблемы собственных значений для матриц, элементы которых заданы приближенно, естественно возникает вопрос об устойчивости полученного решения, иными словами, вопрос о том, как изменяются собственные значения и собственные векторы при изменении элементов данной матрицы в пределах допустимой погрешности.

То, что в отдельных случаях проблема собственных значений не может быть устойчивой, ясно из следующих соображений. Допустим, что данная матрица, если ее численное задание рассматривать как точное, имеет лишь простые собственные значения, однако, при некотором определенном изменении ее элементов в пределах точности задания можно прийти к матрице, имеющей кратное собственное значение, с нелинейным элементарным делителем. В этом случае каноническая форма матрицы при изменении ее элементов в пределах точности задания претерпевает качественное изменение, переходя от чисто диагональной формы к общей канонической форме. В частности, даже число собственных векторов изменяется

скаккообразно. В этих условиях, конечно, полная проблема собственных значений, вместе с определением собственных векторов, просто теряет смысл. В условиях же, близких к описанной ситуации, проблема определения собственных векторов наверное не имеет устойчивого решения.

Пусть A данная матрица и $A + dA$ близкая к ней матрица. Выясним как изменяются собственные значения и собственные векторы матрицы A , когда она получает приращение dA . Проведем подсчет в предположении, что все собственные значения матрицы A различны, отбрасывая величины второго порядка малости, т. е. будем рассматривать dA (и соответственно dX и $d\lambda$) как дифференциалы, а не как конечные приращения.

Пусть

$$AX_i = \lambda_i X_i \quad (i = 1, 2, \dots, n), \quad (1)$$

Тогда

$$(dA) X_i + A dX_i = \lambda_i dX_i + d\lambda_i X_i. \quad (2)$$

Пусть V_1, \dots, V_n — собственные векторы сопряженной матрицы A^* , соответствующие собственным значениям $\bar{\lambda}_1, \dots, \bar{\lambda}_n$. Тогда

$$((dA) X_i, V_j) + (A(dX_i), V_j) = \lambda_i (dX_i, V_j) + d\lambda_i (X_i, V_j). \quad (3)$$

Положив в равенстве (3) $i = j$, получим

$$((dA) X_i, V_i) + (A(dX_i), V_i) = (dX_i, \bar{\lambda}_i V_i) + d\lambda_i (X_i, V_i),$$

откуда

$$d\lambda_i = \frac{((dA) X_i, V_i)}{(X_i, V_i)}, \quad (4)$$

ибо $A^*V_i = \bar{\lambda}_i V_i$.

Положим теперь $i \neq j$. Тогда, в силу равенств $(X_i, V_j) = 0$ и $(A(dX_i), V_j) = \lambda_j (dX_i, V_j)$, получим

$$(\lambda_i - \lambda_j) (dX_i, V_j) = ((dA) X_i, V_j),$$

откуда

$$(dX_i, V_j) = \frac{((dA) X_i, V_j)}{\lambda_i - \lambda_j}.$$

Пусть

$$dX_i = \sum_{j=1}^n \alpha_{ij} X_j. \quad (5)$$

Тогда

$$(dX_i, V_j) = \alpha_{ij} (X_j, V_j),$$

и, следовательно,

$$\alpha_{ij} = \frac{((dA) X_i, V_j)}{(X_j, V_j)(\lambda_i - \lambda_j)} \quad \text{при } i \neq j. \quad (6)$$

Коэффициент α_{ii} остается, естественно, неопределенным, в силу неоднозначности собственного вектора, и без нарушения общности можно считать, что $\alpha_{ii} = 0$.

Перейдем теперь к оценкам. Из формулы (4) получим

$$|d\lambda_i| \leq \frac{\|dA\| \cdot |X_i| \cdot |V_i|}{|(X_i, V_i)|} = c_i \|dA\|,$$

где

$$c_i = \frac{|X_i| \cdot |V_i|}{|(X_i, V_i)|}. \quad (7)$$

Ясно, что $c_i \geq 1$. Если собственные векторы вещественны, то

$$c_i = \frac{1}{|\cos \varphi_i|},$$

где φ_i угол между векторами X_i и V_i .

Число c_i назовем коэффициентом перекоса матрицы A , соответствующим собственному значению λ_i . Таким образом, изменение λ_i при данной $\|dA\|$ может быть тем больше, чем больше соответствующий этому собственному значению коэффициент перекоса c_i . Для нормальных матриц, в частности, для эрмитовых и унитарных матриц

$$|d\lambda_i| \leq \|dA\|,$$

ибо для нормальных матриц $X_i = V_i$. Поэтому для нормальных матриц задача определения собственных значений всегда устойчива. Для произвольных же матриц задача определения собственных значений будет не устойчивой только при большом коэффициенте перекоса.

Что же касается определения собственных векторов, то, как показывают формулы (5) и (6),

$$|dX_i| \leq \|dA\| \cdot |X_i| \cdot \sum_{j=1}^n \frac{c_j}{|\lambda_i - \lambda_j|}, \quad (8)$$

так что задача может быть неустойчивой, только если велик хоть один коэффициент перекоса или если имеются близкие собственные значения.

Приведем теперь результаты, касающиеся изменения собственных значений вещественной матрицы при случайных изменениях элементов матрицы. Пусть элементы матрицы A являются независимыми случайными величинами со средними значениями a_{ij} и с одной и той же дисперсией σ^2 . Тогда любое вещественное собственное значение будет случайной величиной, имеющей в первом приближении дисперсию $c(\lambda) \sigma^2$, где $c(\lambda)$ коэффициент перекоса, соответствующий этому собственному значению¹⁾. В случае, если $a_{ij} = a_{ji}$, т. е. матрица из

¹⁾ Д. К. Фаддеев [4].

средних значений симметрична, но допускаются не симметричные вариации, дисперсия каждого собственного значения равна σ^2 . Для симметричных же матриц при допустимых симметричных изменениях ее элементов такой же результат получается, если предположить, что дисперсия диагональных элементов равна σ^2 , а недиагональных $\frac{1}{2}\sigma^2$. Собственные значения в этом случае являются (в первом приближении) независимыми случайными величинами. Что же касается коэффициентов характеристического полинома, то их дисперсии могут быть довольно большими для каждого отдельного коэффициента, однако при этом коэффициенты будут уже не независимыми и вероятные их изменения связаны так, что они не приводят к большим изменениям собственных значений. Тем не менее неосторожное округление коэффициентов в процессе их вычисления может нарушить их взаимную связь и привести затем к неправильным значениям собственных чисел.

Рассмотрим пример, иллюстрирующий сказанное выше. Возьмем

$$A = \begin{bmatrix} 5 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix}, \quad A + dA = \begin{bmatrix} 5.1 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix}.$$

Как мы видели, матрица A плохо обусловлена. Характеристические полиномы для матриц A и $A + dA$ соответственно будут

$$t^4 - 35t^3 + 146t^2 - 100t + 1$$

и

$$t^4 - 35.1t^3 + 149t^2 - 110.6t + 7.8,$$

и потому собственными значениями первой матрицы будут (с точностью до трех десятичных знаков) числа 0.010, 0.843, 3.858, 30.289, для второй числа 0.079, 0.844, 3.874, 30.303. Мы видим, что результаты полностью согласуются со сказанным выше. Изменение одного элемента матрицы на 0.1 вызвало изменение собственных значений самое большое на 0.069, в то время как коэффициенты характеристического полинома изменились значительно, не только относительно, но и абсолютно, а именно, максимально на 10.6. Однако значительно меньшее изменение коэффициентов характеристического полинома может привести к большему изменению всех или части корней. Так, если „округлить“ коэффициент 35.1, заменив его на 35.0, то, как легко вычислить, наибольший корень полинома $t^4 - 35t^3 + 149t^2 - 110.6t + 7.8$ будет 30.185. Тем самым изменение наибольшего корня превосходит 0.1.

Погрешности, возникающие от возможной неустойчивости, конечно, не зависят от выбора метода численного решения задачи, в отличие

от погрешностей, возникающих от неизбежного округления промежуточных результатов. Вопроса об оценке этих погрешностей мы не будем касаться.

Отметим, что плохая обусловленность матрицы в смысле решения системы никак не связана с устойчивостью проблемы собственных значений. Действительно, плохая обусловленность означает только, что среди собственных значений имеются малые по модулю по сравнению с другими собственными значениями.

§ 42. Метод А. Н. Крылова

Работа А. Н. Крылова [1] явилась первой в большом цикле работ, посвященных приведению векового уравнения к полиномиальному виду.

Идея А. Н. Крылова заключалась в предварительном преобразовании уравнения

$$\varphi(t) = \begin{vmatrix} a_{11} - t & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - t & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - t \end{vmatrix} = 0 \quad (1)$$

в эквивалентное ему, вообще говоря, уравнение вида

$$D(t) = \begin{vmatrix} b_{11} - t & b_{12} & \dots & b_{1n} \\ b_{21} - t^2 & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{n1} - t^n & b_{n2} & \dots & b_{nn} \end{vmatrix} = 0, \quad (2)$$

развертывание которого по степеням t осуществляется, очевидно, значительно проще, при помощи разложения определителя по минорам 1-го столбца.

Для осуществления указанного преобразования А. Н. Крылов вводит в рассмотрение дифференциальное уравнение, связанное с данной матрицей; одновременно он ставит вопрос о нахождении чисто алгебраического преобразования, переводящего уравнение (1) в уравнение (2).

Выяснению алгебранческой сущности преобразования А. Н. Крылова посвящены работы Н. Н. Лузина [1], [2], И. Н. Хлодовского [1], Ф. Р. Гантмахера [1], Д. К. Фаддеева [1]. Мы изложим метод А. Н. Крылова в его алгебраической интерпретации.

Равенство нулю определителя

$$\varphi(t) = \begin{vmatrix} a_{11} - t & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - t & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - t \end{vmatrix} \quad (3)$$

есть необходимое и достаточное условие для того, чтобы система однородных уравнений

$$\begin{aligned} tx_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ tx_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ &\vdots \\ tx_n &= a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n \end{aligned} \quad (4)$$

имела решение x_1, x_2, \dots, x_n , отличное от нулевого.

Преобразуем систему (4) следующим образом. Умножим первое уравнение на t и заменим tx_1, \dots, tx_n их выражениями (4) через x_1, \dots, x_n .

Это дает

$$t^2x_1 = b_{21}x_1 + b_{22}x_2 + \dots + b_{2n}x_n, \quad (5)$$

где

$$b_{2k} = \sum_{s=1}^n a_{1s}a_{sk}. \quad (6)$$

Умножим далее уравнение (5) на t и заменим снова tx_1, tx_2, \dots, tx_n их выражениями через x_1, \dots, x_n . Мы получим

$$t^3x_1 = b_{31}x_1 + b_{32}x_2 + \dots + b_{3n}x_n.$$

Повторяя этот процесс $(n - 1)$ раз, мы перейдем от системы (4) к системе

$$\begin{aligned} tx_1 &= b_{11}x_1 + b_{12}x_2 + \dots + b_{1n}x_n \\ t^2x_1 &= b_{21}x_1 + b_{22}x_2 + \dots + b_{2n}x_n \\ &\vdots \\ t^n x_1 &= b_{n1}x_1 + b_{n2}x_2 + \dots + b_{nn}x_n, \end{aligned} \quad (7)$$

коэффициенты которой b_{ik} будут определяться по рекуррентным формулам

$$\begin{aligned} b_{1k} &= a_{1k}, \\ b_{ik} &= \sum_{s=1}^n b_{i-1,s}a_{sk} \quad (i = 2, \dots, n; k = 1, \dots, n). \end{aligned} \quad (8)$$

Очевидно, что определитель системы (7) будет иметь вид (2).

Система (7) имеет ненулевое решение для всех значений t , удовлетворяющих уравнению $\varphi(t) = 0$. Таким образом, $D(t)$ обращается в нуль при всех t , являющихся корнями уравнения $\varphi(t) = 0$.

Покажем, что

$$\frac{D(t)}{\varphi(t)} = \begin{vmatrix} 1 & 0 & \dots & 0 \\ b_{11} & b_{12} & \dots & b_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n-1,1} & b_{n-1,2} & \dots & b_{n-1,n} \end{vmatrix} = N, \quad (9)$$

т. е. при $N \neq 0$ $D(t)$ отличается от искомого характеристического полинома только численным множителем.

Пусть все корни $\varphi(t)$ различны. Так как все корни $\varphi(t)$ являются корнями $D(t)$, то $D(t)$ делится на $\varphi(t)$. Так как, кроме того, степени $D(t)$ и $\varphi(t)$ одинаковы, частное должно быть постоянным (не зависеть от t). Сравнивая коэффициенты при t^n , получим

$$\frac{D(t)}{\varphi(t)} = N.$$

В случае, если $\varphi(t)$ имеет кратные корни, равенство

$$D(t) = N\varphi(t) \quad (10)$$

сохраняется, что следует хотя бы из соображений непрерывности.

Можно проверить это равенство и непосредственно умножением входящих в него определителей, если при этом использовать соотношения (8).

Из равенства (10) видно, что если $N = 0$, то $D(t)$ тождественно равно нулю. В этом случае указанное преобразование ничего не дает. Однако и при $N = 0$ А. Н. Крылов предлагает особый прием, алгебраическая сущность которого будет выяснена ниже.

Обратимся теперь к коэффициентам b_{ik} , определяющим $D(t)$. Введем в рассмотрение векторы B_i с компонентами $b_{i1}, b_{i2}, \dots, b_{in}$. Равенства

$$b_{ik} = \sum_{s=1}^n b_{i-1,s} a_{sk} \quad (i = 2, \dots, n)$$

показывают, что

$$B_i = A' B_{i-1}, \quad (11)$$

где A' — матрица, транспонированная к данной.

Из равенства (11) следует, что

$$B_i = A'^{i-1} B_1, \quad (i = 2, \dots, n).$$

В свою очередь, $B_1 = A' B_0$, где $B_0 = (1, 0, \dots, 0)'$. Таким образом, окончательно,

$$B_i = A'^i B_0 \quad (i = 1, 2, \dots, n). \quad (12)$$

Очевидно, что преобразовывать систему (4) можно, исходя, например, из второго уравнения этой системы. В этом случае t войдет во второй столбец определителя $D(t)$, а коэффициенты b_{ik} будут определяться по формулам (12), где $B_0 = (0, 1, \dots, 0)'$.

Метод А. Н. Крылова естественным образом обобщается, если ввести в рассмотрение вместо вектора B_0 специального вида произвольный вектор $B_0 = (b_{01}, b_{02}, \dots, b_{0n})'$.

Пусть

$$u = b_{01}x_1 + b_{02}x_2 + \dots + b_{0n}x_n, \quad (13)$$

где x_1, x_2, \dots, x_n решение системы (4).

Тогда, повторяя прежние рассуждения, получим:

$$\begin{aligned} u &= b_{01}x_1 + b_{02}x_2 + \dots + b_{0n}x_n \\ tu &= b_{11}x_1 + b_{12}x_2 + \dots + b_{1n}x_n \\ t^2u &= b_{21}x_1 + b_{22}x_2 + \dots + b_{2n}x_n \\ &\vdots \\ t^n u &= b_{n1}x_1 + b_{n2}x_2 + \dots + b_{nn}x_n. \end{aligned} \quad (14)$$

где $B_i = (b_{i1}, b_{i2}, \dots, b_{in})' = A'^i B_0$.

Рассматривая $(n+1)$ равенство (14) как систему линейных однородных уравнений с $n+1$ неизвестным u, x_1, \dots, x_n , получим, что ненулевое решение возможно в том и только в том случае, когда определитель

$$D(t) = \begin{vmatrix} 1 & b_{01} & \dots & b_{0n} \\ t & b_{11} & \dots & b_{1n} \\ \cdot & \cdot & \ddots & \cdot \\ t^n & b_{n1} & \dots & b_{nn} \end{vmatrix} = 0. \quad (15)$$

Повторяя прежние рассуждения, найдем, что

$$D(t) = \varphi(t) N,$$

где на этот раз

$$N = \begin{vmatrix} b_{01} & b_{02} & \dots & b_{0n} \\ b_{11} & b_{12} & \dots & b_{1n} \\ \cdot & \cdot & \ddots & \cdot \\ b_{n-1,1} & b_{n-1,2} & \dots & b_{n-1,n} \end{vmatrix}. \quad (16)$$

Так же как и для рассмотренного выше частного случая, преобразование ничего не дает, если $N = 0$.

Предположим поэтому сначала, что $N \neq 0$. На основании равенства $D(t) = N\varphi(t)$ коэффициенты p_i характеристического полинома определяются как отношения $\frac{(-1)^{n-i} N_i}{N}$, где N_i суть алгебраические дополнения элементов t^{n-i} в определителе $D(t)$. Определение коэффициентов характеристического полинома через указанные отношения и составляет сущность работы А. Н. Крылова. Однако проведенное исследование дает возможность определить искомые коэффициенты, минуя вычисления миноров, существенно сократив при этом число нужных операций.

Действительно, ввиду того, что элементы строк определителя (16) являются компонентами векторов B_0, B_1, \dots, B_{n-1} , условие $N \neq 0$

равносильно линейной независимости этих векторов. Поэтому при $N \neq 0$ векторы B_0, B_1, \dots, B_{n-1} образуют базис пространства. Следовательно, вектор B_n является их линейной комбинацией:

$$B_n = q_1 B_{n-1} + \dots + q_n B_0. \quad (17)$$

Покажем, что коэффициенты этого соотношения и являются коэффициентами p_i характеристического полинома, записанного в виде:

$$\varphi(t) = (-1)^n [t^n - p_1 t^{n-1} - \dots - p_n].$$

Действительно, отняв от последней строки определителя $D(t)$ линейную комбинацию предыдущих строк с соответствующими коэффициентами q_1, q_2, \dots, q_n , получим, на основании равенства (17), что

$$D(t) = \begin{vmatrix} 1 & b_{01} & \dots & b_{0n} \\ \vdots & \ddots & \ddots & \ddots \\ t^{n-1} & b_{n-1,1} & \dots & b_{n-1,n} \\ t^n - q_1 t^{n-1} - \dots - q_n & 0 & \dots & 0 \end{vmatrix} = (-1)^n [t^n - q_1 t^{n-1} - \dots - q_n] N.$$

Отсюда

$$\varphi(t) = \frac{D(t)}{N} = (-1)^n [t^n - q_1 t^{n-1} - \dots - q_n],$$

что и требовалось доказать.

Равенство (17) позволяет находить коэффициенты $q_1 = p_1, q_2 = p_2, \dots, q_n = p_n$ как решения системы линейных уравнений, эквивалентной этому векторному равенству.

Равенство (17) связывает метод А. Н. Крылова с соотношением Кели — Гамильтона (примененным к матрице A').

Действительно, из соотношения

$$A'^n = p_1 A'^{n-1} + \dots + p_n E$$

следует, что

$$A'^n B_0 = p_1 A'^{n-1} B_0 + \dots + p_n B_0,$$

т. е.

$$B_n = p_1 B_{n-1} + \dots + p_n B_0. \quad (17)$$

Очевидно, что вместо системы (17) для определения коэффициентов p_i можно употреблять систему

$$C_n = p_1 C_{n-1} + \dots + p_n C_0, \quad (17')$$

где векторы C_n определяются равенствами $C_k = A^k C_0$.

Для определения коэффициентов p_i при помощи решения системы (17) или (17') нужно произвести $\frac{3}{2} n^2(n+1)$ умножений и делений.

В первоначальной форме метод А. Н. Крылова требовал $\frac{1}{3}(n^4 + 4n^3 + 2n^2 - n - 3)$ умножений и делений.

В случае, если $N = 0$, система, эквивалентная равенству (17), не дает возможности определить коэффициенты характеристического полинома, так как определитель этой системы как раз равен N .

Алгебраическая сущность упомянутого приема А. Н. Крылова заключается в том, что возможно определить коэффициенты полинома наименьшей степени $\theta(\lambda)$ такого, что $\theta(A)C_0 = 0$, т. е. коэффициенты минимального аннулирующего полинома. Вообще говоря, это будет минимальный полином матрицы, и его корни будут совпадать со всеми корнями характеристического полинома, но будут иметь меньшую кратность. Однако при неудачном выборе вектора C_0 вместо минимального полинома может получиться какой-либо его делитель и тогда часть корней уравнения $|A - tE| = 0$ может быть потеряна. Как показали Н. Н. Лузин и И. Н. Хлодовский,¹⁾ в качестве полинома $\theta(t)$ при специальном выборе вектора C_0 можно получить любой делитель минимального полинома. Этот результат уже был отмечен (§ 8, п. 2).

Если минимальный полином матрицы не совпадает с характеристическим полиномом, то $N = 0$ при любом выборе вектора C_0 . Действительно, в этом случае $\psi(A)C_0 = 0$, и, так как степень полинома $\psi(t)$ меньше n , векторы $C_0, AC_0, \dots, A^{n-1}C_0$ линейно-зависимы. Что же касается дополнительного вырождения, то его можно избежать за счет изменения начального вектора C_0 .

Итак, метод А. Н. Крылова дает возможность определить коэффициенты характеристического полинома, если $N \neq 0$, или некоторого его делителя, вообще говоря, минимального полинома, если $N = 0$.

Практически обстоятельство $N = 0$ обнаружится само собой в процессе прямого хода при решении системы (17) по методу Гаусса. Именно, в части уравнений исключаются все коэффициенты одновременно, так что эти уравнения обратятся в тождества $0 = 0$. Эти уравнения (пусть число их равно $n - m$) нужно отбросить; в оставшейся системе надо отбросить $n - m$ последних столбцов, начиная со столбца свободных членов (т. е. со столбца из компонент вектора C_m). Последний из оставшихся столбцов, составленный из компонент вектора C_m , надо принять за свободный член новой системы. Решение системы дает коэффициенты линейной зависимости C_m от C_0, C_1, \dots, C_{m-1} , т. е. коэффициенты минимального аннулирующего вектора C_0 полинома.

Разберем теперь два примера, оба взятые из статьи А. Н. Крылова [1].

¹⁾ Н. Н. Лузин [1], [2]; И. Н. Хлодовский [1].

В качестве первого примера определим коэффициенты характеристического полинома для матрицы

$$\begin{bmatrix} -5.509882 & 1.870086 & 0.422908 & 0.008814 \\ 0.287865 & -11.811654 & 5.711900 & 0.058717 \\ 0.049099 & 4.308033 & -12.970687 & 0.229826 \\ 0.006235 & 0.269851 & 1.397369 & -17.596207 \end{bmatrix},$$

взятой А. Н. Крыловым из работы Леверье [1]. По вычислениям Леверье

$$\varphi(t) = t^4 + 47.888430t^3 + 797.2789t^2 + 5349.457t + 12296.555;$$

$$\lambda_1 = -17.86303; \lambda_2 = -17.15266; \lambda_3 = -7.57404; \lambda_4 = -5.29870.$$

В табл. IV.1 мы приводим схему вычислений коэффициентов p_i , рассматривая их как решения системы (17').

Таблица IV.1

Вычисление коэффициентов характеристического уравнения по методу А. Н. Крылова

	C_0	C_1	C_2	C_3	C_4	Σ
I	1	-5.509882	30.917951	-179.01251	1100.7201	948.11566
	0	0.287865	-4.705449	66.38829	-967.5973	-905.62659
	0	0.049099	0.334184	-23.08728	576.5226	553.81860
	0	0.006235	0.002224	-0.649152	-4.04004	-4.68073
II		-5.166683	26.548910	-136.3606	705.6054	
III		1	-16.34603	230.62300	-8361.2884	-3146.0115
			1.136758	-34.41064	741.5585	708.28462
			0.104141	-2.087086	16.91759	14.93465
			1	-30.27086	652.3451	623.0742
				1.065352	-51.01828	-49.95292
				1	-47.8887	-46.8887
					-797.287	-796.287
		1			-5349.53	-5348.53
	1				-12296.8	-12295.8

Здесь в части I расположены последовательно вычисляемые компоненты векторов $A^k C_0$ ($k = 0, 1, 2, 3, 4$) и обычные контрольные

суммы. В строке II осуществляется контроль, аналогичный применявшемуся в § 30 при вычислении последовательных итераций. В III части содержится решение полученной системы, которое мы находим по схеме единственного деления.

Окончательным контролем вычисления коэффициентов p_i является сравнение значения p_1 со следом матрицы. Так как $\text{Sp } A = -47.888\ 430$, мы видим, что значение p_1 , найденное из решения системы, достаточно точно.

Сравнение с данными Леверье показывает, однако, что коэффициенты вычислены с меньшей степенью точности. Известная потеря точности является органическим недостатком метода А. М. Крылова и объясняется тем обстоятельством, что коэффициенты системы, определяющей p_i , являются величинами различных порядков, что приводит, как правило, к не очень хорошей обусловленности системы.

Несколько лучший результат можно получить, применяя для решения системы схему главных элементов.

В качестве второго примера возьмем еще одну матрицу из статьи А. Н. Крылова [1]:

$$\begin{bmatrix} 5 & 30 & -48 \\ 3 & 14 & -24 \\ 3 & 15 & -25 \end{bmatrix}. \quad (18)$$

В приведенной ниже таблице три первых строчки содержат коэффициенты системы для определения q_i :

1	5	-29	125
0	3	-15	63
0	3	-15	63
	0	0	0

Уже первый шаг процесса Гаусса показывает, что здесь имеет место случай вырождения. Урезанная система

$$\begin{aligned} q_1 + 5q_2 &= -29 \\ 3q_2 &= -15 \end{aligned}$$

дает в качестве решения $q_2 = -5$, $q_1 = -4$. Таким образом, в этом случае мы определили коэффициенты полинома второй степени $t^2 + 5t + 4$, являющегося лишь делителем характеристического полинома.

На практике при пользовании методом Крылова ситуация точного вырождения может появиться лишь при особых обстоятельствах (например, когда по существу физической задачи, приводящейся к исследованию собственных значений матрицы, эта матрица по физи-

ческому смыслу должна иметь минимальный полином, отличный от характеристического). Чаще встречается случай приближенного вырождения. В этом случае система уравнений (17) (или (17')) оказывается плохо обусловленной, но это не портит дела. Действительно, проводя процесс прямого хода решения системы по методу исключения, нужно лишь остановиться, когда коэффициенты всех уравнений, начиная с некоторого, почти исчезнут в пределах точности вычисления, и далее действовать так, как будто они обратились в нуль точно.

В заключение заметим, что для матрицы вида

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ 0 & a_{32} & \dots & a_{3n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & a_{n-1,n} & a_{nn} \end{bmatrix},$$

при условии, что элементы $a_{21}, a_{32}, \dots, a_{n-1,n}$ отличны от нуля, система уравнений (17), построенная исходя из начального вектора

$$C_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

будет треугольной, ибо у вектора AC_0 все компоненты, кроме первых двух, равны нулю, у вектора A^2C_0 все компоненты, кроме первых трех, равны нулю и т. д. В частности, отмеченное обстоятельство будет иметь место для трехдиагональной матрицы.

§ 43. Определение собственных векторов по методу А. Н. Крылова

Предположим, что методом А. Н. Крылова найдены коэффициенты характеристического полинома и что все собственные числа вычислены и оказались различными. Покажем, как, используя проведенные вычисления, определить собственные векторы матрицы. Пусть C_0 исходный вектор в процессе А. Н. Крылова и пусть X_1, X_2, \dots, X_n собственные векторы матрицы A , принадлежащие $\lambda_1, \lambda_2, \dots, \lambda_n$. Согласно теореме 6.3 векторы X_1, \dots, X_n линейно-независимы.

Разложим вектор C_0 по собственным векторам:

$$C_0 = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n. \quad (1)$$

Тогда

$$\begin{aligned} C_1 &= AC_0 = \alpha_1 \lambda_1 X_1 + \alpha_2 \lambda_2 X_2 + \dots + \alpha_n \lambda_n X_n \\ C_{n-1} &= A^{n-1} C_0 = \alpha_1 \lambda_1^{n-1} X_1 + \alpha_2 \lambda_2^{n-1} X_2 + \dots + \alpha_n \lambda_n^{n-1} X_n. \end{aligned} \quad (2)$$

Векторы C_1, C_2, \dots, C_{n-1} вычислены в процессе нахождения собственных чисел. Покажем, что собственные векторы могут быть получены в виде линейных комбинаций

$$\beta_{10} C_{n-1} + \beta_{11} C_{n-2} + \dots + \beta_{1n-1} C_0 \quad (i = 1, 2, \dots, n)$$

при подходящем выборе коэффициентов β_{ij} . Рассмотрим линейную комбинацию:

$$\begin{aligned} \beta_{10} C_{n-1} + \beta_{11} C_{n-2} + \dots + \beta_{1n-1} C_0 &= \\ &= \alpha_1 (\beta_{10} \lambda_1^{n-1} + \beta_{11} \lambda_1^{n-2} + \dots + \beta_{1n-1}) X_1 + \\ &+ \alpha_2 (\beta_{10} \lambda_2^{n-1} + \beta_{11} \lambda_2^{n-2} + \dots + \beta_{1n-1}) X_2 + \\ &+ \dots + \dots + \dots + \\ &+ \alpha_n (\beta_{10} \lambda_n^{n-1} + \beta_{11} \lambda_n^{n-2} + \dots + \beta_{1n-1}) X_n = \\ &= \alpha_1 \varphi_1(\lambda_1) X_1 + \alpha_2 \varphi_1(\lambda_2) X_2 + \dots + \alpha_n \varphi_1(\lambda_n) X_n, \end{aligned} \quad (3)$$

где

$$\varphi_1(t) = \beta_{10} t^{n-1} + \beta_{11} t^{n-2} + \dots + \beta_{1n-1}. \quad (4)$$

Подберем коэффициенты $\beta_{10}, \dots, \beta_{1n-1}$ так, чтобы

$$\varphi_1(\lambda_1) \neq 0, \quad \varphi_1(\lambda_2) = \dots = \varphi_1(\lambda_n) = 0. \quad (5)$$

Для этого достаточно взять в качестве $\varphi_1(t)$ полином

$$\begin{aligned} \varphi_1(t) &= (t - \lambda_2) \dots (t - \lambda_n) = \\ &= \frac{(t - \lambda_1)(t - \lambda_2) \dots (t - \lambda_n)}{t - \lambda_1} = \frac{(-1)^n \varphi(t)}{t - \lambda_1} = \\ &= \frac{t^n - p_1 t^{n-1} - \dots - p_n}{t - \lambda_1}. \end{aligned} \quad (6)$$

Здесь $\varphi(t)$ — характеристический полином, коэффициенты и корни которого уже вычислены.

Коэффициенты частного (6) легко вычисляются по схеме Хорнера, т. е. по рекуррентным формулам

$$\beta_{10} = 1, \quad \beta_{1j} = \lambda_1 \beta_{1,j-1} - p_j, \quad j = 1, \dots, n-1. \quad (7)$$

Таким образом,

$$\beta_{10} C_{n-1} + \beta_{11} C_{n-2} + \dots + \beta_{1n-1} C_0 = \alpha_1 \varphi_1(\lambda_1) X_1,$$

т. е. составленная нами линейная комбинация есть собственный вектор X_1 с точностью до численного множителя. Конечно, коэффи-

циент α_1 должен быть отличным от нуля; это обеспечивается успешным завершением процесса А. Н. Крылова. Так как собственный вектор определен с точностью до постоянного множителя, мы можем принять за собственный вектор построенную линейную комбинацию.

Аналогично

$$X_i = \sum_{j=0}^{n-1} \beta_{ij} C_{n-i-j}, \quad (8)$$

где

$$\beta_{i0} = 1, \quad \beta_{ij} = \lambda_i \beta_{i,j-1} - p_j, \quad j = 1, \dots, n-1. \quad (9)$$

В качестве примера вычислим собственный вектор матрицы Леверье, принадлежащий собственному числу $\lambda_4 = -5.29870$.

Выпишем характеристическое уравнение с коэффициентами, взятыми из табл. IV.1:

$$t^4 + 47.8887t^3 + 797.287t^2 + 5349.53t + 12296.8 = 0.$$

Вычисляя числа β_{4j} при $j = 0, 1, 2, 3$, получим

$$1; \quad 42.5900; \quad 571.615; \quad 2320.71.$$

Составим линейные комбинации по формуле (8), располагая вычисления в таблице:

Таблица IV.2

Вычисление собственного вектора по методу А. Н. Крылова

$\beta_{43}C_0$	$\beta_{42}C_1$	$\beta_{41}C_2$	$\beta_{40}C_3$	X_4	\tilde{X}_4
2320.71	-3149.53	1316.80	-179.01	308.97	1
0	164.548	-200.405	66.388	30.531	0.098815
0	28.066	14.233	-23.087	19.212	0.062181
0	3.5640	0.0947	-0.6492	3.0095	0.009741

Последний столбец содержит компоненты собственного вектора X_4 , нормированного так, что его первая компонента равна единице. Ниже мы увидим, что значения компонент собственного вектора, вычисленного по методу А. Н. Крылова, хорошо согласуются со значениями, вычисленными другими методами.

§ 44. Метод Хессенберга

В методе Хессенберга¹⁾, так же как и в методе Крылова, разыскивается нулевая линейная комбинация векторов $C_0, AC_0, \dots, A^{n-1}C_0$. В то время как в методе Крылова коэффициенты такой линейной

1) Цурмюль [3].

комбинации находятся посредством решения системы линейных уравнений (17), в методе Хессенберга искомая линейная комбинация получается как последний вектор в рекуррентно строящейся последовательности векторов Z_1, \dots, Z_n, Z_{n+1} при $Z_1 = C_0$ таких, что у вектора Z_{j+1} первые j компонент равны нулю. Каждый последующий вектор получается итерацией предшествующего с последующим „исправлением“ посредством добавления подходящей линейной комбинации всех предшествующих векторов. Иначе говоря,

$$Z_{j+1} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ z_{j+1, j+1} \\ \vdots \\ z_{n, j+1} \end{bmatrix} = AZ_j + h_{1j}Z_1 + \dots + h_{jj}Z_j. \quad (1)$$

В качестве вектора Z_1 удобно взять вектор $(1, 0, \dots, 0)'$. Указанный процесс не всегда возможно осуществить. Естественное течение процесса нарушается, если на каком-либо шагу получится $z_{ii} = 0$.

Рассмотрим основной случай, предполагая $z_{11} \neq 0, \dots, z_{nn} \neq 0$. В этом случае векторы Z_1, \dots, Z_n оказываются линейно-независимыми, так что матрица

$$Z = [Z_1, Z_2, \dots, Z_n]$$

будет неособенной. Ясно, что

$$Z_{j+1} = \varphi_j(A) Z_1, \quad (2)$$

где $\varphi_j(t) = t^j + \dots$ — некоторый полином степени j . В силу линейной независимости векторов Z_1, \dots, Z_n равенство $f(A)Z_1 = 0$ невозможно, если только степень полинома f меньше n . Полином же φ_n аннулирует Z_1 и его степень равна n . Следовательно, полином φ_n совпадает с характеристическим полиномом матрицы A .

Полиномы φ_j , очевидно, связаны рекуррентными соотношениями

$$\varphi_j(t) = (t + h_{jj})\varphi_{j-1}(t) + h_{j-1, j}\varphi_{j-2}(t) + \dots + h_{1j}\varphi_0(t), \quad (3)$$

где $\varphi_0(t) = 1$ ($j = 1, \dots, n$).

Таким образом, коэффициенты характеристического полинома определяются, как только известны все коэффициенты h_{ij} .

Нетрудно видеть, что система векторных равенств

$$Z_2 = AZ_1 + h_{11}Z_1$$

$$Z_3 = AZ_2 + h_{12}Z_1 + h_{22}Z_2$$

$$\dots \dots \dots \dots \dots \dots \dots$$

$$0 = Z_{n+1} = AZ_n + h_{1n}Z_1 + \dots + h_{nn}Z_n$$

равносильно матричному равенству

$$AZ + ZH = 0, \quad (4)$$

где

$$H = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ -1 & h_{22} & \dots & h_{2n} \\ 0 & -1 & \dots & h_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & h_{nn} \end{bmatrix}. \quad (5)$$

Последнее равенство позволяет последовательно определять все элементы матриц H и Z .

Для уяснения вычислительной схемы целесообразно равенство $AZ + ZH = 0$ представить в форме

$$(A|Z) \left(\frac{Z}{H} \right) = 0.$$

Здесь $(A|Z)$ и $\left(\frac{Z}{H} \right)$ прямоугольные матрицы, составленные из матриц A , Z и H в указанном порядке. Составим теперь следующую схему

$$\begin{array}{ccccccccc} a_{11} & a_{12} & \dots & a_{1n} & z_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2n} & z_{21} & z_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & z_{n1} & z_{n2} & \dots & z_{nn} \\ & & & & h_{11} & h_{12} & \dots & h_{1n} \\ & & & & -1 & h_{22} & \dots & h_{2n} \\ & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & 0 & 0 & \dots & -1 & h_{nn} \end{array}$$

Первые n строк этой схемы образованы матрицей $(A|Z)$, последние n столбцов матрицей $\left(\frac{Z}{H} \right)$. В начале процесса нам известна матрица A и первый столбец матрицы Z . Умножение 1-й строки матрицы $(A|Z)$ на 1-й столбец матрицы $\left(\frac{Z}{H} \right)$ позволяет определить элемент h_{11} , умножение остальных строк на 1-й столбец матрицы $\left(\frac{Z}{H} \right)$ — элементы z_{22}, \dots, z_{n2} соответственно. Как только определены эти элементы, умножение матрицы $(A|Z)$ на 2-й столбец матрицы $\left(\frac{Z}{H} \right)$ дает последовательно элементы $h_{12}, h_{22}, z_{33}, \dots, z_{n3}$. Далее производится умножение матрицы $(A|Z)$ на 3-й, ..., n -й столбцы матрицы $\left(\frac{Z}{H} \right)$. Вычисления допускают обычный контроль при помощи составления строчных сумм.

Таблица IV.3

Определение коэффициентов характеристического полинома по методу Хессенберга

	$\left(\frac{A'}{Z}\right)$			Σ	$\left(\frac{Z}{H}\right)$		
-5.509882	0.287865	0.049099	0.006235	-5.166683	1	0	0
1.870086	-11.811654	4.308033	0.269851	-5.363684	0	0.287865	0
0.422908	5.711900	-12.970687	1.397369	-5.438510	0	0.049099	1.136759
0.008814	0.058717	0.229326	-17.596207	-17.299350	0	0.006235	0.10414110
1	0	0	0	1	-5.509882	-0.55915162	-0.48166191
0	0.287865	0.049099	0.006235	0.343199	-1	10.336146	-0.00939009
0	0	1.136759	0.10414110	1.2408990	0	-1	-0.21730602
0	0	0	1.0653609	1.0653609	0	0	-0.20553668

4	3	2	1	0	$\varphi_i(1)$	$k \diagdown i$
0	0	0	0	1	1	0
0	0	0	1	16.346028	5.509882	1
0	0	1	30.270860	264.18531	59.146734	2
0	1	30.270860	797.27877	5349.4555	698.72941	3
1	47.888430	797.27877		12296.551	18492.173	4

При осуществлении вычислений на настольных машинах целесообразно вычисления расположить в форме

$$\begin{bmatrix} A' & Z \\ Z' & H \end{bmatrix}$$

и заменить умножение строк на столбцы умножением столбцов. Двойная запись матрицы Z оправдывается простотой действий.

Коэффициенты характеристического полинома определяются параллельно с определением чисел h_{ii} по рекуррентным соотношениям, вытекающим из соотношений (3). Для контроля рекуррентно вычисляются значения полиномов при $t = 1$, равные суммам их коэффициентов.

В табл. IV.3 определяются коэффициенты характеристического полинома матрицы Леверье.

Для контроля была вычислена левая часть равенства (4). Наибольший по модулю элемент полученной матрицы оказался равным 0.0000004.

Как только собственное значение найдено, легко определяется принадлежащий ему собственный вектор. Действительно, матричное равенство (4) показывает, что $A = -ZHZ^{-1}$, т. е. что матрица A подобна матрице $-H$. Следовательно, собственный вектор X_i матрицы A , принадлежащий собственному значению λ_i , связан соотношением

$$X_i = ZY_i$$

с собственным вектором Y_i матрицы H , принадлежащим собственному значению $-\lambda_i$. Собственные же векторы для матрицы H определяются без труда, так как произвольно задавшись последней компонентой собственного вектора, мы определим остальные его компоненты из системы линейных уравнений:

$$\begin{aligned} -y_1 + (h_{22} + \lambda) y_2 + \dots + h_{2n} y_n &= 0 \\ \vdots &\quad \vdots \\ -y_{n-1} + (h_{nn} + \lambda) y_n &= 0 \end{aligned} \tag{6}$$

с треугольной матрицей. Отброшенное первое уравнение можно использовать для контроля.

В качестве примера определим по табл. IV.4 собственный вектор матрицы Леверье, принадлежащий наименьшему по модулю собственному значению. Приближение к этому собственному значению, вычисленное как корень полинома

$$t^4 + 47.888430t^3 + 797.27877t^2 + 5349.4555t + 12296.551,$$

равно -5.298700 .

Результат подстановки компонент вектора Y_4 в первое уравнение системы (6) равен -0.0007 .

Таблица IV.4

Определение собственного вектора по методу Хессенберга

	$H + \lambda_4 E$				Y_4	X_4	\tilde{X}_4
0.211182	-0.559152	-0.481662	-0.009390	308.95220	308.9522	1.000000	
-1	5.537446	-22.577119	-0.217306	106.05866	30.5306	0.098820	
	-1	8.626132	-0.205537	12.318870	19.2109	0.062181	
		-1	12.318870	1.000000	3.0095	0.009741	

К изложенному методу можно подойти с несколько иной точки зрения. Именно, метод может быть истолкован как метод приведения данной матрицы A преобразованием подобия к специальному виду

$$-\begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ -1 & h_{22} & \dots & h_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -1 & h_{nn} \end{bmatrix},$$

причем преобразующая матрица Z выбирается треугольной. Полиномы $\varphi_1(t)$, $\varphi_2(t)$, ..., $\varphi_n(t)$, как легко видеть, будут характеристическими полиномами для матриц

$$-\begin{bmatrix} h_{11} \end{bmatrix}, -\begin{bmatrix} h_{11} & h_{12} \\ -1 & h_{22} \end{bmatrix}, -\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ -1 & h_{22} & h_{23} \\ 0 & -1 & h_{33} \end{bmatrix} \text{ и т. д.}$$

Действительно, развертывая характеристический определитель такой матрицы по элементам последней строки, мы придем к прежним рекуррентным соотношениям (3).

Рассмотрим теперь исключительные случаи, которые могут встретиться по ходу процесса.

Может случиться, что на i -м шагу процесса $z_{ii} = 0$, но хотя бы одна компонента вектора Z_i отлична от нуля. В этом случае, мы, заполняя столбец для Z_{i+1} , поставим в него нули только на тех местах, на которых фактически можно добиться нулевого значения за счет добавления линейной комбинации предыдущих столбцов. Процесс продолжается дальше таким же образом. Если при этом окажется, что мы построим n ненулевых векторов Z_1, \dots, Z_n , то они автоматически будут линейно-независимыми, и матрица Z (в данном случае уже не треугольная, но получающаяся из треугольной перестановкой столбцов) будет осуществлять подобное преобразование матрицы A в матрицу H . Но может случиться, что на каком-либо шаге мы придем к нулевому вектору. Это, очевидно, произойдет

в том и только в том случае, когда векторы $Z_1, AZ_1, \dots, A^{n-1}Z_1$ линейно-зависимы, т. е. когда минимальный аннулирующий полином для Z_1 есть только делитель характеристического полинома.

В этом случае, дойдя до нулевого вектора мы прекращаем процесс. Последний полином $\varphi_i(t)$ будет минимальным аннулирующим полиномом для вектора Z_1 и его корни будут образовывать лишь часть спектра собственных значений.

Рассмотрим два примера, иллюстрирующих оба возможных исключительных случая (табл. IV.5 и табл. IV.6).

Из табл. IV.5 мы видим, что в первом примере уже $z_{22} = 0$. Однако процесс все же удается довести до конца.

Таблица IV.5

Метод Хессенберга в вырожденном случае

2	0	3	-1	1	0	0	0
1	2	1	1	0	0	0	14
-1	2	1	3	0	3	0	0
-1	6	2	7	0	-1	7/3	0
1	0	0	0	-2	2	7/3	-14
0	0	3	-1	-1	-1/3	-14/9	-14/3
0	0	0	7/3	0	-1	-23/3	-8
0	14	0	0	0	0	-1	-2

Опишем кратко ход процесса. Последовательно определяем: из (1×1) $h_{11} = -2$, из (2×1) $z_{22} = 0$, из (3×1) $z_{32} = 3$. Это позволяет считать $z_{33} = z_{34} = 0$. Далее из (4×1) определяем $z_{42} = -1$, из (1×2) $h_{12} = 2$, из (2×2) $z_{23} = 0$, из (3×2) $h_{22} = -\frac{1}{3}$, из (4×2) $z_{43} = \frac{7}{3}$. Полагаем $z_{44} = 0$. Из (1×3) определяем $h_{13} = \frac{7}{3}$, из (2×3) $z_{24} = 14$, из (3×3) $h_{23} = -\frac{14}{9}$, из (4×3) $h_{33} = -\frac{23}{3}$. Наконец, из (1×4) определяем $h_{14} = -14$, из (2×4) $h_{44} = -2$, из (3×4) $h_{24} = -\frac{14}{3}$, из (4×4) $h_{34} = -8$.

После определения матрицы H определяем характеристический полином, как указано выше.

В качестве второго примера рассмотрим матрицу (18) из § 42. В этом случае (табл. IV.6) $Z_3 = (0, 0, 0)'$, и мы вычисляем лишь делитель характеристического полинома.

Таблица IV.6

Метод Хессенберга в вырожденном случае

5	30	-48	1	0	0
3	14	-24	0	30	0
3	15	-25	0	-48	0
1	0	0	-5	54	
0	30	-48	-1	10	
0	0	0	0	-1	

Именно, имеем $\varphi_0 = 1$, $\varphi_1 = t - 5$, $\varphi_2 = (t + 10)(t - 5) + 54 = t^2 + 5t + 4$.

В заключение этого параграфа заметим, что при вычислении коэффициентов характеристического полинома можно избежать использования рекуррентных соотношений для полиномов φ_i . Вместо этого можно находить характеристический полином φ_n (или его делитель в случае вырождения) непосредственно применения способ А. Н. Крылова¹⁾ к матрице $-H$, что приводит, как мы видели выше, к решению треугольной системы. Так, в последнем примере надо применить метод А. Н. Крылова к матрице

$$B = \begin{bmatrix} 5 & -54 \\ 1 & -10 \end{bmatrix}.$$

Имеем

$$C_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad BC_0 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \quad B^2 C_0 = \begin{bmatrix} -29 \\ -5 \end{bmatrix}.$$

Следовательно, коэффициенты полинома φ_2 определяются из системы

$$p_2 - 5p_1 = -29$$

$$p_1 = -5,$$

откуда $p_1 = -5$, $p_2 = -4$ т. е. $\varphi_2 = t^2 + 5t + 4$.

§ 45. Метод Самуэльсона

Близким к методу А. Н. Крылова является также и метод, предложенный Самуэльсоном [1].

1) Сейбл, и Бержер [1].

Вычислительная схема этого метода такова. Вычисляется прямогородальная матрица

$$\left[\begin{array}{ccccccc|cc} R & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & -a_{11} \\ RM & 0 & 0 & 0 & 0 & \dots & 0 & 1 & -a_{11} & -RS \\ RM^2 & 0 & 0 & 0 & 0 & \dots & 1 & -a_{11} & -RS & -RMS \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ RM^{n-1} & 1 & -a_{11} & -RS & -RMS & \dots & \dots & \dots & \dots & -RM^{n-2}S \end{array} \right], \quad (1)$$

где R, S и M клетки в следующем разбиении данной матрицы

$$A = \left[\begin{array}{c|ccc} a_{11} & a_{12} & \dots & a_{1n} \\ \hline a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{array} \right] = \left[\begin{array}{cc} a_{11} & R \\ S & M \end{array} \right]. \quad (2)$$

Далее, посредством элементарных преобразований (как это делается в задаче исключения § 22) нужно добиться того, чтобы на месте строки RM^{n-1} оказалась нулевая строка. Тогда остальные элементы последней строки дадут, вообще говоря, коэффициенты характеристического полинома. Процесс исключения, как мы видели, очень однообразен и прост. Это является основным достоинством схемы.

Автор выводит указанную схему из преобразования системы линейных дифференциальных уравнений, связанной с матрицей, к одному линейному уравнению порядка n посредством специального приема исключения. Краткое алгебраическое обоснование схемы заключается в следующем.

Пусть

$$X_0 = \begin{bmatrix} x_{10} \\ x_{20} \\ \vdots \\ x_{n0} \end{bmatrix} = \begin{bmatrix} x_{10} \\ Y_0 \end{bmatrix} \quad (3)$$

произвольный вектор.

Пусть далее

$$AX_0 = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} = \begin{bmatrix} x_{11} \\ Y_1 \end{bmatrix}; \quad \dots; \quad A^{n-1}X_0 = \begin{bmatrix} x_{1,n-1} \\ x_{2,n-1} \\ \vdots \\ x_{n,n-1} \end{bmatrix} = \begin{bmatrix} x_{1,n-1} \\ Y_{n-1} \end{bmatrix};$$

$$A^nX_0 = \begin{bmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{nn} \end{bmatrix} = \begin{bmatrix} x_{1n} \\ Y_n \end{bmatrix}. \quad (4)$$

Из построения следует, что

$$x_{1k} = a_{11}x_{1,k-1} + RY_{k-1} \quad (5)$$

$$Y_k = Sx_{1,k-1} + MY_{k-1} \quad (k = 1, \dots, n).$$

Таким образом, мы имеем n^2 соотношений (n и $n(n-1)$) между $n^2 + n$ величинами. Они дают возможность исключить из системы равенств (5) и (6) векторы Y_1, \dots, Y_n , т. е. $n(n-1)$ величин. В результате этого исключения останется n равенств, связывающих $2n$ чисел, именно, компоненты вектора Y_0 и числа x_{10}, \dots, x_{1n} .

Проведем это исключение. Имеем при $k = 1, 2, \dots, n$:

$$\begin{aligned} x_{1k} &= a_{11}x_{1,k-1} + RY_{k-1} = \\ &= a_{11}x_{1,k-1} + RSx_{1,k-2} + RMY_{k-2} = \\ &= a_{11}x_{1,k-1} + RSx_{1,k-2} + RMSx_{1,k-3} + RM^2Y_{k-3} = \\ &= \dots = \\ &= a_{11}x_{1,k-1} + RSx_{1,k-2} + RMSx_{1,k-3} + \\ &\quad + \dots + RM^{k-2}Sx_{10} + RM^{k-1}Y_0, \end{aligned}$$

или

$$RM^{k-1}Y_0 = x_{1k} - a_{11}x_{1,k-1} - RSx_{1,k-2} - \dots - RM^{k-2}Sx_{10}. \quad (7)$$

Коэффициенты этих n равенств образуют, очевидно, матрицу (1).

Исключая из этих n равенств компоненты вектора Y_0 , мы получим одно линейное соотношение между числами x_{10}, \dots, x_{1n} с постоянными коэффициентами, не зависящими от выбора исходного вектора.

С другой стороны, исходя из соотношения Кели — Гамильтона, мы имеем

$$x_{1n} - p_1x_{1,n-1} - \dots - p_nx_{10} = 0.$$

Это равенство тоже является линейной зависимостью между числами x_{10}, \dots, x_{1n} с постоянными коэффициентами, не зависящими от выбора вектора.

Эта зависимость будет совпадать с зависимостью, полученной методом исключения, в том случае, если матрица A такова, что мы вправе считать числа $x_{10}, \dots, x_{1,n-1}$ независимыми переменными, т. е. если мы можем им придавать независимо друг от друга произвольные значения за счет подходящего выбора остальных компонент исходного вектора X_0 или, иными словами, за счет вектора Y_0 .

Более строго обосновать метод Самуэльсона можно посредством следующего соотношения между коэффициентами характеристических полиномов окаймленной и окаймляемой матриц.

Пусть

$$\begin{aligned}\varphi(t) &= (-1)^n (t^n + p_1 t^{n-1} + \dots + p_n) \\ f(t) &= (-1)^n (t^{n-1} + q_1 t^{n-2} + \dots + q_{n-1})\end{aligned}\quad (8)$$

характеристические полиномы матриц A и M . (Мы, вопреки обычной записи полиномов, изменили знак у коэффициентов p_k и q_k).

Тогда справедливы следующие соотношения:

$$\begin{aligned}p_1 &= -a_{11} + q_1 \\ p_2 &= -RS - q_1 a_{11} + q_2 \\ p_3 &= -RMS - q_1 RS - q_2 a_{11} + q_3 \\ &\vdots \\ p_{n-1} &= -RM^{n-3}S - q_1 RM^{n-4}S - \dots + q_{n-1} \\ p_n &= -RM^{n-2}S - q_1 RM^{n-3}S - \dots - q_{n-2} RS - q_{n-1} a_{11}.\end{aligned}\quad (9)$$

Эти соотношения получаются из правила раскрытия окаймленного определителя.

Далее, если применить к матрице M метод А. Н. Крылова, приняв за исходный вектор R' (с компонентами a_{12}, \dots, a_{1n}), то коэффициенты q_1, q_2, \dots, q_{n-1} будут определяться из системы уравнений

$$M'^{n-1}R' + q_1 M'^{n-2}R' + \dots + q_{n-1} R' = 0. \quad (10)$$

Коэффициенты p_1, p_2, \dots, p_n являются в силу соотношений (9) линейными неоднородными формами от q_1, \dots, q_{n-1} и, следовательно, могут быть одновременно вычислены методом исключения (см. § 22). Из двух возможных модификаций метода исключения следует взять ту, в которой компоненты векторов $R', M'R', \dots, M'^{n-1}R'$ располагаются в строки схемы. Тогда эти строки, рассматриваемые как матрицы, суть R, RM, \dots, RM^{n-1} . При этом коэффициенты соотношений (9) окажутся расположенными точно в согласии со схемой Самуэльсона. Из приведенного выше обоснования метода легко выяснить область его применения. Действительно, она совпадает с областью применимости метода А. Н. Крылова для матрицы M , исходя из вектора R' .

В качестве примера возьмем снова матрицу Леверье. Вычисления коэффициентов характеристического полинома произведем согласно описанной схеме (см. табл. IV, 7). Сначала мы вычисляем матрицу (1), располагая ее элементы в первых четырех строках. Далее проводим исключение, как было показано в § 22. Последняя строка дает искомые значения коэффициентов, которые почти в точности совпадают со значениями, вычисленными по методу А. Н. Крылова. Последний столбец, как обычно, есть контрольный столбец.

Таблица IV. 7

Определение коэффициентов характеристического полинома по схеме Самуэльсона

Число операций, нужных для определения коэффициентов характеристического полинома, по методу Самуэльсона немного меньше, чем в методе А. Н. Крылова. Действительно, составление матрицы (1) требует $n(n - 1)^2$ умножений, а процесс исключения в схеме Самуэльсона требует столько же операций сколько решение системы в методе А. Н. Крылова.

§ 46. Метод А. М. Данилевского

Элегантный и весьма эффективный метод вычисления коэффициентов характеристического полинома предложен А. М. Данилевским [1]. Геометрический смысл этого метода состоит в следующем. Данная матрица A рассматривается как матрица оператора в базисе $e_1 = (1, 0, \dots, 0)', e_2 = (0, 1, \dots, 0)', \dots, e_n = (0, 0, \dots, 1)'$.

Предполагается, что векторы $e_1, Ae_1, \dots, A^{n-1}e_1$ линейно-независимы. Тогда

$$A^n e_1 = p_1 A e_1^{n-1} + p_2 A^{n-2} e_1 + \dots + p_n e_1.$$

Ясно, что коэффициенты p_1, p_2, \dots, p_n суть искомые коэффициенты характеристического полинома.

В базисе $e_1, Ae_1, \dots, A^{n-1}e_1$ рассматриваемый оператор, очевидно, будет иметь так называемую матрицу Фробениуса

$$P = \begin{bmatrix} 0 & 0 & \dots & p_n \\ 1 & 0 & \dots & p_{n-1} \\ 0 & 1 & \dots & p_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & p_1 \end{bmatrix},$$

содержащую в явном виде искомые коэффициенты характеристического полинома.

Переход от базиса e_1, e_2, \dots, e_n к базису $e_1, Ae_1, \dots, A^{n-1}e_1$ осуществляется постепенно в $n - 1$ шагов. Каждый шаг состоит в переходе от базиса $e_1, Ae_1, \dots, A^{k-1}e_1, e_{k+1}, \dots, e_n$ к базису $e_1, Ae_1, \dots, A^{k-1}e_1, A^k e_1, e_{k+2}, \dots, e_n$. Для осуществимости всего процесса необходимо считать, что все промежуточные системы векторов действительно являются базисами, т. е. состоят из линейно-независимых векторов. Ниже будет рассмотрено, как следует поступать в случаях вырождения. Пока же мы будем рассматривать лишь невырожденный процесс.

Обозначим через $A^{(k)}$ матрицу, полученную при $k - 1$ -м шаге процесса, так что $A = A^{(1)}, P = A^{(n)}$. Столбцами матрицы $A^{(k)}$ являются координаты векторов $Ae_1, A^2e_1, \dots, A^ke_1, Ae_{k+1}, \dots, Ae_n$ в базисе

$e_1, Ae_1, \dots, A^{k-1}e_1, e_{k+1}, \dots, e_n$. Поэтому первые $k - 1$ столбцов матрицы $A^{(k)}$ будут совпадать с одноименными столбцами матрицы Фробениуса Р. Имеем

$$A^{(k+1)} = S_k^{-1} A^{(k)} S_k,$$

где S_k — соответствующая матрица преобразования координат. Ясно, что

$$S_k = \begin{bmatrix} 1 & \dots & s_{1, k+1} & \dots & 0 \\ 0 & \dots & s_{2, k+1} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & s_{n, k+1} & \dots & 1 \end{bmatrix},$$

где $s_{1, k+1}, s_{2, k+1}, \dots, s_{n, k+1}$ суть координаты вектора $A^k e_1$ в базисе $e_1, Ae_1, \dots, A^{k-1}e_1, e_{k+1}, \dots, e_n$. Эти координаты, как мы видели выше, суть не что иное, как элементы $a_{ik}^{(k)}$ k -го столбца матрицы $A^{(k)}$. Имеем далее

$$S_k^{-1} = \begin{bmatrix} 1 & \dots & -\frac{s_{1, k+1}}{s_{k+1, k+1}} & \dots & 0 \\ 0 & \dots & -\frac{s_{2, k+1}}{s_{k+1, k+1}} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \frac{1}{s_{k+1, k+1}} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & -\frac{s_{n, k+1}}{s_{k+1, k+1}} & \dots & 1 \end{bmatrix}.$$

Вычисление матрицы $A^{(k+1)}$ целесообразно проводить в два приема. Сперва вычисляется вспомогательная матрица $B^{(k)} = S_k^{-1} A^{(k)}$. Это действие, в силу строения матрицы S_k^{-1} состоит в том, что $(k+1)$ -я строка матрицы $A^{(k)}$ умножается на $\frac{1}{a_{k+1, k}^{(k)}}$, а от каждой из остальных строк отнимается полученная $(k+1)$ -я строка матрицы $B^{(k)}$, умноженная на соответствующий элемент k -го столбца матрицы $A^{(k)}$. Очевидно, что в результате этих действий первые $(k-1)$ столбцов не изменятся, в k -м же столбце на $(k+1)$ -м месте окажется единица, а все остальные элементы станут нулями. Вычисления остальных элементов матрицы $B^{(k)}$ будут двухчленными, напоминающими вычисления метода Гаусса. Полученная матрица $B^{(k)}$ умножается затем

справа на S_k . При этом изменяется только один, именно $(k+1)$ -й столбец. Его элементами будут, как нетрудно видеть,

$$\sum_{j=1}^n b_{1j}^{(k)} a_{jk}^{(k)}, \quad \sum_{j=1}^n b_{2j}^{(k)} a_{jk}^{(k)}, \dots, \quad \sum_{j=1}^n b_{nj}^{(k)} a_{jk}^{(k)}.$$

Иначе говоря, $(k+1)$ -й столбец матрицы $A^{(k+1)}$ есть результат итерации k -го столбца матрицы $A^{(k)}$ матрицей $B^{(k)}$.

Таким образом, переход от матрицы $A^{(k)}$ к матрице $A^{(k+1)}$ происходит по формулам

$$\begin{aligned} b_{k+1,j}^{(k)} &= \frac{1}{a_{k+1,k}^{(k)}} \cdot a_{k+1,j}^{(k)} \\ b_{ij}^{(k)} &= a_{ij}^{(k)} - a_{ik}^{(k)} b_{k+1,j}^{(k)} \quad (i \neq k+1) \\ a_{ij}^{(k+1)} &= b_{ij}^{(k)} \quad (j \neq k+1) \\ a_{i,k+1}^{(k+1)} &= \sum_{j=1}^n b_{ij}^{(k)} a_{jk}^{(k)}. \end{aligned}$$

В качестве примера мы снова возьмем матрицу Леверье.

Поясним табл. IV. 8. В графах 2, 3, 4, 5 последовательно записываются матрицы $A^{(1)}$, $B^{(1)}$, $B^{(2)}$ и $B^{(3)}$. Графы 6 и 7 контрольные, графа 6 содержит суммы строк матриц $B^{(k)}$, графа 7 суммы строк матриц $A^{(k)}$. В графу 1 записываются k -е столбцы матрицы $A^{(k)}$, вычисление которых сопровождается обычным контролем. Коэффициенты характеристического полинома располагаются, таким образом, в четырех последних строках графы 1. Так как они вычисляются одновременно, контрольное совпадение p_1 со следом матрицы является вместе с тем и показателем точности вычисления остальных коэффициентов; полученные результаты ближе к данным Леверье, чем результаты, найденные по методу А. Н. Крылова и методу Самуэльсона.

Число операций, нужных для вычисления по методу А. М. Данилевского, значительно меньше, чем по двум указанным методам, и, как мы увидим ниже, меньше, чем по другим методам.

Именно, число умножений и делений на k -м шагу будет

$$n - k + (n - k)(n - 1) + (n - k)n = 2n(n - k),$$

и потому общее число умножений и делений будет $n^3 - n^2$.

Метод А. М. Данилевского позволяет вычислять собственные векторы как самой матрицы A , так и ее транспонированной. Действительно, так как

$$S^{-1}AS = P, \tag{1}$$

где

$$S = S_1 \ S_2 \ \dots \ S_{n-1}, \tag{2}$$

Таблица IV.8

Вычисление коэффициентов характеристического полинома по методу А. М. Данилевского

1	2	3	4	5	6	7
-5.509882	-5.509882	1.870086	0.422908	0.008814	-3.208074	
0.287865	0.287865	-11.811654	5.711900	0.058717	-5.753172	
0.049099	0.049099	4.308093	-12.970687	0.229326	-8.384229	
0.006235	0.006235	0.269851	1.397369	-17.596207	-15.922752	
-59.146733	0	-224.21096	109.75157	1.1326872	-113.32670	51.737524
-16.346028	1	-41.031921	19.842287	0.20397409	-19.985660	4.7002331
1.1367579	0	6.3226593	-13.944923	0.21931108	-7.402953	-12.588854
0.10414109	0	0.52568503	1.2736523	-17.597479	-15.798142	-16.219686
-74.251862	1	-258.39454	116.92259	-16.041507		
-698.72936	0	0	-615.81773	12.543678	-603.27405	-686.18568
-264.18530	1	0	-180.67895	3.3575611	-176.32139	-259.82774
-30.270859	0	1	-12.267276	0.19292681	-11.074349	-29.077932
1.0653607	0	0	2.5511798	-17.617571	-15.066391	-16.552210
-992.12016	1	1	-806.21278	-1.523405		
-12296.550	0	0	0	-11542.147	-11542.147	-12296.550
-5349.4555	1	0	0	-4365.4005	-4364.4005	-5348.4555
-797.27875	0	1	0	-500.38776	-499.38776	-796.27875
-47.888430	0	0	1	-16.536719	-15.536719	-46.888430

то собственный вектор U матрицы A , принадлежащий собственному значению λ , выражается через собственный вектор Y матрицы P по формуле

$$U = SY = S_1 S_2 \dots S_{n-1} Y.$$

Компоненты y_1, \dots, y_n вектора Y находятся без труда как решения системы

$$\begin{aligned} p_n y_n &= \lambda y_1 \\ y_1 + p_{n-1} y_n &= \lambda y_2 \\ y_2 + p_{n-2} y_n &= \lambda y_3 \\ \vdots &\quad \vdots \quad \vdots \\ y_{n-1} + p_1 y_n &= \lambda y_n. \end{aligned}$$

Полагая $y_n = 1$, получим последовательно

$$\begin{aligned} y_{n-1} &= \lambda y_n - p_1 = \lambda - p_1 \\ y_{n-2} &= \lambda y_{n-1} - p_2 = \lambda^2 - p_1 \lambda - p_2 \\ \vdots &\quad \vdots \quad \vdots \\ y_1 &= \lambda y_2 - p_{n-1} = \lambda^{n-1} - p_1 \lambda^{n-2} - \dots - p_{n-1}. \end{aligned}$$

Первое уравнение системы будет удовлетворяться тождественно, так как $\lambda^n - p_1 \lambda^{n-1} - \dots - p_n = 0$.

Для вычисления компонент y_i следует применять рекуррентные формулы. Эти формулы совпадают с формулами схемы Хорнера для деления характеристического полинома на $t - \lambda$.

Обозначим далее

$$Y = Y^{(n)}, Y^{(n-1)} = S_{n-1} Y^{(n)}, Y^{(n-2)} = S_{n-2} Y^{(n-1)}, \dots, Y^{(1)} = S_1 Y^{(2)} = U.$$

Компоненты вектора $Y^{(k)}$ вычисляются по компонентам вектора $Y^{(k+1)}$ по двучленным формулам

$$\begin{aligned} y_i^{(k)} &= y_i^{(k+1)} + y_{k+1}^{(k+1)} a_{ik}^{(k)} \quad (i \neq k+1) \\ y_{k+1}^{(k)} &= y_{k+1}^{(k+1)} a_{k+1,k}^{(k)}. \end{aligned} \tag{3}$$

Формулы для вычисления компонент собственного вектора транспонированной матрицы имеют несколько более простой рисунок. Переходя в равенстве (1) к транспонированным матрицам, получим

$$S' A' (S^{-1})' = P',$$

так что собственный вектор матрицы A' , принадлежащий собственному значению λ , связан с соответствующим собственным вектором Z матрицы P' соотношением

$$V = (S^{-1})' Z = (S_1^{-1})' (S_2^{-1})' \dots (S_{n-1}^{-1})' Z.$$

Компоненты z_1, \dots, z_n вектора Z находятся из системы

$$\begin{aligned} z_2 &= \lambda z_1 \\ z_3 &= \lambda z_2 \\ &\vdots \\ z_n &= \lambda z_{n-1} \\ p_n z_1 + p_{n-1} z_2 + \dots + p_1 z_n &= \lambda z_n. \end{aligned}$$

Полагая $z_1 = 1$, получим последовательно $z_2 = \lambda$, $z_3 = \lambda^2, \dots, z_n = \lambda^{n-1}$. Последнее уравнение выполняется тождественно.

Далее

$$V = Z^{(1)},$$

где

$$Z = Z^{(n)}, Z^{(n-1)} = (S_{n-1}^{-1})' Z^{(n)}, Z^{(n-2)} = (S_{n-2}^{-1})' Z^{(n-1)}, \dots, Z^{(1)} = (S_1^{-1})' Z^{(2)}.$$

На этот раз каждое преобразование $(S_k^{-1})'$ будет менять лишь одну $(k+1)$ -ю компоненту предыдущего вектора, так что компонентами вектора $Z^{(k)}$ будут $z_1, z_2, \dots, z_k, v_{k+1}, \dots, v_n$. При этом

$$\begin{aligned} v_{k+1} &= \frac{1}{a_{k+1, k}^{(k)}} \left(- \sum_{i=1}^k a_{ik}^{(k)} z_i + z_{k+1} - \sum_{i=k+2}^n a_{ik}^{(k)} v_i \right) = \\ &= \sum_{i=1}^{k+1} m_{ik} z_i + \sum_{i=k+2}^n m_{ik} v_i, \end{aligned} \quad (4)$$

где

$$\begin{aligned} m_{ik} &= -\frac{a_{ik}^{(k)}}{a_{k+1, k}^{(k)}} \quad (l \neq k+1) \\ m_{k+1, k} &= \frac{1}{a_{k+1, k}^{(k)}}. \end{aligned}$$

Формулы (3) и (4) для вычисления собственных векторов матрицы A и ее транспонированной различны по своей структуре, но требуют приблизительно одинакового числа вычислительных операций. Для нахождения собственных векторов матрицы A , конечно, можно использовать любую из этих формул, транспонируя при желании матрицу A в начале процесса.

В качестве примера рассмотрим вычисление собственного вектора матрицы Леверье, принадлежащего собственному значению λ_4 , используя данные табл. IV. 8. Для λ_4 получено приближенное значение, равное — 5.298695.

Пользуясь схемой Хорнера

$$\begin{array}{rccccc} 1 & 47.888430 & 797.27875 & 5349.4555 & 12296.550 & || & -5.298695 \\ 1 & 42.589735 & 571.60873 & 2320.6752 & \underline{-0.0000789}, & & \end{array}$$

вычислим компоненты собственного вектора Y матрицы P . Получим $Y = Y^{(4)} = (2320.6752; 571.60873; 42.589735; 1)'$. По формулам (3) находим последовательно

$$Y^{(3)} = (1621.9458; 307.42343; 12.318876; 1.0653607)'$$

$$Y^{(2)} = (893.32453; 106.05874; 14.003580; 2.3482620)'$$

$$Y^{(1)} = U = (308.95339; 30.530599; 19.210958; 3.009538)'.$$

После нормировки к единичной первой компоненте получим

$$\tilde{U} = (1.000000; 0.098819; 0.062181; 0.009741)'.$$

Для нахождения собственного вектора транспонированной матрицы, принадлежащего λ_4 , предварительно вычисляем числа m_{ik} , при $k = 1, 2, 3$, именно

$$\begin{array}{lll} m_{i1} & m_{i2} & m_{i3} \\ 19.140507 & 52.031073 & 655.86178 \\ 3.4738506 & 14.379516 & 247.97733 \\ -0.170562 & 0.87969479 & 28.413718 \\ -0.021659 & -0.09161237 & 0.93864923. \end{array}$$

Далее вычисляем

$$Z = (1, -5.298695, 28.076169, -148.76706)'$$

и

$$V = (1, 0.641980, 0.535599, 0.0138036)'.$$

Отметим, что при вычислении, особенно последней компоненты, происходит значительное уничтожение значащих цифр, что делает результат недостаточно надежным.

Если λ есть кратное собственное значение, то для него легко вычисляется вся „башня“ корневых векторов. Она будет только одна, так как (в силу предположения линейной независимости векторов $e_1, Ae_1, \dots, A^{n-1}e_1$) матрица A имеет взаимно простые элементарные делители.

Корневые векторы матрицы A (или ее транспонированной A') вычисляются по формулам (3) (или (4)), в которых компоненты собственных векторов Y (или Z) должны быть заменены компонентами корневых векторов матрицы P (или P'). Последние находятся без труда.

Как было показано выше, компоненты y_1, y_2, \dots, y_n собственного вектора Y матрицы P суть коэффициенты частного от деления характеристического полинома на $t - \lambda$. Соответственно, компоненты корневых векторов суть коэффициенты от деления характеристического полинома на $(t - \lambda)^2, (t - \lambda)^3, \dots, (t - \lambda)^m$, где m есть кратность собственного значения, причем первой компонентой всегда является свободный член. Эти компоненты находятся последовательно по схеме Хорнера.

Для матрицы же P' корневые векторы суть

$$Z = (1, \lambda, \lambda^2, \dots, \lambda^{n-1})$$

$$Z_1 = (0, -1, 2\lambda, \dots, (n-1)\lambda^{n-2})$$

$$Z_2 = \left(0, 0, 1, \dots, \frac{(n-1)(n-2)}{2} \lambda^{n-3} \right)$$

$$Z_{m-1} \equiv (0, 0, 0, \dots, 1, \dots, C_{n-1}^{m-1} \lambda^{n-m}).$$

В качестве примера рассмотрим матрицу

$$A = \begin{bmatrix} 13 & 16 & 16 \\ -5 & -7 & -6 \\ -6 & -8 & -7 \end{bmatrix},$$

собственные значения которой есть $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = -3$.

Коэффициенты характеристического полинома матрицы вычисляются в табл. IV. 9. Эти вычисления дают, что характеристический полином матрицы равен $t^3 + t^2 - 5t + 3$.

Найдем собственный и корневой вектор, принадлежащие собственному значению $\lambda = 1$.

Применяя схему Хорнера, получим

$$\begin{array}{cccc} 1, & 1, & -5, & 3 \parallel 1 \\ 1, & 2, & -3 \\ 1, & 3 \end{array}$$

так что собственный вектор Y матрицы P равен $(-3, 2, 1)'$, а корневой $\tilde{Y} = (0, 3, 1)'$.

Поэтому собственным вектором матрицы A будет

$$U = S_1 S_2 Y = (16, -4, -8)',$$

а корневым

$$\tilde{U} = S_1 S_2 \tilde{Y} = (32, -9, -14)',$$

Алгебраическое изложение метода А. М. Данилевского, данное в книге В. Н. Фаддеевой [1], эквивалентно переходу от базиса e_1, e_2, \dots, e_n к базису $A'^{n-1}e_n, A'^{n-2}e_n, \dots, A'e_n, e_n$, в результате чего каноническая матрица Фробениуса получается в несколько иной форме.

Таблица IV.9

Определение коэффициентов характеристического полинома по методу А. М. Данилевского

1	2	3	4	5	6
13	13	16	16	45	
-5	-5	-7	-6	-18	
-6	-6	-8	-7	-21	
8.6	0	-2.2	0.4	-1.8	9.0
-1.2	1	1.4	1.2	3.6	1.0
-3.2	0	0.4	0.2	0.6	-3.0
4.2	1	-0.4	1.8		
-3	0	0	0.9375	0.9375	
5	1	0	1.1250	2.1250	
-1	0	1	-0.0625	0.9375	

Отметим одну деталь, связывающую метод Данилевского и метод Хессенберга. Элементы k -го столбца матрицы $A^{(k)}$ являются, по самому построению, коэффициентами в равенстве

$$A^k e_1 = a_{1k}^{(k)} e_1 + a_{2k}^{(k)} A e_1 + \dots + a_{kk}^{(k)} A^{k-1} e_1 + a_{k+1k}^{(k)} e_{k+1} + \dots + a_{nk}^{(k)} e_n,$$

откуда следует, что полином

$$\varphi_k(t) = t^k - a_{kk}^{(k)} t^{k-1} - \dots - a_{1k}^{(k)}$$

обладает тем свойством, что вектор $\varphi_k(A) e_1 = Z_{k+1}$ имеет первые k компонент в исходном базисе равными нулю. Поэтому вектор Z_{k+1} и полином $\varphi_k(t)$ совпадают с соответствующим вектором и соответствующим полиномом в методе Хессенберга. Тем самым все числа, образующие k -й столбец матрицы $A^{(k)}$, встречаются в вычислениях, проводимых по методу Хессенберга — верхние k в качестве коэффициентов (с обратными знаками) полиномов $\varphi_k(t)$, нижние $n-k$ в качестве компонент вектора Z_{k+1} .

Обратимся теперь к рассмотрению возможных вырождений процесса.

Может оказаться, что на некотором шагу $a_{k+1,k}^{(k)} = 0$. Это показывает, что система векторов $e_1, Ae_1, \dots, A^k e_1, e_{k+2}, \dots, e_n$ ли-

нейно-зависима. Если при этом окажется, что хоть одно из чисел $a_{jk}^{(k)} \neq 0$ при $j > k+1$, то в матрице $A^{(k)}$ нужно переставить $k+1$ -ю и j -ю строки и одновременно $k+1$ -й и j -й столбцы. Такая перестановка равносильна переходу от базиса $e_1, Ae_1, \dots, A^{k-1}e_1, e_{k+1}, \dots, e_j, \dots, e_n$ к базису $e_1, Ae_1, \dots, A^{k-1}e_1, e_j, \dots, e_{k+1}, \dots, e_n$. Легко видеть, что если $e_1, Ae_1, \dots, A^{n-1}e_1$ линейно-независимы, то такое j обязательно найдется. После совершенной трансформации процесс продолжается. При вычислении собственных векторов сделанная трансформация, конечно, должна быть учтена. (В надлежащий момент нужно переставить $k+1$ -ю и j -ю компоненты соответствующих векторов.)

Указанная трансформация может быть полезна, даже если $a_{k+1,k}^{(k)} \neq 0$, но среди чисел $a_{jk}^{(k)}, j > k+1$, есть число значительно большее по модулю, чем $a_{k+1,k}^{(k)}$, так как она увеличивает точность вычисления. Если такая трансформация проделывается на каждом шаге, мы приходим к схеме, подобной схеме главных элементов метода Гаусса.

Если же все $a_{jk}^{(k)} = 0$ при $j \geq k+1$, что обозначает, что уже векторы $e_1, Ae_1, \dots, A^k e_1$ линейно-зависимы, то матрица $A^{(k)}$ имеет вид

$$\left[\begin{array}{cccc|c} 0 & 0 & \dots & 0 & a_{k1}^{(k)} \\ 1 & 0 & \dots & 0 & a_{k2}^{(k)} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & a_{kk}^{(k)} \\ \hline & & & 0 & A_2^{(k)} \end{array} \right] C = \left[\begin{array}{cc} A_1^{(k)} & C \\ 0 & A_2^{(k)} \end{array} \right],$$

и, следовательно, характеристический полином матрицы A равен произведению характеристических полиномов матриц $A_1^{(k)}$ и $A_2^{(k)}$.

Матрица $A_1^{(k)}$ уже есть каноническая матрица Фробениуса. Она соответствует индуцированному оператору на инвариантном подпространстве, натянутом на векторы $e_1, \dots, A^{k-1}e_1$.

К матрице $A_2^{(k)}$ нужно снова применить общий процесс приведения.

Таким образом, в рассматриваемом случае процесс нахождения характеристического полинома только упрощается. Однако вычисление канонического базиса (в частности и собственных векторов) несколько усложняется. Мы не будем останавливаться на решении этого вопроса.

Метод А. М. Данилевского допускает следующее обобщение. Приведение к канонической форме можно осуществить посредством перехода от базиса e_1, e_2, \dots, e_n к базису $f, Af, \dots, A^{n-1}f$, где f некоторый вектор, выбираемый, вообще говоря, произвольно, требуется только, чтобы векторы $f, Af, \dots, A^{n-1}f$ были линейно-независимыми,

Если элементарные делители матрицы взаимно просты, такой вектор всегда найдется.

Этот вариант требует уже n шагов, так как добавляется еще один „нулевой“ шаг, состоящий в переходе от базиса (e_1, e_2, \dots, e_n) к базису (f_1, f_2, \dots, f_n) , что осуществляется посредством преобразования подобия

$$A_1 = S_0^{-1} A S_0,$$

где S_0 — матрица, первый столбец которой составлен из компонент вектора f , остальные совпадают со столбцами единичной матрицы. Дальнейший ход процесса ничем не отличается от описанного выше. Бауэр [6] рекомендует выбирать начальный вектор так, чтобы его координаты по собственным векторам, соответствующим большим по модулю собственным значениям, были бы малыми. Методы построения таких векторов будут описаны в гл. IX.

§ 47. Метод Леверье и видоизменение Д. К. Фаддеева

В этом параграфе мы изложим метод, известный под названием метода Леверье [1], требующий большего числа операций, чем все рассмотренные выше методы, но совершенно не чувствительный к частным особенностям матрицы, в частности, к „провалам“ промежуточных определителей.

Пусть

$$\varphi(t) = (-1)^n [t^n - p_1 t^{n-1} - p_2 t^{n-2} - \dots - p_n] \quad (1)$$

характеристический полином матрицы и $\lambda_1, \lambda_2, \dots, \lambda_n$ его корни, среди которых могут быть равные. Обозначим

$$\sum_{i=1}^n \lambda_i^k = s_k. \quad (2)$$

Тогда справедливы соотношения, известные под названием формул Ньютона:

$$kp_k = s_k - p_1 s_{k-1} - \dots - p_{k-1} s_1 \quad (k = 1, \dots, n). \quad (3)$$

Если числа s_k известны, то, решая рекуррентную систему (3), мы сможем найти нужные нам коэффициенты p_k .

Покажем, как определяются числа s_k . Имеем

$$s_1 = \lambda_1 + \lambda_2 + \dots + \lambda_n = \operatorname{Sp} A.$$

Далее, в силу § 1 п. 10, характеристические числа матрицы A^k будут $\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k$. Следовательно,

$$s_k = \lambda_1^k + \lambda_2^k + \dots + \lambda_n^k = \operatorname{Sp} A^k. \quad (4)$$

Таким образом, процесс вычисления сводится к последовательному вычислению степеней матрицы A , затем к вычислению их следов и, наконец, к решению рекуррентной системы (3). Вычисление n

степеней матрицы A (у последней матрицы A^n нужно вычислить только диагональные элементы) требует большого числа, правда однообразных, операций, и потому метод Леверье гораздо более трудоемкий, чем вышеизложенные методы. Его ценность состоит, как это уже упоминалось, в его универсальности. Число необходимых по методу Леверье умножений равно

$$\frac{1}{2}(n-1)(2n^3 - 2n^2 + n + 2).$$

Заметим, что при вычислении степеней матрицы полезно осуществлять контроль при помощи столбца, состоящего из сумм элементов каждой строки матрицы A . Результат умножения матрицы A на этот столбец должен совпадать с аналогичным столбцом матрицы A^2 .

Действительно, пусть Σ_1 столбец сумм матрицы A , Σ_2 столбец сумм матрицы A^2 . Пусть $U = (1, 1, \dots, 1)'$. Тогда

$$\Sigma_1 = AU, \quad \Sigma_2 = A^2U,$$

т. е.

$$\Sigma_2 = A\Sigma_1. \quad (5)$$

Очевидно, сказанное верно и для других степеней.

Изложим теперь видоизменение метода, предложенное Д. К. Фаддеевым ¹⁾, которое кроме упрощений при вычислении коэффициентов характеристического полинома позволит нам определить обратную матрицу и собственные векторы матрицы.

Будем вместо последовательности A, A^2, \dots, A^n вычислять последовательность A_1, A_2, \dots, A_n , построенную следующим образом:

$$\begin{aligned} A_1 &= A, & \text{Sp } A_1 &= q_1, & B_1 &= A_1 - q_1 E \\ A_2 &= AB_1, & \frac{\text{Sp } A_2}{2} &= q_2, & B_2 &= A_2 - q_2 E \\ \cdots &\cdots & \cdots &\cdots & \cdots &\cdots \\ A_{n-1} &= AB_{n-2}, & \frac{\text{Sp } A_{n-1}}{n-1} &= q_{n-1}, & B_{n-1} &= A_{n-1} - q_{n-1} E \\ A_n &= AB_{n-1}, & \frac{\text{Sp } A_n}{n} &= q_n, & B_n &= A_n - q_n E. \end{aligned} \quad (6)$$

Докажем, что а) $q_1 = p_1, q_2 = p_2, \dots, q_n = p_n$;

б) B_n — нулевая матрица;

в) если A неособенная матрица, то

$$A^{-1} = \frac{B_{n-1}}{p_n}.$$

(Если матрица A особенная, то $(-1)^{n-1} B_{n-1}$ будет матрицей, союзной с матрицей A).

1). Д. К. Фаддеев и И. С. Соминский. Сборник задач по высшей алгебре, 1949, задача № 979. См. также Сурьо [1] и Фрейм [1].

Докажем сначала а) методом математической индукции. Очевидно, что $p_1 = \text{Sp } A = q_1$. Предположим, что $q_1 = p_1, q_2 = p_2, \dots, q_{k-1} = p_{k-1}$, и докажем, что $q_k = p_k$. Согласно нашей конструкции:

$$A_k = A^k - q_1 A^{k-1} - \dots - q_{k-1} A = A^k - p_1 A^{k-1} - \dots - p_{k-1} A.$$

Следовательно,

$$\begin{aligned} \text{Sp } A_k &= k q_k = \text{Sp } A^k - p_1 \text{Sp } A^{k-1} - \dots - p_{k-1} \text{Sp } A = \\ &= s_k - p_1 s_{k-1} - \dots - p_{k-1} s_1. \end{aligned}$$

Отсюда, в силу формул Ньютона $k q_k = k p_k$ и, следовательно, $q_k = p_k$.

Далее, в силу соотношения Кели — Гамильтона

$$B_n = A^n - p_1 A^{n-1} - \dots - p_n E = 0. \quad (7)$$

Наконец, из равенства (7) следует, что

$$AB_{n-1} = A_n = B_n + p_n E = p_n E,$$

так что

$$A^{-1} = \frac{1}{p_n} B_{n-1}. \quad (8)$$

Равенство $A_n = p_n E$ может быть использовано для контроля вычисления; очевидно, что отклонение A_n от скалярной матрицы является мерой точности вычислений. Кроме этого окончательного контроля, целесообразно пользоваться частным контролем, составляя для матриц B_i столбцы сумм. При этом справедливо соотношение

$$\Sigma_{i+1} = A \Sigma_i - p_{i+1} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \Sigma_i = A \Sigma_0 - p_i \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad (9)$$

где Σ_i столбец сумм матрицы B_i , Σ_0 — аналогичный столбец матрицы A .

Формула (8) дает алгорифм для обращения матриц. Для матриц не очень высокого порядка в случае, если нужно решить как задачу нахождения собственных чисел, так и задачу обращения матрицы, указанный метод очень удобен.

Число операций, нужных для получения коэффициентов p_i (включая и вычисление матрицы B_n), равно $(n-1) n^3$ умножений.

В табл. IV. 10 на прежнем примере показана схема вычисления по методу Д. К. Фаддеева коэффициентов характеристического полинома и элементов обратной матрицы.

Перейдем теперь к определению собственных векторов матрицы A .

Пусть собственные числа уже вычислены и при этом оказались различными. Построим матрицу

$$Q_k = \lambda_k^{n-1} E + \lambda_k^{n-2} B_1 + \dots + B_{n-1}. \quad (10)$$

Таблица IV. 10

Определение коэффициентов характеристического полинома и элементов обратной матрицы по методу Д. К. Фаддеева

A'				B_1				\sum			
$p_1 = -47.888430$	A_2	B_2	B_3	$p_2 = -797.27876$	A_3	B_4	B_5	$p_3 = -5349.4555$	A_4	A^{-1}	$p_4 = -12296.551$
-5.509882	0.287865	0.049099	0.00623	42.378548	1.870086	0.422908	0.008814	44.68036			
1.870086	-11.811654	4.308033	0.269851	0.287865	36.076776	5.711900	0.058717	42.13526			
0.422908	5.711900	-12.970687	1.397369	0.049099	4.308033	34.917743	0.229326	39.50420			
0.008814	0.058717	0.229326	-17.596207	0.006235	0.269851	1.397369	30.292223	31.96568			
$p_1 = -47.888430$				564.33711	58.98700	23.13088	0.42522	646.8802			
-282.94165	-400.96516	-427.95884	-532.69187	9.07995	396.31360	132.18346	2.39755	539.9746			
-3091.3122	-4094.0411	-4213.8419	-4649.1712	2.68546	99.69550	369.31992	4.22567	475.9266			
-12296.551	-12296.551	-12296.550	-12296.551	0.30081	11.01857	25.74858	264.58689	301.6549			
-0.001	0.001	-0.001	-0.001	70.5604	1255.4144	458.38882	276.1613	6.2599	12998.953		
0.001	0.001	-0.001	-0.001	32.0618	419.6360	556.38396	11.4757	1893.894			
0.001	0.001	-0.001	-0.001	4.4283	52.7398	1135.6136	16.2164	1603.527			
						98.8129	700.2843	856.265			

где B_i матрицы, вычисленные в процессе нахождения коэффициентов характеристического полинома, а λ_k есть k -е собственное число матрицы A .

Можно доказать, в предположении, что все $\lambda_1, \dots, \lambda_n$ различны, что матрица Q_k ненулевая.

Покажем, что каждый столбец матрицы Q_k состоит из компонент собственного вектора, принадлежащего собственному числу λ_k .

Действительно,

$$\begin{aligned} (\lambda_k E - A) Q_k &= (\lambda_k E - A) (\lambda_k^{n-1} E + \lambda_k^{n-2} B_1 + \dots + B_{n-1}) = \\ &= \lambda_k^n E + \lambda_k^{n-1} (B_1 - A) + \lambda_k^{n-2} (B_2 - AB_1) + \dots - AB_{n-1} = \\ &= \lambda_k^n E - p_1 \lambda_k^{n-1} E - p_2 \lambda_k^{n-2} E - \dots - p_n E = 0. \end{aligned}$$

Отсюда следует, что $(\lambda_k E - A) u = 0$, где u любой столбец построенной матрицы Q_k , т. е. что

$$\lambda_k u = Au. \quad (11)$$

Это равенство показывает, что u есть собственный вектор.

Замечание 1. Вычисляя собственные векторы описанным образом, нет необходимости, конечно, находить все столбцы матрицы Q_k . Следует ограничиться вычислением одного столбца; его элементы получаются в виде линейной комбинации с прежними коэффициентами одноименных столбцов матриц B_i .

Замечание 2. Для вычисления столбца u матрицы Q_k , удобно пользоваться рекуррентной формулой:

$$u_0 = e; \quad u_i = \lambda_k u_{i-1} + b_i, \quad (12)$$

где b_i — взятый нами столбец матрицы B_i , а e — одноименный столбец единичной матрицы.

Тогда

$$u = u_{n+1}.$$

В качестве примера вычислим собственный вектор X_4 матрицы Леверье, принадлежащий собственному числу $\lambda_4 = -5.29870$.

Таблица VI. II

Определение собственного вектора по методу Д. К. Фаддеева

I	II	III	IV	V	VI
2258.1433	-2990.2530	1189.8294	-148.7675	308.9522	1
70.5604	-48.1119	8.0822	0	30.5307	0.098820
32.0618	-14.2294	1.3785	0	19.2109	0.062181
4.4283	1.5939	0.1751	0	3.0095	0.009741

В графах I, II, III расположены компоненты 1-го столбца матриц B_i , умноженные на соответствующие степени λ_4 , в графе IV компоненты вектора $\lambda_4^3(1, 0, 0, 0)'$. Графа V содержит компоненты вектора X_4 , графа VI — его компоненты после нормирования.

§ 48. Эскалаторный метод¹⁾

Оригинальный метод определения собственных чисел и собственных векторов матрицы известен под названием эскалаторного метода. Этот метод дает индуктивную конструкцию, посредством которой, зная собственные числа и собственные векторы матрицы A_{k-1} и ее транспонированной, можно составить уравнение для определения собственных чисел матрицы A_k , полученной из A_{k-1} окаймлением, и затем вычислить по несложным формулам компоненты собственных векторов для матрицы A_k и ее транспонированной. Применение эскалаторного метода начинается с отыскания собственных векторов матрицы 2-го порядка. Эта задача решается совсем просто.

Большим достоинством метода является наличие мощного контроля, дающего возможность вычислителю на каждом шагу быть уверенным как в своих вычислениях, так и в отсутствии потери значащих цифр. Кроме того, сама форма уравнения для определения собственных значений оказывается очень удобной при применении метода Ньютона.

Метод основан на использовании свойств ортогональности собственных векторов матрицы и ее транспонированной.

Мы не будем проводить общей индукции от k -го шага к $k+1$ -му, а ограничимся рассмотрением перехода от матрицы 3-го порядка к матрице 4-го порядка. Для удобства компоненты векторов будем обозначать, вопреки общепринятому, различными буквами. Предположим, что все собственные числа матрицы A_3 вещественны и различны.

Итак, пусть λ_r ($r = 1, 2, 3$) собственные числа матриц A_3 и A'_3 , где

$$A_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}. \quad (1)$$

Пусть далее $X_r = (x_r, y_r, z_r)'$ и $X'_r = (x'_r, y'_r, z'_r)'$ ($r = 1, 2, 3$) совокупность собственных векторов этих матриц.

¹⁾ Моррис и Хед [1], [2]; Моррис [5].

Эти собственные векторы могут быть нормированы так, что

$$\begin{aligned} & \begin{bmatrix} x'_1 & x'_2 & x'_3 \\ y'_1 & y'_2 & y'_3 \\ z'_1 & z'_2 & z'_3 \end{bmatrix} \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{bmatrix} = \\ & = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{bmatrix} \begin{bmatrix} x'_1 & x'_2 & x'_3 \\ y'_1 & y'_2 & y'_3 \\ z'_1 & z'_2 & z'_3 \end{bmatrix} = E. \quad (2) \end{aligned}$$

Это следует из установленных в § 10 п. 3 свойств ортогональности собственных векторов матрицы и ее транспонированной.

Пусть A_4 матрица 4-го порядка, полученная из A_3 окаймлением, $X = (x, y, z, u)'$ ее собственный вектор, принадлежащий собственному числу λ .

Имеем

$$\begin{aligned} \lambda x &= a_{11}x + a_{12}y + a_{13}z + a_{14}u \\ \lambda y &= a_{21}x + a_{22}y + a_{23}z + a_{24}u \\ \lambda z &= a_{31}x + a_{32}y + a_{33}z + a_{34}u \\ \lambda u &= a_{41}x + a_{42}y + a_{43}z + a_{44}u. \end{aligned} \quad (3)$$

Умножим первые три уравнения системы (3), соответственно, на x'_r , y'_r , z'_r и сложим. Получим:

$$\begin{aligned} \lambda(xx'_r + yy'_r + zz'_r) &= (a_{11}x'_r + a_{21}y'_r + a_{31}z'_r)x + \\ &+ (a_{12}x'_r + a_{22}y'_r + a_{32}z'_r)y + (a_{13}x'_r + a_{23}y'_r + a_{33}z'_r)z + \\ &+ (a_{14}x'_r + a_{24}y'_r + a_{34}z'_r)u, \end{aligned}$$

откуда в силу того, что $(x'_r, y'_r, z'_r)'$ есть собственный вектор для матрицы A'_3

$$\lambda(xx'_r + yy'_r + zz'_r) = \lambda_r(xx'_r + yy'_r + zz'_r) + (a_{14}x'_r + a_{24}y'_r + a_{34}z'_r)u$$

и, следовательно,

$$xx'_r + yy'_r + zz'_r = -\frac{P'_ru}{\lambda_r - \lambda}, \quad (4)$$

где

$$P'_r = a_{14}x'_r + a_{24}y'_r + a_{34}z'_r. \quad (5)$$

Пусть

$$P_r = a_{41}x_r + a_{42}y_r + a_{43}z_r. \quad (6)$$

Тогда, в силу ортогональных свойств (2), справедливо следующее соотношение:

$$\sum_{r=1}^3 P_r(x'_r x + y'_r y + z'_r z) = P, \quad (7)$$

где

$$P = a_{41}x + a_{42}y + a_{43}z = -(a_{44} - \lambda)u. \quad (8)$$

Действительно,

$$\begin{aligned} \sum_{r=1}^3 (a_{41}x_r + a_{42}y_r + a_{43}z_r)(x'_r x + y'_r y + z'_r z) &= \\ &= a_{41}(x_1 x'_1 + x_2 x'_2 + x_3 x'_3)x + a_{41}(x_1 y'_1 + x_2 y'_2 + x_3 y'_3)y + \\ &\quad + a_{41}(x_1 z'_1 + x_2 z'_2 + x_3 z'_3)z + \dots = a_{41}x + a_{42}y + a_{43}z. \end{aligned}$$

Заменив в уравнении (7) выражение $x'_r x + y'_r y + z'_r z$ на $-\frac{P'_r u}{\lambda_r - \lambda}$ согласно (4), получим следующее уравнение для определения собственных чисел матрицы A_4 :

$$a_{44} - \lambda = \sum_{r=1}^3 \frac{P'_r P'_r}{\lambda_r - \lambda}. \quad (9)$$

Уравнение (9) мы будем называть эскалаторной формой характеристического уравнения или эскалаторным уравнением. Далее, умножая (4) последовательно на x_r , y_r , z_r ($r = 1, 2, 3$) и складывая, получим, снова принимая во внимание свойства ортогональности (2):

$$\begin{aligned} \frac{x}{u} &= -\sum_{r=1}^3 \frac{P'_r x_r}{\lambda_r - \lambda}; \quad \frac{y}{u} = -\sum_{r=1}^3 \frac{P'_r y_r}{\lambda_r - \lambda}; \\ \frac{z}{u} &= -\sum_{r=1}^3 \frac{P'_r z_r}{\lambda_r - \lambda}. \end{aligned} \quad (10)$$

Аналогично

$$\begin{aligned} \frac{x'}{u'} &= -\sum_{r=1}^3 \frac{P_r x'_r}{\lambda_r - \lambda}; \quad \frac{y'}{u'} = -\sum_{r=1}^3 \frac{P_r y'_r}{\lambda_r - \lambda}; \\ \frac{z'}{u'} &= -\sum_{r=1}^3 \frac{P_r z'_r}{\lambda_r - \lambda}. \end{aligned} \quad (11)$$

Таким образом, найдя собственное число λ из уравнения (9), мы находим по формулам (10) и (11) собственные векторы матриц A_4 и A'_4 , принадлежащие этому числу, с точностью до численного множителя. Для возможности продолжения процесса мы должны еще нормировать их в смысле формулы (2).

Нетрудно проверить (снова используя свойства (2)), что

$$\frac{xx' + yy' + zz'}{uu'} = \sum_{r=1}^3 \frac{P_r P'_r}{(\lambda_r - \lambda)^2}.$$

Следовательно,

$$\frac{xx' + yy' + zz' + uu'}{uu'} = 1 + \sum_{r=1}^3 \frac{P_r P'_r}{(\lambda_r - \lambda)^2}.$$

Таким образом, мы удовлетворим условию нормированности при

$$\frac{1}{uu'} = 1 + \sum_{r=1}^3 \frac{P_r P'_r}{(\lambda_r - \lambda)^2}.$$

Заметим, что если левую часть эскалаторного уравнения обозначить через $f(\lambda)$

$$f(\lambda) = -a_{44} + \lambda + \sum_{r=1}^3 \frac{P_r P'_r}{\lambda_r - \lambda}, \quad (12)$$

то

$$f'(\lambda) = 1 + \sum_{r=1}^3 \frac{P_r P'_r}{(\lambda_r - \lambda)^2}.$$

Таким образом,

$$\frac{1}{uu'} = f'(\lambda).$$

Не теряя общности, мы можем считать, что $u = \pm u'$, выбирая знак так, чтобы $\frac{1}{u^2} = \pm f'(\lambda)$ было положительным. Это дает

$$\begin{aligned} u = u' &= \frac{1}{\sqrt{f'(\lambda)}}, & \text{если } f'(\lambda) > 0; \\ u = -u' &= \frac{1}{\sqrt{-f'(\lambda)}}, & \text{если } f'(\lambda) < 0. \end{aligned} \quad (13)$$

В заключение отметим контрольные равенства:

$$\begin{aligned} \sum_{r=1}^3 P_r x'_r &= a_{41}; & \sum_{r=1}^3 P_r y'_r &= a_{42}; & \sum_{r=1}^3 P_r z'_r &= a_{43}; \\ \sum_{r=1}^3 P'_r x_r &= a_{14}; & \sum_{r=1}^3 P'_r y_r &= a_{24}; & \sum_{r=1}^3 P'_r z_r &= a_{34}; \\ \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 &= a_{44}. \end{aligned} \quad (14)$$

Последние равенства показывают, что все „новые“ элементы матрицы A_4 употребляются нами для контроля. Хорошее совпадение контрольных формул гарантирует правильность вычислений на каждом шагу. При окончании процесса полезно проверить для найденных векторов выполнение условия ортогональности.

Мы описали процесс только для матриц четвертого порядка: переход к общему случаю очевиден. При соблюдении контрольных

равенств метод гарантирует очень большую точность как для всех собственных чисел, так и для компонент принадлежащих им собственных векторов.

В случае симметричной матрицы эскалаторный процесс естественно облегчается, так как при этом все величины, отмеченные штрихом (т. е. относящиеся к транспонированной матрице) будут совпадать с соответствующими величинами без штрихов. Из вида эскалаторного уравнения можно в этом случае заключить, что собственные корни последовательно окаймляемых матриц разделяются. Это обстоятельство сильно облегчает определение корней, которые обычно находятся по методу Ньютона.

Заметим, что эскалаторная форма характеристического уравнения оказывается удобнее для применения способа Ньютона, чем развернутая, так как вычисление $f(\lambda)$ и $f'(\lambda)$ осуществляется очень легко.

Не вдаваясь в подробности, отметим, что в случае, когда последовательные эскалаторные уравнения имеют одинаковые или комплексные корни, описанный процесс должен быть некоторым образом¹⁾ изменен.

Найдем при помощи эскалаторного метода собственные числа и собственные векторы для матрицы Леверье.

Решение будет состоять из трех этапов.

I этап. Для матрицы A_2

$$\begin{bmatrix} -5.509882 & 1.870086 \\ 0.287865 & -11.811654 \end{bmatrix}$$

уравнение для определения собственных чисел будет

$$\lambda^2 + 17.321536\lambda + 64.542487 = 0;$$

его корни:

$$\lambda_1 = -11.895952, \quad \lambda_2 = -5.425584.$$

Для контроля образуем $\lambda_1 + \lambda_2 = -17.321536$ и вычислим след матрицы A_2 :

$$\text{Sp } A_2 = -17.321536.$$

Далее, вычисляем собственные векторы матриц A_2 и A'_2 , решая соответствующие системы, и нормируем полученные векторы:

X_1	X_2	X'_1	X'_2	
-0.061767	4.679234	-0.210926	0.210926	
0.210926	0.210926	4.679234	0.061767	I

X_1	X_2	X'_1	X'_2	
0.905643	1.138422	26.638114	0.442009	
				P_i и P'_i

¹⁾ Моррис и Хэд [2].

Первый этап окончен.

II этап. Образуем матрицу

$$\begin{bmatrix} -5.509882 & 1.870086 & 0.422908 \\ 0.287865 & -11.811654 & 5.711900 \\ 0.049099 & 4.308033 & -12.970687 \end{bmatrix}.$$

Выписываем на отдельный листок вновь введенные коэффициенты a_{13} , a_{23} и a_{31} , a_{32} в виде столбцов и, прикладывая их к столбцам собственных векторов матрицы A_2 , находим величины P'_i и P_i (накоплением). Для удобства дальнейших вычислений выписываем их рядом с собственными векторами в схеме (I), располагая в строку.

Теперь мы можем написать эскалаторное уравнение для матрицы A_3 :

$$f(\lambda) = 12.970687 + \lambda + \frac{24.124621}{-11.895952 - \lambda} + \frac{0.503193}{-5.425584 - \lambda} = 0.$$

Определяем его корни по методу Ньютона, располагая вычисление по схеме:

λ	-15	-16.651	-17.3458	-17.3975	-17.397655
$-11.895952 - \lambda$	3.104	4.755	5.4498	5.501548	5.501703
$-5.425584 - \lambda$	9.574	11.225	11.9202	11.971916	11.972071
$12.970687 + \lambda$	-2.029	-3.680	-4.3751	-4.426813	-4.426968
$P_1 P'_1$	7.772	5.074	4.4267	4.385061	4.384936
$-11.895958 - \lambda$	0.053	0.045	0.0422	0.042031	0.042031
$f(\lambda)$	5.796	1.439	0.0938	0.000279	-0.000001
$P_2 P'_2$	2.504	1.067	0.8123	0.797059	0.797014
$(-11.895952 - \lambda)^2$	0.006	0.004	0.0035	0.003511	0.003511
$f'(\lambda)$	3.510	2.071	1.8158	1.800570	1.800525
$\Delta\lambda$	-1.651	-0.6948	-0.0517	-0.000155	0.000000

Таким образом, $\lambda_1 = -17.397655$.

Аналогично находим $\lambda_2 = -7.594378$ и $\lambda_3 = -5.300190$. Контроль:

$$\lambda_1 + \lambda_2 + \lambda_3 = -30.292223, \quad \operatorname{Sp} A_3 = -30.292223.$$

Перейдем теперь к определению компонент собственных векторов матриц A_3 и A'_3 , которые находятся, с точностью до множителя, по формулам, аналогичным формулам (10) и (11). При этом удобно составить вспомогательные схемы (III) и (IV).

$P'_i x_i$	$P'_i y_i$	$P_i x'_i$	$P_i y'_i$	
-1.645356 2.068264	5.618671 0.093231	-0.191024 0.240123	4.237716 0.070318	III
0.422908	5.711902	0.049099	4.308034	

$\frac{1}{\lambda_i + 17.397655}$	$\frac{1}{\lambda_i + 7.594378}$	$\frac{1}{\lambda_i + 5.300190}$	
0.181762 0.083528	-0.232473 0.461086	-0.151613 -7.974863	IV

Схема (III) содержит произведения чисел P_i и P'_i на компоненты соответствующих векторов и получается из схемы (I); последняя строка осуществляет контроль (например, $\sum_{i=1}^2 P'_i x_i = a_{12}$). Схема (IV) содержит множители $\frac{1}{\lambda_i - \lambda}$, где вместо λ последовательно взяты три вычисленных корня.

Далее, нормирующие множители определяются из схемы (II) и двух аналогичных схем, служащих для вычисления двух других корней. Так как $f'(\lambda_i) > 0$,

$$D_i = \frac{1}{\sqrt{f'(\lambda_i)}} \quad \text{и} \quad z_i = z'_i = D_i \quad (i = 1, 2, 3).$$

Вычисляя, получим $D_1 = 0.745248$, $D_2 = 0.644055$, $D_3 = 0.172627$.

Используя предыдущие схемы, без труда находим компоненты собственных векторов матриц A_3 и A'_3 (в окончательную схему мы их вписываем после умножения на соответствующий множитель нормирования).

X_1	X_2	X_3	X'_1	X'_2	X'_3	
0.094129	-0.860553	2.804268	0.010928	-0.099909	0.325572	
-0.766896	0.813572	0.275404	-0.578409	0.613612	0.207717	
0.745248	0.644055	0.172627	0.745248	0.644055	0.172627	
0.835026	1.114160	0.333026	0.137039	0.182847	0.054654	P_i и P'_i

Второй этап окончен.

III этап. Вычислив величины P_i и P'_i , пишем эскалаторное уравнение для матрицы A_4 :

$$17.596207 + \lambda + \frac{0.114431}{-17.397655 - \lambda} + \frac{0.203721}{-7.594378 - \lambda} + \frac{0.018201}{-5.300190 - \lambda} = 0$$

и вычисляем его корни:

$$\lambda_1 = -17.863262, \quad \lambda_2 = -17.152427$$

$$\lambda_3 = -7.574044, \quad \lambda_4 = -5.298698$$

$$\sum_{i=1}^4 \lambda_i = -47.888431; \quad \text{Sp } A_4 = -47.888430.$$

Далее вычисляем собственные векторы матрицы A_4 , нормируя их обычным образом:

X_1	X_2	X_3	X_4
-0.019872	0.032932	-0.351235	1.135218
0.169807	-0.261310	0.328467	0.112183
-0.187215	0.236640	0.260927	0.070591
0.808482	0.586694	0.045005	0.011058

и собственные векторы матрицы A'_4 :

X'_1	X'_2	X'_3	X'_4
-0.014058	0.023297	-0.248476	0.803091
0.780383	-1.200908	1.509559	0.515564
-1.140762	1.441927	1.589924	0.430123
0.808482	0.586694	0.045005	0.011058

Окончательным контролем вычисления является вычисление произведения двух соответствующих матриц, состоящих из компонент собственных векторов. Вместо единичной матрицы была получена матрица

$$\begin{bmatrix} 1.000005 & -0.000004 & 0.000000 & 0.000002 \\ -0.000003 & 1.000004 & -0.000002 & -0.000003 \\ -0.000002 & 0.000000 & 0.999994 & 0.000000 \\ -0.000001 & 0.000000 & 0.000004 & 1.000006 \end{bmatrix}.$$

Наконец, для сравнения представим эскалаторное уравнение в виде обычного полинома:

$$t^4 + 47.888430t^3 + 797.27877t^2 + 5349.4556t + 12296.550 = 0,$$

а собственный вектор, принадлежащий λ_4 , нормируем так, чтобы его первая компонента была равна единице. Это дает $X_4 = (1; 0.098820; 0.062183; 0.009741)$.

Мы видим, что коэффициенты написанного уравнения совпадают с данными Леверье с большей точностью, чем при вычислении другими методами. Соотношения ортогональности также выполнены со значительной точностью.

§ 49. Метод интерполяции

Методы, приведенные нами в предыдущих параграфах, решали задачу приведения к полиномиальному виду векового уравнения. Развитый в этом параграфе метод интерполяции применим к более общей задаче, а именно к развертыванию определителя вида

$$F(t) = \begin{vmatrix} f_{11}(t) & \dots & f_{1n}(t) \\ \vdots & \ddots & \vdots \\ f_{n1}(t) & \dots & f_{nn}(t) \end{vmatrix} \quad (1)$$

($f_{ik}(t)$ данный полином от t), в частности, к развертыванию характеристического определителя $D(t) = |A - tE|$ и определителя $|A - Bt|$, где A и B данные матрицы.

Сущность метода заключается в следующем. Пусть известно, что $F(t)$ есть полином, степень которого не превосходит числа k . Как известно, такой полином вполне определяется своими значениями в $k+1$ -й точке и может быть восстановлен по таким значениям при помощи той или другой интерполяционной формулы.

Поэтому для явного представления $F(t)$ нужно вычислить значение $k+1$ численных определителей

$$F(\lambda_i) = \begin{vmatrix} f_{11}(\lambda_i) & \dots & f_{1n}(\lambda_i) \\ \vdots & \ddots & \vdots \\ f_{n1}(\lambda_i) & \dots & f_{nn}(\lambda_i) \end{vmatrix} \quad (i = 0, 1, \dots, k), \quad (2)$$

где $\lambda_0, \lambda_1, \dots, \lambda_k$ некоторые числа, выбираемые, вообще говоря, произвольно.

Вычисление нужных определителей можно осуществить, например, по схеме, изложенной в § 17.

Для построения полинома $F(t)$ по его значениям наиболее удобно пользоваться интерполяционной формулой Ньютона, применимой для равноотстоящих абсцисс λ_i .

Мы приводим формулу Ньютона для $\lambda_i = t, i = 0, \dots, k$:

$$F(t) = \sum_{i=0}^k \frac{\Delta^i F(0)}{i!} t(t-1)\dots(t-i+1), \quad (3)$$

где $\Delta^i F(l)$ обозначает i -ю разность вычисленных значений полинома $F(t)$, определяемую по рекуррентной формуле

$$\Delta^i F(l) = \Delta^{i-1} F(l+1) - \Delta^{i-1} F(l).$$

Положим

$$\frac{t(t-1)\dots(t-i+1)}{i!} = \sum_{m=1}^i c_{mi} t^m.$$

Тогда формула (3) преобразуется к виду:

$$\begin{aligned} F(t) &= \sum_{i=0}^k \Delta^i F(0) \left(\sum_{m=1}^i c_{mi} t^m \right) = \\ &= F(0) + \sum_{m=1}^k \left(\sum_{i=m}^k c_{mi} \Delta^i F(0) \right) t^m. \end{aligned} \quad (4)$$

Эта формула носит название интерполяционной формулы А. А. Маркова.

В работе Ш. Е. Микеладзе [1], в которой был описан интерполяционный метод раскрытия полиномиальных определителей, в качестве интерполяционной формулы выбрана формула (4).

Таблица коэффициентов интерполяционной формулы А. А. Маркова для $m \leq i \leq 20$ приведена в книге В. Н. Фаддеевой [1].

При пользовании интерполяционной формулой целесообразно для контроля опорных значений определителя (1) вычислить еще хоть одно значение $F(t)$, именно, в нашем случае, $F(k+1)$, ибо $\Delta^{k+1} F(0)$ должно быть равным нулю, а $\Delta^k F(0)$ и $\Delta^k F(1)$ равными между собой.

Метод интерполяции требует большого числа действий. Так, для вычисления коэффициентов характеристического полинома при помощи интерполяционной формулы (4) требуется прежде всего вычислить $(n+1)$ определителей n -го порядка. Это потребует

$$\frac{n+1}{3} (n-1)(n^2+n+3) \text{ умножений и делений.}$$

Если коэффициенты интерполяционной формулы взять из таблицы, то для получения коэффициентов характеристического полинома нужно еще произвести $\frac{n(n+1)}{2}$ умножение.

Таким образом, общее число нужных операций умножений и делений

$$\frac{n+1}{3} (n-1)(n^2+n+3) + \frac{n(n+1)}{2}$$

на много превышает число операций, нужных для вычисления тех же коэффициентов по методу А. М. Данилевского и по методу А. Н. Крылова.

Кроме того, указанный метод не позволяет как-нибудь упростить задачу нахождения собственных векторов матрицы, в то время как при вычислении по методам А. М. Данилевского или А. Н. Крылова задача об определении собственных векторов матрицы сильно упрощается. Тем не менее, метод интерполяции интересен как метод, дающий возможность решать более общие задачи.

В качестве примера снова проведем вычисления для матрицы Леверье.

Опуская утомительное вычисление определителей, найдем, что

$$\begin{aligned}\varphi(0) &= 12296.55, & \varphi(1) &= 18492.17, & \varphi(2) &= 26583.68 \\ \varphi(3) &= 36894.41, & \varphi(4) &= 49771.69.\end{aligned}$$

Далее, составим таблицу разностей:

t	$\varphi(t)$	Δ	Δ^2	Δ^3	Δ^4
0	12296.55				
1	18492.17	6195.62			
2	26583.68	8091.51	1895.89		
3	36894.41	10310.73	2219.82	323.33	
4	49771.69	12877.28	2566.55	347.33	24

(Отметим, что в случае вычисления коэффициентов характеристического полинома вычисление лишних значений $\varphi(t)$ для контроля можно не производить, так как надежным контролем в этом случае является равенство $\Delta^k \varphi(0) = (-1)^k n!$).

Наконец, вычисляем коэффициенты характеристического полинома, располагая вычисления по схеме:

t	$\Delta^i \varphi(0)$	c_{4t}	c_{3t}	c_{2t}	c_{1t}
1	6195.62				1.0000 0000
2	1895.89			0.5000 0000	-0.5000 0000
3	323.33		0.1666 6667	-0.5000 0000	0.3333 3333
4	24.00	0.0416 6667	-0.2500 0000	0.4583 3333	-0.2500 0000
			47.8883	797.280	5349.45

Коэффициент $p_4 = \varphi(0) = 12296.55$.

Метод интерполяции применим всегда; в частности, случай, когда характеристический полином имеет кратные корни, ничем не выделяется среди других случаев.

Если вместо чисел $0, \dots, k$ взять в качестве узлов интерполяции числа $t_i = a + hi$, то формула (4) видоизменится следующим образом:

$$F(t) = F(a) + \sum_{m=1}^k \left(\sum_{i=m}^k c_{mi} h^i \Delta^i F(a) \right) (t - a)^m. \quad (5)$$

Иногда может быть целесообразно в качестве интерполяционных абсцисс брать неравноотстоящие числа. В этом случае можно пользоваться общей интерполяционной формулой Ньютона. Однако в случае неравноотстоящих абсцисс удобнее строить нужный полином методом неопределенных коэффициентов. Именно,

$$F(t) = a_0 t^k + a_1 t^{k-1} + \dots + a_k.$$

Тогда для определения чисел a_j , $j = 0, \dots, k$, мы получим систему алгебраических уравнений

$$F(\lambda_i) = a_0 \lambda_i^k + a_1 \lambda_i^{k-1} + \dots + a_k,$$

которую можно решить каким-либо из изложенных методов.

Метод интерполяции удобно применять, в частности, к раскрытию определителя $|A - Bt|$ в случае, если матрица B имеет малый определитель. Если же определитель B не является малым числом, то определение коэффициентов искомого полинома лучше производить посредством преобразования

$$|A - Bt| = |B| |AB^{-1} - tE|.$$

Матрица AB^{-1} может быть найдена по методу исключения (§ 22).

Интерполяционный метод оказывается полезным для вычисления собственных значений минуя вычисление характеристического полинома.

Существует следующий метод для определения корней полинома¹⁾ $f(t)$ (или какой-либо другой аналитической функции²⁾), являющийся несложным обобщением известного способа ложного положения. Берутся три произвольных значения t_0, t_1, t_2 независимой переменной и вычисляются соответствующие значения функции $y_0 = f(t_0)$, $y_1 = f(t_1)$, $y_2 = f(t_2)$. По найденным значениям строится интерполяционный полином второй степени и находятся его корни. Тот из корней, который оказался ближе к t_2 , чем другой, принимается за следующее приближение t_3 . Далее тем же способом строится t_4 , исходя из тройки чисел t_1, t_2, t_3 и т. д.

Доказательство сходимости этого процесса, при произвольных начальных приближениях, неизвестно. Оно известно лишь в предположении о достаточной близости исходных приближений к вычисляемому корню.¹⁾ Однако практически процесс оказывается всегда сходящимся.

Выбор интерполяционного полинома именно второй степени (а не первой и не более высокой) удобен тем, что при нем без существенных осложнений имеется возможность, в случае необходимости, выхода на комплексную плоскость с вещественной осью, даже если $f(t)$ имеет вещественные коэффициенты и исходное приближение берется вещественным.

Расчетные формулы таковы. Положим $t_1 - t_0 = \Delta_1$, $t_2 - t_1 = \Delta_2$, $t_3 - t_2 = \Delta_3$. Тогда Δ_3 есть корень квадратного уравнения

$$az^2 + bz + c = 0, \quad (6)$$

при

$$\begin{aligned} a &= \Delta_1(y_2 - y_1) + \Delta_2(y_0 - y_1) \\ b &= \Delta_1(\Delta_1 + 2\Delta_2)(y_2 - y_1) + \Delta_2^2(y_0 - y_1) \\ c &= \Delta_1\Delta_2(\Delta_1 + \Delta_2)y_2. \end{aligned} \quad (7)$$

Формулы становятся еще удобнее, если ввести отношения поправок, положив

$$\delta_2 = \frac{\Delta_2}{\Delta_1}; \quad \delta_3 = \frac{\Delta_3}{\Delta_2}. \quad (8)$$

Для δ_3 получим квадратное уравнение

$$\alpha\delta^2 + \beta\delta + \gamma = 0 \quad (9)$$

¹⁾ Мюллэр [Muller D.], Math. Tables and Other Aids Comput., 1956, 10, 208—215.

²⁾ Франк [Frank W. L.], J. Assoc. Comput. Machinery, 1958, 5, № 2, 154—160.

с коэффициентами

$$\begin{aligned}\alpha &= \delta_2(y_2 - y_1) + \delta_2^2(y_0 - y_1) \\ \beta &= (1 + 2\delta_2)(y_2 - y_1) + \delta_2^2(y_0 - y_1) \\ \gamma &= (1 + \delta_2)y_2.\end{aligned}\quad (10)$$

Для решения квадратного уравнения здесь удобна формула

$$\delta = \frac{2\gamma}{-\beta \pm \sqrt{\beta^2 - 4\alpha\gamma}}, \quad (11)$$

причем знак при квадратном корне должен выбираться так, чтобы из двух возможных значений знаменателя получалось большее по модулю.

Новая поправка получится по формуле $\Delta_3 = \Delta_2 \delta_3$, новое приближение

$$t_3 = t_2 + \Delta_3. \quad (12)$$

При применении этого приема к вычислению собственных значений¹⁾ нет нужды в вычислении коэффициентов характеристического полинома. Вычисление значений характеристического полинома при фиксированных значениях t сводится к вычислению численных определителей $|A - tE|$, что возможно производить без особого труда, например, по методу исключения.

Описанный прием особенно хорошо приложим, если матрица A имеет вид, удобный для вычисления определителей $|A - tE|$, например, если матрица почти треугольна или трехдиагональна. Эти соображения дают основания рекомендовать применение квадратичной интерполяции после преобразования исходной матрицы к почти треугольной форме (например, методом Хессенберга или методом вращений Гивенса, см. § 51) или к трехдиагональной (например, биортогональным алгорифмом, см. § 63, или методом вращений в симметричном случае).

В качестве примера вычислим два собственных значения матрицы (4) § 51. В табл. IV. 13 эта матрица подобным преобразованием приведена к трехдиагональной форме.

Возьмем $t_0 = 0$, $t_1 = 0.5$, $t_2 = 1$.

Тогда $y_0 = 0.28615247$, $y_1 = -0.01927552$, $y_2 = -0.07370353$. Проведя вычисления по формулам (10), (11), (12), получим $t_3 = -1.26691227$. Вычисляя определитель $|A - t_3 E|$, получим $y_3 = -0.31978253$. Снова применяем формулы (10), (11), (12), увеличив в них все индексы на единицу. Получим $t_4 = 0.84456137$.

После шести шагов процесс стабилизируется на $t_8 = 0.79670667$. Это значение совпадает в пределах точности с λ_2 .

¹⁾ Франк [1].

Вычисляем теперь другой корень, исходя из тех же начальных приближений $t_0 = 0$, $t_1 = 0.5$, $t_2 = 1$. Значения y_i находим по формуле

$$y_i = \frac{1}{t_i - \lambda_2} \cdot |A - t_i E|.$$

После трех шагов процесса получим

$$\lambda_3 = 0.63828382$$

с удовлетворительной точностью.

§ 50. Метод ортогонализации последовательных итераций

Излагаемый метод, подобно методу А. Н. Крылова и методу Хессенберга, имеет целью отыскание равной нулю линейной комбинации последовательности итераций произвольного вектора матрицей A . В то время как в методе Крылова это делается при помощи решения линейной системы, а в методе Хессенберга постепенным наращиванием нулевых компонент в „исправленных“ итерациях, в этом параграфе мы для той же цели применим процесс ортогонализации.

Именно, исходя из вектора X_1 , строим его итерацию AX_1 и ортогонализуем ее с вектором X_1 , т. е. строим вектор $X_2 = AX_1 + g_{11}X_1$ так, что $(X_1, X_2) = 0$. Это будет выполнено, если

$$g_{11} = -\frac{(AX_1, X_1)}{(X_1, X_1)}.$$

Далее, строим вектор AX_2 и ортогонализуем его с векторами X_1 и X_2 . В результате придем к вектору

$$X_3 = AX_2 + g_{21}X_1 + g_{22}X_2,$$

где

$$g_{21} = -\frac{(AX_2, X_1)}{(X_1, X_1)}, \quad g_{22} = -\frac{(AX_2, X_2)}{(X_2, X_2)}.$$

Процесс естественно продолжается по формулам

$$X_{i+1} = AX_i + g_{ii}X_1 + g_{i2}X_2 + \dots + g_{ii}X_i,$$

$$g_{ik} = -\frac{(AX_k, X_k)}{(X_k, X_k)} \quad (k = 1, 2, \dots, i)$$

до тех пор, пока мы не придем к нулевому вектору. Это во всяком случае произойдет на n -м шагу процесса, но может случиться и ранее, если минимальный аннулирующий X_1 полином не является характеристическим полиномом. Ясно, что

$$X_{i+1} = \varphi_i(A)X_1,$$

где полиномы $\varphi_i(t)$ связаны друг с другом рекуррентными соотношениями

$$\varphi_i(t) = (t + g_{ii})\varphi_{i-1}(t) + \dots + g_{ii}\varphi_0(t).$$

Таким образом, вычислив все коэффициенты g_{ij} , мы можем последовательно вычислить все полиномы $\varphi_0 = 1$, $\varphi_1(t), \dots, \varphi_n(t) = \varphi(t)$. В случае, если процесс оканчивается раньше времени, мы получим минимальный аннулирующий вектор X_1 полином.

Заметим, что в нормальном случае $AX + XG = 0$, где X есть неособенная матрица со столбцами X_1, \dots, X_n , а

$$G = \begin{bmatrix} g_{11} & \cdots & g_{1n} \\ -1 & \cdots & g_{2n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & -1 & g_{nn} \end{bmatrix}.$$

Таким образом, матрица A подобна матрице $-G$. Это обстоятельство позволяет определить собственные векторы матрицы A в точности тем же процессом, что и в методе Хессенберга.

В прилагаемой табл. IV. 12 дается численная иллюстрация метода на примере матрицы Леверье.

Таблица состоит из пяти частей. В первой части помещены взаимно ортогональные векторы X_1, \dots, X_4 и их итерации. Части II, III и IV содержат (X_i, X_k) , (AX_i, X_k) и g_{ik} соответственно. В части V вычисляются коэффициенты характеристического полинома по рекуррентным формулам. При вычислении итераций производится обычный контроль по суммам. Из табл. IV. 12 видно, что коэффициенты характеристического полинома вычислены с достаточной степенью точности. В качестве контроля можно также вычислить теоретически нулевую матрицу $AX + GX$.

Метод ортогонализации итераций в общем случае довольно трудоемок. Объем вычислительной работы при его применении больше, чем, например, при применении метода Хессенберга.

Однако в случае, если матрица A симметрична, картина чрезвычайно упрощается, именно, в этом случае матрица G будет трехдиагональной.

Действительно, матрица X , столбцы которой попарно ортогональны, может быть представлена в виде

$$X = UD,$$

где U — ортогональная матрица, $D = [d_1, \dots, d_n]$, $d_i = \sqrt{X_i \cdot X_i}$. Из равенства $AX + XG = 0$ следует, что $DGD^{-1} = -U^{-1}AU$. Если A симметрична, то матрица $-U^{-1}AU$ тоже симметрична. Следовательно, $d_i g_{ij} d_j^{-1} = d_j g_{ji} d_i^{-1}$. Так как $g_{ij} = 0$ при $i - j > 1$, то $g_{ji} = 0$ при $i - j > 1$, т. е. G действительно есть трехдиагональная матрица.

Метод ортогонализации итераций в применении к симметричной матрице носит название метода минимальных итераций. Он был впервые описан Ланцошем в его широко известных работах [2] — [5]. Среди вычислительных методов линейной алгебры метод

Taganrog IV, 12

Определение коэффициентов характеристического полинома методом ортогонализации итераций

минимальных итераций занимает исключительно важное положение ввиду многочисленных связей его с другими современными методами, точными и итерационными.

Методу минимальных итераций и его обобщениям будет посвящена гл. VI и частично гл. VII.

§ 51. Преобразование симметричной матрицы к трехдиагональному виду посредством вращений

Вращением мы будем называть преобразование координат с элементарной матрицей вращения

$$T_{ij} = \begin{bmatrix} 1 & & & & \\ & c & \dots & -s & \\ & \vdots & \ddots & & \\ & & & 1 & \\ & s & \dots & c & \\ & & & & \ddots \\ & & & & & 1 \end{bmatrix} \quad (1)$$

при $c^2 + s^2 = 1$.

Геометрически вращение может быть интерпретировано как поворот базисных векторов e_i и e_j на некоторый угол, осуществляемый в плоскости, натянутой на векторы e_i и e_j . Матрица T_{ij} ортогональна.

Мы покажем,¹⁾ что любую симметричную матрицу можно привести к трехдиагональной форме посредством цепочки вращений, т. е. цепочки преобразований подобия с матрицами вида T_{ij} .

Произведем необходимые подсчеты. Пусть A — симметричная матрица, $B = AT_{ij}$, $C = T'_{ij}B = T'_{ij}AT_{ij}$. Легко видеть, что все столбцы матрицы B совпадают со столбцами матрицы A , за исключением i -го и j -го столбцов, которые получаются из соответствующих столбцов матрицы A по формулам:

$$\begin{aligned} B_i &= cA_i + sA_j \\ B_j &= -sA_i + cA_j. \end{aligned} \quad (2)$$

¹⁾ Гивенс [1], [2].

В свою очередь строки матрицы C совпадают со строками матрицы B , за исключением i -й и j -й, которые получаются из соответствующих строк матрицы B по таким же формулам:

$$\begin{aligned} C^i &= cB^i + sB^j \\ C^j &= -sB^i + cB^j. \end{aligned} \quad (3)$$

При этом для построения строк C^i и C^j нужно вычислить только четыре элемента c_{ii} , c_{ij} , c_{ji} , c_{jj} (причем c_{ji} только для контроля, так как $c_{ji} = c_{ij}$, но они получаются неодинаковыми вычислениями). Остальные элементы строк C^i и C^j не только теоретически равны соответствующим элементам столбцов B_i и B_j , но при их вычислении выполняются одинаковые действия.

Пусть $1 < i < j$. Покажем, что c и s можно выбрать так, чтобы $c_{i-1,j} = 0$. Действительно, $c_{i-1,j} = b_{i-1,j} = -sa_{i-1,i} + ca_{i-1,j}$, так что достаточно взять $\frac{s}{c} = \frac{a_{i-1,j}}{a_{i-1,i}}$ и, следовательно,

$$s = \frac{a_{i-1,j}}{\pm \sqrt{a_{i-1,i}^2 + a_{i-1,j}^2}}, \quad c = \frac{a_{i-1,i}}{\pm \sqrt{a_{i-1,i}^2 + a_{i-1,j}^2}}.$$

Выбор знака знаменателя безразличен.

Весь процесс приведения симметричной матрицы к трехдиагональному виду выглядит так. За счет преобразований посредством T_{23}, \dots, T_{2n} аннулируются по очереди элементы первой строки, начиная с третьего. Затем за счет T_{34}, \dots, T_{3n} аннулируются элементы второй строки, начиная с четвертого. Ясно, что при этом элементы первой строки больше меняться не будут. Действительно, первые два элемента первой строки не будут меняться при преобразованиях посредством T_{34}, \dots, T_{3n} . Оставшиеся, равные нулю элементы, будут подвергаться линейным однородным преобразованиям и потому останутся равными нулю. Далее, за счет преобразований посредством T_{45}, \dots, T_{4n} аннулируются элементы третьей строки, начиная с пятого и т. д.

Из сказанного выше ясно, что каждое последующее преобразование не будет изменять ранее аннулированные элементы. Таким образом, самое большое через $\frac{(n-1)(n-2)}{2}$ преобразований, мы перейдем от данной симметричной матрицы A к трехдиагональной матрице S .

Вычислительную схему метода проиллюстрируем на примере матрицы

$$A = \begin{bmatrix} 1.00 & 0.42 & 0.54 & 0.66 \\ 0.42 & 1.00 & 0.32 & 0.44 \\ 0.54 & 0.32 & 1.00 & 0.22 \\ 0.66 & 0.44 & 0.22 & 1.00 \end{bmatrix}, \quad (4)$$

уже встречавшейся ранее в примерах решения системы линейных уравнений.

Характеристический полином этой матрицы равен

$$t^4 - 4t^3 + 4.752t^2 - 2.111856t + 0.28615248.$$

Здесь все коэффициенты вычислены точно.

Собственные значения матрицы (4), вычисленные с точностью до $5 \cdot 10^{-9}$, суть

$$\begin{aligned}\lambda_1 &= 2.32274880, \quad \lambda_2 = 0.79670669, \quad \lambda_3 = 0.63828380, \\ \lambda_4 &= 0.24226071.\end{aligned}$$

Процесс трехдиагонализации проведен в табл. IV.13.

Таблица состоит из четырех частей. В части II наряду с данной матрицей A расположены результаты последовательных преобразований подобия посредством T_{23} , T_{24} , T_{34} . Последняя матрица является искомой матрицей S . Части I и IV вспомогательные, часть III — контрольная.

Дадим описание одного шага заполнения таблицы (заключающегося в осуществлении преобразования посредством матрицы T_{ij}). В части II переписываются все элементы предыдущей матрицы, кроме элементов, лежащих в двух строках и двух столбцах с номерами i и j . Затем i -й и j -й столбцы преобразуются по формулам (1) и элементы преобразованных столбцов, кроме элементов с индексами ii , ij , ji , jj записываются на надлежащие места в части II. Выделенные четыре элемента вносятся в часть I. Образованная матрица заполняется далее по симметрии. Оставшиеся четыре элемента строятся затем по формулам (2) над числами,ложенными в части I. В четвертую часть записываются коэффициенты s и s , определяющие матрицы вращений (в последней строке) и необходимые для их вычисления числа. Контроль (часть III) осуществляется при помощи вычисления соответствующих столбцовых сумм (для контроля операций над столбцами) и при помощи вычисления следов построенных матриц, которые должны быть равны между собой.

После того, как построена трехдиагональная матрица, подобная исходной матрице A , отыскание собственных значений может быть осуществлено различными способами.

Наиболее прямым путем является построение характеристического полинома $\varphi(t)$ для матрицы S (а следовательно, и подобной ей матрицы A) по рекуррентным формулам

$$\varphi_0 = 1, \quad \varphi_i(t) = (t - s_{ii})\varphi_{i-1}(t) - s_{i-1i}^2\varphi_{i-2}(t), \quad \varphi_n(t) = (-1)^n \varphi(t)$$

и затем нахождение его корней.

Таблица IV.13

Приведение симметричной матрицы к трехдиагональной при помощи цепочки вращений

I								
II	1.00 0.42 0.54 0.66	0.42 1.00 0.32 0.44	0.54 0.32 1.00 0.22	0.66 0.44 0.22 1.00	1 0 0 0.66	0.68410525 0.68410525 —0.07876923 0.44379135	0 1.31015383 0.68984614 —0.21224804	0.59289121 0.36134790 0.44379135 1
III	4.00000000	2.18	2.08			3.99999997 2.98024313	2.3592812 —0.44379135	1.8915433
IV		0.4680 0.61394061	0.68410526 0.78935221			0.90959999 0.71967235		0.95057876 0.69431385
I		1.25101197 1.01369821		—0.59027359 0.41154187			0.51483019 0.31656653	—0.46952426 0.63863277
II	1 0 0.00000000	0.95057877 1.60414343 —0.20405479 —0.13906436	0 —0.20405479 0.68984614 —0.09805848	0.00000000 —0.13906436 —0.09805848 0.70601043	1 —0.13906436 0 0	0.95057877 1.60414343 —0.24693573 0	0 —0.24693573 0.60370643 —0.02833380	0 0 —0.02833380 0.79215011
III	4.00000000	3.01123416	0.38773287	0.46888759 —0.277679020	3.99999997		0.58446099	0.16910852
IV			0.060977254 0.8263477	0.244693573 0.56316014				

Коэффициенты полиномов $\varphi_i(t)$ удобно вычислять следующим образом. Коэффициенты располагаются согласно схеме

φ_0	φ_1	\dots	φ_{n-1}	φ_n	
			1	t^n	
1			$\pi_{n-1}^{(n)}$	t^{n-1}	
.			.	.	
.			.	.	
1	\dots		$\pi_1^{(n-1)}$	$\pi_1^{(n)}$	t
1	$\pi_0^{(1)}$	\dots	$\pi_0^{(n-1)}$	$\pi_0^{(n)}$	t^0

Схема затем заполняется слева направо по рекуррентной формуле

$$\pi_j^{(i+1)} = -s_{i-1}^2 \pi_j^{(i-1)} - s_{ii} \pi_j^{(i)} + \pi_{j-1}^{(i)}.$$

Участвующие в формуле коэффициенты, очевидно, входят в схему в следующем расположении

$$\begin{matrix} \pi_j^{(i-1)} & \pi_j^{(i)} & \pi_j^{(i+1)} \\ & \pi_{j-1}^{(i)} & \end{matrix}$$

Гивенс¹⁾ рекомендует другой прием, позволяющий обойти вычисление коэффициентов характеристического полинома. Этот прием использует то обстоятельство, что полиномы $\varphi_0, \varphi_1, \dots, \varphi_n$ образуют ряд Штурма.

Наконец, как мы видели в § 49, здесь удобно проводить вычисление корней методом квадратичной интерполяции.

Вычисление собственных векторов для матрицы S может быть осуществлено так же, как в методе Хессенберга, т. е. посредством решения соответствующей треугольной системы

$$\begin{aligned} (s_{11} - \lambda_i) v_1 + s_{12} v_2 &= 0 \\ s_{12} v_1 + (s_{22} - \lambda_i) v_2 + s_{23} v_3 &= 0 \\ \vdots &\vdots \\ s_{n-1,n} v_{n-1} + (s_{nn} - \lambda_i) v_n &= 0 \end{aligned} \tag{5}$$

для компонент v_1, \dots, v_n собственного вектора V_i , принадлежащего λ_i .

Однако здесь удобно задаться первой компонентой (а не последней, как в методе Хессенберга) и затем последовательно вычислять вторую, третью и т. д.

¹⁾ Гивенс [2].

Оказывается возможным дать и явные формулы для компонент собственного вектора, принадлежащего собственному значению λ_i . Именно,

$$v_k = \frac{1}{s_{12}s_{23} \dots s_{k-1,k}} \varphi_{k-1}(\lambda_i). \quad (6)$$

Действительно, подстановка этих значений для компонент в левую часть k -го уравнения системы (5) дает

$$\begin{aligned} & s_{k-1,k} \frac{1}{s_{12} \dots s_{k-2,k-1}} \varphi_{k-2}(\lambda_i) + (s_{kk} - \lambda_i) \frac{1}{s_{12} \dots s_{k-1,k}} \varphi_{k-1}(\lambda_i) + \\ & + s_{k,k+1} \frac{1}{s_{12} \dots s_{k,k+1}} \varphi_k(\lambda_i) = - \frac{1}{s_{12} \dots s_{k-1,k}} [(\lambda_i - s_{kk}) \varphi_{k-1}(\lambda_i) - \\ & - \varphi_k(\lambda_i) - s_{k-1,k}^2 \varphi_{k-2}(\lambda_i)] = 0 \quad \text{при } 2 \leq k \leq n-1. \end{aligned}$$

Таким образом, уравнения от второго до $n-1$ -го удовлетворяются. Очевидно, удовлетворяется и первое уравнение. Последнее же является следствием остальных.

Однако пользоваться явными формулами (6) менее целесообразно, чем численно решать систему (5), как по объему вычислений, так и по надежности результата.

Для перехода от собственных векторов матрицы S к собственным векторам матрицы A нужно использовать соотношение

$$S = [T_{23} \dots T_{n-1,n}]' A T_{23} \dots T_{n-1,n},$$

из которого следует, что каждый собственный вектор U матрицы A выражается через соответствующий собственный вектор V матрицы S по формуле

$$U = T_{23} T_{24} \dots T_{n-1,n} V. \quad (7)$$

т. е. U получается из V посредством цепочки умножений на матрицы поворотов T_{ij} . При каждом отдельном умножении будут меняться только две компоненты предыдущего вектора — i -я и j -я — по формулам

$$\begin{aligned} v'_i &= cv_i - sv_j \\ v'_j &= sv_i + cv_j. \end{aligned} \quad (8)$$

Здесь через v_i и v_j обозначены компоненты предыдущего вектора, через v'_i и v'_j — следующего.

Хотя количество умножений в этом процессе весьма значительно, ошибки округления накапливаются медленно, так как они умножаются на коэффициенты c и s , по модулю меньшие единицы.

Наконец, отметим, что если окажется, что один или несколько элементов $s_{k-1,k}$ равны нулю, то матрица S разобьется на два или несколько якобиевых ящиков и задача вычисления собственных значений и собственных векторов только облегчится. Это

явление наверное будет иметь место, если исходная матрица имеет кратные собственные значения.

Мы закончим этот параграф вычислением характеристического полинома для матрицы, приведённой нами к трехдиагональному виду в табл. IV.13, определением собственных значений этой матрицы и вычислением двух собственных векторов, принадлежащих наибольшему и наименьшему собственным значениям.

Таблица IV.14

Вычисление коэффициентов характеристического полинома по рекуррентным формулам

$k \backslash l$	0	1	2	3	4
4					1
3				1	-3.99999997
2			1	-3.20784986	4.75199990
1		1	-2.60414343	2.21170431	-2.11185592
0	1	-1	0.70054343	-0.36194532	0.28615247
	1 0.90360000	1.60414343 0.060977255	0.60370643 0.0008030309		

Здесь в первой части таблицы расположены коэффициенты последовательных полиномов φ_i (по столбцам), во второй записаны коэффициенты рекуррентных формул s_{ii} и $s_{i-1,i}^2$ (вычисленные по данным табл. IV.13). Таким образом находим, что искомый характеристический полином будет

$$t^4 - 3.99999997t^3 + 4.75199990t^2 - 2.11185592t + 0.28615247.$$

Его наибольший и наименьший корни будут

$$\lambda_1 = 2.32274880 \text{ и } \lambda_4 = 0.24226070.$$

Для определения принадлежащих им собственных векторов найдем сначала соответствующие собственные векторы матрицы S , решая систему (5).

Получим

$$V_1 = (1, 1.3915194, -0.19994896, 0.00370089)'$$

$$V_4 = (1, -0.79713467, -0.54680289, -0.02817925)'.$$

Далее вычисляем последовательно

$$T_{34}V_1 = (1, 1.3915194, -0.16731157, -0.10954506)'$$

$$T_{24}T_{34}V_1 = (1, 1.0774967, -0.16731157, 0.88731461)'$$

$$T_{23}T_{24}T_{34}V_1 = U_1 = (1, 0.793587, 0.747805, 0.887315)'$$

и

$$T_{34}V_4 = (1, -0.79713467, -0.43597992, 0.33122345)'$$

$$T_{24}T_{34}V_4 = (1, -0.34370275, -0.43597992, -0.79183400)'$$

$$T_{23}T_{24}T_{34}V_4 = U_4 = (1, 0.133129, -0.538968, -0.791834)'$$

Отметим, что указанный процесс можно применять и к несимметричной матрице, только при этом в результате вместо трехдиагональной матрицы получится почти треугольная матрица

$$\begin{bmatrix} s_{11} & s_{12} & 0 & \dots & 0 & 0 \\ s_{21} & s_{22} & s_{23} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ s_{n-1,1} & s_{n-1,2} & s_{n-1,3} & \dots & s_{n-1,n-1} & s_{n-1,n} \\ s_{n1} & s_{n2} & s_{n3} & \dots & s_{n,n-1} & s_{nn} \end{bmatrix}.$$

Таким образом, в случае несимметричной матрицы цепочка вращений приводит ее почти к такому же виду, как и в методе Хессенберга и в методе ортогонализации итераций. Проблема собственных значений решается аналогично указанным методам.

§ 52. Уточнение полной проблемы собственных значений

Пусть A данная матрица, собственные значения которой попарно различны. Пусть мы располагаем приближенными значениями $\lambda_1, \dots, \lambda_n$ для собственных чисел матрицы A , приближенными собственными векторами U_1, \dots, U_n матрицы A , так же как и приближенными собственными векторами V_1, \dots, V_n сопряженной матрицы A^* . Ставится задача об уточнении всей совокупности перечисленных величин.

Будем искать уточненные значения в виде

$$\begin{aligned} \tilde{\lambda}_i &= \lambda_i + \Delta\lambda_i \\ \tilde{U}_i &= U_i + \Delta U_i \\ \tilde{V}_i &= V_i + \Delta V_i \end{aligned} \tag{1}$$

считая числа $\Delta\lambda_i$, так же как и компоненты векторов ΔU_i и ΔV_i , малыми.

Без нарушения общности можно считать, что

$$\Delta U_i = \sum_{\substack{j=1 \\ j \neq i}}^n h_{ij} U_j. \quad (2)$$

Действительно, собственные векторы определены с точностью до скалярного множителя и потому мы вправе считать, что i -я координата уточненного собственного вектора по отношению к базису U_1, \dots, U_n равна единице. На том же основании мы вправе считать, что

$$\Delta V_i = \sum_{\substack{j=1 \\ j \neq i}}^n k_{ij} V_j. \quad (3)$$

Очевидно, что коэффициенты h_{ij} и k_{ij} будут малыми числами.

Выразим поправки $\Delta \lambda_i$ и коэффициенты h_{ij} и k_{ij} через невязки известного нам решения полной проблемы собственных значений, т. е. через

$$\begin{aligned} r_i &= AU_i - \lambda_i U_i \\ r_i^* &= A^* V_i - \bar{\lambda}_i V_i. \end{aligned} \quad (4)$$

Уравнение

$$A \tilde{U}_i = \tilde{\lambda}_i \tilde{U}_i$$

перепишем в виде

$$AU_i + A\Delta U_i = \lambda_i U_i + \lambda_i \Delta U_i + \Delta \lambda_i U_i + \Delta \lambda_i \Delta U_i.$$

Введя в это равенство невязку r_i и отбросив последний член правой части равенства, получим

$$A\Delta U_i - \lambda_i \Delta U_i \approx -r_i + \Delta \lambda_i U_i. \quad (5)$$

Составим теперь скалярные произведения с векторами V_j ($j = 1, \dots, n$). Тогда

$$(A\Delta U_i, V_j) - \lambda_i (\Delta U_i, V_j) = -(r_i, V_j) + \Delta \lambda_i (U_i, V_j). \quad (6)$$

Но, с точностью до малых 2-го порядка,

$$(A\Delta U_i, V_j) = (\Delta U_i, A^* V_j) \approx (\Delta U_i, \bar{\lambda}_j V_j) = \lambda_j (\Delta U_i, V_j).$$

Поэтому

$$(\lambda_j - \lambda_i) (\Delta U_i, V_j) \approx -(r_i, V_j) + \Delta \lambda_i (U_i, V_j). \quad (7)$$

Положив $i = j$, получим

$$\Delta \lambda_i \approx \frac{(r_i, V_i)}{(U_i, V_i)}. \quad (8)$$

Будем теперь считать, что $i \neq j$. Так как с точностью до малых 2-го порядка

$$\begin{aligned} (\Delta U_i, V_j) &\approx h_{ij}(U_j, V_j) \\ \Delta \lambda_i(U_i, V_j) &\approx 0, \end{aligned}$$

то из равенства (7) получим

$$h_{ij} = \frac{(r_i, V_j)}{(\lambda_i - \lambda_j)(U_j, V_j)}. \quad (9)$$

Аналогично

$$k_{ij} = \frac{(r_i^*, U_j)}{(\bar{\lambda}_i - \bar{\lambda}_j)(U_j, V_j)}. \quad (10)$$

Из вида этих формул мы заключаем, что для уточнения собственных значений и собственных векторов используются по существу лишь результаты контрольных вычислений, произведенных после получения исходного приближения к решению полной проблемы собственных значений.

Отметим, что при этом коэффициенты h_{ij} и k_{ij} связаны друг с другом легко проверяемым соотношением

$$(U_j, V_j) h_{ij} + (U_i, V_j) \bar{k}_{ij} = -(U_i, V_j),$$

так что для вычисления коэффициентов h_{ij} даже нет необходимости вычислять невязки r_i^* .

Уточненные значения собственных чисел могут быть вычислены также по формуле

$$\tilde{\lambda}_i = \lambda_i + \Delta \lambda_i = \lambda_i + \frac{(r_i, V_i)}{(U_i, V_i)} = \frac{(AU_i, V_i)}{(U_i, V_i)}, \quad (11)$$

из которой видно, что для получения уточненного значения для собственного числа достаточно лишь знать приближенные значения для собственных векторов U_i и V_i .

Отметим, что указанный процесс есть не что иное, как применение одного шага метода Ньютона к нелинейной системе

$$A\bar{U}_i = \tilde{\lambda}_i \bar{U}_i,$$

$$A^*\bar{V}_i = \tilde{\lambda}_i \bar{V}_i.$$

Для симметричной матрицы можно считать $U_i = V_i$ и потому

$$\begin{aligned} \Delta \lambda_i &= \frac{(r_i, U_i)}{(U_i, U_i)} \\ h_{ij} &= \frac{(r_i, U_j)}{(\lambda_i - \lambda_j)(U_j, U_j)}. \end{aligned}$$

Приведем результаты уточнения полной проблемы собственных значений для матрицы Леверье. В качестве исходных приближений возьмем данные эскалаторного метода, округленные до трех десятичных знаков. Имеем

$$\lambda_1 = -17.863, \quad \lambda_2 = -17.152, \quad \lambda_3 = -7.574, \quad \lambda_4 = -5.299.$$

Имеем также

U_1	U_2	U_3	U_4	V_1	V_2	V_3	V_4
-0.020	0.033	-0.351	1.135	-0.014	0.023	-0.248	0.803
0.170	-0.261	0.328	0.112	0.780	-1.201	1.510	0.516
0.187	0.237	0.261	0.071	-1.141	1.442	1.590	0.430
0.808	0.587	0.045	0.011	0.808	0.587	0.045	0.011
0.999111	1.000543	0.999343	0.999848				

В последней строке размещены соответствующие скалярные произведения (U_j, V_j) при $j = 1, 2, 3, 4$. Вычисления проведем для $t = 1$. Имеем

r_1	(r_1, V_j)	$(\lambda_1 - \lambda_j)(U_j, V_j)$	h_{1j}	ΔU_1	\tilde{U}_1	$c\tilde{U}_1$
-0.00110982	-0.00026259		0	0.00013616	-0.01986384	-0.019873
0.00228956	-0.00014959	-0.711386	0.00021028	-0.00027565	0.16972435	0.169806
0.00181651	0.00662120	-10.282240	-0.00064394	-0.00012429	-0.18712429	-0.187214
0.00001071	0.00107144	-12.562090	-0.00008529	0.00009352	0.80809352	0.808482

В последнем столбце помещается собственный вектор, принадлежащий собственному значению λ_1 , нормированный так, чтобы его последняя компонента совпадала с последней компонентой собственного вектора, вычисленного эскалаторным методом.

По формуле (8) находим

$$\Delta \lambda_1 = -0.000263; \quad \tilde{\lambda}_1 = -17.863263.$$

Приводим также результаты вычислений для $t = 2, 3, 4$.

\tilde{U}_2	\tilde{U}_3	\tilde{U}_4
0.032933	-0.351235	1.135218
-0.261309	0.328466	0.112182
0.236640	0.260925	0.070589
0.586694	0.045006	0.011058

$$\tilde{\lambda}_2 = -17.152428, \quad \tilde{\lambda}_3 = -7.574043, \quad \tilde{\lambda}_4 = -5.208696.$$

Полученные значения для собственных чисел и компонент собственных векторов верны уже с точностью до $2 \cdot 10^{-6}$.

ГЛАВА V

ЧАСТИЧНАЯ ПРОБЛЕМА СОБСТВЕННЫХ ЗНАЧЕНИЙ

Настоящая глава посвящена частичной проблеме собственных значений, которая состоит, как было сказано выше, в определении одного или нескольких, как правило немногих, собственных значений матрицы и принадлежащих им собственных векторов. Своевобразие частичной проблемы заключается в том, что методы для ее решения должны основываться на косвенных соображениях, использующих те или другие свойства собственных значений и собственных векторов. Все методы для решения частичной проблемы являются итерационными методами.

Для построения этих методов используются две, по существу различные, основные идеи.

Первую идею мы поясним в предположении, что в пространстве существует базис из собственных векторов. Исходя из некоторого вектора, вообще говоря, произвольного, строят бесконечную последовательность векторов так, чтобы в этой последовательности все более преобладала одна составляющая в разложении по собственным векторам. Тогда построенная последовательность будет сходиться по направлению к выделенному собственному вектору. Попутно определяется и собственное значение. Процессы, основанные на этой идее могут быть применены и при отсутствии базиса из собственных векторов. В этом случае при их обосновании можно использовать разложение по каноническому базису. При этом некоторое видоизменение методов позволяет вычислять несколько векторов из канонического базиса.

Вторая идея основывается на экстремальных свойствах собственных значений и применима только к симметричным матрицам. Эта идея близка к идеи релаксации для решения линейной системы уравнений. Методы, основанные на этой идее, дают последовательность векторов, все лучше реализующих максимум (или минимум) отношения $\frac{(AX, X)}{(X, X)}$.

Выбор поправок для перехода от предыдущего вектора к следующему может осуществляться различно. Важнейшая группа методов, использующих эту идею, в которых поправки берутся в на-

правлении градиента функционала $\frac{(AX, X)}{(X, X)}$, будет рассмотрена в главе VII. В этой же главе будут вкратце рассмотрены методы, аналогичные методам координатной релаксации, простой и групповой.

§ 53. Определение наибольшего по модулю собственного значения матрицы при помощи последовательных итераций

В настоящем параграфе мы изложим метод, позволяющий вычислять наибольшее по модулю собственное значение матрицы и принадлежащий ему собственный вектор при помощи вычисления последовательности итераций произвольного вектора. Излагаемый метод называется степенным и является простейшим итерационным процессом для решения частичной проблемы собственных значений. Он применим для произвольной матрицы, хотя ход итерационного процесса существенно зависит от того, как входит наибольшее по модулю собственное значение матрицы в ее каноническую форму Жордана. В связи с этим приходится различать несколько возможных случаев. Мы, однако, не будем исследовать проблему во всей общности и ограничимся рассмотрением лишь важнейших частных случаев.

Для упрощения изложения мы будем предполагать, что все собственные значения матрицы, кроме, может быть, наибольшего по модулю, имеют линейные элементарные делители, хотя выводы, которые мы сделаем, имеют место и без этого предположения. Будем также считать, что элементы исследуемых матриц вещественны.

1. Наибольшее по модулю собственное значение вещественное и простое. В этом случае наибольшему по модулю собственному значению соответствует один линейный элементарный делитель, так что в силу только что сформулированного соглашения, все элементарные делители матрицы линейны. Поэтому существует базис из собственных векторов U_1, U_2, \dots, U_n , принадлежащих собственным значениям $\lambda_1, \lambda_2, \dots, \lambda_n$, расположенным в порядке убывания модулей, причем $|\lambda_1| > |\lambda_2|$, но среди остальных могут быть равные. Возьмем произвольный вектор Y_0 и образуем последовательность его итераций матрицей A

$$AY_0, A^2Y_0, \dots, A^kY_0, \dots$$

Напишем разложение вектора Y_0 по собственным векторам

$$Y_0 = a_1U_1 + a_2U_2 + \dots + a_nU_n. \quad (1)$$

Среди чисел a_i некоторые могут равняться нулю. Предположим, однако, что $a_1 \neq 0$.¹⁾

¹⁾ Так как $a_1 = (V_1, Y_0)$, где V_1 первый собственный вектор транспонированной матрицы, то требование $a_1 \neq 0$ будет выполнено в случае, если вектор Y_0 не будет ортогонален к вектору V_1 .

Очевидно, что

$$AY_0 = a_1\lambda_1 U_1 + a_2\lambda_2 U_2 + \dots + a_n\lambda_n U_n$$

$$A^k Y_0 = a_1\lambda_1^k U_1 + a_2\lambda_2^k U_2 + \dots + a_n\lambda_n^k U_n.$$
(2)

Обозначим $A^k Y_0 = Y_k = (y_{1k}, y_{2k}, \dots, y_{nk})'$ и выясним структуру компонент вектора Y_k . Пусть

$$U_1 = (u_{11}, u_{21}, \dots, u_{n1})', \quad U_2 = (u_{12}, u_{22}, \dots, u_{n2})', \dots, \quad U_n = \\ = (u_{1n}, u_{2n}, \dots, u_{nn})'.$$

Тогда из (2) получим

$$v_{ik} = a_1 u_{i1} \lambda_1^k + a_2 u_{i2} \lambda_2^k + \dots + a_n u_{in} \lambda_n^k.$$

Коэффициент при λ_1^k по крайней мере в одной из компонент не равен нулю, так как $a_1 \neq 0$ по предположению и вектор U_1 не нулевой. Пусть y_k (первый индекс опускаем) какая-либо из компонент вектора Y_k , для которой коэффициент при λ_1^k отличен от нуля. Тогда

$$y_k = c_1 \lambda_1^k + c_2 \lambda_2^k + \dots + c_n \lambda_n^k. \quad (3)$$

причем коэффициент c_i не зависит от индекса k и $c_1 \neq 0$.

Рассмотрим отношение компонент двух соседних итераций.

$$\frac{y_{k+1}}{y_k} = \frac{c_1 \lambda_1^{k+1} + c_2 \lambda_2^{k+1} + \dots + c_n \lambda_n^{k+1}}{c_1 \lambda_1^k + c_2 \lambda_2^k + \dots + c_n \lambda_n^k} = \\ = \lambda_1 \frac{1 + b_2 x_2^{k+1} + b_3 x_3^{k+1} + \dots + b_n x_n^{k+1}}{1 + b_2 x_2^k + b_3 x_3^k + \dots + b_n x_n^k}, \quad (4)$$

где

$$b_i = \frac{c_i}{c_1}, \quad \alpha_i = \frac{\lambda_i}{\lambda_1}. \quad (5)$$

Произведя деление и удерживая члены до порядка α_2^{2k} и α_3^k включительно, получим

$$\frac{y_{k+1}}{y_k} = \lambda_1 [1 - b'_2 \alpha_2^k - b'_3 \alpha_3^k + b_2 b'_2 \alpha_2^{2k}] + o(\alpha_3^k + \alpha_2^{2k}). \quad (6)$$

где

$$b'_2 = b_2(1 - \alpha_2), \quad b'_3 = b_3(1 - \alpha_3). \quad (7)$$

Отсюда мы видим, что если k достаточно велико, то

$$\lambda_1 \approx \frac{y_{k+1}}{y_k}. \quad (8)$$

Так как обычно все компоненты вектора U_1 отличны от нуля, то в качестве y_k может быть взята, как правило, любая компонента

вектора Y_k . Таким образом, первое собственное значение приближенно равно отношению любых соответствующих компонент двух соседних достаточно высоких итераций произвольного вектора матрицы A .

При практическом выполнении итераций следует вычислять отношения $\frac{y_{k+1}}{y_k}$ для нескольких компонент. Хорошее совпадение этих отношений будет показывать, что в выражении (6) различие значений коэффициентов b'_2, b'_3 уже перестало играть заметную роль.

Быстрота сходимости процесса в рассматриваемом случае определяется величиной отношения $\frac{\lambda_2}{\lambda_1}$ и может быть медленной, если это отношение близко к единице.

Для того чтобы избежать роста компонент, иногда целесообразно при вычислении итераций тем или другим способом нормировать на каждом шагу получаемые векторы. Удобными нормировками являются деление вектора на его первую компоненту, или на наибольшую компоненту, или, наконец, нормировка к единичной длине. При этом вместо последовательности Y_k мы получим последовательность $\tilde{Y}_k = \gamma_k Y_k$, где γ_k нормирующие множители, и для получения λ_1 надо брать отношения компонент векторов $A\tilde{Y}_k$ и \tilde{Y}_k .

Может случиться, хотя это и маловероятно, что начальный вектор Y_0 выбран неудачно, именно так, что коэффициент a_1 равен нулю, или очень близок к нулю. В этом случае не будет ясной картины сходимости итераций по направлению. Действительно, на первых шагах итерации преобладающим будет член, зависящий от λ_2 (если $a_2 \neq 0$). Однако в дальнейшем, если даже a_1 точно равно нулю, то после нескольких шагов итерации слагаемое, зависящее от λ_1 , появится, благодаря ошибкам округления, сначала с очень малым коэффициентом; по мере дальнейших итераций это слагаемое будет довольно быстро возрастать по сравнению с остальными. «Борьба за преобладание» членов, зависящих от λ_1 и λ_2 , вызывает неясность в ходе процесса. При неудачном выборе, в указанном смысле, начального вектора, его необходимо изменить.

Описанный процесс дает возможность определить также и все компоненты собственного вектора, принадлежащего наибольшему собственному числу. Именно, отношения компонент вектора Y_k стремятся к отношениям компонент этого собственного вектора.

Действительно, при $a_1 \neq 0$

$$\begin{aligned} Y_k = A^k Y_0 = \lambda_1^k [a_1 U_1 + a_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k U_2 + \dots + a_n \left(\frac{\lambda_n}{\lambda_1}\right)^k U_n] = \\ = a_1 \lambda_1^k [U_1 + O\left(\frac{\lambda_2}{\lambda_1}\right)^k]. \end{aligned} \quad (9)$$

Пример 1. Попытаемся определить первое собственное значение матрицы Леверье.

Возьмем вектор $(1, 0, 0, 0)'$ за исходный и образуем 20 итераций, нормируя их на каждом шагу посредством деления на первую компоненту.

Приведем только две последние итерации:

\tilde{Y}_{19}	$A\tilde{Y}_{19}$	Отношения компонент
1.000000	—17.4655	—17.466
—8.20321	143.3809	—17.479
8.17013	—143.0881	—17.514
—7.95957	149.2676	—18.753
<hr/>		
—6.99265	132.0949	

Из приведенных данных мы видим, что отношения различных компонент еще далеки друг от друга; это показывает, что процесс еще не установился. Действительно, с точностью до трех знаков, $\lambda_1 = -17.863$. Процесс итерации сходится медленно из-за того, что второе собственное значение $\lambda_2 = -17.152$ мало отличается по модулю от первого.

Пример 2. Определим первое собственное значение и принадлежащий ему собственный вектор для матрицы

$$\begin{bmatrix} -5.509882 & 1.870086 & 0.422908 \\ 0.287865 & -11.811654 & 5.711900 \\ 0.049099 & 4.308033 & -12.970687 \end{bmatrix}.$$

Возьмем в качестве начального вектор $(1, 0, 0)'$. Приведем таблицу итераций, начиная с 12-й итерации:

\tilde{Y}_{12}	$A\tilde{Y}_{12}$	\tilde{Y}_{13}	$A\tilde{Y}_{13}$	\tilde{Y}_{14}	$A\tilde{Y}_{14}$
1.0000000	—17.351783	1.0000000	—17.378482	1.0000000	—17.389552
—8.1139091	141.126754	—8.1332710	141.483894	—8.1413264	141.632991
7.8783245	—137.093170	7.9009117	—137.468256	7.9102568	—137.625469
<hr/>					
0.7644154	—13.318201	0.7675407	—13.362844	0.7689304	—13.382030

Найдем отношения соответствующих компонент для 12-й, 13-й, 14-й и 15-й итераций:

$$\begin{aligned} &—17.351783 \quad —17.378482 \quad —17.389552 \\ &—17.393189 \quad —17.395694 \quad —17.396795 \\ &—17.401310 \quad —17.399257 \quad —17.398356. \end{aligned}$$

Три последних отношения позволяют нам считать, что $\lambda_1 = -17.39$ или $\lambda_1 = -17.40$. Как мы видели в § 48, с точностью до четырех знаков, $\lambda_1 = -17.3977$.

Далее, для компонент собственного вектора находим следующие значения

$$\begin{array}{ccc} 1.00000 & 1.00000 & 1.00000 \\ -8.13327 & -8.14133 & -8.14472 \\ 7.90081 & 7.91026 & 7.91426. \end{array}$$

Мы видим, что последний результат уже довольно близко подходит к точному значению, так как в § 48 было найдено, что $U_1 = (0.094129, -0.766896, 0.745248)'$ или после соответствующего нормирования $U_1 = (1.00000, -8.14729, 7.91730)'$.

Пример 3. Найдем первое собственное значение и принадлежащий ему вектор для матрицы

$$\begin{bmatrix} 0.22 & 0.02 & 0.12 & 0.14 \\ 0.02 & 0.14 & 0.04 & -0.06 \\ 0.12 & 0.04 & 0.28 & 0.08 \\ 0.14 & -0.06 & 0.08 & 0.26 \end{bmatrix}.$$

В табл. III.1 были вычислены 14 итераций, исходя из вектора $(0.76, 0.08, 1.12, 0.68)'$.

Вычисляя отношения компонент 14-й и 13-й итераций, находим для λ_1 значение 0.4800.

(Мы игнорируем вторую компоненту итераций из-за ее малости по сравнению с остальными компонентами).

Отношения компонент 7-й и 6-й итераций дают для λ_1 значения

$$0.4800, \quad 0.4792, \quad 0.4808.$$

Для компонент собственного вектора из 14-й итерации находим следующие значения:

$$1.0000, \quad 0.0000, \quad 1.0000, \quad 1.0000.$$

Нетрудно проверить, что точное значение $\lambda_1 = 0.48$ и принадлежащий ему собственный вектор имеет компоненты 1, 0, 1, 1.

2. Наибольшее собственное значение вещественное, кратное, но соответствующие ему элементарные делители линейны. В этом случае формула (3) остается верной, но несколько первых членов можно соединить вместе, так что

$$y_k = c_1 \lambda_1^k + c_{r+1} \lambda_{r+1}^k + \dots + c_n \lambda_n^k,$$

где r кратность λ_1 .

Все дальнейшие рассуждения остаются в силе и потому

$$\frac{y_{k+1}}{y_k} = \lambda_1 + O \left(\frac{\lambda_{r+1}}{\lambda_1} \right)^k. \quad (10)$$

Таким образом, и в этом случае, при условии $a_1 \neq 0$, отношение $\frac{y_{k+1}}{y_k}$ дает приближенное значение наибольшего собственного числа.

Вопрос о кратности корня не может быть решен без более детального исследования. Мы еще вернемся к этому вопросу ниже.

Векторы $Y_k = A^k Y_0$, так же как и в предыдущем случае, сходятся по направлению к одному из собственных векторов, принадлежащих λ_1 , именно к собственному вектору, лежащему в циклическом подпространстве, порожденном вектором Y_0 . Исходя из различных начальных векторов, мы приедем, вообще говоря, к различным собственным векторам.

Пример 4. Определим первое собственное значение матрицы

1.022551	0.116069	-0.287028	-0.429969
0.228401	0.742521	-0.176368	-0.283720
0.326141	0.097221	0.197209	-0.216487
0.433864	0.148965	-0.193686	0.006472

Решая характеристическое уравнение

$$\lambda^4 - 1.968753\lambda^3 + 1.391184\lambda^2 - 0.415291\lambda + 0.044360 = 0,$$

получим для собственных чисел значения:

$$\lambda_1 = \lambda_2 = 0.667483, \quad \lambda_3 = 0.346148, \quad \lambda_4 = 0.287639.$$

Определим λ_1 при помощи степенного метода, взяв за начальный вектор $(1, 1, 1, 1)^T$.

Приведем таблицу итераций, начиная с 9-й итерации:

\tilde{Y}_9	$A\tilde{Y}_9$	\tilde{Y}_{10}	$A\tilde{Y}_{10}$	\tilde{Y}_{11}	$A\tilde{Y}_{11}$
1.000000	0.666160	1.000000	0.666822	1.000000	0.667151
1.844723	1.230507	1.847165	1.232545	1.848387	1.233563
0.676506	0.449420	0.674643	0.449211	0.673660	0.449088
0.875250	0.583298	0.875613	0.584025	0.875834	0.584399
4.396479	2.929385	4.397421	2.932603	4.397881	2.934201

Вычислим отношения компонент этих итераций:

0.666160	0.666822	0.667151
0.667042	0.667263	0.667373
0.664325	0.665850	0.666639
0.666466	0.666990	0.667249

Последние четыре отношения дают для λ_1 значение 0.667, верное с точностью до третьего знака.

3. Два наибольшие по модулю собственные значения вещественны и противоположны по знаку. Из равенства (3) мы видим, что в этом случае четные и нечетные итерации имеют различные коэффициенты при соответствующих степенях λ_1 , так как

$$y_{2k} = (c_1 + c_2)\lambda_1^{2k} + c_3\lambda_3^{2k} + \dots + c_n\lambda_n^{2k},$$

$$y_{2k+1} = (c_1 - c_2)\lambda_1^{2k+1} + c_3\lambda_3^{2k+1} + \dots + c_n\lambda_n^{2k+1},$$

и потому две соседние итерации не могут быть использованы для определения λ_1 . Однако мы можем определить λ_1^2 по одной из следующих формул:

$$\lambda_1^2 = \frac{y_{2k+2}}{y_{2k}} \quad \text{или} \quad \lambda_1^2 = \frac{y_{2k+1}}{y_{2k-1}}. \quad (11)$$

Для нахождения собственных векторов, принадлежащих λ_1 и $\lambda_2 = -\lambda_1$, целесообразно построить векторы $Y_k + \lambda_1 Y_{k-1}$ и $Y_k - \lambda_1 Y_{k-1}$. Отношения компонент этих векторов будут стремиться, соответственно, к отношению компонент векторов U_1 и U_2 , принадлежащих собственным числам λ_1 и λ_2 .

Действительно, в силу равенства

$$Y_k = a_1\lambda_1^k U_1 + a_2(-\lambda_1)^k U_2 + a_3\lambda_3^k U_3 + \dots \quad (12)$$

имеем

$$Y_k + \lambda_1 Y_{k-1} = 2a_1\lambda_1^k U_1 + a_3(\lambda_3 + \lambda_1)\lambda_3^{k-1} U_3 + \dots = \\ = \lambda_1^k \left[2a_1 U_1 + O\left(\frac{\lambda_3}{\lambda_1}\right)^k \right]$$

$$Y_k - \lambda_1 Y_{k-1} = 2a_2(-\lambda_1)^k U_2 + a_3(\lambda_3 - \lambda_1)\lambda_3^{k-1} U_3 + \dots = \\ = (-\lambda_1)^k \left[2a_2 U_2 + O\left(\frac{\lambda_3}{\lambda_1}\right)^k \right]. \quad (13)$$

Пример 5. Нетрудно вычислить, что собственные значения матрицы

$$A = \begin{bmatrix} 4.2 & -3.4 & 0.3 \\ 4.7 & -3.9 & 0.3 \\ -5.6 & 5.2 & 0.1 \end{bmatrix}$$

суть $\lambda_1 = -\lambda_2 = 0.5$, $\lambda_3 = 0.4$, причем собственному значению 0.5 принадлежит собственный вектор $(1, 1, -1)'$, собственному значению -0.5 принадлежит собственный вектор $(-\frac{2}{3}, -\frac{5}{6}, 1)' \approx (-0.667, -0.833, 1)'$.

Проведя вычисления получим

Y_0	Y_{23}	Y_{24}	Y_{25}
0.2	$0.25972708 \cdot 10^{-6}$	$0.22548439 \cdot 10^{-6}$	$0.65159766 \cdot 10^{-7}$
0.4	$0.23588520 \cdot 10^{-6}$	$0.23740533 \cdot 10^{-6}$	$0.59199296 \cdot 10^{-7}$
0.6	$-0.21119890 \cdot 10^{-6}$	$-0.24898850 \cdot 10^{-6}$	$-0.53103718 \cdot 10^{-7}$
	$0.28441339 \cdot 10^{-6}$	$0.21390121 \cdot 10^{-6}$	$0.71255344 \cdot 10^{-7}$

Для отношений соответствующих компонент векторов $Y_{25} \parallel Y_{23}$ получим

$$0.25087782, \quad 0.25096655, \quad 0.25143936,$$

откуда для λ_1 находим три приближенных значения

$$0.500877, \quad 0.500966, \quad 0.501437,$$

так что, с точностью до трех десятичных знаков, $\lambda_1 \approx 0.501$.

Далее находим

$Y_{25} + 0.501 Y_{23}$	\tilde{U}_1	$Y_{25} - 0.501 Y_{24}$	\tilde{U}_2
0.17	1.000	-0.0479	-0.669
0.17	1.000	-0.0597	-0.834
-0.1778	-0.998	0.0716	1.000.

Таким образом, компоненты собственных векторов определены с точностью до $2 \cdot 10^{-3}$.

4. Наибольшие по модулю собственные значения образуют простую комплексную пару. Пусть λ_1 и λ_2 комплексно-сопряженные, наибольшие по модулю собственные значения и $|\lambda_3| < |\lambda_1|$. Согласно формуле (3)

$$y_k = c_1 \lambda_1^k + c_2 \lambda_2^k + \dots + c_n \lambda_n^k,$$

причем в этом случае c_1 и c_2 комплексно сопряжены. Пусть

$$\begin{aligned} c_1 &= Re^{i\alpha}, & c_2 &= Re^{-i\alpha} \\ \lambda_1 &= re^{i\theta}, & \lambda_2 &= re^{-i\theta}, \end{aligned} \tag{14}$$

тогда

$$y_k = 2Rr^k \cos(k\theta + \alpha) + c_3 \lambda_3^k + \dots \tag{15}$$

Присутствие множителя $\cos(k\theta + \alpha)$ будет причиной того, что значения y_k будут сильно колебаться как по величине, так и по знаку. Таким образом, наличие комплексных корней, наибольших по модулю, сразу обнаруживается при составлении итераций. Положим

$$\begin{aligned} p &= -(\lambda_1 + \lambda_2) = -2r \cos \theta \\ q &= \lambda_1 \lambda_2 = r^2. \end{aligned} \tag{16}$$

Тогда λ_1 и λ_2 будут корнями квадратного уравнения

$$t^2 + pt + q = 0.$$

Коэффициенты p и q могут быть определены из следующих соображений. Пусть k настолько велико, что $y_k \approx c_1\lambda_1^k + c_2\lambda_2^k$. Тогда

$$\begin{aligned} y_{k+1} + py_k + qy_{k-1} &\approx c_1\lambda_1^{k-1}[\lambda_1^2 + p\lambda_1 + q] + \\ &+ c_2\lambda_2^{k-1}[\lambda_2^2 + p\lambda_2 + q] = 0. \end{aligned} \quad (17)$$

Здесь приближенное равенство справедливо с точностью до $O(|\lambda_3|^k)$.

Аналогичное приближенное равенство

$$z_{k+1} + pz_k + qz_{k-1} \approx 0 \quad (18)$$

будет справедливо для любой другой компоненты z_k вектора $Y_k = A^k Y_0$. В качестве z_k можно также взять $z_k = y_{k+1}$ или, более обще, любую компоненту вектора $Z_k = A^k Z_0$, где Z_0 произвольный начальный вектор.

Из равенств (17) и (18) получим, вообще говоря,

$$\begin{aligned} p &\approx -\frac{y_{k-1}z_{k+1} - z_{k-1}y_{k+1}}{y_{k-1}z_k - z_{k-1}y_k} \\ q &\approx \frac{y_kz_{k+1} - z_ky_{k+1}}{y_{k-1}z_k - z_{k-1}y_k}. \end{aligned} \quad (19)$$

В частности, если взять $z_k = y_{k+1}$, получим

$$\begin{aligned} p &\approx -\frac{y_{k-1}y_{k+2} - y_ky_{k+1}}{y_{k-1}y_{k+1} - y_k^2} \\ q &\approx \frac{y_ky_{k+2} - y_{k+1}^2}{y_{k-1}y_{k+1} - y_k^2}. \end{aligned} \quad (20)$$

Дадим более строгое обоснование формул (19) и (20). Это позволит выяснить условия их применимости и оценить погрешность.

Пусть

$$y_k = c_1\lambda_1^k + c_2\lambda_2^k + O(|\lambda_3|^k)$$

$$z_k = d_1\lambda_1^k + d_2\lambda_2^k + O(|\lambda_3|^k).$$

Тогда

$$\begin{aligned} y_{k-1}z_{k+1} - z_{k-1}y_{k+1} &= \\ &= (c_1\lambda_1^{k-1} + c_2\lambda_2^{k-1} + O(|\lambda_3|^k))(d_1\lambda_1^{k+1} + d_2\lambda_2^{k+1} + O(|\lambda_3|^k)) - \\ &- (c_1\lambda_1^{k+1} + c_2\lambda_2^{k+1} + O(|\lambda_3|^k))(d_1\lambda_1^{k-1} + d_2\lambda_2^{k-1} + O(|\lambda_3|^k)) = \\ &= (c_1d_2 - c_2d_1)(\lambda_2^2 - \lambda_1^2)\lambda_1^{k-1}\lambda_2^{k-1} + O(|\lambda_1^k\lambda_3^k|). \end{aligned}$$

Аналогично

$$\begin{aligned}y_{k-1}z_k - z_{k-1}y_k &= (c_1d_2 - c_2d_1)(\lambda_2 - \lambda_1)\lambda_1^{k-1}\lambda_2^{k-1} + O(|\lambda_1^k\lambda_2^k|) \\y_kz_{k+1} - z_{k-1}y_{k+1} &= (c_1d_2 - c_2d_1)(\lambda_2 - \lambda_1)\lambda_1^k\lambda_2^k + O(|\lambda_1^k\lambda_2^k|).\end{aligned}$$

Поэтому

$$\begin{aligned}-\frac{y_{k-1}z_{k+1} - z_{k-1}y_{k+1}}{y_{k-1}z_k - z_{k-1}y_k} &= -(\lambda_1 + \lambda_2) + O\left(\left|\frac{\lambda_3}{\lambda_2}\right|^k\right) = p + O\left(\left|\frac{\lambda_3}{\lambda_2}\right|^k\right) \\-\frac{y_kz_{k+1} - z_{k-1}y_{k+1}}{y_{k-1}z_k - z_{k-1}y_k} &= \lambda_1\lambda_2 + O\left(\left|\frac{\lambda_3}{\lambda_2}\right|^k\right) = q + O\left(\left|\frac{\lambda_3}{\lambda_2}\right|^k\right),\end{aligned}\quad (19')$$

если только $c_1d_2 - c_2d_1 \neq 0$. Аналогично

$$\begin{aligned}-\frac{y_{k-1}y_{k+2} - y_ky_{k+1}}{y_{k-1}y_{k+1} - y_k^2} &= p + O\left(\left|\frac{\lambda_3}{\lambda_2}\right|^k\right) \\-\frac{y_ky_{k+2} - y_{k+1}^2}{y_{k-1}y_{k+1} - y_k^2} &= q + O\left(\left|\frac{\lambda_3}{\lambda_2}\right|^k\right).\end{aligned}\quad (20')$$

Отметим, что для формул (20') условие $c_1d_2 - c_2d_1 \neq 0$ всегда выполняется, ибо $d_1 = \lambda_1 c_1$, $d_2 = \lambda_2 c_2$, так что $c_1d_2 - c_2d_1 = c_1\lambda_2(\lambda_2 - \lambda_1) \neq 0$.

Собственные значения λ_1 и λ_2 можно определять минуя вычисление p и решение квадратного уравнения. Именно, определив $q = r^2$ по одной из формул (19) или (20), найдем r и вычислим выражение¹⁾

$$\begin{aligned}\mu_k &= \frac{1}{2}[ry_{k-1} + r^{-1}y_{k+1}] \approx \\&\approx \frac{1}{2}[r(c_1\lambda_1^{k-1} + c_2\lambda_2^{k-1}) + r^{-1}(c_1\lambda_1^{k+1} + c_2\lambda_2^{k+1})] = \\&= \frac{1}{2}[c_1r^k e^{ik\theta} + c_2r^k e^{-ik\theta}][e^{-i\theta} + e^{i\theta}] = y_k \cos \theta.\end{aligned}\quad (21)$$

Отсюда находим

$$\cos \theta \approx \frac{\mu_k}{y_k} \quad (22)$$

с точностью до величин порядка $\left(\frac{\lambda_3}{\lambda_1}\right)^k$.

После того как собственные значения λ_1 и λ_2 определены, соответствующие им собственные векторы легко определяются. Именно, из приближенных равенств:

$$Y_k \approx a_1\lambda_1^k U_1 + a_2\lambda_2^k U_2$$

$$Y_{k+1} \approx a_1\lambda_1^{k+1} U_1 + a_2\lambda_2^{k+1} U_2$$

1) Эйткен [5].

находим

$$\begin{aligned} Y_{k+1} - \lambda_2 Y_k &\approx a_1 \lambda_1^k (\lambda_1 - \lambda_2) U_1 \\ Y_{k+1} - \lambda_1 Y_k &\approx a_2 \lambda_2^k (\lambda_2 - \lambda_1) U_2, \end{aligned} \quad (23)$$

откуда следует, что $Y_{k+1} - \lambda_2 Y_k$ и $Y_{k+1} - \lambda_1 Y_k$, с точностью до малых слагаемых, являются собственными векторами, соответствующими собственным значениям λ_1 и λ_2 .

Пример 6. Собственные значения матрицы

$$A = \begin{bmatrix} 26 & -54 & 4 \\ 13 & -28 & 3 \\ 26 & -56 & 5 \end{bmatrix}$$

суть $\lambda_1 = 1 + 5i$, $\lambda_2 = 1 - 5i$ и $\lambda_3 = 1$. Собственный вектор, принадлежащий λ_1 (нормированный соответствующим образом) есть $U_1 = (1, 0.53974564 - 0.09141494i, 1.03656599 + 0.01589827i)^T$.

Вычислим собственные значения λ_1 и λ_2 и принадлежащие им собственные векторы степенным методом. Имеем

Y_0	Y_8	Y_9	Y_{10}
0.2	1293880.4	3669538.0	-26301835
0.4	654932.2	2528583.4	-11971081
0.6	1348746.6	3708420.2	-27650581
	3297559.2	9906541.6	-65923497

Беря за y_k первую компоненту и за z_k вторую компоненту, получим

$$p \approx -\frac{17367716 \cdot 10^5}{86838591 \cdot 10^4} = -1.9999997$$

$$q \approx \frac{22578047 \cdot 10^6}{86838591 \cdot 10^4} = 26.000015,$$

откуда

$$\lambda_1 \approx 0.9999999 + 5.0000015i$$

$$\lambda_2 \approx 0.9999999 - 5.0000015i.$$

Далее

$$(Y_{10} - \lambda_2 Y_9)^T =$$

$$= (-22632297 + 18347696i, -9442498 + 12642921i, -23942161 + 18542107i)^T,$$

так что после соответствующей нормировки

$$\tilde{U}_1 = (1, 0.5250 - 0.1330i, 1.0391 + 0.0231i)^T.$$

Используя Y_{18} , Y_{19} и Y_{20} , получим

$$\lambda_1 = 1.00000000 + 5.00000000i,$$

$$U_1 = (1, 0.5397452 - 0.0914155i, 1.0365660 + 0.0158983i)^T.$$

5. Наибольшее по модулю собственное значение вещественно и находится в жордановом ящике второго порядка. Нами уже отмечалось, что ход итерационного степенного процесса существенно зависит от структуры канонической формы Жордана, связанной с данной матрицей. В этом пункте мы на простейшем примере покажем характер тех изменений, которые возникают, если наибольшему по модулю собственному значению соответствует нелинейный элементарный делитель.

Именно, рассмотрим случай, когда λ_1 вещественно и принадлежит в канонической форме Жордана ящику $\begin{bmatrix} \lambda_1 & 0 \\ 1 & \lambda_1 \end{bmatrix}$, а следующее собственное значение λ_2 по модулю меньше, чем λ_1 . Для простоты выкладок мы будем считать, как и прежде, что всем остальным собственным значениям соответствуют линейные элементарные делители.

В рассматриваемом случае вместо базиса из собственных векторов мы берем канонический базис U_1, U_2, \dots, U_n . Воздействие матрицы A на векторы этого базиса происходит по формулам

$$AU_1 = \lambda_1 U_1 + U_2$$

$$AU_2 = \lambda_1 U_2$$

$$AU_3 = \lambda_3 U_3$$

$$\vdots \quad \vdots \quad \vdots$$

$$AU_n = \lambda_n U_n,$$

и, следовательно,

$$\begin{aligned} A^k U_1 &= \lambda_1^k U_1 + k\lambda_1^{k-1} U_2 \\ A^k U_2 &= \lambda_1^k U_2 \\ A^k U_3 &= \lambda_3^k U_3 \\ &\vdots \quad \vdots \quad \vdots \\ A^k U_n &= \lambda_n^k U_n. \end{aligned} \tag{24}$$

Пусть Y_0 начальный вектор. Мы будем предполагать, что проекция вектора Y_0 в корневое подпространство, соответствующее собственному значению λ_1 , отлична от нуля и не является собственным вектором. Примем ее за первый вектор U_1 канонического базиса. Тогда

$$Y_0 = U_1 + a_3 U_3 + \dots + a_n U_n.$$

и, в силу (24),

$$Y_k = A^k Y_0 = \lambda_1^k U_1 + k \lambda_1^{k-1} U_2 + \lambda_3^k a_3 U_3 + \dots + \lambda_n^k a_n U_n.$$

Любая компонента вектора Y_k будет иметь вид (мы по-прежнему опускаем первый индекс)

$$y_k = c_1 \lambda_1^k + c_2 k \lambda_1^{k-1} + c_3 \lambda_3^k + \dots + c_n \lambda_n^k. \quad (25)$$

Отношение $\frac{y_{k+1}}{y_k}$ по-прежнему стремится к λ_1 , но медленнее, чем любая геометрическая прогрессия из-за наличия множителя k во втором слагаемом. Именно:

$$\frac{y_{k+1}}{y_k} = \lambda_1 \left(1 + O\left(\frac{1}{k}\right) \right).$$

Практически определить λ_1 из отношения $\frac{y_{k+1}}{y_k}$ становится почти невозможным¹⁾.

Для определения собственного значения λ_1 следует поступать так же, как при определении комплексной пары собственных значений, т. е. искать коэффициенты $p = -2\lambda_1$ и $q = \lambda_1^2$ квадратного уравнения, двойным корнем которого является λ_1 . Итак, пусть

$$y_k = c_1 \lambda_1^k + c_2 k \lambda_1^{k-1} + c_3 \lambda_3^k + \dots$$

Тогда

$$\begin{aligned} y_{k+1} + p y_k + q y_{k-1} &= c_1 \lambda_1^{k+1} (\lambda_1 + p \lambda_1 + q) + \\ &+ c_2 \lambda_1^{k-2} [(k+1) \lambda_1^2 + p k \lambda_1 + q (k-1)] + O(\lambda_3^k) = O(\lambda_3^k). \end{aligned}$$

Аналогично

$$z_{k+1} + p z_k + q z_{k-1} = O(\lambda_3^k),$$

где z_k определяется так же, как и в предыдущем пункте.

Из полученных приближенных равенств находим

$$\begin{aligned} p &\approx -\frac{y_{k-1} z_{k+1} - z_{k-1} y_{k+1}}{y_{k-1} z_k - z_{k-1} y_k} \\ q &\approx \frac{y_k z_{k+1} - z_k y_{k+1}}{y_{k-1} z_k - z_{k-1} y_k}. \end{aligned} \quad (19'')$$

¹⁾ Отметим, что если ящик, к которому принадлежит λ_1 в канонической форме Жордана, имеет более сложную структуру, то в выражении (6) появляются и другие степени λ_1 , умноженные на соответствующие биномиальные коэффициенты:

$$y_k = c_1 \lambda_1^k + c_2 k \lambda_1^{k-1} + c_3 \frac{k(k-1)}{2} \lambda_1^{k-2} + \dots + c_n \lambda_n^k.$$

Отношение $\frac{y_{k+1}}{y_k}$ стремится к λ_1 еще медленнее.

Легко проверить, что эти равенства будут справедливы с точностью до величин порядка $\left(\frac{\lambda_3}{\lambda_1}\right)^k$. Это делается в точности так же, как в предыдущем случае.

Для определения собственного значения λ_1 , очевидно, достаточно определение одного из коэффициентов p или q . Однако совпадение чисел $-\frac{p}{2}$ и $V\bar{q}$ служит контролем правильности гипотезы о входжении собственного значения λ_1 в канонический ящик.

Сделанная гипотеза может быть подтверждена и другими средствами. Именно, найдя $\lambda_1 \approx V\bar{q}$, можно построить так называемые λ -разности

$$\begin{aligned} \Delta y_k &= y_{k+1} - \lambda_1 y_k \\ \Delta^2 y_k &= \Delta y_{k+1} - \lambda_1 \Delta y_k. \end{aligned} \quad (26)$$

Легко вычислить, что

$$\begin{aligned} \Delta y_k &= c_1 \lambda_1^{k+1} + c_2 (k+1) \lambda_1^k + c_3 \lambda_3^{k+1} + \dots + c_n \lambda_n^{k+1} - c_1 \lambda_1^{k+1} - \\ &- c_2 k \lambda_1^k - c_3 \lambda_3^k \lambda_1 - \dots - c_n \lambda_n^k \lambda_1 = c_2 \lambda_1^k + O(\lambda_3^k), \\ \frac{\Delta y_{k+1}}{\Delta y_k} &= \lambda_1 + O\left(\frac{\lambda_3}{\lambda_1}\right)^k, \end{aligned} \quad (27)$$

т. е. $\frac{\Delta y_{k+1}}{\Delta y_k}$ стремится к λ_1 достаточно быстро. Совпадение предела $\frac{\Delta y_{k+1}}{\Delta y_k}$ с вычисленным ранее значением для λ_1 и факт быстрой сходимости $\frac{\Delta y_{k+1}}{\Delta y_k}$ к λ_1 служит подтверждением предположения о том, что λ_1 входит в ящик 1-го порядка.

Далее ясно, что

$$\Delta^2 y_k = O(\lambda_3^k).$$

т. е. вторая λ -разность мала по сравнению с самой компонентой y_k .

Собственный вектор U_2 , соответствующий собственному значению λ_1 , легко определяется. Именно из равенства

$$Y_k = \lambda_1^k U_1 + k \lambda_1^{k-1} U_2 + O(\lambda_3^k)$$

следует

$$Y_{k+1} - \lambda_1 Y_k = \lambda_1^k U_2 + O(\lambda_3^k), \quad (28)$$

т. е. вектор $Y_{k+1} - \lambda_1 Y_k = \tilde{U}_2$ приближенно равен собственному вектору, соответствующему λ_1 . После нормировки точность приближенного равенства будет порядка $\left(\frac{\lambda_3}{\lambda_1}\right)^k$. Корневой вектор, соответствующий собственному значению λ_1 , определен с точностью до слагаемого, пропорционального собственному вектору U_2 . За одно

из возможных приближенных значений корневого вектора может быть взят сам вектор

$$Y_k \approx c_1 \lambda_1^k U_1 + c_2 k \lambda_1^{k-1} U_2.$$

Однако при больших значениях k этот корневой вектор, благодаря множителю k во втором слагаемом, сильно „вытянут“ в направлении собственного вектора U_2 . Целесообразнее взять в качестве приближенного корневого вектора

$$k \lambda_1 Y_{k-1} - (k-1) Y_k = Y_k - k(Y_k - \lambda_1 Y_{k-1}) = Y_k - k \tilde{U}_2 = \lambda_1^k U_1. \quad (29)$$

Полученный вектор лишь скалярным множителем отличается от проекции U_1 начального вектора Y_0 на корневое подпространство, соответствующее собственному значению λ_1 .

Пример 7. Собственные значения матрицы

$$A = \begin{bmatrix} -9 & -2 & -9 \\ -13 & -2 & -12 \\ 16 & 4 & 16 \end{bmatrix}$$

суть $\lambda_1 = \lambda_2 = 2$; $\lambda_3 = 1$. Собственный вектор, соответствующий $\lambda_1 = 2$ есть $U_2 = \left(-\frac{2}{3}, -\frac{5}{6}, 1 \right)' = (-0.666667, -0.833333, 1)'$.

Приведем вычисления степенным методом. За начальный вектор Y_0 возьмем вектор $(1, 0, 1)'$. Его проекция на корневое подпространство есть (после нормировки)

$$\left(-\frac{5}{7}, -\frac{6}{7}, 1 \right)' \approx (-0.714286, -0.857143, 1)'.$$

Имеем

$Y_0^{(0)}$	$Y^{(18)}$	$Y^{(19)}$	$Y^{(20)}$
1	8126470	17301510	36700166
0	10223622	21757958	46137350
-1	-12320776	-26214408	-55574536
	6029316	12845060	27262980.

Сходимость процесса оказывается медленной, что дает основание предполагать наличие кратных собственных значений.

Для вычисления p и q берем за y_k первую компоненту, за z_k вторую. Тогда

$$p = -\frac{-27483387 \cdot 10^4}{-68705321 \cdot 10^3} = -4.0001831, \quad -\frac{p}{2} = 2.000092$$

$$q = \frac{-27484802 \cdot 10^4}{-68705321 \cdot 10^3} = 4.0003891, \quad \sqrt{q} = 2.000097.$$

Если за y_k взять первую компоненту, а за z_k третью, то совпадение $-\frac{p}{2}$ и \sqrt{q} будет еще лучше. Именно,

$$p = -4.0000648, \quad -\frac{p}{2} = 2.000032$$

$$q = 4.0001373, \quad \sqrt{q} = 2.000034.$$

Таким образом, можно считать $\lambda_1 = 2.00003$.

Далее

$\tilde{U}_2 = Y_{20} - 2.00003Y_{19}$	$\tilde{U}_1 = Y_{20} - 20\tilde{U}_2$	U_2	U_1
2096627	-5232374	-0.666668	-0.714401
2620781	-6278270	-0.833334	-0.857202
-3144934	7324144	1.000000	1.000000

Сравнивая найденные значения для \tilde{U}_2 и U_1 с точными, мы видим хорошее совпадение для \tilde{U}_2 и несколько худшее для U_1 . Последнее можно объяснить не очень удачным выбором начального вектора, проекция которого на соответствующее корневое подпространство оказывается довольно близкой к собственному вектору.

6. Имеются два близких по модулю наибольших по модулю собственных значения. В пунктах 4 и 5 для определения собственных значений мы применяли по сути дела один и тот же прием, и основанием для применения этого приема служило нарушение сходимости последовательности $\frac{y_{k+1}}{y_k}$. Из результатов пункта 1 следует,

что причиной плохой сходимости этой последовательности может быть не только равенство наибольших модулей собственных значений, но и их близость. В этом случае тоже может быть применен прием, заключающийся в использовании формул (19) или (20), однако при этом корни составленного квадратного уравнения уже будут вещественными и неизбежно близкими по модулю. Решив квадратное уравнение, мы их определим с точностью до величин порядка $\left(\frac{\lambda_3}{\lambda_2}\right)^k$.

Заметим, что если последовательности отношений компонент сходятся быстро, указанный прием не позволяет определить одновременно λ_1 и λ_2 с удовлетворительной точностью, так как формулы для определения коэффициентов p и q будут содержать в числителе и знаменателе числа, близкие к нулю.

Процесс будет иметь плохую сходимость также, если $|\lambda_2|$ не достаточно превосходит $|\lambda_3|$.

Пример 8. В качестве иллюстрации указанного приема вычислим наибольшее по модулю собственное значение матрицы Леверье.

Используем для этого нормированные итерации вектора $(1, 0, 0, 0)'$, вычисленного нами для примера 1.

Имеем

\tilde{Y}_{18}	\tilde{Y}_{19}	\tilde{Y}_{20}
1.0000000	1.0000000	1.0000000
-8.1970024	-8.2032008	-8.2093658
8.1476812	8.1701265	8.1926032
-7.3754981	-7.9595730	-8.5464157
<hr/>		
-6.4248193	-6.9926473	-7.5631783.

Выпишем также нормирующие множители

$$\rho_{18} = -17.450861, \quad \rho_{19} = -17.458270, \quad \rho_{20} = -17.465517.$$

Очевидно, что

$$p = \rho_{20}\tilde{p}, \quad q = \rho_{19}\tilde{q},$$

где \tilde{p} и \tilde{q} составляются из компонент нормированных итераций по формулам (19').

Если за y_k взять первую компоненту, за z_k вторую, то

$$\tilde{p} = -\frac{0.0123634}{0.0061984} = -1.994611, \quad p = 34.836912$$

$$\tilde{q} = \frac{0.0061650}{0.0061984} = 0.994611, \quad q = 303.27451.$$

Таким образом, для определения λ_1 и λ_2 имеем уравнение

$$t^2 + 34.836912t + 303.27451 = 0,$$

откуда

$$\lambda_1 = -17.418456 - \sqrt{0.12810} \approx -17.4185 - 0.3579 = -17.7764$$

$$\lambda_2 = -17.0606.$$

Принимая за y_k первую компоненту, за z_k сначала третью, а затем четвертую компоненты получим

$$\lambda_1 = -17.831, \quad \lambda_2 = -17.125;$$

$$\lambda_1 = -17.866, \quad \lambda_2 = -17.148.$$

Сравнивая полученные значения для λ_1 , мы видим, что все три значения более близки друг к другу, чем значения, полученные в примере 1. Последнее значение $\lambda_1 = -17.866$ с точностью до $3 \cdot 10^{-3}$ совпадает с точным значением. При этом λ_2 определяется примерно с такой же точностью, как λ_1 .

§ 54. Ускорение сходимости степенного метода

В этом параграфе будут изложены два приема ускорения сходимости степенного метода, в случае если наибольшее собственное значение вещественное и простое.

1. Скалярное произведение. Этот прием особенно удобен в применении к симметричной матрице; однако мы изложим его без этого предположения.

Пусть наряду с последовательностью итераций вектора матрицей A

$$Y_0, \quad Y_1 = AY_0, \quad Y_2 = A^2Y_0, \quad \dots, \quad Y_k = A^kY_0, \quad \dots \quad (1)$$

вычислена также последовательность итераций матрицей A'

$$Z_0, \quad Z_1 = A'Z_0, \quad Z_2 = A'^2Z_0, \quad \dots, \quad Z_k = A'^kZ_0, \quad \dots \quad (1')$$

Введем базисы U_1, \dots, U_n и V_1, \dots, V_n , составленные из собственных векторов матриц A и A' , причем будем предполагать, что эти базисы удовлетворяют условиям ортогональности и нормированности в смысле § 10 п. 3.

Пусть

$$\begin{aligned} Y_0 &= a_1U_1 + a_2U_2 + \dots + a_nU_n \\ Z_0 &= b_1V_1 + b_2V_2 + \dots + b_nV_n. \end{aligned} \quad (2)$$

Тогда

$$\begin{aligned} (Y_k, Z_k) &= (A^kY_0, A'^kZ_0) = (A^{2k}Y_0, Z_0) = \\ &= (a_1\lambda_1^{2k}U_1 + a_2\lambda_2^{2k}U_2 + \dots + a_n\lambda_n^{2k}U_n, b_1V_1 + b_2V_2 + \dots + b_nV_n). \end{aligned}$$

Далее, в силу свойств ортогональности и нормированности системы векторов U_1, \dots, U_n и V_1, \dots, V_n , имеем

$$(Y_k, Z_k) = a_1b_1\lambda_1^{2k} + a_2b_2\lambda_2^{2k} + \dots + a_nb_n\lambda_n^{2k}. \quad (3)$$

Аналогично

$$(Y_{k-1}, Z_k) = a_1b_1\lambda_1^{2k-1} + a_2b_2\lambda_2^{2k-1} + \dots + a_nb_n\lambda_n^{2k-1}. \quad (4)$$

Из равенств (3) и (4) получаем:

$$\begin{aligned} \frac{(Y_k, Z_k)}{(Y_{k-1}, Z_k)} &= \frac{a_1b_1\lambda_1^{2k} + a_2b_2\lambda_2^{2k} + \dots + a_nb_n\lambda_n^{2k}}{a_1b_1\lambda_1^{2k-1} + a_2b_2\lambda_2^{2k-1} + \dots + a_nb_n\lambda_n^{2k-1}} = \\ &= \lambda_1 + O\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}. \end{aligned} \quad (5)$$

Из этой оценки видно, что образование скалярного произведения сокращает число шагов итерации, нужных для определения λ_1 с данной точностью, почти вдвое. Однако при этом требуется дополнительное вычисление последовательности (1').

В случае симметричной матрицы, при $Z_0 = Y_0$ последовательности (1) и (1') совпадают, и поэтому в этом случае применение метода скалярного произведения особенно целесообразно. Начиная с некоторого шага процесса, нужно вычислять соответствующие скалярные произведения и определять λ_1 через их отношения. Именно,

$$\lambda_1 \approx \frac{(A^k Y_0, A^k Y_0)}{(A^{k-1} Y_0, A^k Y_0)}. \quad (6)$$

Так, в примере 3 § 53 мы легко вычисляем

$$(A^7 Y_0, A^7 Y_0) = 0.00007528987$$

$$(A^6 Y_0, A^6 Y_0) = 0.00015685433,$$

что дает для λ_1 значение 0.479999 (вместо значений 0.4800, 0.4792, 0.4808, найденных из отношений компонент $A^7 Y_0$ и $A^6 Y_0$). В качестве второго примера рассмотрим матрицу

$$\begin{bmatrix} 1.0000000 & 0 & 1.0000000 & 0 \\ 1.0000000 & 0.7777778 & 0.3333333 & 0.3333333 \\ 0 & -0.0252525 & 0.5555556 & -0.0252525 \\ 0 & -0.8888889 & -8.6444444 & 0.1111111 \end{bmatrix}, \quad (7)$$

собственные числа которой есть $1, \frac{2}{3}, \frac{4}{9}$ и $\frac{1}{3}$.

Для определения λ_1 образуем итерации $A^k Y_0$, беря в качестве Y_0 вектор $(1, 1, 1, 1)'$.

Приведем 17-ю, 18-ю, 19-ю и 20-ю итерации:

$A^{17} Y_0$	$A^{18} Y_0$	$A^{19} Y_0$	$A^{20} Y_0$
4.6731097	4.6760089	4.6779433	4.6792336
8.3733415	8.3912886	8.4032694	8.4112637
0.0028992	0.0019344	0.0012903	0.0008605
-8.3861607	-8.3998278	-8.4089592	-8.4150555
4.6631897	4.6694041	4.6735438	4.6763023.

Отношения компонент этих итераций будут

$$\begin{array}{lll} 1.000620 & 1.000414 & 1.000276 \\ 1.002143 & 1.001428 & 1.000951 \\ 0.667219 & 0.667029 & 0.666899 \\ 1.001630 & 1.001087 & 1.000725. \end{array}$$

Здесь отношения третьих компонент сильно отличаются от остальных, в силу исчезновения значащих цифр. Последний столбец дает

$\lambda_1 \approx 1.001$; найденное значение совпадает с точным с точностью до одной единицы третьего знака.

Покажем теперь, как можно уточнить это значение применением способа скалярного произведения.

Для этого образуем итерации вектора $(1, 1, 1, 1)'$, транспонированной матрицей A' .

Вычисляя, получим

$$A'^{20}Y_0 = (0.7961118, -0.0002189, 3.9939022, -0.1134904)'.$$

Далее,

$$(A^{20}Y_0, A'^{20}Y_0) = 4.681817 \quad \text{и} \quad (A^{19}Y_0, A'^{20}Y_0) = 4.681816.$$

Отношение

$$\frac{(A^{20}Y_0, A'^{20}Y_0)}{(A^{19}Y_0, A'^{20}Y_0)} = 1.000000$$

дает в качестве λ_1 значение 1.000000, верное с точностью до шести знаков после запятой.

Замечание. Если, находя итерации, мы их нормируем, то для определения λ_1 нужно пользоваться одной из формул

$$\lambda_1 = \frac{(A\tilde{Y}_{k-1}, \tilde{Z}_k)}{(\tilde{Y}_{k-1}, \tilde{Z}_k)} \quad (8)$$

или

$$\lambda_1 = \frac{(\tilde{Y}_k, A'\tilde{Z}_{k-1})}{(\tilde{Y}_k, \tilde{Z}_{k-1})}. \quad (9)$$

Здесь через \tilde{Y}_k и \tilde{Z}_k обозначены нормированные итерации $A^k Y_0$ и $A'^k Z_0$. Способ нормировки при использовании этими формулами безразличен.

2. δ^2 -процесс ¹⁾. Этот прием применим только в случае, когда $|\lambda_1| > |\lambda_2| > |\lambda_3|$, так что λ_1 и λ_2 вещественны.

Предположим, что мы определили ряд величин

$$y_k, \quad y_{k+1}, \quad y_{k+2}, \quad \dots \quad (10)$$

относительно которых известно, что

$$y_k = c_1 \lambda_1^k + c_2 \lambda_2^k + \dots + c_n \lambda_n^k. \quad (11)$$

В качестве y_k можно взять, например, любую компоненту вектора $Y_k = A^k Y_0$, скалярное произведение соответствующих итераций и т. д. Тогда, как было показано в § 53 и в п. 1 § 54, можно приблизенно определить первое собственное число λ_1 как отношение $a_k = \frac{y_{k+1}}{y_k}$.

1) Эйткен [5].

Далее, в § 53 было показано, что

$$u_k = \lambda_1 [1 - b'_2 \alpha_2^k - b'_3 \alpha_3^k + b_2 b'_2 \alpha_2^{2k}] + o(\alpha_3^k + \alpha_2^{2k}), \quad (12)$$

где $b'_2 = \frac{c_2}{c_1}(1 - \alpha_2)$, $b'_3 = \frac{c_3}{c_1}(1 - \alpha_3)$ и $\alpha_i = \frac{\lambda_i}{\lambda_1}$.

Если сходимость последовательности u_k , u_{k+1} , u_{k+2} , ... недостаточно быстрая, то ее можно сильно улучшить следующим приемом, который носит название δ^2 -процесса. Образуем

$$P(u_k) = \frac{\begin{vmatrix} u_k & u_{k+1} \\ u_{k+1} & u_{k+2} \end{vmatrix}}{u_k - 2u_{k+1} + u_{k+2}}. \quad (13)$$

Покажем, что

$$P(u_k) = \lambda_1 + O\left(\frac{\lambda_2}{\lambda_1}\right)^{2k} + O\left(\frac{\lambda_3}{\lambda_1}\right)^k. \quad (14)$$

С этой целью положим

$$u_k = \lambda_1(1 + \varepsilon_k),$$

тогда

$$\begin{vmatrix} u_k & u_{k+1} \\ u_{k+1} & u_{k+2} \end{vmatrix} = \lambda_1^2 \begin{vmatrix} 1 + \varepsilon_k & 1 + \varepsilon_{k+1} \\ 1 + \varepsilon_{k+1} & 1 + \varepsilon_{k+2} \end{vmatrix}.$$

Разбивая последний определитель на сумму четырех, получим

$$\begin{vmatrix} u_k & u_{k+1} \\ u_{k+1} & u_{k+2} \end{vmatrix} = \lambda_1^2 \left[\varepsilon_k - 2\varepsilon_{k+1} + \varepsilon_{k+2} + \begin{vmatrix} \varepsilon_k & \varepsilon_{k+1} \\ \varepsilon_{k+1} & \varepsilon_{k+2} \end{vmatrix} \right].$$

Но

$$u_k - 2u_{k+1} + u_{k+2} = \lambda_1 [\varepsilon_k - 2\varepsilon_{k+1} + \varepsilon_{k+2}].$$

Таким образом,

$$P(u_k) = \lambda_1 \left[1 + \frac{\begin{vmatrix} \varepsilon_k & \varepsilon_{k+1} \\ \varepsilon_{k+1} & \varepsilon_{k+2} \end{vmatrix}}{\varepsilon_k - 2\varepsilon_{k+1} + \varepsilon_{k+2}} \right].$$

Вычисляя, получим, что

$$\frac{\begin{vmatrix} \varepsilon_k & \varepsilon_{k+1} \\ \varepsilon_{k+1} & \varepsilon_{k+2} \end{vmatrix}}{\varepsilon_k - 2\varepsilon_{k+1} + \varepsilon_{k+2}} \approx A\alpha_3^k + B\alpha_2^{2k},$$

где

$$A = \frac{-b'_3(\alpha_2 - \alpha_3)^2}{(1 - \alpha_2)^2}, \quad B = b_2 b'_2 \alpha_2^2.$$

Таким образом,

$$P(u_k) = \lambda_1 + O\left(\frac{\lambda_3}{\lambda_1}\right)^k + O\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}.$$

Отсюда следует, что ошибка в определении λ_1 при помощи δ^2 -процесса может быть намного меньше, чем при нахождении его прямо из последовательности u_k, u_{k+1}, \dots

Отметим, что при фактическом проведении δ^2 -процесса по формуле (13) как в числителе, так и в знаменателе происходит уничтожение значащих цифр. Этого явления можно избежать следующим приемом. Пусть $u_k = c + l_k$, где c — число, составленное из уставившихся десятичных знаков в последовательности u_k . Тогда, как нетрудно проверить,

$$P(u_k) = c + P(l_k),$$

причем второе слагаемое играет роль малой поправки по отношению к первому.

Замечание. При нахождении первого собственного числа симметричной матрицы δ^2 -процесс надо применять не к компонентам итераций, а к соответствующим скалярным произведениям.

Покажем применение δ^2 -процесса на примерах § 53.

В примере 2 § 53 для приведенных отношений применение δ^2 -процесса дает для λ_1 следующие значения:

$$\begin{aligned} & -17.3974 \\ & -17.3977 \\ & -17.3977. \end{aligned}$$

С точностью до четырех знаков $\lambda_1 = -17.3977$.

Аналогично, в примере 3 § 53 применение δ^2 -процесса к приведенным отношениям дает для λ_1 следующие значения:

$$\begin{aligned} & 0.66748 \\ & 0.66748 \\ & 0.66748 \\ & 0.66748. \end{aligned}$$

Таким образом, λ_1 определяется уже с точностью до пяти знаков. (Точное значение $\lambda_1 = 0.667483$).

δ^2 -процесс можно применять также и к определению компонент первого собственного вектора. При этом мы укажем два различных варианта этого процесса, в зависимости от того, можно ли считать известным λ_1 с достаточной степенью точности или нет.

1) Пусть мы знаем только последовательность итераций $A^k Y_0$, причем для удобства вычислений пусть каждая итерация нормирована делением на компоненту z_k с фиксированным номером.

Обозначим любую другую компоненту вектора Y_k через y_k .

Если

$$y_k = c_1 \lambda_1^k + c_2 \lambda_2^k + \dots + c_n \lambda_n^k,$$

$$z_k = b_1 \lambda_1^k + b_2 \lambda_2^k + \dots + b_n \lambda_n^k,$$

то указанное деление сводит z_k к единице, а y_k к v_k , причем

$$\begin{aligned} v_k = \frac{y_k}{z_k} &= \frac{c_1 \lambda_1^k + c_2 \lambda_2^k + \dots + c_n \lambda_n^k}{b_1 \lambda_1^k + b_2 \lambda_2^k + \dots + b_n \lambda_n^k} = \\ &= \frac{c_1}{b_1} + \frac{c_2 b_1 - b_2 c_1}{b_1^2} \left(\frac{\lambda_2}{\lambda_1} \right)^k + O \left(\frac{\lambda_3}{\lambda_1} \right)^k. \end{aligned}$$

Несложное вычисление показывает, что

$$P(v_k) = \frac{c_1}{b_1} + O \left(\frac{\lambda_3}{\lambda_1} \right)^k + O \left(\frac{\lambda_3}{\lambda_1} \right)^{2k}.$$

Таким образом, если δ^2 -процесс применен ко всем компонентам нормированного вектора $A^k Y_0$, то мы найдем отношения коэффициентов, стоящих при степенях λ_i в выражениях y_k . Коэффициенты же эти пропорциональны компонентам собственного вектора.

Так, для собственного вектора примера 2 предыдущего параграфа применение δ^2 -процесса дает для компонент вектора значения

$$\begin{array}{r} 1.00000 \\ -8.14718 \\ 7.91721, \end{array}$$

которые значительно ближе к точным, чем значения, вычисленные непосредственно из тех же итераций.

2) Если λ_i известно достаточно точно, процесс улучшения можно построить следующим образом. Умножим все компоненты векторов $A^{k-1} Y_0$, $A^k Y_0$, $A^{k+1} Y_0$, на λ_1 , 1, λ_1^{-1} соответственно и затем применим к ним δ^2 -процесс. Так как

$$y_{k-1} \lambda_1 = \lambda_1^k c_1 + \lambda_2^{k-1} \lambda_1 c_2 + \dots + \lambda_n^{k-1} \lambda_1 c_n,$$

$$y_k = \lambda_1^k c_1 + \lambda_2^k c_2 + \dots + \lambda_n^k c_n,$$

$$y_{k+1} \lambda_1^{-1} = \lambda_1^k c_1 + \frac{\lambda_2^{k+1}}{\lambda_1} c_2 + \dots + \frac{\lambda_n^{k+1}}{\lambda_1} c_n,$$

то

$$\begin{aligned} \begin{vmatrix} \lambda_1 y_{k-1} & y_k \\ y_k & \lambda_1^{-1} y_{k+1} \end{vmatrix} &= \\ &\approx [c_1 c_2 \lambda_1^{k-1} \lambda_2^{k-1} (\lambda_1 - \lambda_2)^2 + \dots + c_1 c_n \lambda_1^{k-1} \lambda_n^{k-1} (\lambda_1 - \lambda_n)^2] \times \\ &\quad \times \left[1 + O \left(\frac{\lambda_3}{\lambda_1} \right)^k \right]. \end{aligned}$$

Далее,

$$\begin{aligned} \lambda_1 y_{k-1} - 2 y_k + \lambda_1^{-1} y_{k+1} &= c_2 \frac{\lambda_2^{k-1}}{\lambda_1} (\lambda_1 - \lambda_2)^2 + \dots \\ &\quad \dots + c_n \frac{\lambda_n^{k-1}}{\lambda_1} (\lambda_1 - \lambda_n)^2. \end{aligned}$$

Отсюда

$$\frac{\begin{vmatrix} \lambda_1 y_{k-1} & y_k \\ y_k & \lambda_1^{-1} y_{k+1} \end{vmatrix}}{\lambda_1 y_{k-1} - 2y_k + \lambda_1^{-1} y_{k+1}} = c_1 \lambda_1^k \left[1 + O\left(\frac{\lambda_2}{\lambda_1}\right)^k \right].$$

Отношения полученных чисел, вычисленных для разных компонент, дают отношения компонент собственного вектора.

§ 55. Модификации степенного метода

1. Степенной метод со сдвигом. Известно, что собственные значения λ_i матрицы A связаны с собственными значениями μ_i матрицы $B = A - cE$ простыми соотношениями:

$$\lambda_i = \mu_i + c,$$

а собственные векторы обоих матриц совпадают. Это обстоятельство дает возможность определять собственные значения матрицы A , применяя степенной метод к матрице $A - cE$. Такое видоизменение степенного метода мы будем называть степенным методом со сдвигом. Сдвиг меняет взаимоотношение между модулями собственных значений, причем при наличии комплексных корней, даже при вещественном c , изменение этого взаимоотношения может быть довольно сложным. Если же все собственные значения вещественны, то за счет сдвига можно сделать наибольшим по модулю как алгебраически наибольшее, так и алгебраически наименьшее. Именно, (см. рис. 2) при $c < c_0 = \frac{\lambda_1 + \lambda_n}{2}$ наибольшим по модулю будет μ_1 , при $c > c_0$ наибольшим по модулю будет μ_n .

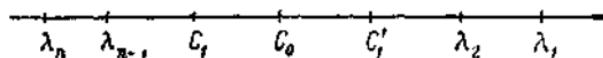


Рис. 2.

Оптимальным значением c для вычисления λ_1 является, как не трудно видеть, $c_1 = \frac{\lambda_n + \lambda_1}{2}$, так как при таком выборе c сходимость степенных итераций для определения μ_1 будет наибыстрейшей. Соответственно, оптимальным значением для вычисления λ_n будет $c'_1 = \frac{\lambda_{n-1} + \lambda_1}{2}$. Конечно, высказанные соображения имеют лишь теоретическое значение, ибо, как правило, мы не знаем, даже грубо, ни λ_1 ни λ_n . Однако все-таки они дают возможность сделать некоторые рекомендации для целесообразного выбора c . Так, если матрица A положительно-определенная, то имеет смысл сделать сдвиг при некотором положительном c и, попробовав вычислить итерации, по поведению их судить о целесообразности сделанного сдвига.

Иногда все же удается воспользоваться грубыми оценками для λ_2 и λ_n .

Сравним ход степенного процесса без сдвига с ходом степенного процесса со сдвигом, близким к оптимальному для матрицы (4) § 51.

Здесь

$$\frac{\lambda_2 + \lambda_4}{2} = \frac{0.7967 + 0.2423}{2} = 0.5185.$$

Мы проведем процесс со сдвигом, полагая $c = 0.5$.

Итерируя матрицей A вектор $(0.3, 0.5, 0.7, 0.9)'$, получим

Y_0	Y_{10}	Y_{11}	Y_{12}
0.3	$0.31005932 \cdot 10^4$	$0.72018993 \cdot 10^4$	$0.16728203 \cdot 10^5$
0.5	$0.24605904 \cdot 10^4$	$0.57153328 \cdot 10^4$	$0.13275282 \cdot 10^5$
0.7	$0.23186358 \cdot 10^4$	$0.53856100 \cdot 10^4$	$0.12509420 \cdot 10^5$
0.9	$0.27512042 \cdot 10^4$	$0.63903554 \cdot 10^4$	$0.14843190 \cdot 10^5$
<hr/>			
	$1.06310236 \cdot 10^4$	$2.46931975 \cdot 10^4$	$0.57356095 \cdot 10^5$

Для отношений компонент получим соответственно

$$\begin{array}{ll} 2.3227488 & 2.3227488 \\ 2.3227485 & 2.3227487 \\ 2.3227494 & 2.3227489 \\ 2.3227484 & 2.3227889. \end{array}$$

В последнем столбце стабилизируется уже седьмой знак. Ту же точность в отношениях компонент мы получим из 9-й и 8-й итераций матрицей $A_1 = A - 0.5E$. Действительно,

Y_0	Y_8	Y_9
0.3	$0.82648864 \cdot 10^2$	$0.15064813 \cdot 10^3$
0.5	$0.65589056 \cdot 10^2$	$0.11955238 \cdot 10^3$
0.7	$0.61805179 \cdot 10^2$	$0.11265531 \cdot 10^3$
0.9	$0.73335608 \cdot 10^2$	$0.13367238 \cdot 10^3$
<hr/>		
	$2.83378707 \cdot 10^2$	$0.51652820 \cdot 10^3$

так что для $0.5 + (y_i)_9 / (y_i)_8$ получим значения

$$\begin{array}{l} 2.3227489 \\ 2.3227489 \\ 2.3227487 \\ 2.3227486. \end{array}$$

Для определения наименьшего собственного значения оптимальным будет сдвиг на

$$\frac{\lambda_1 + \lambda_5}{2} \approx \frac{2.32 + 0.64}{2} = 1.48.$$

Приведем результат вычислений при $c = 1.5$. Именно

$1.5 + (y_i)_{20}/(y_i)_{19}$	$1.5 + (y_i)_{46}/(y_i)_{44}$
0.2434512	0.2422607
0.2492658	0.2422606
0.2406113	0.2422607
0.2409250	0.2422607.

2. Возведение матрицы в степень. Для построения высоких итераций вектора иногда целесообразно предварительно возвести данную матрицу в степень. Наиболее просто вычисляются последовательные степени матрицы A , A^2 , A^4 , A^8 , A^{16} , ... Однако возведение матрицы в квадрат, очевидно, по объему работ равносильно образованию n итераций вектора, так что вычисление матрицы A^{2^k} равносильно построению kn итераций. Соответственно, вычисление $A^{2^k}Y_0$ равносильно вычислению $kn+1$ итерации и потому выигрыш в объеме работы получается, если $kn+1 < 2^k$, т. е. если число итераций, необходимых для получения нужной точности, превосходит $n \lg_2 n$.

Можно ограничиться вычислением некоторой фиксированной степени матрицы A и затем составлять итерации посредством вычисленной степени. Например, вычислив A^8 , можно найти A^8Y_0 , затем $A^8(A^8Y_0) = A^{16}Y_0$ и, наконец, $A^{17}Y_0 = A(A^{16}Y_0)$.

Степени матрицы могут быть использованы и непосредственно для определения наибольшего по модулю собственного значения, в случае, если оно простое и вещественное.

Именно,

$$|\lambda_1| \approx \sqrt[m]{\operatorname{Sp} A^m}. \quad (1)$$

Последнее следует из того, что

$$\operatorname{Sp} A^m = \lambda_1^m + \lambda_2^m + \dots + \lambda_n^m,$$

и, следовательно,

$$\sqrt[m]{\operatorname{Sp} A^m} = \lambda_1 \sqrt[m]{1 + \left(\frac{\lambda_2}{\lambda_1}\right)^m + \dots + \left(\frac{\lambda_n}{\lambda_1}\right)^m} = \lambda_1 + O\left(\frac{1}{m} \left(\frac{\lambda_2}{\lambda_1}\right)^m\right).$$

Этот прием по существу равносителен применению метода Лобачевского к отысканию наибольшего по модулю корня характеристического уравнения матрицы.

Несколько более удобной, чем формула (1), является формула

$$\lambda_1 \approx \frac{\text{Sp}A^{2^k+1}}{\text{Sp}A^{2^k}},$$

так как дополнительное вычисление $\text{Sp}A^{2^k+1}$ эквивалентно (по объему работы) одной итерации вектора матрицей A^{2^k} . Метод следов может быть распространен на случай кратных и комплексных корней¹⁾.

§ 56. Применение степенного метода к отысканию нескольких собственных значений

В § 53 мы рассмотрели несколько случаев, когда наибольшее по модулю собственное значение не изолировано, т. е. когда имелось другое собственное значение, равное или близкое по модулю. Примененный там прием заключался в вычислении коэффициентов квадратного уравнения, корнями которого являются два наибольших по модулю собственных значения. Этот прием может быть естественным образом обобщен. Допустим, что элементарные делители матрицы взаимно просты и что $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_r| > |\lambda_{r+1}| \geq \dots \geq |\lambda_n|$. Обозначим через U_1, \dots, U_n канонический базис пространства. Допустим, что начальный вектор Y_0 взят таким, что все компоненты в его разложении по векторам канонического базиса отличны от нуля, т. е.

$$Y_0 = a_1 U_1 + a_2 U_2 + \dots + a_r U_r + a_{r+1} U_{r+1} + \dots + a_n U_n \\ a_i \neq 0 \quad (i = 1, 2, \dots, n). \quad (1)$$

Пусть $t^r + b_1 t^{r-1} + \dots + b_r = (t - \lambda_1) \dots (t - \lambda_r)$ полином, корнями которого являются $\lambda_1, \dots, \lambda_r$. Тогда, как нетрудно видеть,

$$Y_{k+r} + b_1 Y_{k+r-1} + \dots + b_r Y_k \approx 0 \quad (2)$$

с точностью (в каждой компоненте) до величин порядка $|\lambda_{r+1} + \varepsilon|^k$.

Действительно, в написанной линейной комбинации исчезают все составляющие по векторам U_1, \dots, U_r , а коэффициенты при U_{r+1}, \dots, U_n умножаются не более чем на λ_{r+1}^{k+2} , и, быть может, на величину порядка некоторой степени k (если у матрицы имеются нелинейные элементарные делители). Векторное равенство (2) равносильно системе n равенств для соответствующих компонент. Взяв какие-либо r из них, получим систему r линейных уравнений относительно b_1, \dots, b_r . Будем, для простоты, считать, что компоненты перенумерованы так, что выбранными будут r первых компонент векторов Y_{k+r}, \dots, Y_k .

¹⁾ Фрезер, Дункан и Коллар [1].

Решая полученную систему

$$\begin{aligned} y_{1k+r} + b_1 y_{1k+r-1} + \dots + b_r y_{1k} &= 0 \\ \vdots &\quad \vdots \\ y_{rk+r} + b_1 y_{rk+r-1} + \dots + b_r y_{rk} &= 0 \end{aligned} \tag{3}$$

по формулам Крамера, получим для коэффициентов b_1, \dots, b_r приближенные значения

$$b_1 = \begin{vmatrix} -y_{1,k+r} & y_{1,k+r-2} & \cdots & y_{1k} \\ \vdots & \ddots & \ddots & \vdots \\ -y_{r,k+r} & y_{r,k+r-2} & \cdots & y_{rk} \\ y_{1,k+r-1} & y_{1,k+r-2} & \cdots & y_{1k} \\ \vdots & \ddots & \ddots & \vdots \\ y_{r,k+r-1} & y_{r,k+r-2} & \cdots & y_{rk} \end{vmatrix}, \dots, b_r = \begin{vmatrix} y_{1,k+r-1} & \cdots & -y_{1,k+r} \\ \vdots & \ddots & \vdots \\ y_{r,k+r-1} & \cdots & -y_{rk+r} \\ y_{1,k+r-1} & \cdots & y_{1k} \\ \vdots & \ddots & \vdots \\ y_{r,k+r-1} & \cdots & y_{rk} \end{vmatrix}. \quad (4)$$

Можно показать, что эти равенства справедливы с точностью до величин порядка $\left(\frac{|\lambda_{r+1}| + \epsilon}{|\lambda_r|}\right)^k$. Для $k=2$ эти оценки были проведены выше.

Определив коэффициенты b_1, \dots, b_r , находим собственные значения как корни полинома $t^r + b_1 t^{r-1} + \dots + b_r$. Если среди них окажутся равные, это будет свидетельствовать о наличии нелинейных элементарных делителей у матрицы A . Заметим, что для определения коэффициентов b_1, \dots, b_r можно брать вместо r различных компонент векторов Y_{k+r}, \dots, Y_k какую-либо одну компоненту векторов $Y_{k+r}, \dots, Y_k; Y_{k+r+1}, \dots, Y_{k+1}; \dots; Y_{k+2r-1}, \dots, Y_{k+r-1}$.

При практическом вычислении нет необходимости на самом деле вычислять определители. Можно написанную систему решать численно, одним из описанных выше способов.

Заметим, что сколько-нибудь удовлетворительный результат получается, если определяемые r собственных значений близки по модулю, а следующее $(r+1)$ -е сильно отрывается от них. Если же это обстоятельство не имеет места, то полученная система (3) будет очень плохо обусловлена и определители в формулах (4) будут очень близки к нулю.

Теоретически говоря, указанным процессом можно построить весь характеристический полином (вернее минимальный анулирующий вектор Y_0 полином), приняв $r = n$. В этом случае мы приедем к методу Крылова, выполненному исходя из начального вектора $A^k Y_0$. Конечно, здесь целесообразно считать $k = 0$, чтобы не вычислять лишних итераций. К тому же с увеличением k падает обусловленность системы метода Крылова.

В случае, если $|\lambda_1| > |\lambda_2| > \dots > |\lambda_r| > |\lambda_{r+1}| \geq \dots \geq |\lambda_n|$, можно несколько изменить описанный процесс. Именно, в этом случае достаточно вычислять лишь свободные члены последовательных полиномов $(t - \lambda_1)(t - \lambda_2), \dots, (t - \lambda_1)(t - \lambda_2) \dots (t - \lambda_r)$, так как

эти числа с точностью до знака равны произведению последовательных собственных значений. Выпишем соответствующие формулы¹⁾:

$$\lambda_1 \approx \frac{y_{1k+1}}{y_{1k}}$$

$$\lambda_1 \lambda_2 \approx \frac{\begin{vmatrix} y_{1k+2} & y_{1k+1} \\ y_{2k+2} & y_{2k+1} \end{vmatrix}}{\begin{vmatrix} y_{1k+1} & y_{1k} \\ y_{2k+1} & y_{2k} \end{vmatrix}}$$

$$\dots$$

$$\lambda_1 \lambda_2 \dots \lambda_r \approx \frac{\begin{vmatrix} y_{1k+r} & \dots & y_{1k+1} \\ \dots & \dots & \dots \\ y_{rk+r} & \dots & y_{rk+1} \end{vmatrix}}{\begin{vmatrix} y_{1k+r-1} & \dots & y_{1k} \\ \dots & \dots & \dots \\ y_{rk+r-1} & \dots & y_{rk} \end{vmatrix}}.$$

Заметим, что определение произведения даже двух собственных значений наталкивается на препятствие в виде исчезновения значащих цифр, так как при достаточно большом числе итераций строки определятелей становятся почти пропорциональными (в случае, если первое собственное значение сильно отрывается от второго). Поэтому, как правило, даже второе собственное значение определяется при помощи степенного метода с гораздо меньшей степенью точности, чем первое.

В § 53 мы уже рассмотрели примеры на вычисление коэффициентов уравнений, корнями которых являются наибольшие по модулю собственные значения для $r = 2$. Мы ограничимся этими примерами и проиллюстрируем здесь лишь второй описанный прием. Именно, определим второе собственное значение для матрицы (7) § 54. Взяв $k = 18$, получим, исходя из первых двух компонент соответствующих векторов,

$$\lambda_1 \lambda_2 \approx \frac{\begin{vmatrix} 4.6792336 & 4.6779433 \\ 8.4112637 & 8.4032694 \end{vmatrix}}{\begin{vmatrix} 4.6779433 & 4.6760089 \\ 8.4032694 & 8.3912886 \end{vmatrix}} = \frac{-0.0265541}{-0.0397902} = 0.667353.$$

Исходя из вторых и четвертых компонент тех же итераций, получим

$$\lambda_1 \lambda_2 \approx 0.666110.$$

В § 54 мы вычислили, используя эти же итерации, что $\lambda_1 = -1.001$. Это дает для λ_2 значения 0.6667 или 0.6654. Точное значение $\lambda_2 = 0.66666\dots$

¹⁾ Эйткен [5].

Наконец, определим λ_2 через отношение определителей, составленных из одноименных компонент соседних итераций. Используя первые компоненты 17-й, 18-й, 19-й и 20-й итераций, получим

$$\lambda_1 \lambda_2 \approx \frac{\begin{vmatrix} 4.6792336 & 4.6779433 \\ 4.6779433 & 4.6760089 \end{vmatrix}}{\begin{vmatrix} 4.6779433 & 4.6760089 \\ 4.6760089 & 4.6731097 \end{vmatrix}} = \frac{-0.0030156}{-0.0045170} = 0.667611,$$

так что $\lambda_2 \approx 0.6669$.

§ 57. Ступенчатый степенной метод

Определение двух и более собственных значений при помощи степенного метода наталкивается, как мы только что видели, на две трудности. Это, во-первых, возможное уничтожение значащих цифр при составлении нужных линейных комбинаций и, во-вторых, отсутствие критерия, по которому можно было бы судить о достижении удовлетворительной точности. От обоих этих недостатков свободен несколько более трудоемкий ступенчатый степенной метод, к изложению которого мы сейчас переходим. Мы его разберем подробно в применении к задаче определения двух наибольших по модулю собственных значений и принадлежащих им собственных векторов (или собственного и корневого, если наибольшее по модулю собственное значение входит в канонический ящик второго порядка) и лишь коснемся его обобщения на задачу определения первых r собственных значений. Мы рассмотрим три модификации метода. При этом мы будем считать, что $|\lambda_1| \geq |\lambda_2| > |\lambda_3| \geq \dots \geq |\lambda_n|$. Для простоты изложения будем также предполагать, что собственные значения, начиная с λ_3 , имеют линейные элементарные делители.

1. Вполне стабилизирующийся ступенчатый метод. Пусть X_0 и Y_0 — два произвольных вектора. Образуем векторы AX_0 и AY_0 и построим такие их линейные комбинации X_1 и Y_1 , что первые две компоненты векторов X_1 и Y_1 , образуют матрицу

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

Для этого нужно умножить прямоугольную матрицу, составленную из компонент векторов AX_0 , AY_0 на матрицу второго порядка, обратную к матрице, составленной из первых двух компонент векторов AX_0 , AY_0 .

Это можно сделать, например, так. Построим вектор \tilde{X}_1 , поделив все компоненты вектора AX_0 на первую компоненту. Вектор \tilde{Y}_1 строится посредством вычитания из вектора AY_0 вектора \tilde{X}_1 , умноженного на первую компоненту вектора AY_0 и деления всех компо-

нент полученного вектора на его вторую компоненту. Наконец, вектор X_1 строится посредством вычитания из вектора \tilde{X}_1 вектора Y_1 , умноженного на вторую компоненту вектора \tilde{X}_1 .

Далее процесс повторяется. Именно, векторы X_k и Y_k строятся как линейные комбинации векторов $A^k X_{k-1}$ и $A^k Y_{k-1}$ такие, что их первые две компоненты образуют единичную матрицу второго порядка.

Теорема 57.1. Пусть собственные значения матрицы A удовлетворяют неравенствам

$$|\lambda_1| \geq |\lambda_2| > |\lambda_3| \geq \dots \geq |\lambda_n|.$$

Тогда, если 1) отличен от нуля определитель из первых двух компонент собственных векторов U_1 и U_2 , принадлежащих собственным значениям λ_1 и λ_2 , или собственного вектора U_1 и корневого U_2 , принадлежащих собственному значению $\lambda_1 = \lambda_2$; 2) не равен нулю определитель $c_1 d_2 - c_2 d_1$ из коэффициентов разложения

$$X_0 = c_1 U_1 + c_2 U_2 + \dots + c_n U_n$$

$$Y_0 = d_1 U_1 + d_2 U_2 + \dots + d_n U_n$$

векторов X_0 и Y_0 по собственным (корневым) векторам; 3) все определители, составленные из первых двух компонент векторов $A^k X_0$ и $A^k Y_0$ отличны от нуля, то последовательности векторов X_k и Y_k имеют пределы X и Y и эти предельные векторы лежат в инвариантном подпространстве, натянутом на векторы U_1 и U_2 .

Доказательство. Из процесса построения векторов X_k и Y_k ясно, что векторы X_k и Y_k являются линейными комбинациями векторов $A^k X_0$ и $A^k Y_0$. Пусть

$$F_k = \begin{bmatrix} x_{1k} & y_{1k} \\ x_{2k} & y_{2k} \\ \vdots & \vdots \\ \vdots & \vdots \\ x_{nk} & y_{nk} \end{bmatrix}$$

— двухстолбцевая матрица, составленная из компонент векторов $A^k X_0$ и $A^k Y_0$,

$$\Phi_k = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \xi_{3k} & \eta_{3k} \\ \vdots & \vdots \\ \vdots & \vdots \\ \xi_{nk} & \eta_{nk} \end{bmatrix}$$

— матрица, составленная из компонент векторов X_k и Y_k . Из построения следует, что матрица Φ_k получается из матрицы F_k линейным комбинированием столбцов, что равносильно умножению справа на некоторую матрицу второго порядка. В качестве этой матрицы, очевидно, нужно взять матрицу

$$\begin{bmatrix} x_{1k} & y_{1k} \\ x_{2k} & y_{2k} \end{bmatrix}^{-1},$$

так что

$$\Phi_k = F_k \begin{bmatrix} x_{1k} & y_{1k} \\ x_{2k} & y_{2k} \end{bmatrix}^{-1}.$$

Третье условие теоремы обеспечивает существование матриц $\begin{bmatrix} x_{1k} & y_{1k} \\ x_{2k} & y_{2k} \end{bmatrix}^{-1}$, т. е. обеспечивает бесконечную продолжимость процесса.

Перейдем теперь к оценкам компонент векторов X_k и Y_k или, что то же самое, к оценкам элементов матрицы Φ_k . Для этого выразим элементы матрицы Φ_k через x_{ik} и y_{ik} . Имеем

$$\begin{bmatrix} x_{1k} & y_{1k} \\ x_{2k} & y_{2k} \end{bmatrix}^{-1} = \frac{1}{\begin{vmatrix} x_{1k} & y_{1k} \\ x_{2k} & y_{2k} \end{vmatrix}} \begin{bmatrix} y_{2k} & -y_{1k} \\ -x_{2k} & x_{1k} \end{bmatrix},$$

и, следовательно,

$$\xi_{ik} = \frac{x_{ik}y_{2k} - y_{ik}x_{2k}}{x_{1k}y_{2k} - x_{2k}y_{1k}} \quad \text{при } i \geq 3 \quad (1)$$

$$\eta_{ik} = \frac{y_{ik}x_{1k} - x_{ik}y_{1k}}{x_{1k}y_{2k} - x_{2k}y_{1k}} \quad \text{при } i \geq 3. \quad (2)$$

Допустим сначала, что U_1 и U_2 — собственные векторы, отвечающие, может быть, равным собственным значениям λ_1 и λ_2 . Из разложений

$$\begin{aligned} X_0 &= c_1 U_1 + c_2 U_2 + \dots + c_n U_n \\ Y_0 &= d_1 U_1 + d_2 U_2 + \dots + d_n U_n \end{aligned} \quad (3)$$

следует

$$A^k X_0 = c_1 \lambda_1^k U_1 + c_2 \lambda_2^k U_2 + \dots + c_n \lambda_n^k U_n$$

$$A^k Y_0 = d_1 \lambda_1^k U_1 + d_2 \lambda_2^k U_2 + \dots + d_n \lambda_n^k U_n.$$

Следовательно,

$$x_{ik} = c_1 \lambda_1^k u_{1i} + c_2 \lambda_2^k u_{2i} + \dots + c_n \lambda_n^k u_{ni} \quad (4)$$

$$y_{ik} = d_1 \lambda_1^k u_{1i} + d_2 \lambda_2^k u_{2i} + \dots + d_n \lambda_n^k u_{ni}$$

и потому

$$\begin{aligned} x_{ik}y_{2k} - x_{2k}y_{ik} &= (c_1\lambda_1^k u_{1i} + c_2\lambda_2^k u_{2i} + \dots + c_n\lambda_n^k u_{ni})(d_1\lambda_1^k u_{12} + \\ &+ d_2\lambda_2^k u_{22} + \dots + d_n\lambda_n^k u_{n2}) - (c_1\lambda_1^k u_{12} + c_2\lambda_2^k u_{22} + \dots + c_n\lambda_n^k u_{n2}) \times \\ &\times (d_1\lambda_1^k u_{1i} + \dots + d_n\lambda_n^k u_{ni}) = (c_1d_2 - c_2d_1)(u_{22}u_{1i} - u_{12}u_{2i})\lambda_1^k\lambda_2^k [1 + \\ &+ O\left(\left|\frac{\lambda_3}{\lambda_2}\right|^k\right)]. \end{aligned}$$

В силу второго условия теоремы $c_1d_2 - c_2d_1 \neq 0$. Поэтому

$$\xi_{ik} = \frac{u_{1i}u_{22} - u_{2i}u_{12}}{u_{11}u_{22} - u_{12}u_{21}} + O\left(\left|\frac{\lambda_3}{\lambda_2}\right|^k\right). \quad (5)$$

В силу первого условия теоремы $u_{11}u_{22} - u_{12}u_{21} \neq 0$. Аналогично

$$\eta_{ik} = \frac{u_{11}u_{2i} - u_{21}u_{1i}}{u_{11}u_{22} - u_{12}u_{21}} + O\left(\left|\frac{\lambda_3}{\lambda_2}\right|^k\right). \quad (6)$$

В случае, если $\lambda_1 = \lambda_2$ и это собственное значение лежит в каноническом ящике второго порядка, будем иметь, что

$$\xi_{ik} = \frac{u_{1i}u_{22} - u_{2i}u_{12}}{u_{11}u_{22} - u_{12}u_{21}} + O\left(k\left|\frac{\lambda_3}{\lambda_2}\right|^k\right) \quad (5')$$

$$\eta_{ik} = \frac{u_{11}u_{2i} - u_{21}u_{1i}}{u_{11}u_{22} - u_{12}u_{21}} + O\left(k\left|\frac{\lambda_3}{\lambda_2}\right|^k\right). \quad (6')$$

Переходя к пределу в равенствах (5) и (6) (или (5') и (6')) при $k \rightarrow \infty$, получим

$$\xi_{ik} \rightarrow x_i = \frac{u_{1i}u_{22} - u_{2i}u_{12}}{u_{11}u_{22} - u_{12}u_{21}}$$

$$\eta_{ik} \rightarrow y_i = \frac{u_{11}u_{2i} - u_{21}u_{1i}}{u_{11}u_{22} - u_{12}u_{21}}.$$

Таким образом, предельные векторы (с компонентами x_i и y_i) равны

$$\begin{aligned} X &= \frac{u_{22}}{u_{11}u_{22} - u_{12}u_{21}} U_1 - \frac{u_{12}}{u_{11}u_{22} - u_{12}u_{21}} U_2 \\ Y &= -\frac{u_{21}}{u_{11}u_{22} - u_{12}u_{21}} U_1 + \frac{u_{11}}{u_{11}u_{22} - u_{12}u_{21}} U_2. \end{aligned} \quad (7)$$

Теорема доказана. В процессе доказательства даны и оценки быстроты сходимости

Замечание. Условиям 2 и 3 теоремы можно всегда удовлетворить за счет подходящего выбора начальных векторов X_0 и Y_0 . Выполнения же условия 1 можно добиться за счет изменения, в случае

надобности, нумерации компонент векторов пространства, что равносильно одновременному изменению строк и столбцов матрицы A . Нетрудно доказать далее, что если условия 1 и 2 выполнены, то условие 3 будет выполняться при достаточно больших k .

На доказанной теореме и основывается рассматриваемая модификация ступенчатого степенного метода.

Последовательность векторов X_k и Y_k строится до тех пор, пока не окажутся выполненными, с достаточной степенью точности, равенства

$$\begin{aligned} X_{k+1} &\approx X_k \\ Y_{k+1} &\approx Y_k. \end{aligned}$$

Такая стабилизация наступит, если условия теоремы 57.1 выполнены. В силу теоремы, подпространство, натянутое на предельные векторы X и Y , совпадает с инвариантным подпространством, натянутым на векторы U_1 и U_2 . Поэтому наша задача сводится теперь к решению полной проблемы собственных значений в этом двумерном подпространстве. Примем за базис этого подпространства предельные векторы X и Y . Тогда

$$\begin{aligned} AX &= \alpha X + \beta Y \\ AY &= \gamma X + \delta Y, \end{aligned}$$

так что матрица

$$L = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \quad (8)$$

есть матрица индуцированного оператора в рассматриваемом подпространстве.

Числа $\alpha, \beta, \gamma, \delta$ легко определяются. Действительно, α и β — суть первые две компоненты вектора AX , γ и δ — первые две компоненты вектора AY . Приближенные значения для них уже были нами получены как соответствующие компоненты векторов AX_k и AY_k . Искомые собственные значения совпадают с собственными значениями матрицы L . Координаты же векторов U_1 и U_2 относительно базиса X и Y равны компонентам собственных (или собственного и корневого) векторов матрицы L .

Так, если $\lambda_1 \neq \lambda_2$, легко вычислим, что

$$\begin{aligned} U_1 &= X + \frac{\lambda_1 - \alpha}{\gamma} Y \\ U_2 &= X + \frac{\lambda_2 - \alpha}{\gamma} Y. \end{aligned} \quad (9)$$

Мы не будем приводить примера, иллюстрирующего рассматриваемую модификацию ступенчатого степенного метода, так как она нами рассмотрена лишь с целью теоретического обоснования последующих модификаций, более пригодных для численного осуществления.

2. Ступенчатые итерации Баузера¹⁾. Исходя из двух векторов Z_0 и Y_0 , образуем две последовательности векторов Z_k и Y_k следующим образом. Вектор Z_{k+1} получается из вектора AZ_k делением на первую его компоненту. Вектор Y_{k+1} получается как линейная комбинация векторов AY_k и AZ_k , такая, что первая компонента вектора Y_{k+1} равна нулю, вторая равна единице. Иными словами, векторы Y_k совпадают с одноименными векторами предыдущей модификации при $X_0 = Z_0$, а векторы Z_k совпадают с нормированными к единичной первой компоненте векторами обычного степенного метода.

Поэтому в условиях предыдущей теоремы векторы Y_k стабилизируются. Поведение же векторов Z_k зависит от взаимного расположения первых двух собственных значений, которое a priori может быть неизвестным. Именно, последовательность Z_k будет стабилизироваться, если $|\lambda_1| > |\lambda_2|$, но, быть может, очень медленно, если $|\lambda_1|$ близок к $|\lambda_2|$ или если $\lambda_1 = \lambda_2$, и не будет стабилизироваться, если $|\lambda_1| = |\lambda_2|$, но $\lambda_1 \neq \lambda_2$, т. е. если $\lambda_1 = -\lambda_2$ и если λ_1, λ_2 образуют комплексную пару.

Тем не менее, используя векторы Z_k и Y_k при достаточно большом k , можно определить λ_1 и λ_2 , так же как и принадлежащие им собственные (или собственный и корневой) векторы, не сложнее, чем используя вполне стабилизирующиеся итерации X_k и Y_k предыдущей модификации.

Ясно, что (при $Z_0 = X_0$) векторы Z_k и векторы X_k связаны соотношением

$$X_k = Z_k - \varepsilon Y_k,$$

где ε_k есть вторая компонента вектора Z_k . Поэтому, при достаточно большом k , можно считать, что

$$Z_k - \varepsilon_k Y \approx X,$$

где X и Y —предельные векторы стабилизирующегося процесса, т. е. можно считать, что вектор Z_k попадает в подпространство, натянутое на векторы X и Y .

Примем векторы Z_k и Y за базис этого подпространства. Выясним, как действует матрица A на этот базис. Имеем при всех k соотношения

$$AZ_k = \rho_k Z_{k+1}$$

$$AY_k = \sigma_k Z_{k+1} + \tau_k Y_{k+1}.$$

Коэффициенты ρ_k , σ_k и τ_k легко определяются из сравнения первых двух компонент в этих соотношениях.

¹⁾ Баузэр [7].

При выбранном нами достаточно большом k мы вправе заменить векторы \bar{Y}_k и \bar{Y}_{k+1} на Y , вектор Z_{k+1} на $Z_k + (\varepsilon_{k+1} - \varepsilon_k) Y$, так что

$$\begin{aligned} AZ_k &= p_k Z_k + \rho_k (\varepsilon_{k+1} - \varepsilon_k) Y \\ AY &= \sigma_k Z_k + [\tau_k + \sigma_k (\varepsilon_{k+1} - \varepsilon_k)] Y. \end{aligned} \quad (10)$$

Следовательно, подлежащие определению собственные значения матрицы A равны собственным значениям матрицы

$$M_k = \begin{bmatrix} p_k & \sigma_k \\ \rho_k (\varepsilon_{k+1} - \varepsilon_k) & \tau_k + \sigma_k (\varepsilon_{k+1} - \varepsilon_k) \end{bmatrix}. \quad (11)$$

а собственные векторы (или собственный и корневой векторы) имеют координаты в базисе Z_k и Y , равные компонентам собственных векторов (или собственного и корневого векторов) матрицы M_k .

Легко вычислить, что характеристический полином матрицы M_k есть

$$t^2 - [p_k + \tau_k + \sigma_k (\varepsilon_{k+1} - \varepsilon_k)] t + \rho_k \tau_k. \quad (12)$$

Отметим, что если в процессе стабилизируется не только второй столбец, но и первый, то матрица M_k стабилизируется и предельная матрица M имеет треугольную форму

$$\begin{bmatrix} p & \sigma \\ 0 & \tau \end{bmatrix}.$$

3. Нестабилизирующиеся итерации. Здесь V_0 и W_0 некоторые начальные векторы, векторные последовательности строятся по формулам

$$\begin{aligned} V_{k+1} &= AV_k \\ W_{k+1} &= AW_k - \eta_k V_{k+1}, \end{aligned} \quad (13)$$

где η_k есть отношение первых компонент векторов AW_k и V_{k+1} , так что первая компонента вектора W_{k+1} равна нулю при $k \geq 0$.

Ясно, что векторы V_k и W_k лишь нормировкой отличаются от векторов Z_k и Y_k предыдущей модификации. Именно,

$$\begin{aligned} V_k &= a_k Z_k \\ W_k &= c_k Y_k, \end{aligned} \quad (14)$$

где a_k — первая компонента вектора V_k , c_k — вторая компонента вектора W_k . Поэтому векторы V_k и W_k также можно принять за базис инвариантного подпространства, натянутого на векторы U_1 и U_2 , при достаточно большом k .

Легко вывести связи между числами p_k , σ_k и τ_k предыдущей модификации с первыми компонентами векторов V_k , V_{k+1} , W_k и W_{k+1} .

Именно,

$$\rho_k = \frac{a_{k+1}}{a_k}, \quad \tau_k = \frac{c_{k+1}}{c_k}, \quad \sigma_k = \tau_k \frac{a_{k+1}}{c_k}. \quad (15)$$

Наконец, $\varepsilon_k = \frac{b_k}{a_k}$, где b_k — вторая компонента вектора V_k .

Соотношения (10) для векторов Z_k и Y_k , имеющие место при достаточно больших k , превращаются в соотношения

$$AV_k = \frac{a_{k+1}}{a_k} V_k + \frac{1}{c_k} \left(b_{k+1} - b_k \frac{a_{k+1}}{a_k} \right) W_k$$

$$AW_k = \eta_k \frac{a_{k+1}}{a_k} V_k + \left[\frac{c_{k+1}}{c_k} + \frac{\eta_k}{c_k} \left(b_{k+1} - b_k \frac{a_{k+1}}{a_k} \right) \right] W_k.$$

Поэтому искомые собственные значения матрицы A равны собственным значениям матрицы

$$N_k = \begin{bmatrix} \frac{a_{k+1}}{a_k} & \eta_k \frac{a_{k+1}}{a_k} \\ \frac{1}{c_k} \left(b_{k+1} - b_k \frac{a_{k+1}}{a_k} \right) & \frac{c_{k+1}}{c_k} + \frac{\eta_k}{c_k} \left(b_{k+1} - b_k \frac{a_{k+1}}{a_k} \right) \end{bmatrix}, \quad (16)$$

т. е. корням квадратного уравнения

$$t^2 - \left[\frac{a_{k+1}}{a_k} + \frac{c_{k+1}}{c_k} + \frac{\eta_k}{c_k} \left(b_{k+1} - b_k \frac{a_{k+1}}{a_k} \right) \right] t + \frac{a_{k+1} c_{k+1}}{a_k c_k} = 0. \quad (17)$$

Собственные векторы определяются аналогично предыдущей модификации.

В рассматриваемой модификации, при выполнении условий теоремы 57.1, векторы W_k сходятся лишь по направлению, а векторы V_k могут сходиться по направлению быстро или медленно, или совсем не сходиться.

Критерием, позволяющим окончить процесс, может служить стабилизация свободного члена уравнения (17).

В случае, если $|\lambda_1| > |\lambda_2|$, векторы V_k сходятся по направлению, так что $b_{k+1} \approx b_k \frac{a_{k+1}}{a_k}$, и потому за собственные значения можно принять $\lambda_1 = \frac{a_{k+1}}{a_k}$, $\lambda_2 = \frac{c_{k+1}}{c_k}$.

Указанная модификация несколько проще предыдущей в процессе составления последовательностей векторов, но менее удобна в заключительных операциях. Поэтому целесообразно, начав процесс в этой модификации, перейти затем соответствующим нормированием к предыдущей.

В качестве примера определим первые два собственных значения матрицы Леверье и принадлежащие им собственные векторы при помощи третьей модификации ступенчатого степенного метода.

Возьмем $V_0 = (1, 1, 1, 1)^T$, $W_0 = (0, 1, 1, 1)^T$. Тогда

$V_1 = AV_0$	AW_0	W_1	V_{17}	W_{17}
-3.208074	2.301808	-0.00000001	-0.10419891 · 10 ⁹	0
-5.753172	-6.041037	-10.168965	0.72583945 · 10 ⁹	-0.15619598 · 10 ⁹
-8.384229	-8.433328	-14.449051	-0.37864164 · 10 ⁹	0.57277646 · 10 ⁹
-15.922752	-15.928987	-27.353636	-0.11545164 · 10 ¹¹	-0.14996613 · 10 ⁹
-33.268227	-28.101544			
-0.71750365			-0.91870763 · 10	
V_{17}	W_{17}	V_{18}	W_{18}	
0.16696157 · 10 ¹⁰	0	-0.26335495 · 10 ¹¹	0	
-0.1114017 · 10 ¹¹	0.29882041 · 10 ⁹	0.17832314 · 10 ⁹	-0.57619870 · 10 ¹¹	
0.53854629 · 10 ¹¹	-0.10951099 · 10 ¹¹	-0.72561121 · 10 ¹¹	0.21119162 · 10 ¹¹	
0.20281721 · 10 ¹²	0.28675616 · 10 ¹¹	-0.35613659 · 10 ¹⁰	-0.55281857 · 10 ¹¹	
-0.10903620 · 10		-0.13126160 · 10		

Собственные значения матрицы N_{17} определяются из уравнения

$$t^2 + 35.015145t + 306.38882 = 0,$$

собственные значения матрицы N_{18} из уравнения

$$t^2 + 35.015455t + 306.39423 = 0.$$

Мы видим, что процесс достаточно стабилизировался.

Для собственных значений получаем

$$\lambda_1 = -17.8629, \quad \lambda_2 = -17.1522 \quad \text{при } k = 17$$

или

$$\lambda_1 = -17.8631, \quad \lambda_2 = -17.1523 \quad \text{при } k = 18.$$

Для определения собственных векторов, принадлежащих найденным собственным значениям, найдем предварительно (при $k = 18$) собствен-

ные векторы матрицы N_k . Нетрудно подсчитать, что $X_1 = (1, -9.4429)', X_2 = (1, -6.0357)'$. Теперь имеем

$$\begin{array}{lll} U_1 = V_{18} - 9.4429 W_{18} & \tilde{U}_1 & U_2 = V_{18} - 6.0357 W_{18} \\ 0.166962 \cdot 10^{10} & -0.019861 & 0.166962 \cdot 10^{10} \\ -0.142657 \cdot 10^{11} & 0.169702 & -0.132477 \cdot 10^{11} \\ 0.157293 \cdot 10^{11} & -0.187112 & 0.119970 \cdot 10^{11} \\ -0.679638 \cdot 10^{11} & 0.808482 & 0.297402 \cdot 10^{11} \end{array} \quad \begin{array}{l} \tilde{U}_3 \\ 0.032937 \\ -0.261341 \\ 0.236668 \\ 0.586694. \end{array}$$

Все три модификации двухступенчатого степенного метода могут быть обобщены в форме r -ступенчатого степенного метода, позволяющего уже, вообще говоря, определять r собственных значений и принадлежащих им векторов канонического базиса. Первая модификация дает стабилизирующийся процесс, если $|\lambda_r| > |\lambda_{r+1}|$. Применение первой модификации позволяет построить инвариантное подпространство, натянутое на векторы U_1, \dots, U_r канонического базиса, соответствующие собственным значениям $\lambda_1, \dots, \lambda_r$. Тем самым применение первой модификации сводит частичную проблему для матрицы n -го порядка к решению полной проблемы для матрицы порядка r . При $r = n$, очевидно, первая модификация теряет содержательность. Поэтому она может применяться лишь при r значительно меньших n .

Иначе обстоит дело со второй модификацией и мало отличающейся от нее третьей. Здесь при $r = n$ мы приходим к так называемому треугольному степенному методу, который решает полную проблему собственных значений для матрицы и будет нами изложен в § 78 гл. VIII. При больших r , в частности при $r = n$, метод дает хорошие результаты лишь в случае, если все собственные значения, подлежащие определению, вещественны и различны. Наличие комплексных корней и собственных значений, принадлежащих ящикам Жордана высших порядков, сильно затрудняет проведение процесса, в чем мы убедились даже при рассмотрении случая $r = 2$.

§ 58. Метод λ -разности

Метод λ -разности дает возможность, зная наибольшее по модулю собственное значение λ_1 , находить следующее собственное значение λ_2 и принадлежащий ему собственный вектор при условии, что

$$|\lambda_1| > |\lambda_2| > |\lambda_3|.$$

Метод состоит в следующем.

Пусть вычислена последовательность

$$y_1, y_2, \dots, y_m, \dots, y_k, \dots \quad (1)$$

и из нее определено $\lambda_1 \approx \frac{y_{k+1}}{y_k}$. Здесь y_k — любая компонента вектора $Y_k = A^k Y_0$.

Образуем разность

$$\Delta y_k = y_{k+1} - \lambda_1 y_k = c_2(\lambda_2 - \lambda_1) \lambda_2^k + \dots + c_n(\lambda_n - \lambda_1) \lambda_n^k. \quad (2)$$

Так как λ_2 по модулю больше, чем все остальные собственные значения, и $c_2 \neq 0$, то первый член этой разности будет преобладать, и мы сможем определить λ_2 аналогично тому, как мы определяли λ_1 . Именно,

$$\lambda_2 \approx \frac{y_{k+1} - \lambda_1 y_k}{y_k - \lambda_1 y_{k-1}}. \quad (3)$$

Однако при таком определении λ_2 мы наталкиваемся на исчезновение значащих цифр, так как в числителе и знаменателе отношения (3) нам приходится вычитать величины, близкие друг другу. Практически целесообразно, найдя λ_1 из отношения y_{k+1} и y_k , вернуться назад и определить λ_2 из отношения

$$\lambda_2 \approx \frac{y_{m+1} - \lambda_1 y_m}{y_m - \lambda_1 y_{m-1}}, \quad m < k, \quad (4)$$

беря в качестве m наименьшее из чисел, при котором преобладание λ_2 над следующими собственными числами уже начинает сказываться. Указанный прием дает для λ_2 довольно грубые значения, однако часто достаточные для нужд практики. Теоретически возможно при помощи аналогичного процесса определять и следующие собственные числа.

Очевидно, что для определения второго собственного вектора процесс составления λ -разности надо произвести в последовательности $AY_0, A^2Y_0, \dots, A^kY_0, \dots$ Действительно, разность

$$A^{k+1}Y_0 - \lambda_1 A^k Y_0 = a_2(\lambda_2 - \lambda_1) \lambda_2^k X_2 + \dots + a_n(\lambda_n - \lambda_1) \lambda_n^k X_n$$

показывает, что компоненты вектора X_2 могут быть найдены аналогично тому, как мы определяли компоненты вектора X_1 в § 53.

Для примера определим второе собственное число матрицы (7) § 54.

В качестве λ_1 будем брать как значение, полученное непосредственно из отношений компонент 20-й и 19-й итераций ($\lambda_1 \approx 1.001$), так и уточненное при помощи скалярного произведения значение ($\lambda_1 \approx 1.000000$).

Принимая за y_k первую компоненту вектора $A^k Y_0$, получим (при $\lambda_1 \approx 1.000000$), учитывая 17-ю, 18-ю и 19-ю итерации ($m = 18$):

$$\lambda_2 \approx \frac{y_{m+1} - \lambda_1 y_m}{y_m - \lambda_1 y_{m-1}} \approx \frac{4.677943 - 4.676009}{4.676009 - 4.673110} = \frac{0.001934}{0.002899} = 0.6671.$$

Аналогично, принимая за y_k четвертую компоненту вектора $A^k Y_0$ получим

$$\lambda_2 \approx \frac{-0.009131}{-0.013667} \approx 0.6681.$$

Таким образом, знание достаточно точного значения λ_1 дало возможность определить и λ_2 относительно точно (три знака после запятой) (точно $\lambda_2 = 0.666 \dots$).

Если в качестве λ_1 мы возьмем более грубое значение $\lambda_1 \approx 1.001$, то, вычисляя прежнее отношение, мы столкнемся с описанным явлением исчезновения значащих цифр. В этом случае в качестве m надо взять число значительно меньшее чем 20.

Так, рассматривая 9-ю, 10-ю и 11-ю итерации вектора $A^k Y_0$

$A^9 Y_0$	$A^{10} Y_0$	$A_{11} Y_0$
4.4665336	4.5365193	4.5841480
7.1243407	7.5407651	7.8281626
0.0699857	0.0476287	0.0321941
— 7.4707539	— 7.7678185	— 7.9777169
4.1901061	4.3570946	4.4667878

получим, вычисля величины $y_{m+1} - \lambda y_m$:

$m = 9$	$m = 10$
0.06552	0.04309
0.40930	0.27986
— 0.02242	— 0.01548
— 0.28959	— 0.20213.

Отношения этих величин дают для λ_2 значения

0.658
0.684
0.690
0.698.

Таким образом, знание весьма грубого значения для λ_1 позволило нам все же, используя ранние итерации, получить для λ_2 значение, верное с точностью до трех единиц второго знака.

Приближенные значения для компонент второго собственного вектора мы можем получить как соответствующие отношения компонент вектора $A^{m+1} Y_0 - \lambda_1 A^m Y_0$. Взяв $m = 9$, мы получим, используя ранее вычисленные компоненты вектора $A^{10} Y_0 - \lambda_1 A^9 Y_0$, следующие значения для компонент собственного вектора (после нормирования):

$$1.00; 6.49; -0.36; -4.69.$$

Для рассматриваемой матрицы второй собственный вектор имеет компоненты

$$1, \frac{31}{5} = 6.2, -\frac{1}{3} = -0.333 \dots, -\frac{71}{15} = -4.733 \dots,$$

так что результаты вычисления достаточно хорошо согласуются с точными данными.

Метод λ -разности может быть обобщен следующим образом. Пусть известны λ_1 и λ_2 и требуется определить λ_3 , причем известно также, что $|\lambda_3| > |\lambda_4|$. В этом случае мы составляем „вторую $\lambda_1\lambda_2$ -разность“, т. е.

$$\begin{aligned}\Delta^2 y_k = (y_{k+2} - \lambda_1 y_{k+1}) - \lambda_2 (y_{k+1} - \lambda_1 y_k) = \\ = y_{k+2} - (\lambda_1 + \lambda_2) y_{k+1} + \lambda_1 \lambda_2 y_k.\end{aligned}$$

Легко видеть, что

$$\Delta^2 y_k = c_3 (\lambda_3 - \lambda_1) (\lambda_3 - \lambda_2) \lambda_3^k + \dots + c_n (\lambda_n - \lambda_1) (\lambda_n - \lambda_2) \lambda_n^k$$

и, при достаточно большом k ,

$$\lambda_3 \approx \frac{\Delta^2 y_{k+1}}{\Delta^2 y_k}.$$

Пропадание знаков при составлении второй $\lambda_1\lambda_2$ -разности будет еще значительнее, чем при использовании первой λ_1 -разности, так что в случае вещественных λ_1 и λ_2 метод почти не применим. Относительно хорошие результаты получаются лишь в случае, когда λ_1 и λ_2 образуют комплексную пару. В этом случае вторую $\lambda_1\lambda_2$ -разность целесообразно искать в виде

$$\Delta^2 y_k = y_{k+2} + p y_{k+1} + q y_k,$$

где p и q коэффициенты квадратного трехчлена $(t - \lambda_1)(t - \lambda_2)$ и могут быть вычислены по формулам (19) § 53.

Более того, если за простым вещественным корнем λ_1 или комплексной парой λ_1 и λ_2 следует комплексная пара, то используя λ -разности вместо итераций в формулах (19) § 53, мы сможем получить коэффициенты квадратного трехчлена $(t - \lambda_2)(t - \lambda_3)$ или $(t - \lambda_3)(t - \lambda_4)$ соответственно.

§ 59. Метод исчерпывания

Методы исчерпывания и понижения (§ 60) дают возможность определять последующее собственное значение и принадлежащий ему собственный вектор, после того как предшествующие собственные значения и принадлежащие им собственные векторы известны с достаточной степенью точности. В отличие от методов § 56 и § 58, направленных к той же цели, применение методов исчерпывания и понижения не влечет потери точности. Поэтому эти методы дают возможность, в частности, решить полную проблему собственных значений при помощи цепочки решений частичных проблем.

Для простоты изложения будем предполагать, что все собственные значения матрицы A вещественны.

Для проведения одного шага метода исчерпывания для матрицы A нужно знать предварительно не только какое-либо ее собственное значение λ_1 (не обязательно наибольшее по модулю) и принадлежащий ему собственный вектор $U_1 = (u_1, u_2, \dots, u_n)'$, но также и собственный вектор $V_1 = (v_1, v_2, \dots, v_n)'$ матрицы A' , принадлежащий собственному значению λ_1 . Будем предполагать, кроме того, что все собственные значения матрицы A попарно различны, так что существуют базисы U_1, \dots, U_n и V_1, \dots, V_n , состоящие из собственных векторов матриц A и A' , удовлетворяющих условиям нормированности $(U_i, V_i) = 1$ при $i = 1, 2, \dots, n$.

Составим матричное произведение $U_1 V_1'$, где V_1' строка, составленная из компонент вектора V_1 .

Это будет квадратная матрица

$$U_1 V_1' = \begin{bmatrix} u_1 v_1 & u_1 v_2 & \dots & u_1 v_n \\ u_2 v_1 & u_2 v_2 & \dots & u_2 v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n v_1 & u_n v_2 & \dots & u_n v_n \end{bmatrix}. \quad (1)$$

Заметим, что матричное произведение $V_1' U_1$ равно числу 1, так как оно равно скалярному произведению (U_1, V_1) .

Далее образуем матрицу

$$A_1 = A - \lambda_1 U_1 V_1'.$$

Докажем, что матрица A_1 обладает теми же собственными числами и векторами, что и матрица A , за исключением первого собственного числа, вместо которого появляется собственное число, равное нулю. Действительно,

$$A_1 U_1 = A U_1 - \lambda_1 (U_1 V_1') U_1 = A U_1 - \lambda_1 U_1 (V_1' U_1) = A U_1 - \lambda_1 U_1 = 0$$

$$A U_i = A U_i - \lambda_1 (U_1 V_1') U_i = A U_i - \lambda_1 U_1 (V_1' U_i) = \lambda_i U_i,$$

так как $V_1' U_1 = 1$, $V_1' U_i = (V_1, U_i) = 0$ в силу ортогональных свойств векторов U_1, U_2, \dots, U_n и V_1, V_2, \dots, V_n .

Указанное свойство матрицы A дает возможность, исходя из векторной последовательности $A_1 Y_0, \dots, A_1^m Y_0, \dots$ определить λ_2 и U_2 аналогично тому, как мы определяли λ_1 и U_1 из последовательности $A Y_0, \dots, A^k Y_0, \dots$, так как собственное число λ_2 будет первым собственным числом для матрицы A_1 . Мы будем называть этот процесс процессом исчерпывания.

Покажем, что

$$A_1^m Y_0 = A^m Y_0 - \lambda_1^m U_1 V_1' Y_0, \quad (2)$$

т. е. что для практического применения указанного процесса нет надобности вычислять матрицу A_1 на самом деле и образовывать

ряд векторов $A_1 Y_0, \dots, A_1^m Y_0, \dots$, а достаточно вычислить лишь два соседних вектора $A_1^{m+1} Y_0$ и $A_1^m Y_0$ по формуле (2).

Для установления равенства (2) введем так называемое билинейное разложение матрицы A .

На основании ортогональных свойств системы собственных векторов матрицы A и ее транспонированной верно матричное равенство

$$E = U_1 V'_1 + U_2 V'_2 + \dots + U_n V'_n.$$

Умножая это равенство слева на A и заменяя AU_i на $\lambda_i U_i$ ($i=1, 2, \dots, n$), получим, что

$$A = \lambda_1 U_1 V'_1 + \lambda_2 U_2 V'_2 + \dots + \lambda_n U_n V'_n.$$

Процесс исчерпывания уничтожает первое слагаемое в этом разложении, так что

$$A_1 = \lambda_2 U_2 V'_2 + \dots + \lambda_n U_n V'_n.$$

Далее

$$A^m = \lambda_1^m U_1 V'_1 + \lambda_2^m U_2 V'_2 + \dots + \lambda_n^m U_n V'_n.$$

Аналогично

$$A_1^m = \lambda_2^m U_2 V'_2 + \dots + \lambda_n^m U_n V'_n.$$

Поэтому

$$A_1^m = A^m - \lambda_1^m U_1 V'_1,$$

откуда вытекает равенство (2).

Таким образом, применять метод исчерпывания можно в двух вариантах. В одном из них надо вычислить вектор $U_1 V'_1 Y_0$, образовать векторы $A_1^{m+1} Y_0$ и $A_1^m Y_0$ по формуле (2) и затем определить λ_2 и U_2 обычным для степенного метода образом.

В этом варианте происходит значительное уничтожение значащих цифр и потому в качестве m приходится брать число, значительно меньшее, чем число итераций, применявшихся для определения числа λ_1 и компонент собственных векторов U_1 и V_1 . Точность при использовании этого варианта получается невысокой.

Второй вариант заключается в фактическом построении матрицы A_1 и вычислении итераций посредством A_1 . Этот вариант требует большего объема вычислений, но обеспечивает значительно лучшие результаты в смысле точности.

В обоих вариантах возможно применение приемов улучшения сходимости.

Пример. Рассмотрим снова матрицу (7) § 54.

Метод исчерпывания требует для определения второго собственного числа знания как первого собственного числа, так и принадлежащих ему собственных векторов как матрицы A , так и матрицы A' . Поэтому при пользовании этим методом необходимо наряду

с вычислением последовательности итераций $A^k Y_0$ вычислять и последовательность итераций $A'^k Y_0$. Таким образом, при определении λ_1 мы всегда можем уточнить значение λ_1 при помощи метода скалярного произведения.

В нашем примере, используя двадцать итераций вектора $Y_0 = (1, 1, 1, 1)^T$ матрицей A и матрицей A' , мы получили для λ_1 значение $\lambda_1 = 1.000000$ (см. п. 1, § 54).

Для компонент собственных векторов матриц A и A' мы получаем, нормируя компоненты векторов $A^{20}Y_0$ и $A'^{20}Y_0$ значения:

$$\begin{array}{cc} 1.00000 & 1.00000 \\ 1.79757 & -0.00027 \\ 0.00018 & 5.01676 \\ -1.79838 & -0.14256 \end{array}$$

Точные значения компонент первого собственного вектора матрицы A суть 1, 1.8, 0, -1.8. Следуя теории, нужно прежде всего нормировать векторы U_1 и Y_1 так, чтобы $(U_1, U_1) = 1$. Вычисляя множитель нормирования, получим $c = 0.795678$. Таким образом, для компонент первых собственных векторов матриц A и A' мы получим значения:

$$\begin{array}{cc} 1.00000 & 0.79568 \\ 1.79757 & -0.00021 \\ 0.00018 & 3.99173 \\ -1.79838 & -0.11343 \end{array}$$

Теперь мы можем образовать матричное произведение $U_1 V'_1$. Именно:

$$U_1 V'_1 = \begin{bmatrix} 0.79568 & -0.00021 & 3.99173 & -0.11343 \\ 1.43029 & -0.00038 & 7.17541 & -0.20390 \\ 0.00014 & 0 & 0.00072 & -0.00002 \\ -1.43093 & 0.00038 & -7.17865 & 0.20399 \end{bmatrix}.$$

Далее образуем матрицу A_1 .

$$A_1 = A - \lambda_1 U_1 V'_1 =$$

$$= \begin{bmatrix} 0.20432 & 0.00021 & -2.99173 & 0.11343 \\ -0.43029 & 0.77816 & -6.84208 & 0.53723 \\ -0.00014 & -0.02525 & 0.55484 & -0.02523 \\ 1.43093 & -0.88927 & -1.46579 & -0.09288 \end{bmatrix}.$$

и образуем итерации вектора $Y_0 = (1, 1, 1, 1)^T$ этой матрицей.

Приведем 17-ю и 18-ю итерации матрицей A_1 .

$A_1^{17}Y_0$	$A_1^{18}Y_0$
-0.00869	-0.00580
-0.05388	-0.03597
0.00290	0.00193
0.04107	0.02741
<hr/>	<hr/>
-0.01861	-0.01242.

Отношения компонент 17-й и 18-й итераций дадут для λ_2 значения:

0.667
0.668
0.666
0.667.

Из 18-й итерации получим посредством нормирования к единичной первой компоненте $U_2 = (1.00, 6.20, -0.333, -4.73)'$.

Мы видим, что как само второе собственное число λ_2 , так и компоненты второго собственного вектора определяются методом исчерпывания более точно, чем методом λ -разности. Однако этот метод требует в случае несимметричной матрицы много дополнительной работы. В случае симметричной матрицы метод исчерпывания может быть рекомендован.

При вычислении второго собственного числа и компонент второго собственного вектора можно пользоваться описанной модификацией метода исчерпывания, при которой вектор $A_1^k Y_0$ вычисляется, минуя итерации матрицей A_1 , по формуле (2). Вычисляя, получим

$$U_1 V_1' Y_0 = (4.67377, 8.40142, 0.00084, -8.40521)'.$$

Теперь вычисляем $A_1^k Y_0$ по формуле (2) при $k = 9$ и 10. Это дает:

$A_1^9 Y_0$	$A_1^{10} Y_0$
-0.20724	-0.13725
-1.27708	-0.86065
0.06915	0.04679
0.93446	0.63739.

Отношения компонент равны:

0.662

0.674

0.677

0.682.

так что $\lambda_2 \approx 0.67$ с точностью до 10^{-2} . Далее, нормируя $A_1^{10}Y_0$, получим $U_2 = (1.00, 6.27, -0.34, -4.64)'$.

Метод исчерпывания почти без изменений может быть перенесен на матрицы с комплексными элементами и вещественные матрицы с комплексными собственными значениями. При его обосновании нужно рассмотреть вместо строк V_i' строки V_i^* , составленные из чисел комплексно-сопряженных с компонентами собственных векторов V_i матрицы A^* , принадлежащих собственным значениям $\bar{\lambda}_i$.

При применении метода исчерпывания в случае вещественной матрицы, для которой определена пара комплексно-сопряженных значений с соответствующими им собственными векторами, следует проводить процесс исчерпывания по формуле

$$A_1 = A - \lambda_1 U_1 V_1^* - \overline{\lambda_1 U_1 V_1^*} = A - 2\operatorname{Re}(\lambda_1 U_1 V_1^*).$$

Это позволяет оставаться в классе вещественных матриц.

§ 60. Метод понижения

Пусть для матрицы A вычислено первое собственное значение λ_1 и принадлежащий ему собственный вектор $U_1 = (u_1, \dots, u_n)'$. Рассмотрим матрицу

$$P = \begin{bmatrix} u_1 & 0 & \dots & 0 \\ u_2 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ u_n & 0 & \dots & 1 \end{bmatrix}.$$

Нетрудно проверить, что

$$P^{-1} = \begin{bmatrix} \frac{1}{u_1} & 0 & \dots & 0 \\ -\frac{u_2}{u_1} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{u_n}{u_1} & 0 & \dots & 1 \end{bmatrix},$$

Матрица $P^{-1}AP$ подобна матрице A и потому собственные числа обеих матриц одинаковы.

Но

$$\begin{aligned} P^{-1}AP &= \begin{bmatrix} \lambda_1 & \frac{a_{12}}{u_1} & \dots & \frac{a_{1n}}{u_1} \\ 0 & a_{22} - \frac{u_2}{u_1} a_{12} & \dots & a_{2n} - \frac{u_2}{u_1} a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} - \frac{u_n}{u_1} a_{12} & \dots & a_{nn} - \frac{u_n}{u_1} a_{1n} \end{bmatrix} = \\ &= \begin{bmatrix} \lambda_1 & b_{12} & \dots & b_{1n} \\ 0 & \ddots & & \vdots \\ \vdots & & B & \\ 0 & & & \end{bmatrix}. \end{aligned}$$

Таким образом,

$$|P^{-1}AP - tE| = (\lambda_1 - t) |B - tE|.$$

Следовательно, собственными значениями матрицы A кроме λ_1 будут собственные значения матрицы $n-1$ -го порядка B . Если нормировать вектор U_1 так, что $u_1 = 1$, то мы будем иметь

$$B = \begin{bmatrix} a_{22} - u_2 a_{12} & \dots & a_{2n} - u_2 a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n2} - u_n a_{12} & \dots & a_{nn} - u_n a_{1n} \end{bmatrix}.$$

Для нахождения λ_2 нам, очевидно, нужно построить последовательность BY_0, \dots, B^kY_0 и находить λ_2 как отношения любых компонент построенных векторов.

Далее, пусть Z некоторый собственный вектор матрицы B соответствующий собственному значению λ . Тогда матрица $P^{-1}AP$ будет иметь собственный вектор $\begin{bmatrix} z_1 \\ Z \end{bmatrix}$. Определим z_1 .

Имеем

$$\begin{bmatrix} \lambda_1 & a_{12} & \dots & a_{1n} \\ 0 & \ddots & & \vdots \\ \vdots & & B & \\ 0 & & & \end{bmatrix} \begin{bmatrix} z_1 \\ Z \end{bmatrix} = \begin{bmatrix} \lambda_1 z_1 + a_{12} z_2 + \dots + a_{1n} z_n \\ BZ \end{bmatrix} = \lambda \begin{bmatrix} z_1 \\ Z \end{bmatrix}.$$

Приравнивая первые компоненты в этом векторном равенстве, получим

$$\lambda_1 z_1 + a_{12} z_2 + \dots + a_{1n} z_n = \lambda z_1,$$

откуда

$$z_1 = \frac{a_{12} z_2 + \dots + a_{1n} z_n}{\lambda - \lambda_1}.$$

Наконец, собственный вектор U матрицы A определится по формуле

$$U = P \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} z_1 \\ u_2 z_1 + z_2 \\ \vdots \\ u_n z_1 + z_n \end{bmatrix}.$$

Пример. Определим второе собственное число и компоненты принадлежащего ему собственного вектора для матрицы (7) § 54. Как мы видели, для этой матрицы было определено первое собственное число $\lambda \approx 1.001$ (§ 54, п. 1) и компоненты принадлежащего ему собственного вектора 1, 1.79757, 0.00018, -1.79838 (§ 59).

Для определения λ_2 вычислим матрицу

$$B = \begin{bmatrix} 0.77778 & -1.46424 & 0.33333 \\ -0.02525 & 0.55538 & -0.02525 \\ -0.88889 & -6.84606 & 0.11111 \end{bmatrix}.$$

Далее образуем итерации вектора $Y_0 = (1, 1, 1)'$ матрицей B . Приведем результат 15-й и 16-й итераций:

$$\begin{array}{rcc} B^{15} Y_0 & & B^{16} Y_0 \\ \hline -0.09565 & & -0.06388 \\ 0.00725 & & 0.00484 \\ 0.06339 & & 0.04243 \\ \hline -0.02502 & & -0.01661 \end{array}$$

Определим λ_2 из отношения компонент 16-й и 15-й итераций. Это дает:

$$\begin{array}{c} 0.668 \\ 0.668 \\ 0.669. \end{array}$$

Далее определим компоненты второго собственного вектора, считая $Z = B^{16} Y_0$. Для этого сначала определим

$$z_1 = \frac{a_{12} z_2 + a_{13} z_3 + a_{14} z_4}{\lambda_2 - \lambda_1} = \frac{0.00484}{-0.333} \approx -0.0145.$$

Далее вычисляем компоненты u_2 , u_3 , u_4 вектора U_2 : $u_2 = -0.0900$, $u_3 = 0.00484$, $u_4 = 0.0685$. Таким образом, компоненты вектора U_2 суть -0.0145 ; -0.0900 ; 0.00484 ; 0.0685 , так что после нормирования $U_2 = (1, 6.21, -0.333, -4.72)'$.

§ 61. Координатная релаксация

Метод, описанный в этом параграфе, применим только в случае, если матрица A симметрична. В некотором роде он близок к итерационным методам для решения линейных систем, основанным на релаксации того или другого функционала. В данном случае роль такого функционала играет отношение Релея.

Вычислим прежде всего, как изменяется отношение $\mu(X) = \frac{(AX, X)}{(X, X)}$ при изменении X в определенном направлении. Пусть

$$X' = X + \alpha Y, \quad (1)$$

где Y — некоторый фиксированный вектор, определяющий направление изменения вектора X . Тогда

$$\begin{aligned} (AX', X') &= (AX + \alpha AY, X + \alpha Y) = (AX, X) + \alpha(AX, Y) + \\ &+ \alpha(AY, X) + \alpha^2(AY, Y) = (AX, X) + 2\alpha(AX, Y) + \alpha^2(AY, Y) \end{aligned}$$

и, следовательно,

$$\mu(X') = \frac{(AX', X')}{(X', X')} = \frac{(AX, X) + 2\alpha(AX, Y) + \alpha^2(AY, Y)}{(X, X) + 2\alpha(X, Y) + \alpha^2(Y, Y)}. \quad (2)$$

Подберем теперь множитель α так, чтобы отношение $\mu(X')$ достигало наибольшего значения. Вычисляя производную $\mu(X')$ по α , получим

$$\begin{aligned} \frac{d\mu(X')}{d\alpha} &= \frac{[2(AX, Y) + 2\alpha(AY, Y)][(X, X) + 2\alpha(X, Y) + \alpha^2(Y, Y)] - }{[(X, X) + 2\alpha(X, Y) + \alpha^2(Y, Y)]^2} \\ &- \frac{[2(X, Y) + 2\alpha(Y, Y)][(AX, X) + 2\alpha(AX, Y) + \alpha^2(AY, Y)]}{[(X, X) + 2\alpha(X, Y) + \alpha^2(Y, Y)]^2} \end{aligned}$$

и потому для определения α будем иметь уравнение

$$\begin{aligned} [(AX, Y)(Y, Y) - (AY, Y)(X, Y)]\alpha^2 + [(AX, X)(Y, Y) - (AY, Y)(X, X)]\alpha + \\ + (AX, X)(X, Y) - (AX, Y)(X, X) = a\alpha^2 + b\alpha + c = 0. \quad (3) \end{aligned}$$

Исследование уравнения (3) показывает, что его корни всегда вещественны и различны. Правда, может случиться, что коэффициент при α^2 обращается в нуль. Тогда следует условно считать, что один из корней уравнения равен ∞ . Это будет, если один из экстремумов достигается на векторе Y . Вторым исключением является случай, когда $\mu(X')$ оказывается не зависящим от α . В этом случае

уравнение (3) превращается в тождество $0 = 0$. Геометрический смысл сказанного очень прост. Для пояснения мы предположим, что матрица A положительно определена, что не нарушает общности, так как любую симметричную матрицу можно сделать положительно-определенной за счет добавления скалярной матрицы, причем при таком преобразовании собственные векторы не меняются, а все значения отношения Релея (в частности, все собственные значения) изменяются на постоянное слагаемое. В этом предположении отношение

Релея $\mu(X) = \frac{(AX, X)}{(X, X)}$ есть, очевидно, $\frac{1}{\rho^2}$,

где ρ — длина вектора, исходящего из начала координат в направлении вектора X до пересечения с эллипсоидом $(AX, X) = 1$. Векторы $X' = X + \alpha Y$ лежат в плоскости (двумерном подпространстве), натянутой на векторы X и Y . Эта плоскость пересекает эллипсоид $(AX, X) = 1$ по эллипсу, причем отношение Релея достигает максимума на векторе, направленном по малой оси этого эллипса и минимума на векторе, направленном по большой оси эллипса (см. рис. 3). Может случиться что вектор Y направлен по одной из осей этого эллипса. Вот именно в этом случае один из корней уравнения (3) обращается в бесконечность. Наконец, может случиться, что эллипс вырождается в окружность.

Тогда уравнение (3) обращается в тож-

дество, и α может быть взято произвольным. Заметим сразу, что если при некоторых X и Y получается не круговое сечение и мы сделаем достаточно малую деформацию векторов X и Y , то направление осей эллипса изменится сколь угодно мало.

Несложное вычисление дает, что экстремальное значение μ' равно

$$\frac{\delta + \alpha\alpha}{d}, \quad (4)$$

где $d = (X, X)(Y, Y) - (X, Y)^2$, $\delta = (AX, X)(Y, Y) - (AX, Y)(X, Y)$ и $a = (AX, Y)(Y, Y) - (AY, Y)(X, Y)$ есть коэффициент при α^2 в уравнении (3). Так как $d > 0$, то из (4) следует, что для получения максимума μ' нужно брать больший корень уравнения (3) при $\alpha > 0$ и меньший корень уравнения при $\alpha < 0$.

Координатный релаксационный метод заключается в том, что за вектор Y на каждом шагу процесса берется один из координатных векторов e_i .

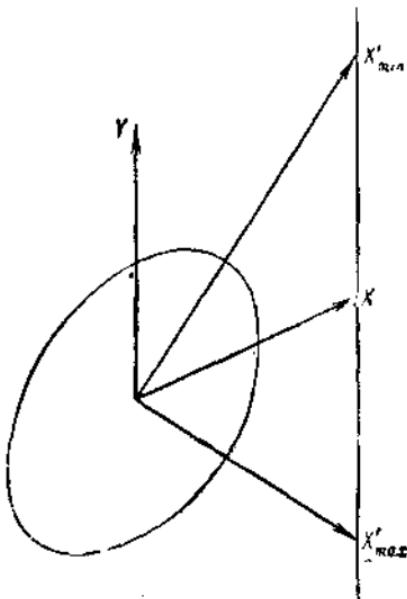


Рис. 3.

Пусть $X = (x_1, \dots, x_n)^T$. Введем обозначение

$$\begin{aligned} F &= AX = (f_1, \dots, f_n)^T \\ p &= (AX, X) \\ q &= (X, X). \end{aligned} \quad (5)$$

Тогда

$$\begin{aligned} (X, AY) &= (X, Ae_i) = (AX, e_i) = f_i \\ (AY, Y) &= (Ae_i, e_i) = a_{ii} \\ (Y, Y) &= (e_i, e_i) = 1 \\ (X, Y) &= (X, e_i) = x_i. \end{aligned} \quad (6)$$

Квадратное уравнение для определения α перейдет в уравнение

$$\alpha^2 [f_i - a_{ii}x_i] + \alpha [p - a_{ii}q] + px_i - f_iq = 0. \quad (7)$$

Выбрав соответствующий корень, определим следующее приближение

$$X' = X + \alpha e_i. \quad (8)$$

Для этого приближения будем иметь

$$F' = AX' = AX + \alpha Ae_i = F + \alpha A_i, \quad (9)$$

где A_i есть i -й столбец матрицы A . Далее

$$\begin{aligned} p' &= (AX', X') = (F', X') = p + 2\alpha f_i + \alpha^2 a_{ii}, \\ q' &= (X', X') = q + 2\alpha x_i + \alpha^2. \end{aligned} \quad (10)$$

Таким образом, величины p' , q' и компоненты вектора F' , нужные для проведения следующего шага, легко вычисляются.

Если на следующем шаге менять j -ю компоненту $j \neq i$, то в квадратном уравнении нужно заменить a_{ii} на a_{jj} , x_i на x'_j , f_i на f'_j , q на q' и p на p' .

Выбор номеров изменяемых компонент может осуществляться различно. Простейшей возможностью является циклическое чередование индексов. Ввиду возможного накопления ошибок округления время от времени нужно вычислять величины q , p и f_i ($i = 1, 2, \dots, n$) непосредственно по определяющим их формулам.

Отметим также соотношение

$$\mu' = \frac{p - f_i x_i + \alpha a}{q - x_i^2} = \frac{p'}{q'}, \quad (11)$$

которое можно использовать для контроля вычислений.

Максимизация отношения Релея за счет изменения вектора в данном направлении может осуществляться несколько иначе, чем было описано выше.

По существу нам нужно найти максимум отношения Релея $\frac{(AX', X')}{(X', X')}$ в плоскости $X' = c_1 X + c_2 Y$, натянутой на данные векторы X и Y .

Считая X' нормированным вектором ($(X', X') = 1$), мы придем к задаче об отыскании максимума квадратичной формы

$$(AX, X) c_1^2 + 2(AX, Y) c_1 c_2 + (AY, Y) c_2^2 \quad (12)$$

при условии

$$(X, X) c_1^2 + 2(X, Y) c_1 c_2 + (Y, Y) c_2^2 = 1. \quad (13)$$

Решая эту задачу методом множителей Лагранжа, мы придем к системе уравнений для определения коэффициентов c_1 и c_2

$$\begin{aligned} [(AX, X) - \mu(X, X)] c_1 + [(AX, Y) - (X, Y) \mu] c_2 &= 0, \\ [(AX, Y) - (X, Y) \mu] c_1 + [(AY, Y) - (Y, Y) \mu] c_2 &= 0, \end{aligned} \quad (14)$$

откуда следует, что множитель Лагранжа μ является корнем квадратного уравнения

$$\left| \begin{array}{cc} (AX, X) - (X, X) \mu & (AX, Y) - (X, Y) \mu \\ (AX, Y) - (X, Y) \mu & (AY, Y) - (Y, Y) \mu \end{array} \right| = 0. \quad (15)$$

Больший корень этого квадратного уравнения будет давать значение искомого максимума μ' , коэффициенты же c_1 и c_2 определяются из системы (14) после подстановки в нее вместо μ найденного значения μ' .

Для отношения c_1 и c_2 получаем

$$\alpha = \frac{c_2}{c_1} = \frac{(AX, X) - (X, X) \mu'}{(X, Y) \mu' - (AX, Y)} = \frac{(X, Y) \mu' - (AX, Y)}{(AY, Y) - (Y, Y) \mu'}. \quad (16)$$

Так как, в конце концов, нормировать вектор X' нет необходимости, можно полагать

$$X' = X + \alpha Y.$$

В случае координатной релаксации, когда $Y = e_i$, получим в прежних обозначениях, что μ' есть больший корень квадратного уравнения

$$(q - x_i^2) \mu'^2 - (a_{ii}q + p - 2x_i f_i) \mu + a_{ii}p - f_i^2 = 0 \quad (17)$$

и

$$X' = X + \alpha e_i, \quad (18)$$

где

$$\alpha = \frac{q\mu' - p}{f_i - x_i \mu'} = \frac{f_i - x_i \mu'}{\mu' - a_{ii}}. \quad (19)$$

Переход к следующему шагу осуществляется так же, как в прежней схеме. Хорошим контролем является выполнение равенства $\mu' = \frac{p'}{q'}$.

Циклический координатный процесс оказывается сходящимся почти всегда, однако доказательство сходимости мы дадим при некоторых ограничениях.

Теорема 61.1. Пусть матрица A симметрична и ее наибольшее собственное значение λ_1 простое. Пусть начальный вектор X_0 выбран так, что 1) $\frac{(AX_0, X_0)}{(X_0, X_0)} > \lambda_2$, где λ_2 — второе, в порядке убывания, собственное значение матрицы A и что 2) $\frac{(AX_0, X_0)}{(X_0, X_0)} > > (Ae_i, e_i) = a_{ii}$ при всех $i = 1, 2, \dots, n$. При этих условиях последовательные приближения X_k в циклическом координатном процессе сходятся по направлению к собственному вектору, отвечающему собственному значению λ_1 .

Доказательство. Будем предполагать, что на каждом шаге процесса последовательные приближения нормируются к единичной длине и из двух взаимно противоположных нормированных векторов выбирается один в некоторой фиксированной полусфере. Обозначим через D совокупность векторов длины единица, удовлетворяющих условиям 1) и 2) теоремы и лежащих в выбранной полусфере. Из условия теоремы следует, что начальное приближение принадлежит области D . Все последующие приближения будут также принадлежать области D , ибо по ходу процесса отношение Релея не убывает.

Рассмотрим подробнее отдельный шаг циклического координатного процесса. Пусть X — некоторый вектор из области D . Обозначим через $\varphi_i(X)$ нормированный вектор, взятый в выбранной полусфере, на котором достигается максимум (AX, X) в подпространстве, натянутом на векторы X и e_i . В силу второго условия теоремы $(AX, X) > (Ae_i, e_i)$ и потому вектор $\varphi_i(X)$ определяется однозначно. Из приведенных выше геометрических соображений очевидно, что нелинейный оператор φ_i непрерывен в области D .

В новых обозначениях процесс может быть описан следующим образом

$$X_1 = \varphi_1(X_0), \quad X_2 = \varphi_2(X_1), \quad \dots, \quad X_n = \varphi_n(X_{n-1})$$

$$X_{n+1} = \varphi_1(X_n), \quad \dots, \quad X_{nk+j} = \varphi_j(X_{nk+j-1}), \quad \dots \quad (j = 1, 2, \dots, n)$$

Через μ_0, μ_1, \dots обозначим значения $(AX_0, X_0), (AX_1, X_1), \dots$ функционала (AX, X) . Эти значения удовлетворяют неравенствам $\mu_0 \leq \mu_1 \leq \mu_2 \leq \dots$ и ограничены сверху числом λ_1 . Следовательно, существует $\lim_{k \rightarrow \infty} \mu_k = \mu$.

Рассмотрим последовательность векторов $X_1, X_{n+1}, X_{2n+1}, \dots$ Все векторы этой последовательности находятся в области D . В силу ограниченности и замкнутости единичной сферы из этой последовательности можно извлечь сходящуюся последовательность X_{nk_j+1} . Пусть Y есть предел этой последовательности. Вектор Y находится внутри области D , ибо $(AY, Y) = \mu > \mu_0 > \max(\lambda_2, (Ae_i, e_i))$.

В силу непрерывности оператора φ_2 в области D имеем

$$\tilde{Y} = \varphi_2(Y) = \lim_{j \rightarrow \infty} \varphi_2(X_{nk_j+1}) = \lim_{j \rightarrow \infty} X_{nk_j+2}.$$

Следовательно, $(AY, \tilde{Y}) = \lim (AX_{nk_j+2}, X_{nk_j+2}) = \mu$. Но по определению оператора φ_2 на векторе \tilde{Y} достигается максимум (AZ, Z) в плоскости, натянутой на Y и e_2 и этот максимум равен $\mu = (AY, Y) > (Ae_2, e_2)$. Следовательно, $\tilde{Y} = Y$. Точно так же доказывается, что $\varphi_3(Y) = Y, \dots, \varphi_n(Y) = Y, \varphi_1(Y) = Y$, так что вектор Y оказывается неподвижным для всех операторов φ_i ($i = 1, 2, \dots, n$).

Это значит, что за счет изменения каждой отдельной координаты вектора Y отношение Релея не может быть увеличено. Следовательно, все частные производные отношения Релея в точке Y равны нулю и, следовательно, Y есть собственный вектор. Но, в силу условий 1) и 2) теоремы, это может быть только собственный вектор, соответствующий наибольшему собственному значению λ_1 .

Итак, единственной предельной точкой последовательности $X_1, X_{n+1}, X_{2n+1}, \dots$ является собственный вектор, отвечающий наибольшему собственному значению.

К тому же вектору сходятся и остальные последовательности $X_i, X_{n+i}, X_{2n+i}, \dots$, а потому и вся последовательность X_1, X_2, \dots . Теорема доказана.

Замечание. Условие 2) теоремы не является существенным, ибо оно, вообще говоря, будет выполняться автоматически после проведения первого цикла процесса при любом выборе начального вектора. Если же снять первое ограничение, то, повторяя доказательство, мы получим, что каждая предельная точка последовательности X_0, X_1, X_2, \dots является собственным вектором матрицы A . Если допустить, что собственные значения матрицы A различны, то, в силу равенства отношения Релея на всех предельных точках, мы можем заключить, что предельная точка единственна, т. е. последовательность X_0, X_1, \dots сходится к некоторому собственному вектору. Однако этот собственный вектор не обязан принадлежать наибольшему собственному значению, что можно проиллюстрировать на следующем примере. Пусть

$$A = \begin{bmatrix} 35 & -22 & 8 \\ -22 & 38 & 14 \\ 8 & 14 & 53 \end{bmatrix}.$$

Если взять в качестве начального вектора $X_0 = \left(\frac{2}{3}, -\frac{1}{3}, \frac{2}{3} \right)'$, то окажется, что все последовательные приближения будут равны X_0 .

Легко проверить, что X_0 есть собственный вектор, отвечающий собственному значению $\lambda_2 = 54$. Наибольшим же собственным значением является $\lambda_1 = 63$.

Более внимательный анализ показывает, что последовательность X_0, X_1, \dots может сходиться к собственному вектору, соответствующему не наибольшему собственному значению, только если она, начиная с некоторого вектора, стабилизируется.

Использование координатной одношаговой релаксации для нахождения наибольшего собственного значения, вообще говоря, невыгодно, ибо по ходу процесса приходится на каждом малом шаге решать квадратное уравнение (7).

Мы все же приведем пример (матрица (4) § 51), показывающий ход процесса, опуская промежуточные вычисления.

X_0	0.34	0.20	0.90	0.75	1.8300504
X_1	0.59995498	0.41211685	0.76621642	0.70021179	2.2138031
X_2	0.69781651	0.51354408	0.67966297	0.67841258	2.2977743
X_3	0.73541175	0.56051428	0.62985366	0.67161815	2.3165637
X_4	0.75062330	0.58274350	0.60231592	0.67072572	2.3211276
X_5	0.75694325	0.59351423	0.58729670	0.67163815	2.3223053
X_6	0.75958696	0.59882656	0.57913506	0.67281146	2.3226235
X_7	0.76068465	0.60148118	0.57469926	0.67377537	2.3227126
X_8	0.76113138	0.60282111	0.57228443	0.67445635	2.3227381
X_9	0.76130659	0.60350310	0.57096708	0.67490251	2.3227456
X_{10}	0.76137102	0.60386377	0.57024369	0.67517902	2.3227477
X_{11}	0.76139227	0.60403891	0.56985092	0.67535101	2.3227485
\tilde{X}_{11}	1.0000	0.7933	0.7484	0.8870	

Здесь в последнем столбце приведены значения $\rho(X_i)$.

Групповая координатная релаксация заключается в следующем. Ищется максимум (или минимум) отношения $\frac{(AX, X)}{(X, X)}$ за счет одновременного изменения нескольких координат предыдущего приближения. От шага к шагу выбираемые группы координат изменяются.

Дадим описание одного шага процесса¹⁾. Пусть X некоторое приближение к собственному вектору, принадлежащему наибольшему по модулю собственному значению λ_1 , найденное в предыдущих k шагах. Положим для определенности, что в рассматриваемом шаге изменяемые координаты имеют номера $1, \dots, r$. Натянем на векторы X ,

¹⁾ Хестинс [1].

e_1, \dots, e_r подпространство Q и будем искать максимум $\mu(X')$ в этом подпространстве. Пусть

$$X' = c_0 X + c_1 e_1 + \dots + c_r e_r. \quad (20)$$

Тогда

$$(AX', X') = \sum_{i,j=0}^r \gamma_{ij} c_i e_j, \quad (21)$$

где

$$\gamma_{00} = (AX, X)$$

$$\gamma_{0i} = (AX, e_i) = (X, Ae_i) = (X, A_i) \quad (22)$$

$$\gamma_{ij} = (Ae_i, e_j) = a_{ij} \quad (i, j = 1, 2, \dots, r).$$

Здесь через A_i обозначен i -й столбец матрицы A . В свою очередь

$$(X', X') = \sum_{i,j=0}^r \beta_{ij} c_i c_j, \quad (23)$$

где

$$\beta_{00} = (X, X) = x_1^2 + \dots + x_n^2$$

$$\beta_{0i} = (X, e_i) = x_i \quad (24)$$

$$\beta_{ij} = (e_i, e_j) = \delta_{ij} \quad (i, j = 1, 2, \dots, r),$$

δ_{ij} — символ Кронекера. Обозначим

$$\begin{aligned} \Gamma &= (\gamma_{ij}) \\ B &= (\beta_{ij}). \end{aligned} \quad (25)$$

Легко видеть, что максимум $\mu(X')$ равен наибольшему корню μ' уравнения

$$|\Gamma - tB| = 0. \quad (26)$$

а вектор, реализующий этот максимум, есть решение линейной однородной системы с матрицей $\Gamma - \mu' B$. Полученное решение целесообразно нормировать так, чтобы $c_0 = 1$.

Таким образом, один шаг групповой координатной релаксации равносителен решению частичной обобщенной проблемы собственных значений для матрицы $r+1$ -го порядка. Однако обобщенная проблема в данном случае легко сводится к обычной.

Для этого в подпространстве, натянутом на X, e_1, \dots, e_r выбираем ортонормальный базис $Z = \frac{\tilde{X}}{|\tilde{X}|}, e_1, \dots, e_r$, где $\tilde{X} = (0, 0, \dots, 0, x_{r+1}, \dots, x_n)'$.

Будем рассматривать вектор, реализующий максимум отношения Релея в виде

$$X' = s_0 Z + s_1 e_1 + \dots + s_r e_r. \quad (27)$$

Тогда

$$\frac{(AX', X')}{(X', X')} = \frac{\sum_{i,j=0}^r a_{ij} s_i s_j}{s_0^2 + \dots + s_r^2}, \quad (28)$$

где

$$\begin{aligned} a_{00} &= (AZ, Z) \\ a_{0i} &= a_{i0} = (Z, A_i) \\ a_{ij} &= a_{ij} \text{ при } i, j = 1, \dots, r. \end{aligned} \quad (29)$$

Поэтому числа s_0, \dots, s_r являются компонентами собственного вектора матрицы

$$S = \begin{bmatrix} a_{00} & a_{01} & \dots & a_{0r} \\ a_{10} & a_{11} & \dots & a_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r0} & a_{r1} & \dots & a_{rr} \end{bmatrix}, \quad (30)$$

соответствующего наибольшему собственному значению. Это последнее и равно искомому максимуму.

Таким образом, в выбранном базисе каждый отдельный шаг метода групповой релаксации требует решения частичной проблемы собственных значений для матрицы порядка $r+1$.

Метод групповой координатной релаксации имеет смысл применять лишь для матриц большого порядка.

§ 62. Уточнение отдельного собственного значения и принадлежащего ему собственного вектора

1. Метод Дерведюэ.¹⁾ Пусть λ и U суть приближенные значения для некоторого собственного значения матрицы A и принадлежащего этому собственному значению собственного вектора. Обозначим через $\lambda^* = \lambda + \Delta\lambda$ и $U^* = U + \Delta U$ точные значения того и другого. Пусть

$$AU - \lambda U = r. \quad (1)$$

Для определения $\Delta\lambda$ и ΔU имеем нелинейное уравнение

$$AU^* = \lambda^* U^*$$

или

$$A(U + \Delta U) = (\lambda + \Delta\lambda)(U + \Delta U). \quad (2)$$

Отбрасывая члены 2-го порядка малости, получим

$$AU + A\Delta U = \lambda U + \Delta\lambda U + \lambda \Delta U$$

или, принимая во внимание (1),

$$r + A\Delta U - \Delta\lambda U - \lambda \Delta U = 0. \quad (3)$$

¹⁾ Дерведюэ [2].

Без нарушения общности можно считать, что первая компонента вектора ΔU равна 0, ибо собственный вектор U^* определен с точностью до постоянного множителя. Записав равенство (3) в компонентах, мы получим систему n линейных уравнений с n неизвестными $\Delta\lambda, \Delta u_2, \dots, \Delta u_n$. Эта система имеет вид

$$\begin{aligned} -u_1 \Delta\lambda + a_{12} \Delta u_2 + \dots + a_{1n} \Delta u_n &= -r_1 \\ -u_2 \Delta\lambda + (a_{22} - \lambda) \Delta u_2 + \dots + a_{2n} \Delta u_n &= -r_2 \\ \vdots &\quad \vdots \\ -u_n \Delta\lambda + a_{n2} \Delta u_2 + \dots + (a_{nn} - \lambda) \Delta u_n &= -r_n. \end{aligned} \quad (4)$$

Найдя $\Delta\lambda, \Delta u_2, \dots, \Delta u_n$, можно повторить процесс.

В качестве примера уточним собственное значение и принадлежащий ему собственный вектор для матрицы примера 2 § 53. На странице 333 было найдено, что $\lambda_1 = 17.39$ и $U_1 = (1.00000, -8.14472, 7.91426)'$. Вычисляя невязку, получим $r = (-0.00420498, 0.0592605, -0.0630314)'$. Ниже приведено решение системы для определения $\Delta\lambda, \Delta u_2$ и Δu_3 по методу единственного деления

-1	1.870086	0.422908	0.00420498	1.297199
8.14472	5.578346	5.711900	-0.05926046	19.375706
-7.91426	4.308033	4.419313	0.06303143	0.8761174
1	-1.870086	-0.422908	-0.00420498	-1.297199
	20.809673	9.156367	-0.02501208	29.941029
	-10.492314	1.072309	0.02975214	-9.390253
	1	0.4400053	-0.00120195	1.438804
		5.6889803	0.01714090	5.706124
	1	1	0.00301300	1.0030130
1			-0.0025277	0.9974724
			-0.0076578	0.9923423

Таким образом,

$$\Delta\lambda = -0.007658, \quad \Delta u_2 = -0.00253, \quad \Delta u_3 = 0.00301,$$

и потому мы получаем уточненные значения

$$\lambda_1 = -17.397658, \quad U_1 = (1.00000, -8.14725, 7.91727)'.$$

Полученные значения уже значительно ближе к точным (см. § 53).

2. Метод Виландта. Для уточнения отдельного собственного значения и принадлежащего ему собственного вектора можно применить следующее видоизменение степенного метода, впервые предложенное Виландтом¹⁾.

Пусть μ есть приближение к собственному значению λ матрицы A , которому соответствует собственный вектор U . Тогда число $v = \frac{1}{\lambda - \mu}$ будет собственным значением для матрицы $B = (A - \mu E)^{-1}$, которому будет соответствовать, как это не трудно проверить, собственный вектор U . Остальными собственными значениями для матрицы B будут $v_i = \frac{1}{\lambda_i - \mu}$ и если μ достаточно близко к λ , то v будет значительно преобладать над остальными собственными значениями матрицы B . Поэтому применение степенного метода с матрицей B уже при небольшом количестве итераций даст хорошее приближение для числа v . Собственное значение λ найдется по формуле

$$\lambda = \mu + \frac{1}{v}. \quad (5)$$

Итерации матрицей $B = (A - \mu E)^{-1}$ требуют либо фактического обращения матрицы $A - \mu E$, либо менее трудоемкого решения системы линейных уравнений

$$(A - \mu E) U_i = U_{i-1}, \quad (6)$$

дающее сразу итерацию вектора U_{i-1} матрицей B . За вектор U_0 следует брать известное приближение к собственному вектору.

Определитель матрицы $A - \mu E$ будет близок к нулю и система (6) окажется плохо обусловленной. Однако характер плохой обусловленности здесь будет таким, что с малой точностью определяются абсолютные значения компонент вектора U_i , но отношения компонент этого вектора будут, как правило, определяться с хорошей точностью. Для вычисления же компонент собственного вектора только это и нужно, так как последние определяются лишь с точностью до скалярного множителя.

Возможна еще одна модификация процесса, в которой итерации получаются в результате решения системы

$$(A - \mu_i E) U_i = U_{i-1}, \quad (7)$$

где μ_i есть приближение к собственному значению λ , уточненное на предыдущем шаге. Это дает ускорение сходимости процесса. Правда оно мало существенно, так как процесс и без него сходится очень быстро.

Заметим еще, что при μ очень близких к λ решение системы

$$(A - \mu E) U_i = U_{i-1} \quad (8)$$

¹⁾ Бодевиг [6], стр. 293–294.

почти не отличается от решения однородной системы

$$(A - \mu E) U = 0 \quad (9)$$

для координат собственного вектора.

Собственное значение ν определяется как отношение соответствующих компонент векторов U_i и U_{i-1} . В силу указанной выше плохой обусловленности системы, эти отношения определяются с невысокой точностью, но это не портит дела, так как число $\frac{1}{\nu}$ играет роль малой поправки в формуле $\lambda = \mu + \frac{1}{\nu}$ для определения собственного значения матрицы.

Метод Виландта проиллюстрируем на примере предыдущего пункта.

Пусть $\mu = -17.39$, $U_0 = (1.00; -8.14; 7.91)'$. Решая по методу единственного деления соответствующую систему

11.880118	1.870086	0.422908	1.00	15.173112
0.287865	5.578346	5.711900	-8.14	3.438111
0.049099	4.308033	4.419313	7.91	16.686445
1	0.15741308	0.03559796	0.08417425	1.2771853
	5.5330323	5.7016526	-8.1642308	3.0704541
	4.3003042	4.4175652	7.9058671	16.6237365
	1	1.0304751	-1.4755437	0.5549315
		-0.01379120	14.251154	14.237362
1		1	-1033.3512	-1032.3512
			1063.3671	1064.3671
1			-130.5185	-129.5185

получим, что $U_1 = (1.0000, -8.14725, 7.91728)'$ и $\frac{1}{\nu} = -0.007655$ (или $-0.007655, -0.007662$). Это дает $\lambda = -17.397655$. Вектор U_1 также значительно ближе к собственному вектору, чем U_0 .

3. Метод возмущений. Метод возмущений разработан для ограниченных операторов в Гильбертовом пространстве¹⁾. В применении к матрицам он в основном заключается в следующем.

¹⁾ F. Rellich. Math. Ann., 1936, 113 (4), 600—619. М. К. Гавурин. Вестн. Ленингр. ун-та, 1952, 9, 77—95.

Пусть симметричная матрица A_0 имеет простое собственное значение λ_0 и соответствующий ему собственный вектор X_0 . Обозначим через R матрицу, определенную условиями $R(A_0 - \lambda_0 E)Y = Y$ при любом Y , ортогональном к X_0 , и $RX_0 = 0$. Матрица R существует, ибо оператор с матрицей $A_0 - \lambda_0 E$ будет невырожденным на подпространстве, ортогональном к вектору X_0 . Пусть A некоторая другая симметричная матрица, λ ее собственное значение и X соответствующий собственный вектор, причем $(X, X_0) \neq 0$, что позволяет считать, без нарушения общности, что $X = X_0 + Z$, при $(Z, X_0) = 0$.

Положим $B = A - A_0$, $\mu = \lambda - \lambda_0$. Тогда имеют место соотношения

$$((\mu E - B)X, X_0) = 0 \quad (10)$$

$$(E - R(\mu E - B))X = X_0. \quad (11)$$

Действительно, $(\mu E - B)X = (\lambda E - A)X - (\lambda_0 E - A_0)X = (A_0 - \lambda_0 E)X$, так что $((\mu E - B)X, X_0) = (X, (A_0 - \lambda_0 E)X_0) = 0$. Далее,

$$(E - R(\mu E - B))X = X - R(A_0 - \lambda_0 E)X = X - Z = X_0.$$

Если матрица $E - R(\mu E - B)$ невырожденная (что будет иметь место, например, если $|\mu|$ и $\|B\|$ достаточно малы), то формула (11) эквивалентна формуле

$$X = (E - R(\mu E - B))^{-1}X_0. \quad (12)$$

При малых $|\mu|$ и $\|B\|$ правую часть равенства (12) можно разложить в ряд. Ограничиваюсь членами второго порядка и принимая во внимание, что $RX_0 = 0$, получим приближенную формулу

$$X \approx X_0 - RBX_0 + RBRBX_0 - \mu R^2 BX_0, \quad (13)$$

которая вместе с (10) дает

$$\mu \approx \frac{(BX_0, X_0) - (RBX_0, BX_0) + (RBRBX_0, RBX_0)}{\|X_0\|^2 + \|BRX_0\|^2}. \quad (14)$$

Для применения изложенного к задаче об уточнении отдельного собственного значения симметричной матрицы воспользуемся приемом „ложного возмущения“, также предложенного М. К. Гавуриным¹⁾.

Пусть для симметричной матрицы A известен приближенный нормированный собственный вектор X_0 и приближенное собственное значение $\lambda_0 = (AX_0, X_0)$. Построим матрицу

$$A_0 = A - X_0 r' - r X_0'$$

где $r = AX_0 - \lambda_0 X_0$. Ясно, что A_0 симметрична, и легко проверяется, что для нее вектор X_0 является собственным вектором, принадлежащим собственному значению λ_0 .

¹⁾ М. К. Гавурин. Успехи матем. наук, 1957, 12, № 1, 173—175.

Приближенные формулы (13) и (14) превращаются в

$$X \approx X_0 - Rr - \mu R^2 r \quad (15)$$

$$\mu \approx -\frac{(Rr, r)}{1 + \|Rr\|^2}. \quad (16)$$

Уточнив по формулам (15) и (16) собственный вектор и собственное значение, можно повторять процесс до получения требуемой точности.

Векторы Rr и R^2r находятся соответственно из систем линейных уравнений

$$(A_0 - \lambda_0 E) Z = r \quad (17)$$

$$(Z, X_0) = 0$$

и

$$(A_0 - \lambda_0 E) Z = Rr \quad (18)$$

$$(Z, X_0) = 0.$$

Так как $|A_0 - \lambda_0 E| = 0$, одно из n первых уравнений в системах (17) и (18) может быть отброшено и использовано лишь для контроля.

Если в формуле (15) отбросить член $\mu R^2 r$ второго порядка малости, то отпадает необходимость в решении второй системы.

Отметим, что метод может применяться, начиная с довольно грубых приближений для X_0 .

На описанный здесь прием реализации метода возмущений обратил внимание авторов Л. А Руховец.

Для матрицы (4) § 51 имеем $\lambda = 0.24226071$, $X = (0.718846, 0.095699, -0.387435, -0.569207)'$.

Исходя из приближения $X_0 = (0.704361, 0.176090, -0.440225, -0.528271)'$, полученного нормированием вектора $(0.8, 0.2, -0.5, -0.6)'$, и $\lambda_0 = (AX_0, X_0) = 0.248992$, получим, после однократного применения формул (15) и (16) и нормирования, $X_1 = (0.718839, 0.095717, -0.387445, -0.569206)'$, $\lambda_1 = 0.24226072$.

При проведении итерационного процесса посредством n -кратного повторного применения формул (15) и (16), быстрота сходимости имеет порядок q^n , если же пользоваться упрощенной формулой (15) с отброшенным членом $\mu R^2 r$, то быстрота сходимости будет иметь порядок q^{3n} . Здесь $q = \frac{\|r\|}{\tau - 2\|r\|}$, где r невязка начального приближения, τ расстояние от уточняемого собственного значения до ближайшего соседнего. Эти оценки нам сообщены М. К. Гавуриным.

ГЛАВА VI

МЕТОД МИНИМАЛЬНЫХ ИТЕРАЦИЙ И ДРУГИЕ МЕТОДЫ, ОСНОВАННЫЕ НА ИДЕЕ ОРТОГОНАЛИЗАЦИИ

Методы, которые будут описаны в этой и следующей главах, применимы к решению системы линейных уравнений и к решению проблемы собственных значений — полной (гл. VI) и частичной (гл. VII). Хотя исторически некоторые методы, описанные в гл. VII (в частности, метод наискорейшего спуска), появились ранее методов гл. VI, мы излагаем сначала эти последние, ибо появление их позволило упростить и развить дальше первые.

Несмотря на то, что методы гл. VI являются точными, а методы гл. VII итерационными, их вычислительные схемы по существу имеют один рисунок.

§ 63. Метод минимальных итераций

1. Построение базисных векторов в невырожденном случае. Пусть матрица A симметрична. Метод минимальных итераций¹⁾ для решения полной проблемы собственных значений есть не что иное, как метод ортогонализации последовательных итераций некоторого начального вектора (§ 50). Однако мы изложим его независимо от результатов § 50, ввиду возникновения многих важных особенностей метода, наличие которых позволяет обобщить метод и расширить область его применения.

Будем в этом пункте предполагать, что симметричная матрица A имеет попарно различные собственные значения $\lambda_1, \lambda_2, \dots, \lambda_n$. Пусть U_1, U_2, \dots, U_n соответствующие им собственные векторы, которые мы будем считать нормированными.

Выберем некоторый начальный вектор X . Он может быть единственным образом представлен в виде

$$X = a_1 U_1 + a_2 U_2 + \dots + a_n U_n. \quad (1)$$

Будем далее пока предполагать, что все коэффициенты $a_i \neq 0$. Тогда система векторов $X, AX, \dots, A^{n-1}X$ будет линейно-независимой.

¹⁾ Ланцош [2].

Теорема 63.1. Если векторы $X, AX, \dots, A^{n-1}X$ линейно-независимы и векторы p_0, p_1, \dots, p_{n-1} получены из них при помощи процесса ортогонализации, то эти векторы определяются по трехчленным рекуррентным формулам

$$\begin{aligned} p_{i+1} &= Ap_i - \alpha_i p_i - \beta_i p_{i-1} \quad (i = 1, 2, \dots, n-2) \\ p_0 &= X, \quad p_1 = Ap_0 - \alpha_0 p_0, \end{aligned} \quad (2)$$

где

$$\begin{aligned} \alpha_i &= \frac{(Ap_i, p_i)}{(p_i, p_i)} \quad (i = 0, 1, \dots, n-2) \\ \beta_i &= \frac{(Ap_i, p_{i-1})}{(p_{i-1}, p_{i-1})} = \frac{(p_i, Ap_{i-1})}{(p_{i-1}, p_{i-1})} = \frac{(p_i, p_i)}{(p_{i-1}, p_{i-1})} \quad (i = 1, 2, \dots, n-2). \end{aligned} \quad (3)$$

Доказательство. Пусть $X, AX, \dots, A^{n-1}X$ линейно-независимы, и векторы p_0, p_1, \dots, p_{n-1} получены из них процессом ортогонализации. Тогда $p_i = A^i X + c_1^{(i)} A^{i-1} X + \dots + c_n^{(i)} X$.

Отсюда следует, что вектор $p_{i+1} - Ap_i$ принадлежит подпространству, натянутому на векторы $X, AX, \dots, A^i X$, или, что то же самое, подпространству, натянутому на векторы p_0, p_1, \dots, p_i . Итак, вектор p_{i+1} выражается через предшествующие по формуле

$$p_{i+1} = Ap_i - \gamma_i^{(i)} p_i - \dots - \gamma_0^{(i)} p_0,$$

где $\gamma_1^{(i)}, \dots, \gamma_0^{(i)}$ некоторые числа. Из соотношений ортогональности вектора p_{i+1} к предшествующим векторам получим

$$\gamma_j^{(i)} = \frac{(Ap_i, p_j)}{(p_j, p_j)}.$$

Но при $j = 0, 1, \dots, i-2$ имеют место равенства $(Ap_i, p_j) = 0$, ибо $(Ap_i, p_j) = (p_i, Ap_j)$, а вектор Ap_j есть линейная комбинация векторов p_{j+1}, p_j, \dots, p_0 , каждый из которых ортогонален к вектору p_i при $j \leq i-2$. Следовательно, ненулевыми остаются только два коэффициента

$$\alpha_i = \gamma_i^{(i)} = \frac{(Ap_i, p_i)}{(p_i, p_i)}$$

и

$$\beta_i = \gamma_{i-1}^{(i)} = \frac{(Ap_i, p_{i-1})}{(p_{i-1}, p_{i-1})} = \frac{(p_i, Ap_{i-1})}{(p_{i-1}, p_{i-1})}.$$

Далее,

$$(p_i, Ap_{i-1}) = (p_i, p_i + \alpha_{i-1} p_{i-1} + \beta_{i-1} p_{i-2}) = (p_i, p_i) \quad (4)$$

и, следовательно,

$$\beta_i = \frac{(p_i, p_i)}{(p_{i-1}, p_{i-1})}.$$

Из последней формулы для β_i следует, что $\beta_i > 0$. Для положительно-определенной матрицы положительными будут и коэффициенты α_i .

Очевидно, что векторы p_i представляются в виде

$$p_i = p_i(A)X = p_i(A)p_0 \quad (i = 1, \dots, n-1), \quad (5)$$

где $p_i(t) = t^i + \dots$ некоторый полином степени i .

Полиномы $p_i(t)$ могут вычисляться параллельно с вычислением векторов p_i по формулам

$$p_{i+1}(t) = (t - \alpha_i)p_i(t) - \beta_i p_{i-1}(t), \quad (6)$$

в которых коэффициенты α_i и β_i имеют прежние значения.

Заметим, что вектор $p_n = Ap_{n-1} - \alpha_{n-1}p_{n-1} - \beta_{n-1}p_{n-2}$, где $\alpha_{n-1} = \frac{(Ap_{n-1}, p_{n-1})}{(p_{n-1}, p_{n-1})}$ и $\beta_{n-1} = \frac{(p_{n-1}, p_{n-1})}{(p_{n-2}, p_{n-2})}$, будет ортогонален к n векторам p_0, p_1, \dots, p_{n-1} и, следовательно, вектор p_n равен нулю. Соответственно, полином $p_n(t) = (t - \alpha_{n-1})p_{n-1}(t) - \beta_{n-1}p_{n-2}(t)$ совпадает с характеристическим полиномом. Тем самым получен удобный и простой алгорифм для вычисления коэффициентов характеристического полинома.

Трехчленность рекуррентных соотношений, связывающих векторы p_i и полиномы $p_i(t)$, была уже нами установлена двумя способами, только что и выше в § 50. Осветим это обстоятельство еще с одной точки зрения. Из равенства (5) и разложения

$$p_0 = a_1 U_1 + a_2 U_2 + \dots + a_n U_n$$

следует, что

$$p_i = a_1 p_i(\lambda_1) U_1 + a_2 p_i(\lambda_2) U_2 + \dots + a_n p_i(\lambda_n) U_n.$$

Отсюда, принимая во внимание условие $|U_i| = 1$, получим

$$(p_j, p_i) = a_1^2 p_j(\lambda_1) p_i(\lambda_1) + \dots + a_n^2 p_j(\lambda_n) p_i(\lambda_n) = 0.$$

Последнее равенство эквивалентно равенству

$$\int_m^M p_j(t) p_i(t) dF(t) = 0. \quad (7)$$

Здесь весовая функция интеграла Стильеса определена следующим образом

$$\begin{aligned} F(t) &= 0 & m \leq t \leq \lambda_1 \\ F(t) &= a_1^2 & \lambda_1 < t \leq \lambda_2 \\ F(t) &= a_1^2 + a_2^2 & \lambda_2 < t \leq \lambda_3 \\ &\dots & \dots \\ F(t) &= a_1^2 + a_2^2 + \dots + a_n^2 & \lambda_n < t \leq M, \end{aligned} \quad (8)$$

где m и M — границы спектра матрицы A (рис. 4).

Условие (7) определяет ортогональную по интегральному весу $F(t)$ систему полиномов $p_i(t)$. Но из теории ортогональных полиномов известно¹⁾, что эти полиномы связаны трехчленными рекуррентными соотношениями.

Отметим ряд свойств векторов p_0, \dots, p_{n-1} и полиномов $p_0(t), p_1(t), \dots, p_{n-1}(t), p_n(t)$.

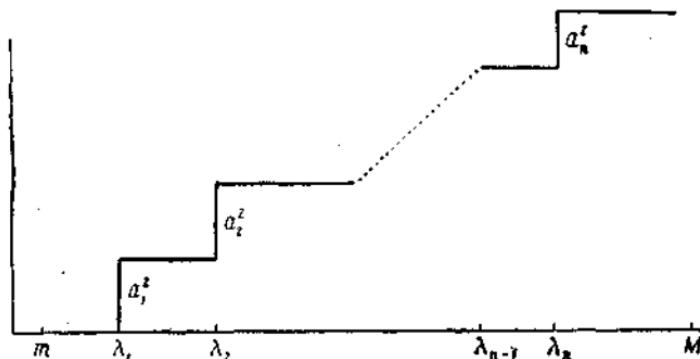


Рис. 4.

Теорема 63.2. Корни полиномов $p_n(t), \dots, p_0(t)$ вещественны и разделяются.

Действительно, в силу соотношений (6) и положительности коэффициентов β_i последовательность полиномов $p_n(t), p_{n-1}(t), \dots, p_0(t)$ есть последовательность Штурма с положительными старшими коэффициентами.

Теорема 63.3. Векторы p_0, \dots, p_{i-1} образуют ортогональный базис подпространства P_i , натянутого на векторы $X, AX, \dots, A^{i-1}X$.

Доказательство. Векторы p_0, \dots, p_{i-1} принадлежат подпространству P_i , линейно-независимы и ортогональны.

Следующая теорема описывает свойства векторов p_i , дающее право называть их минимальными итерациями. Именно это свойство было положено в основу построения системы векторов p_0, \dots, p_{n-1} Ланцошем [2], автором первой публикации, посвященной методу минимальных итераций.

Теорема 63.4. Среди всех векторов вида $Z = A^i X + Y$, где $Y \in P_i$, вектор p_i имеет наименьшую длину.

Действительно, $p_i = A^i X + Y_0$, где Y_0 некоторый определенный вектор из P_i , и, следовательно, $Z = p_i + Y - Y_0$. Но

$$|Z|^2 = (Z, Z) = (p_i + Y - Y_0, p_i + Y - Y_0) = |p_i|^2 + |Y - Y_0|^2,$$

¹⁾ И. П. Натансон. Конструктивная теория функций, 1949, ч. II, гл. IV.

ибо $(p_i, Y - Y_0) = 0$. Таким образом, $|Z|^2 \geq |p_i|^2$ и знак равенства возможен, только если $Y - Y_0 = 0$, т. е. $Z = p_i$.

Теорема 63.5. Оператор с матрицей A имеет в базисе p_0, p_1, \dots, p_{n-1} трехдиагональную матрицу Якоби

$$J = \begin{bmatrix} \alpha_0 & \beta_1 & & & \\ 1 & \alpha_1 & \beta_2 & & \\ & 1 & \alpha_2 & \ddots & \\ & & \ddots & \ddots & \\ & & & \ddots & \beta_{n-1} \\ & & & & 1 & \alpha_{n-1} \end{bmatrix}. \quad (9)$$

Действительно,

$$\begin{aligned} Ap_0 &= \alpha_0 p_0 + p_1 \\ Ap_1 &= \beta_1 p_0 + \alpha_1 p_1 + p_2 \\ &\vdots \\ Ap_{n-2} &= \beta_{n-2} p_{n-3} + \alpha_{n-2} p_{n-2} + p_{n-1} \\ Ap_{n-1} &= \beta_{n-1} p_{n-2} + \alpha_{n-1} p_{n-1}, \end{aligned}$$

т. е. координаты векторов $Ap_0, Ap_1, \dots, Ap_{n-1}$ в базисе p_0, p_1, \dots, p_{n-1} совпадают со столбцами матрицы J . Положительность чисел β_i уже отмечалась. Тем самым теорема доказана.

Следствие.

$$\alpha_0 + \alpha_1 + \dots + \alpha_{n-1} = \operatorname{Sp} A. \quad (10)$$

Действительно, матрицы A и J подобны, и, следовательно, их следы одинаковы.

При фактическом осуществлении метода минимальных итераций равенство (10) дает хороший заключительный контроль.

Теорема 63.6.

$$p_i(t) = \begin{bmatrix} t - \alpha_0 & \beta_1 & & & \\ 1 & t - \alpha_1 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \beta_{i-1} \\ & & & 1 & t - \alpha_{i-1} \end{bmatrix}$$

Доказательство. Полиномы

$$\tilde{p}_i(t) = \begin{vmatrix} t - \alpha_0 & \beta_1 \\ 1 & t - \alpha_1 \beta_2 \\ & \ddots & \ddots & \ddots \\ & & \ddots & \ddots & \beta_{i-1} \\ & & & 1 & t - \alpha_{i-1} \end{vmatrix}$$

связаны такими же рекуррентными соотношениями, что и полиномы $p_i(t)$, в чем легко убедиться разложением определителя по элементам последнего столбца. То, что $\tilde{p}_1(t) = p_1(t)$ и $\tilde{p}_2(t) = p_2(t)$ ясно непосредственно. Следовательно, $\tilde{p}_i(t) = p_i(t)$ при всех $i = 1, 2, \dots, n$.

Теорема 63.7. Векторы p_0, \dots, p_{n-1} удовлетворяют соотношениям

$$(Ap_i, p_j) = 0 \quad \text{при } |i - j| > 1. \quad (11)$$

Действительно, для $j = 0, \dots, i-2$ справедливость теоремы уже была установлена при доказательстве теоремы 63.1. Для $i < j$ утверждение теоремы следует из равенств

$$(Ap_i, p_j) = (p_i, Ap_j).$$

В заключение отметим, что вектор p_1 совпадает с градиентом ξ функционала $\mu(X) = \frac{(AX, X)}{(X, X)}$ в точке p_0 . Два основных свойства градиента

$$(X, \xi) = 0$$

$$(\xi, AX) = (\xi, \xi)$$

являются частным случаем свойств

$$\begin{aligned} (p_i, p_j) &= 0 \\ (Ap_{i-1}, p_i) &= (p_i, p_i) \end{aligned} \quad (12)$$

системы векторов p_0, p_1, \dots, p_{n-1} .

2. Достройка системы базисных векторов в случае вырождения. В предыдущем пункте мы предполагали, во-первых, что все собственные значения матрицы A различны и, во-вторых, что все коэффициенты a_i в разложении выбранного нами начального вектора по собственным векторам отличны от нуля. Эти условия обеспечивали линейную независимость векторов $X, AX, \dots, A^{n-1}X$.

Если хотя бы одно из них не выполнено, то векторы $X, AX, \dots, A^{n-1}X$ уже не будут линейно-независимыми. Если проводить над ними процесс ортогонализации, то на некотором шагу процесс оборвется, именно, мы получим, что $p_r = 0$ при некотором $r < n$.

В этом случае полином $p_r(t)$ будет, очевидно, минимальным аннулирующим вектор p_0 полиномом.

Покажем, как в вырожденном случае достроить систему векторов p_0, \dots, p_{r-1} до ортогонального базиса всего пространства.

Возьмем снова произвольный вектор Y и образуем вектор

$$p_0^{(1)} = Y - \sum_{i=0}^{r-1} c_i p_i, \quad (13)$$

определяя коэффициенты c_i из условия ортогональности вектора $p_0^{(1)}$ с построенными ранее векторами p_0, \dots, p_{r-1} . Это дает

$$c_i = \frac{(Y, p_i)}{(p_i, p_i)} \quad (i = 0, \dots, r-1). \quad (14)$$

Вектор $p_0^{(1)}$ ортогонален по построению ко всем векторам p_0, \dots, p_{r-1} , т. е. принадлежит к подпространству, ортогонально-дополнительному к P_r . Следовательно, и все его итерации будут ортогональны к P_r , ибо P_r инвариантно и, по теореме 11.1, его ортогональное дополнение тоже инвариантно. Поэтому, применяя метод минимальных итераций к вектору $p_0^{(1)}$, мы построим систему векторов $p_0^{(1)}, \dots, p_{l-1}^{(1)}$ попарно ортогональных не только друг к другу, но и ко всем векторам p_0, \dots, p_{r-1} . Если $r+l < n$, мы продолжаем процесс достройки до тех пор, пока не дойдем до базиса всего пространства. При этом все пространство естественным образом разобьется в прямую сумму нескольких попарно ортогональных инвариантных подпространств P_r, P_l, \dots . Характеристический полином матрицы будет равен произведению минимальных полиномов, аннулирующих последовательные начальные векторы.

3. Определение собственных векторов. Рассмотрим сначала невырожденный случай. Пусть мы уже построили ортогональный базис пространства p_0, \dots, p_{n-1} , систему ортогональных полиномов $p_0(t), \dots, p_{n-1}(t), p_n(t)$ и нашли корни характеристического полинома $p_n(t)$. Пусть U_i собственный вектор, принадлежащий собственному значению λ_i . Тогда

$$U_i = c_{i0} p_0 + \dots + c_{in-1} p_{n-1}. \quad (15)$$

В силу ортогональности векторов p_0, \dots, p_{n-1}

$$c_{ij} = \frac{(U_i, p_j)}{(p_j, p_j)}. \quad (16)$$

Таким образом, для определения вектора U_i надо вычислить лишь постоянные (U_i, p_j) . Это делается без труда на основании следующих соображений. Из разложения начального вектора p_0 по собственным векторам матрицы

$$p_0 = a_1 U_1 + \dots + a_n U_n \quad (a_i \neq 0, |U_i| = 1)$$

получим

$$p_j = a_1 p_j(\lambda_1) U_1 + \dots + a_n p_j(\lambda_n) U_n,$$

и, следовательно,

$$(U_i, p_j) = a_i p_j(\lambda_i).$$

Таким образом, с точностью до постоянного множителя

$$U_i = \frac{p_0(\lambda_i)}{(p_0, p_0)} p_0 + \dots + \frac{p_{n-1}(\lambda_i)}{(p_{n-1}, p_{n-1})} p_{n-1}. \quad (17)$$

Отметим, что значения $p_j(\lambda_i)$ можно вычислять непосредственно по рекуррентным соотношениям для полиномов $p_j(t)$.

В вырожденном случае векторы, вычисленные по формуле

$$U_i = \frac{p_0(\lambda_i)}{(p_0, p_0)} p_0 + \dots + \frac{p_{r-1}(\lambda_i)}{(p_{r-1}, p_{r-1})} p_{r-1}, \quad (17')$$

где λ_i корни полинома $p_r(t)$, будут собственными векторами, лежащими в инвариантном подпространстве P_r . Полная система собственных векторов определится, если применить описанный прием ко всем подпространствам P_r, P_1, \dots , в прямую сумму которых разбивается пространство R в процессе дстройки ортогонального базиса.

В качестве примера определим все собственные значения и принадлежащие им собственные векторы для матрицы (4) § 51

$$A = \begin{bmatrix} 1.00 & 0.42 & 0.54 & 0.66 \\ 0.42 & 1.00 & 0.32 & 0.44 \\ 0.54 & 0.32 & 1.00 & 0.22 \\ 0.66 & 0.44 & 0.22 & 1.00 \end{bmatrix}.$$

Вычисление ортогонального базиса и коэффициентов характеристического полинома дано в табл. VI. 1. В качестве начального вектора взят вектор $p_0 = (1, 1, 1, 1)'$.

В первой части таблицы помещены компоненты последовательно вычисляемых векторов $p_0, Ap_0, p_1, Ap_1, \dots, p_4$. Во второй части записаны результаты контрольной проверки выполнения условий ортогональности для векторов p_i и обычного контроля сумм для векторов Ap_i . В третьей части таблицы помещены скалярные произведения $(p_0, p_0), (p_0, Ap_0), (p_1, p_1), \dots$, причем $(p_i, p_i) = (Ap_{i-1}, p_i)$ вычисляются двумя способами, в четвертой находятся коэффициенты α_i и β_i . В последней части таблицы записаны коэффициенты полиномов $p_i(t)$, которые вычисляются рекуррентно, так же как в § 51. В результате вычислений получим

$$\begin{aligned} \varphi(t) = p_4(t) = t^4 - 4.00000005t^3 + \\ + 4.75200016t^2 - 2.11185609t + 0.28615248. \end{aligned}$$

Tagana VI. I

Вычисление базисных векторов и коэффициентов характеристического полинома по методу минимальных итераций

	P_0	Ap_0	p_1	Ap_1	p_2	Ap_2	p_3	Ap_3	p_4	Ap_4
I	1	2.62	0.32	0.1640	-0.02929952	-0.00018850	-0.001685076	-0.001540834	-0.52 · 10 ⁻⁹	
	1	2.18	-0.12	-0.0472	-0.03163768	-0.01495082	0.00717941	0.004809803	0.39 · 10 ⁻⁹	
	1	2.08	-0.22	-0.0812	-0.01816908	-0.02671150	-0.006295284	-0.004731764	0.81 · 10 ⁻⁹	
	1	2.32	0.02	0.1300	0.07910628	0.04185082	0.000801219	0.001462796	0.31 · 10 ⁻⁹	
II		9.20	0		0.1656	0	-0.16 · 10 ⁻⁸	0	0.2 · 10 ⁻⁸	
					-0.016 · 10 ⁻⁸		0.16 · 10 ⁻⁸		0.12 · 10 ⁻⁹	
III	4.0	9.20	0.1656	0.078608	0.0084473237	0.0042745163	0.000094662737	0.0000668690305		
			0.1656		0.0084473229		0.000094662885			
IV	2.3 0.0414		0.47468599 0.05101041		0.50602042 0.01120624		0.71929364		4.000000005	
V							1		1	
							-3.28070641		-4.000000005	
							2.77468599		4.75200016	
							1.05037778		-2.11185609	
							-2.3		0.28615248	

Полученные значения для коэффициентов хорошо согласуются с точными (см. § 51). Вычисляя корни последнего полинома получим для собственных значений

$$\lambda_1 = 2.32274880; \quad \lambda_2 = 0.79670672; \quad \lambda_3 = 0.63828385; \\ \lambda_4 = 0.24226068.$$

Для нахождения принадлежащих им собственных векторов предварительно, в табл. VI. 2, находим коэффициенты разложения.

Таблица VI.2
Вычисление коэффициентов разложения

t	1	2	3	4
λ_i	2.32274880	0.79670672	0.63828385	0.24226068
$p_0(\lambda_i)$	1	1	1	1
$p_1(\lambda_i) = \lambda_i - a_0$	0.02274880	-1.5032933	-1.6617162	-2.0577393
$\lambda_i - a_1$	1.84806281	0.32202073	0.16359786	-0.23242531
$p_2(\lambda_i)$	0.00064121125	0.52549161	-0.31325321	0.43687069
$\lambda_i - a_2$	1.81672838	0.29068630	0.13226343	-0.26375974
$p_3(\lambda_i)$	0.0000044810604	-0.076069604	0.043332881	-0.010262774
$\lambda_i - a_3$	1.60345516	0.07741308	-0.08100979	-0.47703296
$p_4(\lambda_i)$	$-0.3877 \cdot 10^{-8}$	$0.2760 \cdot 10^{-8}$	$0.3062 \cdot 10^{-8}$	$0.3658 \cdot 10^{-8}$
$p'_4(\lambda_i)$	5.348	-0.134	0.106	-0.457
$p_0(\lambda_i)/(p_0, p_0)$	0.25	0.25	0.25	0.25
$p_1(\lambda_i)/(p_1, p_1)$	0.13737198	-9.0778581	-10.034518	-12.425962
$p_2(\lambda_i)/(p_2, p_2)$	0.075907030	-62.208059	-37.083131	51.717053
$p_3(\lambda_i)/(p_3, p_3)$	0.047337111	-803.58551	457.76070	-108.41408

Наконец, по формулам (17) вычисляем все собственные векторы матрицы A и нормируем их к единичной первой норме.

Таблица VI.3
Собственные векторы матрицы A

U_1	U_2	U_3	U_4
1.000000	0.061916	-0.447443	1.000000
0.793587	-0.291847	1.000000	0.133129
0.747804	1.000000	0.042210	-0.538970
0.887315	-0.651533	-0.425676	-0.791834

В качестве второго примера рассмотрим матрицу

$$A = \begin{bmatrix} 10 & 6 & 3 \\ 6 & 5 & 2 \\ 3 & 2 & 2 \end{bmatrix}.$$

За начальный вектор возьмем $p_0 = (1, 1, 1)'$.

Приводим вычисления по описанной выше схеме (мы не отступаем от схемы, хотя в контрольных вычислениях здесь нет необходимости, так как вычисления проводятся точно).

Таблица VI.4
Метод минимальных итераций в случае вырождения

	p_0	Ap_0	p_1	Ap_1	p_2
I	1	19	6	42	0
	1	13	0	24	0
	1	7	-6	6	0
II		39	0	72	
III	3	39	72	216	
IV	13 24		3		
V	1		1 -13		1 -16 15

Из приведенной таблицы видно, что в данном случае имеет место вырождение, ибо уже $p_2 = 0$.

Отметим, что сумма $\alpha_1 + \alpha_2$ равна следу оператора, индуцированного на подпространстве, натянутом на векторы p_0 и p_1 , и не совпадает со следом матрицы A .

Для достройки базиса пространства возьмем новый начальный вектор

$$Y = (1, 0, 0)'.$$

Тогда в качестве $p_0^{(1)}$ нужно взять вектор

$$p_0^{(1)} = Y - \frac{(Y, p_0)}{(p_0, p_0)} p_0 - \frac{(Y, p_1)}{(p_1, p_1)} p_1 = \left(\frac{1}{6}, -\frac{1}{3}, \frac{1}{6}\right)'$$

и провести вычисления по прежней схеме. Мы сразу получаем, что $Ap_0^{(1)} = p_0^{(1)}$, так что $p_0^{(1)}$ есть собственный вектор, принадлежащий собственному значению $\lambda_3 = 1$.

Нетрудно вычислить, что $\lambda_1 = 15$, $\lambda_2 = 1$. Принадлежащие им собственные векторы (в подпространстве, натянутом на p_0 и p_1) вычисляются по формулам (17')

$$U_1 = \frac{1}{3}p_0 + \frac{1}{36}p_1 = \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right)'$$

$$U_2 = \frac{1}{3}p_0 - \frac{1}{6}p_1 = \left(-\frac{2}{3}, \frac{1}{3}, \frac{4}{3}\right)'.$$

4. Решение линейной системы. Знание базиса p_0, \dots, p_{n-1} позволяет также найти решение линейной системы

$$AX = F. \quad (18)$$

Действительно, пусть

$$X = \sum_{i=0}^{n-1} a_i p_i \quad (19)$$

разложение искомого решения по базисным векторам. Наша задача заключается в определении коэффициентов a_i . Пусть

$$F = b_0 p_0 + \dots + b_{n-1} p_{n-1}.$$

Тогда

$$b_i = \frac{(F, p_i)}{(p_i, p_i)}.$$

Из уравнения (18) имеем

$$\sum_{i=0}^{n-1} a_i Ap_i = \sum_{i=0}^{n-1} \frac{(F, p_i)}{(p_i, p_i)} p_i. \quad (20)$$

Но на основании рекуррентного соотношения (2) имеем

$$Ap_i = p_{i+1} + \alpha_i p_i + \beta_i p_{i-1}.$$

Подставляя это выражение в (20) и приравнивая коэффициенты при векторах p_i ($i = 0, 1, \dots, n-1$), получим для определения коэффициентов a_i систему с трехдиагональной матрицей

$$\begin{aligned} \alpha_0 a_0 + \beta_1 a_1 &= \frac{(F, p_0)}{(p_0, p_0)} \\ a_0 + \alpha_1 a_1 + \beta_2 a_2 &= \frac{(F, p_1)}{(p_1, p_1)} \\ &\vdots \\ a_{n-2} + \alpha_{n-1} a_{n-1} &= \frac{(F, p_{n-1})}{(p_{n-1}, p_{n-1})}. \end{aligned} \quad (21)$$

$$a_{n-2} + \alpha_{n-1} a_{n-1} = \frac{(F, p_{n-1})}{(p_{n-1}, p_{n-1})}.$$

Решив ее относительно a_0, \dots, a_{n-1} , получим искомое решение по формуле (19).

Однако, как мы увидим ниже, небольшое изменение метода позволит избежать решения трехдиагональной системы и получить явные формулы для коэффициентов разложения решения по некоторой другой системе базисных векторов.

В вырожденном случае матрица вспомогательной системы разбьется на несколько трехдиагональных ящиков, и это только облегчит решение системы.

§ 64. Биортогональный алгорифм

Биортогональный алгорифм можно рассматривать как обобщение метода минимальных итераций на случай несимметричной матрицы. При этом разнообразие в возможных причинах вырождения значительно увеличивается по сравнению с симметричным случаем, что несколько усложняет теорию метода.

1. Нормальное течение процесса. Пусть — A несимметричная матрица, p_0 и \tilde{p}_0 — два произвольных начальных вектора. Строим две последовательности векторов

$$\begin{aligned} p_1 &= Ap_0 - \alpha_0 p_0 & \tilde{p}_1 &= A'\tilde{p}_0 - \alpha_0 \tilde{p}_0 \\ p_{i+1} &= Ap_i - \alpha_i p_i - \beta_i p_{i-1} & \tilde{p}_{i+1} &= A'\tilde{p}_i - \alpha_i \tilde{p}_i - \beta_i \tilde{p}_{i-1}, \end{aligned} \quad (1)$$

где

$$\alpha_i = \frac{(Ap_i, \tilde{p}_i)}{(p_i, \tilde{p}_i)} = \frac{(p_i, A'\tilde{p}_i)}{(p_i, \tilde{p}_i)} \quad (2)$$

$$\beta_i = \frac{(Ap_i, \tilde{p}_{i-1})}{(p_{i-1}, \tilde{p}_{i-1})} = \frac{(p_i, A'\tilde{p}_{i-1})}{(p_{i-1}, \tilde{p}_{i-1})} = \frac{(\tilde{p}_i, Ap_{i-1})}{(p_{i-1}, \tilde{p}_{i-1})} = \frac{(A'\tilde{p}_i, p_{i-1})}{(p_{i-1}, \tilde{p}_{i-1})} = \frac{(p_i, \tilde{p}_i)}{(p_{i-1}, \tilde{p}_{i-1})}.$$

Ясно, что $p_i = p_i(t) p_0$, $\tilde{p}_i = p_i(t) \tilde{p}_0$, где $p_i(t)$ — полиномы, связанные рекуррентными соотношениями

$$p_{i+1}(t) = (t - \alpha_i) p_i(t) - \beta_i p_{i-1}(t). \quad (3)$$

Из формул (2) видно, что процесс оборвется, как только в первый раз будет иметь место равенство $(p_r, \tilde{p}_r) = 0$, так что до обрыва процесса $(p_i, \tilde{p}_i) \neq 0$. Непосредственным вычислением можно проверить, что для построенных систем векторов выполняются условия биортогональности, т. е., что $(p_i, \tilde{p}_j) = 0$ при $i \neq j$. Поэтому как векторы p_0, p_1, \dots, p_{r-1} , так и векторы $\tilde{p}_0, \tilde{p}_1, \dots, \tilde{p}_{r-1}$ линейно-независимы. Действительно, если

$$\gamma_0 p_0 + \dots + \gamma_{r-1} p_{r-1} = 0,$$

то, составляя скалярные произведения с \tilde{p}_i , получим

$$\gamma_i(p_i, \tilde{p}_i) = 0$$

и так как $(p_i, \tilde{p}_i) \neq 0$, то $\gamma_i = 0$, $i = 0, 1, \dots, r-1$. Таким же образом убеждаемся в линейной независимости векторов $\tilde{p}_0, \tilde{p}_1, \dots, \tilde{p}_{r-1}$. Отсюда следует, что $r \leq n$, т. е. процесс построения векторов p_0, p_1, \dots и $\tilde{p}_0, \tilde{p}_1, \dots$ должен оборваться не позже чем на n -м шаге.

Будем считать, что биортогональный алгорифм протекает нормально, если он обрывается на n -м шаге, и является вырожденным, если он обрывается раньше.

При нормальном течении процесса как вектор p_n , так и вектор \tilde{p}_n равны нулю. Действительно, вектор p_n ортогонален к n линейно-независимым векторам $\tilde{p}_0, \tilde{p}_1, \dots, \tilde{p}_{n-1}$, а вектор \tilde{p}_n ортогонален к n линейно-независимым векторам p_0, p_1, \dots, p_{n-1} .

Покажем, как решается полная проблема собственных значений при нормальном течении биортогонального алгорифма.

Сначала рассмотрим простейший случай, когда собственные значения матрицы A различны.

Пусть U_1, \dots, U_n — собственные векторы матрицы A , V_1, \dots, V_n — собственные векторы матрицы A' и пусть

$$p_0 = a_1 U_1 + \dots + a_n U_n$$

$$\tilde{p}_0 = \tilde{a}_1 V_1 + \dots + \tilde{a}_n V_n.$$

Легко видеть, что как коэффициенты a_1, \dots, a_n , так и коэффициенты $\tilde{a}_1, \dots, \tilde{a}_n$ все отличны от нуля, ибо если бы среди коэффициентов a_1, \dots, a_n (или $\tilde{a}_1, \dots, \tilde{a}_n$) нашлись нулевые, то векторы p_0, \dots, p_{n-1} (или $\tilde{p}_0, \dots, \tilde{p}_{n-1}$) были бы линейно-зависимы, так как все они укладывались бы в подпространство меньшей чем n размерности.

В силу предположения о нормальном течении биортогонального алгорифма мы построим две системы базисных векторов p_0, p_1, \dots, p_{n-1} и $\tilde{p}_0, \tilde{p}_1, \dots, \tilde{p}_{n-1}$ (двойственные, с точностью до нормирования базиса) и одновременно систему полиномов $p_0(t), p_1(t), \dots, p_{n-1}(t), p_n(t)$, причем полином $p_n(t)$ будет характеристическим. Пусть его корни найдены. Будем искать собственные векторы для матриц A и A' в виде

$$U_{i0} = c_{i0} p_0 + \dots + c_{in-1} p_{n-1} \quad (4)$$

$$V_{i0} = \tilde{c}_{i0} \tilde{p}_0 + \dots + \tilde{c}_{in-1} \tilde{p}_{n-1}.$$

Таблица VI.12

Решение линейной системы методом А-минимальных итераций

	q_0	Aq_0	q_1	Aq_1	q_2	Aq_2	q_3	Aq_3	X
I	1	2.62	0.302	0.11684	-0.06303198	-0.01323917	0.000105619	-0.001164594	-1.2577936
I	1	2.18	-0.138	-0.08644	-0.01622351	-0.00529574	0.007640798	0.004960303	0.0434874
I	1	2.08	-0.238	-0.11684	0.00841478	-0.01345978	-0.006534421	-0.004349251	1.0391662
I	1	2.32	0.002	0.08824	0.07888289	0.03199469	-0.001440546	0.000553541	1.4823928
II		9.20	0	0	$0.78 \cdot 10^{-8}$ $0.38 \cdot 10^{-9}$	$1 \cdot 10^{-9}$	$-0.98 \cdot 10^{-10}$ $-0.13 \cdot 10^{-9}$ $-0.59 \cdot 10^{-10}$	$-1 \cdot 10^{-9}$	
III	9.2	21.3256	0.075627200	0.042985206	0.0033309791 0.0033309789	0.0014081463	0.000065400106 0.000065400117		
IV	2.318 0.008220348		0.56838288 0.04404472		0.42274246				
V	0.24 0.26096957		-0.1432 -1.8934986		0.049863598 14.969652			-0.0020185014 -30.863885	

В силу ортогональных свойств базисных векторов

$$\begin{aligned} c_{ik} &= \frac{(U_i, \tilde{p}_k)}{(\tilde{p}_k, \tilde{p}_k)} = \tilde{a}_i \frac{p_k(\lambda_i)}{(\tilde{p}_k, \tilde{p}_k)} \\ \tilde{c}_{ik} &= \frac{(V_i, p_k)}{(p_k, \tilde{p}_k)} = a_i \frac{p_k(\lambda_i)}{(p_k, \tilde{p}_k)}. \end{aligned} \quad (5)$$

Отбрасывая неравные нулю множители, получаем

$$U_i = \frac{p_0(\lambda_i)}{(p_0, \tilde{p}_0)} \tilde{p}_0 + \dots + \frac{p_{n-1}(\lambda_i)}{(p_{n-1}, \tilde{p}_{n-1})} \tilde{p}_{n-1} \quad (6)$$

и

$$V_i = \frac{p_0(\lambda_i)}{(p_0, \tilde{p}_0)} \tilde{p}_0 + \dots + \frac{p_{n-1}(\lambda_i)}{(p_{n-1}, \tilde{p}_{n-1})} \tilde{p}_{n-1}.$$

В качестве иллюстрации решим полную проблему собственных значений для матрицы Леверье. Вычисление двойственных базисов и коэффициентов характеристического полинома дано в табл. VI.5, структура которой совпадает со структурой табл. VI.1 метода минимальных итераций. Конечно, для контроля кроме выполнения условий биортогональности целесообразно вычислять теоретически равные значения (Ap_i, \tilde{p}_i) и $(p_i, A'\tilde{p}_i)$, а также (p_i, \tilde{p}_i) по различным формулам. Заключительный контроль осуществляется вычислением

$$\sum_{i=0}^{n-1} \alpha_i = \text{Sp } A. \quad (7)$$

Из табл. VI.5 видно, что характеристический полином равен

$$\begin{aligned} \varphi(t) = p_4(t) = t^4 + 47.888430t^3 + 797.278777t^2 + \\ + 5349.4556t + 12296.551. \end{aligned}$$

Таким образом, значения для коэффициентов характеристического полинома, найденные по биортогональному алгорифму, почти совпадают с соответствующими значениями, найденными по эскалаторному методу (см. § 48). Имеем далее

$$\begin{aligned} \lambda_1 &= -17.863262; \quad \lambda_2 = -17.152427; \quad \lambda_3 = -7.574044; \\ \lambda_4 &= -5.298698. \end{aligned}$$

Для вычисления собственных векторов предварительно вычисляем множители разложения в табл. VI.6, которая не требует пояснения.

Далее вычисляем собственные векторы матриц A и A' , нормируя их в соответствии с нормировкой эскалаторного метода (табл. VI.7 и VI.8).

Таблица VI.6

Вычисление множителей разложения

t	1	2	3	4
λ_4	-17.863262	-17.152427	-7.574044	-5.298698
$p_0(\lambda_i)$	1	1	1	1
$\lambda_i - \alpha_0 = p_1(\lambda_i)$	-10.359641	-9.648806	-0.070423	2.204923
$\lambda_i - \alpha_1$	-4.403603	-3.692768	5.885615	8.160961
$p_2(\lambda_i)$	43.857113	33.868169	-2.177116	16.231657
$\lambda_i - \alpha_2$	-10.285383	-9.574548	0.003835	2.279181
$p_3(\lambda_i)$	-234.24261	-122.30678	1.465722	-9.157839
$\lambda_i - \alpha_3$	1.484009	2.194844	11.773227	14.048573
$p_4(\lambda_i)$	0.000000250	-0.0000468	0.000125	0.0000109
$p'_4(\lambda_i)$	-91.8	80.7	-224	338
$p_0(\lambda_i)(p_0, \tilde{p}_0)$	1	1	1	1
$p_1(\lambda_i)(p_1, \tilde{p}_1)$	-5.8773660	-5.4740858	-0.03995329	1.2509255
$p_2(\lambda_i)(p_2, \tilde{p}_2)$	1.1887052	0.9179644	-0.05900865	0.4399436
$p_3(\lambda_i)(p_3, \tilde{p}_3)$	0.8010094	0.4182368	-0.00501214	0.0313159

Таблица VI.7

Собственные векторы матрицы A

U_1	U_2	U_3	U_4
-0.019873	0.032933	-0.351235	1.135218
0.169806	-0.261308	0.328468	0.112182
-0.187214	0.236639	0.260923	0.070589
0.808482	0.586694	0.045007	0.011058

Сравнение с данными эскалаторного метода показывает хорошее совпадение.

Биортогональный алгорифм не усложняется, если собственные значения матрицы комплексны.

Для примера рассмотрим в табл. VI.9 матрицу $A = \begin{bmatrix} 4 & 3 \\ -3 & 4 \end{bmatrix}$, собственные значения которой $\lambda_1 = 4 + 3i$, $\lambda_2 = 4 - 3i$.

Таблица VI. 8

Собственные векторы матрицы A'

V_1	V_2	V_3	V_4
-0.014059	0.023297	-0.248476	0.803091
0.780381	-1.200904	1.509558	0.515562
-1.140762	1.441927	1.589924	0.430126
0.808484	0.586694	0.045006	0.011058

Таблица VI. 9

Вычисление двойственной пары базисов и коэффициентов характеристического полинома

	p_0	\tilde{p}_0	Ap_0	$A\tilde{p}_0$	p_1	\tilde{p}_1	Ap_1	$A\tilde{p}_1$	p_2	\tilde{p}_2
I	1	1	7	4	0	-3	-18	-21	0	0
	1	0	1	3	-6	3	-24	3	0	0
II			8	7	0	0	-42	-18		
III	1		7	7	-18	-18	-18	-18		
IV	7				1					
V	-18									
					1				1	
					-7				-8	
									25	

На основании табл. VI. 9 заключаем, что $p_2(t) = t^2 - 8t + 25$, откуда $\lambda_1 = 4 + 3i$, $\lambda_2 = 4 - 3i$. Коэффициенты разложения для U_1 и U_2 будут 1, $\frac{-3+3i}{-18}$ и 1, $\frac{-3-3i}{-18}$. Отсюда

$$U_1 = (1, t)', \quad U_2 = (1, -t)'.$$

Будем теперь считать, что среди собственных значений матрицы A (следовательно, и A') есть равные. Соответствующие элементарные

делители должны быть взаимно просты, ибо в противном случае минимальный полином матрицы A (и A') не совпадал бы с характеристическим и это привело бы к вырождению биортогонального алгорифма не позже, чем на m -м шаге (где m — степень минимального полинома), что противоречит предположению о невырожденном течении алгорифма. Биортогональный алгорифм позволяет определить весь канонический базис пространства. Именно, пусть $\lambda_1, \dots, \lambda_s$ — различные собственные значения матрицы с кратностями n_1, n_2, \dots, n_s , $n_1 + \dots + n_s = n$. Тогда в канонической форме Жордана будет s ящиков с порядками n_1, \dots, n_s . Каждому ящику соответствует канонический базис $U_i^{(0)}, \dots, U_i^{(n_i-1)}$ при $i = 1, 2, \dots, s$, причем

$$\begin{aligned} AU_i^{(0)} - \lambda_i U_i^{(0)} &= 0 \\ AU_i^{(1)} - \lambda_i U_i^{(1)} &= U_i^{(0)} \\ \cdots &\cdots \cdots \cdots \\ AU_i^{(n_i-1)} - \lambda_i U_i^{(n_i-1)} &= U_i^{(n_i-2)}. \end{aligned}$$

Вектор $U_i = U_i^{(0)}$ есть собственный вектор, принадлежащий собственному значению λ_i .

Проведем биортогональный алгорифм. Ясно, что последний полином $p_n(t)$, полученный по ходу биортогонального алгорифма, будет характеристическим полиномом матрицы A (и A'). Собственные векторы матрицы A (и A') также будут определяться по прежним формулам (6), которые, однако, требуют обоснования. Легче всего доказать их справедливость непосредственной проверкой. Пусть

$$U_i = \frac{p_0(\lambda_i)}{(p_0, p_0)} p_0 + \dots + \frac{p_{n-1}(\lambda_i)}{(p_{n-1}, p_{n-1})} p_{n-1}.$$

В силу соотношения

$$Ap_i = p_{i+1} + \alpha_i p_i + \beta_i p_{i-1}$$

имеем

$$\begin{aligned} AU_i &= \frac{p_0(\lambda_i)}{(p_0, p_0)} (p_1 + \alpha_0 p_0) + \frac{p_1(\lambda_i)}{(p_1, p_1)} (p_2 + \alpha_1 p_1 + \beta_1 p_0) + \dots + \\ &+ \frac{p_{n-1}(\lambda_i)}{(p_{n-1}, p_{n-1})} (\alpha_{n-1} p_{n-1} + \beta_{n-1} p_{n-2}) = \left[\frac{\alpha_0 p_0(\lambda_i)}{(p_0, p_0)} + \frac{\beta_1 p_1(\lambda_i)}{(p_1, p_1)} \right] p_0 + \\ &+ \left[\frac{p_0(\lambda_i)}{(p_0, p_0)} + \frac{\alpha_1 p_1(\lambda_i)}{(p_1, p_1)} + \frac{\beta_2 p_2(\lambda_i)}{(p_2, p_2)} \right] p_1 + \dots + \\ &+ \left[\frac{p_{n-2}(\lambda_i)}{(p_{n-2}, p_{n-2})} + \frac{\alpha_{n-1} p_{n-1}(\lambda_i)}{(p_{n-1}, p_{n-1})} \right] p_{n-1}. \end{aligned}$$

Hg

$$\frac{\alpha_0 p_0(\lambda_i)}{(p_0, \tilde{p}_0)} + \frac{\beta_1 p_1(\lambda_i)}{(p_1, \tilde{p}_1)} = \frac{\alpha_0 + p_1(\lambda_i)}{(p_0, \tilde{p}_0)} = \frac{\alpha_0 + \lambda_i - \alpha_0}{(p_0, \tilde{p}_0)} = \lambda_i \frac{1}{(p_0, \tilde{p}_0)} = \lambda_i \frac{p_0(\lambda_i)}{(p_0, \tilde{p}_0)}$$

$$\frac{p_0(\lambda_i)}{(p_0, \tilde{p}_0)} + \frac{\alpha_1 p_1(\lambda_i)}{(p_1, \tilde{p}_1)} + \frac{\beta_2 p_2(\lambda_i)}{(p_2, \tilde{p}_2)} = \frac{\beta_1 p_0(\lambda_i) + \alpha_1 p_1(\lambda_i) + p_2(\lambda_i)}{(p_1, \tilde{p}_1)} = \lambda_i \frac{p_1(\lambda_i)}{(p_1, \tilde{p}_1)}$$

$$\dots$$

$$\frac{p_{n-2}(\lambda_i)}{(p_{n-2}, \tilde{p}_{n-2})} + \frac{\alpha_{n-1} p_{n-1}(\lambda_i)}{(p_{n-1}, \tilde{p}_{n-1})} = \frac{\beta_{n-1} p_{n-2}(\lambda_i) + \alpha_{n-1} p_{n-1}(\lambda_i)}{(p_{n-1}, \tilde{p}_{n-1})} = \lambda_i \frac{p_{n-1}(\lambda_i)}{(p_{n-1}, \tilde{p}_{n-1})}.$$

Следовательно, $AU_i = \lambda_i U_i$, т. е. U_i действительно есть собственный вектор, принадлежащий собственному значению λ_i .

Точно так же верны формулы

$$V_i = \frac{p_0(\lambda_i)}{(p_0, \tilde{p}_0)} \tilde{p}_0 + \dots + \frac{p_{n-1}(\lambda_i)}{(p_{n-1}, \tilde{p}_{n-1})} \tilde{p}_{n-1} \quad (i=1, 2, \dots, s)$$

для собственных векторов матрицы A' .

Покажем теперь, что векторы

являются корневыми векторами, принадлежащими собственному значению λ_i , образующими канонический базис соответствующего подпространства.

Проверим первую из этих формул. Для упрощения выкладки введем обозначения

$$d_1 = \frac{p'_1(\lambda_i)}{(p_1, \tilde{p}_1)}, \dots, d_{n-1} = \frac{p'_{n-1}(\lambda_i)}{(p_{n-1}, \tilde{p}_{n-1})}.$$

Имеем

$$\begin{aligned} AU_i^{(1)} - \lambda_i U_i^{(1)} &= d_1 A p_1 + \dots + d_{n-1} A p_{n-1} - \lambda_i d_1 p_1 - \dots - \lambda_i d_{n-1} p_{n-1} = \\ &= d_1 \beta_1 p_0 + (d_1 \alpha_1 + d_2 \beta_2 - \lambda_i d_1) p_1 + (d_1 + d_2 \alpha_2 + d_3 \beta_3 - \lambda_i d_2) p_2 + \\ &\quad + \dots + (d_{n-2} + d_{n-1} \alpha_{n-1} - \lambda_i d_{n-1}) p_{n-1} = \\ &= b_0 p_0 + b_1 p_1 + \dots + b_{n-1} p_{n-1}. \end{aligned}$$

Но

$$\begin{aligned} b_0 &= d_1 \beta_1 = \frac{p'_1(\lambda_i)}{(p_1, \tilde{p}_1)} \frac{(p_1, \tilde{p}_1)}{(p_0, \tilde{p}_0)} = \frac{1}{(p_0, \tilde{p}_0)} = \frac{p_0(\lambda_i)}{(p_0, \tilde{p}_0)} \\ b_1 &= d_1 \alpha_1 + d_2 \beta_2 - \lambda_i d_1 = \frac{p'_2(\lambda_i)}{(p_2, \tilde{p}_2)} \frac{(p_2, \tilde{p}_2)}{(p_1, \tilde{p}_1)} + \frac{(\alpha_1 - \lambda_i) p'_1(\lambda_i)}{(p_1, \tilde{p}_1)} = \\ &= \frac{1}{(p_1, \tilde{p}_1)} [p'_2(\lambda_i) + (\alpha_1 - \lambda_i) p'_1(\lambda_i)] = \\ &= \frac{1}{(p_1, \tilde{p}_1)} [p'_2(\lambda_i) + (\alpha_1 - \lambda_i)] = \frac{p_1(\lambda_i)}{(p_1, \tilde{p}_1)}, \end{aligned}$$

ибо $p'_2(t) = t - \alpha_1 + p_1(t)$. Далее,

$$\begin{aligned} b_2 &= d_1 + d_2 \alpha_2 + d_3 \beta_3 - \lambda_i d_2 = \\ &= \frac{1}{(p_2, \tilde{p}_2)} [\beta_2 + (\alpha_2 - \lambda_i) p'_2(\lambda_i) + p'_3(\lambda_i)] = \frac{p_2(\lambda_i)}{(p_2, \tilde{p}_2)}, \end{aligned}$$

ибо $p'_3(t) = (t - \alpha_2) p'_2(t) - \beta_2 + p_2(t)$.

Аналогично получаем

$$b_k = \frac{p_k(\lambda_i)}{(p_k, \tilde{p}_k)} \quad (k = 3, \dots, n-2).$$

Наконец,

$$\begin{aligned} b_{n-1} &= d_{n-2} + d_{n-1} \alpha_{n-1} - \lambda_i d_{n-1} = \frac{(\alpha_{n-1} - \lambda_i) p'_{n-1}(\lambda_i)}{(p_{n-1}, \tilde{p}_{n-1})} + \frac{p'_{n-2}(\lambda_i)}{(p_{n-2}, \tilde{p}_{n-2})} = \\ &= \frac{1}{(p_{n-1}, \tilde{p}_{n-1})} [(\alpha_{n-1} - \lambda_i) p'_{n-1}(\lambda_i) + \beta_{n-1} p'_{n-2}(\lambda_i)] = \frac{p_{n-1}(\lambda_i)}{(p_{n-1}, \tilde{p}_{n-1})}, \end{aligned}$$

ибо

$$p'_n(t) = p_{n-1}(t) + (t - \alpha_{n-1}) p'_{n-1}(t) - \beta_{n-1} p'_{n-2}(t) \quad \text{и} \quad p'_n(\lambda_i) = 0,$$

так как λ_i есть n_i -кратный корень полинома $p_n(t)$, $n_i \geq 2$. Таким образом,

$$AU_i^{(1)} - \lambda_i U_i^{(1)} = U_i^{(0)},$$

что и требовалось доказать.

Остальные формулы проверяются аналогично.

В качестве примера рассмотрим матрицу $A = \begin{bmatrix} 4 & 5 & -2 \\ -2 & -2 & 1 \\ -1 & -1 & 1 \end{bmatrix}$,

нормальная форма которой состоит из одного канонического ящика, принадлежащего трехкратному собственному значению $\lambda_1 = 1$.

Таблица VI. 10

Вычисление двойственной пары базисов и коэффициентов характеристического полинома

	p_0	\tilde{p}_0	Ap_0	$A'\tilde{p}_0$	p_1	\tilde{p}_1	Ap_1	$A'\tilde{p}_1$	p_2	\tilde{p}_2	Ap_2	$A'\tilde{p}_2$	p_3	\tilde{p}_3
I	1	1	4	4	0	0	-8	-8	0	0	0	0	0	0
	0	0	-2	5	-2	5	3	-8	-1/4	1/8	-1/8	0	0	0
	0	0	-1	-2	-1	-2	1	3	-5/8	-1/4	-3/8	-1/8	0	0
II			1	7	0		-4	-13	0	0	-4/8	-1/8		
									0	0				
III	1	4	4	-8	-8	13	13	1/8	1/8	5/64	5/64			
IV	4			-13/8				5/8						
	-8			-1/64										
V									1				1	
	1				1				-19/8				-3	
					-4				3/2				3	
													-1	

Согласно табл. VI.10, $p_3(t) = t^3 - 3t^2 + 3t - 1$, $\lambda_1 = \lambda_2 = \lambda_3 = 1$.
Далее вычисляем

$$p_0(1) = 1 \quad p'_0(1) = 0 \quad p''_0(1) = 0$$

$$p_1(1) = -3 \quad p'_1(1) = 1 \quad p''_1(1) = 0$$

$$p_2(1) = \frac{1}{8} \quad p'_2(1) = -\frac{3}{8} \quad p''_2(1) = 2$$

$$U_1^{(0)} = \frac{1}{1} p_0 + \frac{-3}{-8} p_1 + \frac{\frac{1}{8}}{\frac{1}{8}} p_2 = (1, -1, -1)'$$

$$U_1^{(1)} = \frac{1}{-8} p_1 + \frac{-\frac{3}{8}}{\frac{1}{8}} p_2 = (0, 1, -2)'$$

$$U_1^{(2)} = \frac{2}{2! \frac{1}{8}} p_2 = (0, -2, -5)'.$$

Аналогично

$$V_1^{(0)} = \tilde{p}_0 + \frac{3}{8} \tilde{p}_1 + \tilde{p}_2 = (1, -2, -1)'$$

$$V_1^{(1)} = -\frac{1}{8} \tilde{p}_1 - 3\tilde{p}_2 = (0, -1, 1)'$$

$$V_1^{(2)} = 8\tilde{p}_2 = (0, 1, -2)'.$$

В качестве второго примера рассмотрим матрицу, каноническая форма которой состоит из двух ящиков (один из которых первого порядка), а именно матрицу

$$A = \begin{bmatrix} 13 & 16 & 16 \\ -5 & -7 & -6 \\ -6 & -8 & -7 \end{bmatrix}.$$

Ее собственными значениями будут $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = -3$.

Таблица VI.II

Вычисление двойственной пары базисов и коэффициентов характеристического полинома

	p_0	\tilde{p}_0	Ap_0	$A'\tilde{p}_0$	p_1	\tilde{p}_1	Ap_1	$A'\tilde{p}_1$	p_2	\tilde{p}_2	Ap_2	$A'\tilde{p}_2$	p_3	\tilde{p}_3
I	1	1	13	13	0	0	-176	-176	0	0	0	0	0	0
	0	0	-5	16	-5	16	71	-240	16/11	-192/11	-16/11	64/11	0	0
	0	0	-6	16	-6	16	82	-208	-16/11	160/11	-16/11	32/11	0	0
II			2	45	0	0	-23	-624	0	0	-32/11	96/11		
III	1		13	13	-176	-176	2448	2448	-512/11	-512/11	512/121	512/121		
IV	13			-153/11					-1/11					
	-176			32/121										
V	1			1					1				1	1
				-13					10/11				-5	3
									-53/11					

Согласно табл. VI.11, $p_3(t) = t^3 + t^2 - 5t + 3$, $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = -3$.

Далее вычисляем

$$p_0(\lambda_1) = 1 \quad p'_0(\lambda_1) = 0 \quad p_0(\lambda_2) = 1$$

$$p_1(\lambda_1) = -12 \quad p'_1(\lambda_1) = 1 \quad p_1(\lambda_2) = -16$$

$$p_2(\lambda_1) = -\frac{32}{11} \quad p'_2(\lambda_1) = \frac{32}{11} \quad p_2(\lambda_2) = \frac{16}{11}$$

и

$$U_1^{(0)} = \left(1, -\frac{1}{4}, -\frac{1}{2}\right)'; \quad U_1^{(1)} = (0, 1, -2)'; \quad U_2^{(0)} = \left(1, -\frac{1}{2}, -\frac{1}{2}\right)'.$$

2. Вырожденное течение алгорифма. Изучим возможные причины обрыва алгорифма. Допустим, что алгорифм обрывается на r -м шагу, так что $(p_i, \tilde{p}_i) \neq 0$, $i = 0, \dots, r-1$ и $(p_r, \tilde{p}_r) = 0$. Как мы видели, системы векторов p_0, \dots, p_{r-1} и $\tilde{p}_0, \dots, \tilde{p}_{r-1}$ линейно-независимы. Ввиду того, что векторы $p_0, Ap_0, \dots, A^{r-1}p_0$ линейно выражаются через p_0, p_1, \dots, p_{r-1} и обратно, заключаем, что векторы $p_0, Ap_0, \dots, A^{r-1}p_0$ линейно-независимы. Точно так же линейно-независимыми будут и векторы $\tilde{p}_0, A'\tilde{p}_0, \dots, A'^{r-1}\tilde{p}_0$.

Обозначим через P_r подпространство, натянутое на векторы p_0, p_1, \dots, p_{r-1} (или, что то же самое, на векторы $p_0, Ap_0, \dots, A^{r-1}p_0$), через \tilde{P}_r подпространство, натянутое на векторы $\tilde{p}_0, \tilde{p}_1, \dots, \tilde{p}_{r-1}$, через Q_r и \tilde{Q}_r их ортогональные дополнения.

Тогда $P_r \cap \tilde{Q}_r = 0$ и $\tilde{P}_r \cap Q_r = 0$. Действительно, из условий $(p_i, \tilde{p}_j) = 0$, при $i \neq j$ и $(p_i, p_i) \neq 0$, следует, что в P_r не существует ни одного вектора, кроме нулевого, ортогонального ко всем векторам из \tilde{P}_r , и в \tilde{P}_r не существует ни одного вектора, кроме нулевого, ортогонального ко всем векторам из P_r .

При обрыве алгорифма на r -м шаге возможны четыре случая.

1. $p_r = \tilde{p}_r = 0$ (двусторонний обрыв алгорифма). В этом случае $p_r = p_r(A)p_0 = 0$, но в силу линейной независимости векторов $p_0, Ap_0, \dots, A^{r-1}p_0$ ни при каком полиноме $\omega(f)$ ниже r -й степени невозможно $\omega(A)p_0 = 0$. Следовательно, полином $p_r(t)$ есть минимальный полином, аннулирующий вектор p_0 . Так же устанавливается, что $\tilde{p}_r(t)$ есть минимальный полином, аннулирующий вектор \tilde{p}_0 (по отношению к матрице A').

Подпространство P_r будет инвариантным относительно A подпространством, именно циклическим подпространством, порожденным вектором p_0 . Подпространство \tilde{P}_r будет инвариантным относительно A' . Соответственно \tilde{Q}_r будет инвариантным для A , Q_r — инвариантным для A' . Размерности подпространств P_r и \tilde{P}_r одинаковы, следовательно сумма размерностей P_r и \tilde{Q}_r равна размерности всего пространства и, так как $P_r \cap \tilde{Q}_r = 0$, получаем, что все пространство есть прямая сумма инвариантных (относительно A) подпространств P_r и \tilde{Q}_r . Таким же образом, все пространство есть прямая сумма инвариантных (относительно A') подпространств \tilde{P}_r и Q_r .

Итак, в случае двустороннего обрыва алгорифм дает минимальный аннулирующий векторы p_0 и \tilde{p}_0 полином. При этом оказы-

вается возможным разложение пространства в прямую сумму двух инвариантных подпространств, из которых одно циклическое.

2. $p_r = 0, \tilde{p}_r \neq 0$ (односторонний обрыв). В этом случае $p_r(t)$ будет минимальным аннулирующим полиномом для p_0 , но не будет минимальным аннулирующим полиномом для \tilde{p}_0 . Подпространство \tilde{P}_r не будет инвариантным для A' и \tilde{Q}_r не будет инвариантным для A . Хотя по-прежнему $R = P_r + \tilde{Q}_r$, но это разложение не является разложением в прямую сумму инвариантных подпространств.

Аналогичный результат имеет место в случае

3. $p_r \neq 0, \tilde{p}_r = 0$.

В последнем случае

4. $p_r \neq 0, \tilde{p}_r \neq 0$, но $(p_r, \tilde{p}_r) = 0$ (тупиковый обрыв). алгорифм как бы заходит в тупик, так как в результате проведения алгорифма мы не получаем ни какого-либо делителя характеристического полинома, ни инвариантного подпространства.

Можно показать, что односторонние и тупиковые обрывы алгорифма являются исключительными и их можно избежать посредством надлежащего выбора начальных векторов p_0 и \tilde{p}_0 . Именно, векторы p_0 и \tilde{p}_0 всегда можно выбрать так, что в результате проведения алгорифма будет определен минимальный полином матрицы A и канонический базис циклического подпространства, порожденного начальным вектором.

В случае двустороннего обрыва всегда можно осуществить до-стройку полученных систем векторов p_0, \dots, p_{r-1} и $\tilde{p}_0, \dots, \tilde{p}_{r-1}$ до биортогональных базисов пространства. Достройка делается со-вершенно аналогично тому, как это делается в методе минимальных итераций.

§ 65. Метод A -минимальных итераций

Предположим, что матрица A положительно определена. Тогда, как мы видели (теорема 11.19), любая система линейно-независимых векторов может быть подвергнута процессу A -ортогонализации. Если процесс A -ортогонализации провести для системы векторов $q_0, Aq_0, \dots, A^{n-1}q_0$, мы приедем к методу A -минимальных итераций.

Этот метод может быть применен для решения полной проблемы собственных значений, аналогично описанному выше методу минимальных итераций. В применении же к решению системы уравнений с матрицей A метод A -минимальных итераций оказывается более удобным.

Действительно, пусть q_0, \dots, q_{n-1} полученный в процессе A -ортогонализации базис. Если решение системы

$$AX = F \quad (1)$$

искать в виде

$$X = \sum_{i=0}^{n-1} a_i q_i, \quad (2)$$

то из уравнения

$$\sum_{i=0}^{n-1} a_i A q_i = F$$

получим для коэффициентов a_i явные формулы

$$a_i = \frac{(F, q_i)}{(A q_i, q_i)}. \quad (3)$$

Дадим формулы для вычисления обратной матрицы. Условия A -ортогональности векторов q_0, \dots, q_{n-1} могут быть записаны в матричной форме

$$Q' A Q = \Lambda, \quad (4)$$

где Q — матрица со столбцами q_0, \dots, q_{n-1} , $\Lambda = [(q_0, A q_0), \dots, (q_{n-1}, A q_{n-1})]$. Из формулы (4) непосредственно следует, что

$$A^{-1} = Q \Lambda^{-1} Q'. \quad (5)$$

Легко видеть, что система A -ортогональных (сопряженных) векторов строится по трехчленным рекуррентным соотношениям

$$q_{i+1} = A q_i - \gamma_i q_i - \delta_i q_{i-1}, \quad q_1 = A q_0 - \gamma_0 q_0, \quad (6)$$

где

$$\gamma_i = \frac{(A q_i, A q_i)}{(A q_i, q_i)} = \frac{(A q_i, A q_i)}{(A q_i, A q_{i-1})}, \quad (7)$$

$$\delta_i = \frac{(A q_i, A q_{i-1})}{(A q_{i-1}, q_{i-1})} = \frac{(A q_i, q_i)}{(A q_{i-1}, q_{i-1})}.$$

Этот процесс построения векторов q_i может оборваться только если некоторый вектор q_r окажется нулевым. Как правило, это происходит при $r = n$. Преждевременный обрыв свидетельствует о линейной зависимости последовательных итераций $q_0, A q_0, \dots, A^r q_0$, т. е. о том, что минимальный аннулирующий вектор q_0 полином не совпадает с характеристическим. В силу построения

$$q_i = q_i(A) q_0, \quad (8)$$

где $q_i(t)$ полиномы, удовлетворяющие рекуррентным соотношениям

$$q_0 = 1, \quad q_1(t) = t - \gamma_0, \quad q_{i+1}(t) = (t - \gamma_i) q_i(t) - \delta_i q_{i-1}(t). \quad (9)$$

Полиномы $q_i(t)$ будут, очевидно, удовлетворять соотношениям ортогональности

$$\int_m^M q_i(t) q_j(t) t dF(t) = 0 \quad (i \neq j).$$

где $F(t)$ весовая функция для полиномов $p_i(t)$ в методе минимальных итераций.

Векторы q_0, \dots, q_{n-1} обладают рядом свойств, аналогичных свойствам векторов p_0, \dots, p_{n-1} .

В частности, корни полиномов $q_0(t), \dots, q_{n-1}(t), q_n(t)$ вещественны и разделяются; векторы q_0, \dots, q_{i-1} составляют A -ортогональный базис подпространства Q_i , натянутого на векторы $q_0, Aq_0, \dots, A^{i-1}q_0$; среди всех векторов вида $A^i q_0 + Z$, где $Z \in Q_i$, вектор q_i имеет наименьшую A -длину.

Далее, если процесс протекает без вырождения, то оператор с матрицей A имеет в базисе q_0, \dots, q_{n-1} трехдиагональную матрицу Якоби.

$$J = \begin{bmatrix} -\gamma_0 & \delta_0 & & & \\ 1 & \gamma_1 & \delta_1 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots & \delta_{n-1} \\ & & & 1 & \gamma_n & - \end{bmatrix}. \quad (10)$$

Отсюда следует, что

$$\sum_{i=0}^{n-1} \gamma_i = \operatorname{Sp} A. \quad (11)$$

Наконец, $(Aq_i, Aq_j) = 0$ при $|i - j| > 1$.

В случае преждевременного обрыва процесса всегда можно достроить систему A -ортогональных векторов до базиса всего пространства аналогично тому, как это делалось в методе минимальных итераций.

Такую достройку нужно производить, если ставится задача об обращении матрицы или о решении большой серии уравнений с данной матрицей коэффициентов A , но с различными свободными членами.

При решении же одной системы достройка оказывается ненужной, если за начальный вектор q_0 взят свободный член системы, так как при таком выборе начального вектора преждевременный обрыв процесса только сокращает объем вычислений.

Действительно, пусть процесс A -минимальных итераций обрывается на r -м шагу. Тогда $q_r(t) = t^r + d_1 t^{r-1} + \dots + d_{r-1} t + d_r$ есть минимальный аннулирующий вектор F полином, так что

$$A^r F + d_1 A^{r-1} F + \dots + d_{r-1} A F + d_r F = 0,$$

откуда $X = A^{-1} F = \frac{1}{d_r} [-A^{r-1} F - \dots - d_{r-1} F]$. Поэтому X принадлежит подпространству Q_r .

Векторы q_0, \dots, q_{r-1} образуют базис подпространства Q_r , так что решение X представляется в виде

$$X = \sum_{i=0}^{r-1} a_i q_i.$$

Для коэффициентов a_i сохраняются прежние формулы

$$a_i = \frac{(F, q_i)}{(Aq_i, q_i)} \quad (i = 0, 1, \dots, r-1).$$

С тем же успехом вместо свободного члена F в качестве начального вектора q_0 можно взять любую невязку $r_0 = F - AX_0$, где X_0 произвольный вектор. Действительно, система $AX = F$ равносильна системе $A(X - X_0) = r_0$, свободным членом которой является r_0 . В этом случае решение системы получается в форме

$$X = X_0 + \sum_{i=0}^{r-1} a_i q_i,$$

где

$$a_i = \frac{(r_0, q_i)}{(Aq_i, q_i)}.$$

Если матрица A симметрична, но не положительно-определенна, то построение векторов q_i также возможно. Однако в этом случае может произойти „тупиковое“ окончание процесса, в котором на некотором шагу может обратиться в нуль скалярное произведение (Aq_i, q_i) , хотя $q_i \neq 0$. Так же как в биортогональном алгоритме тупиковое окончание связано с неудачным выбором начального вектора и может быть устранено его заменой.

Рассмотрим примеры на решение системы и обращение матрицы. Отметим, что вычислительная схема для построения векторов q_0, \dots, q_{n-1} почти ничем не отличается от вычислительной схемы метода минимальных итераций. Хорошим заключительным контролем при невырожденном течении процесса является выполнение неравенства (11)

$$\sum_{i=0}^{n-1} \gamma_i = \text{Sp } A.$$

В табл. VI.12 приведено решение системы (9) § 23.

Части I—IV табл. VI.12 аналогичны соответствующим частям табл. VI.1. В части V записываются (F, q_i) и коэффициенты a_i .

Полученное решение (ср. § 23) верно с точностью до $2 \cdot 10^{-7}$ в каждой компоненте.

Таблица VI.12

Решение линейной системы методом \hat{A} -минимальных итераций

	q_0	Aq_0	q_1	Aq_1	q_2	Aq_2	q_3	Aq_3	X
I	1	2.62	0.302	0.11684	-0.06303198	-0.01323917	0.000105619	-0.001164594	-1.2577936
	1	2.18	-0.138	-0.08644	-0.01622351	-0.00529574	0.007640798	0.004960303	0.0434874
	1	2.08	-0.238	-0.11864	0.08841478	-0.01345978	-0.006534421	-0.004349251	1.0391662
	1	2.32	0.002	0.08824	0.07888289	0.03199469	-0.001440546	0.000553541	1.4823928
II									
		9.20	0	0	$0.78 \cdot 10^{-8}$	$1 \cdot 10^{-9}$	$-0.98 \cdot 10^{-10}$	$-1 \cdot 10^{-9}$	
					$0.38 \cdot 10^{-9}$		$-0.13 \cdot 10^{-9}$		
III	9.2	21.3256	0.075627200	0.042985206	0.0033309791	0.0014081463	0.000065400106		
					0.0033309789		0.000065400117		
IV	2.318		0.56838288		0.42274246				
	0.008220348		0.04404472						
V	0.24		-0.1432		0.049865598		-0.0020185014		
	0.26086957		-1.8934986		14.969652		-30.863885		

Вычислим также обратную матрицу для матрицы A . Целесообразно предварительно вычислить матрицу $\tilde{Q} = Q\Delta^{-1}$ с элементами $\tilde{q}_{ij} = \frac{q_{ij}}{(Aq_i, q_j)}$, где q_{ij} — есть j -я компонента вектора q_i . Имеем

$$\tilde{Q} = \begin{bmatrix} 0.10869565 & 3.9932723 & -18.922959 & 1.6149668 \\ 0.10869565 & -1.8247403 & -4.8704929 & 116.83158 \\ 0.10869565 & -3.1470159 & 2.5262182 & -99.914532 \\ 0.10869565 & 0.0264455 & 23.681593 & -22.0266662 \end{bmatrix}.$$

Тогда $A^{-1} = \tilde{Q}\tilde{Q}'$, согласно формуле (5). Вычисляя, получим, что

$$A^{-1} = \begin{bmatrix} 2.507586 & -0.123039 & -1.011489 & -1.378342 \\ -0.123039 & 1.322213 & -0.261427 & -0.447454 \\ -1.011489 & -0.261427 & 1.531827 & 0.445608 \\ -1.378342 & -0.447454 & 0.445608 & 2.008551 \end{bmatrix}.$$

Расхождение с данными § 23 наблюдается лишь в трех элементах, причем в каждом элементе не превосходит $1 \cdot 10^{-6}$.

В случае нормального течения алгорифма все собственные значения матрицы находятся как корни полинома $q_n(t)$. Для собственных векторов справедлива формула

$$U_i = \frac{q_0(\lambda_i)}{(q_0, Aq_0)} q_0 + \dots + \frac{q_{n-1}(\lambda_i)}{(Aq_{n-1}, q_{n-1})} q_{n-1}. \quad (12)$$

В случае обрыва алгорифма (не тупикового) последний полином даст минимальный аннулирующий q_0 полином, а собственные векторы, найденные по формуле

$$U_i = \frac{q_0(\lambda_i)}{(Aq_0, q_0)} q_0 + \dots + \frac{q_{r-1}(\lambda_i)}{(Aq_{r-1}, q_{r-1})} q_{r-1}. \quad (12')$$

образуют базис из собственных векторов подпространства Q_r . Процессом достройки можно определить все собственные значения и полную систему собственных векторов.

Определим коэффициенты характеристического полинома и один из собственных векторов рассмотренной выше матрицы. Приводим вычисления, проведенные на основе таблицы VI.12.

Прежде всего

$$(Aq_3, Aq_3) = 0.000045183277$$

$$\gamma_3 = 0.69087467$$

$$\delta_3 = 0.01963390,$$

Далее, вычисляем рекуррентно коэффициенты характеристического полинома:

			1
			—4.00000001
	1		—3.30912534
	1		4.75200003
	—2.88638288		2.48544305
1	—2.318	1.30929117	—0.45139731
			0.28615248

Таким образом,

$$\varphi(t) = q_4(t) = t^4 - 4.00000001t^3 + 4.75200003t^2 - 2.111856000t + 0.28615248.$$

Один из корней этого полинома есть $\lambda_3 = 0.63828371$. Определим собственный вектор, принадлежащий этому собственному значению. Предварительно вычислим коэффициенты разложения по схеме:

	λ_3	0.63828371
	$q_0(\lambda_3)$	1
$q_1(\lambda_3) = \lambda_3 - \gamma_0$	—1.6797163	
$\lambda_3 - \gamma_1$	0.06990083	
$q_2(\lambda_3)$	—0.12563391	
$\lambda_3 - \gamma_2$	0.21554125	
$q_3(\lambda_3)$	0.046903344	
$\lambda_3 - \gamma_3$	—0.05259096	
$q_4(\lambda_3)$	—0.8263 · 10 ⁻⁸	
$q'_4(\lambda_3)$	0.106	
$q_0(\lambda_3)/(q_0, q_0)$	0.10869565	
$q_1(\lambda_3)/(q_1, q_1)$	—22.210478	
$q_2(\lambda_3)/(q_2, q_2)$	—37.716811	
$q_3(\lambda_3)/(q_3, q_3)$	717.17535.	

Теперь собственный вектор находится по формуле (12). Именно,
 $U_3 = (-4.145756, 9.265433, 0.391085, -3.944060)' =$
 $= 9.265433(-0.447443, 1.000000, 0.042209, -0.425675)'.$

Наибольшее расхождение с данными § 63 не превосходит $1 \cdot 10^{-6}$.

Если матрица системы $AX = F$ не симметрична, то решение системы уравнений все же возможно найти посредством метода А-минимальных итераций, который нужно применять к равносильной системе, полученной первой или второй трансформацией Гаусса.

В обоих случаях расчетные формулы можно преобразовать так, чтобы избежать действительного вычисления соответствующих матриц $A'A$ или AA' .

При первой трансформации имеем

$$\begin{aligned} q_{i+1} &= A' A q_i - \gamma_i q_i - \delta_i q_{i-1} = A' v_i - \gamma_i q_i - \delta_i q_{i-1} \\ v_i &= A q_i, \\ \gamma_i &= \frac{(A' A q_i, A' A q_i)}{(A' A q_i, q_i)} = \frac{(A' v_i, A' v_i)}{(v_i, v_i)}, \\ \delta_i &= \frac{(A' A q_i, q_i)}{(A' A q_{i-1}, q_{i-1})} = \frac{(v_i, v_i)}{(v_{i-1}, v_{i-1})}, \\ X &= \sum_{i=0}^{n-1} \frac{(A' F, q_i)}{(A' A q_i, q_i)} q_i = \sum_{i=0}^{n-1} \frac{(F, v_i)}{(v_i, v_i)} q_i. \end{aligned} \quad (13)$$

Описанный метод будем называть методом $A'A$ -минимальных итераций.

Для контроля можно использовать попарную ортогональность векторов v . Действительно,

$$(v_i, v_j) = (A q_i, A q_j) = (A' A q_i, q_j) = 0.$$

Отметим, что

$$A^{-1} = V \Lambda^{-1} Q', \quad (14)$$

где матрицы V и Q составлены из столбцов v_0, v_1, \dots, v_{n-1} и q_0, q_1, \dots, q_{n-1} соответственно, $\Lambda = [(v_0, v_0), \dots, (v_{n-1}, v_{n-1})]$.

Аналогично вторая трансформация Гаусса приводит к методу AA' -минимальных итераций.

Расчетные формулы метода

$$\begin{aligned} w_i &= A' q_i \\ q_{i+1} &= Aw_i - \gamma_i q_i - \delta_i q_{i-1} \\ \gamma_i &= \frac{(Aw_i, Aw_i)}{(w_i, w_i)} \\ \delta_i &= \frac{(w_i, w_i)}{(w_{i-1}, w_{i-1})} \\ X &= A' \sum_{i=0}^{n-1} a_i q_i = \sum_{i=0}^{n-1} a_i w_i \\ a_i &= \frac{(F, q_i)}{(w_i, w_i)}. \end{aligned} \quad (15)$$

Легко видеть, что $(w_i, w_j) = 0$.

В табл. VI. 13 и VI. 14 приведено решение системы (по данным табл. II. 1) методами $A'A$ и AA' -минимальных итераций. Во второй части таблиц проверяется выполнение контрольных соотношений ортогональности.

Таблица VI. 13

решение системы уравнений методом АЛ-минимальных итераций

	q_6	v_6	$A'v_6$	q_1	v_1	$A'v_1$	q_2	v_2	$A'v_2$	q_3	v_3	$A'v_3$	X	
I	1	1.46	3.5474	0.6194927	0.45089331	0.34115826	0.01469831	-0.03114716	-0.03875003	-0.00027375	-0.03919682	0.44089864		
I	1	1.82	3.2484	0.3274927	0.26342271	0.16392188	-0.033065285	0.03809338	0.0472345	0.138603752	0.03886164	-0.3630309		
I	1	0.50	2.1963	-0.7311073	-0.13031765	-0.19253928	0.06783548	0.06653546	0.10756822	-0.07316131	-0.02858813	1.1667982		
I	1	2.34	2.1760	-0.7519073	-0.53614308	-0.31252088	-0.04308823	-0.02479610	-0.11624169	0.01788792	0.00039434	0.39305674		
II		6.12	11.1896			$-2 \cdot 10^{-8}$	$-2 \cdot 10^{-8}$			$-0.2 \cdot 10^{-8}$			$0.2 \cdot 10^{-8}$	
III		11.1696	32.708553		0.63981296	0.27800772			$-5 \cdot 10^{-8}$				$1 \cdot 10^{-8}$	
IV	2.9279073			0.43451405				0.010168550	0.029823601			0.0038797981		
V	0.057281636			0.015893004			2.6355666							
V	0.340986725	3.894	-0.40132318	-0.25677177		4.4707328	0.04546087					-3.1374083	-0.012172511	

Tablica VI. 14

Решение системы уравнений методом АА'-минимальных итераций

	q_0	w_0	Aw_0	q_1	w_1	Aw_1	q_2	w_2	Aw_2	q_3	w_3	Aw_3	X
I	1	0.45	0.3099	-0.45860226	-0.435668002	-0.19902223	0.1703978	0.05128257	0.025854965	0.009691839	0.023239767	0.4408884	
I	0	-0.26	-0.3100	-0.31000000	-0.43491841	-1.04409661	-0.23907255	-0.10313637	-0.038447305	0.009101047	0.010441822	-0.3630310	
I	0	-0.61	-0.4101	-0.41010000	-0.42292962	-0.80751843	0.25745049	0.06879630	0.014310165	-0.0174886523	-0.008946124	1.1667983	
I	-1	-0.46	-0.5439	0.22466226	0.37963104	-0.19902223	0.03929508	0.01722995	0.025854972	0.016521269	0.025784034	0.3935672	
II		-0.88	-0.9541		-0.26.10 ⁻⁸	-2.24965990			-0.73.10 ⁻⁸	0.027572797		0.143.10 ⁻⁸	
II									0.55.10 ⁻⁸			0.753.10 ⁻⁸	
II		0.8538	0.65614723		0.70140446	1.8214434			0.018295543	0.0030195349		-0.105.10 ⁻⁹	
IV	0.76860226			2.58668518				0.16506397				0.0015619834	
IV	0.82156909			0.026084155									
V	-0.6	-0.70274059		-0.37750864	-0.53821819		0.14710157	8.0402954		0.0073852512	4.7284873		

§ 66. A-биортогональный алгорифм

Наряду с описанными выше приемами симметризации для решения систем с несимметричной матрицей можно использовать видоизменение биортогонального алгорифма — A-биортогональный алгорифм. Сущность его заключается в построении двух систем векторов q_0, \dots, q_{n-1} и $\tilde{q}_0, \dots, \tilde{q}_{n-1}$ таких, что $(Aq_i, \tilde{q}_j) = 0$ при $i \neq j$ и $(Aq_i, \tilde{q}_i) \neq 0$ (двойственная пара сопряженных базисов).

Системы векторов строятся по рекуррентным соотношениям

$$\begin{aligned} q_1 &= Aq_0 - \gamma_0 q_0 & \tilde{q}_1 &= A'\tilde{q}_0 - \gamma_0 \tilde{q}_0 \\ q_{i+1} &= Aq_i - \gamma_i q_i - \delta_i q_{i-1} & \tilde{q}_{i+1} &= A'\tilde{q}_i - \gamma_i \tilde{q}_i - \delta_i \tilde{q}_{i-1}, \end{aligned} \quad (1)$$

где

$$\begin{aligned} \gamma_i &= \frac{(Aq_i, A'\tilde{q}_i)}{(Aq_i, \tilde{q}_i)} \\ \delta_i &= \frac{(Aq_i, \tilde{q}_i)}{(Aq_{i-1}, \tilde{q}_{i-1})}. \end{aligned} \quad (2)$$

После построения двойственной пары сопряженных базисов решение линейной системы вычисляется по формуле

$$X = \sum_{i=0}^{n-1} \frac{(F, \tilde{q}_i)}{(Aq_i, \tilde{q}_i)} q_i. \quad (3)$$

Мы не будем вдаваться в исследование возможных вырождений процесса, так как это было бы почти дословным повторением п. 2 § 64. Отметим только, что одностороннего и туликового окончания процесса можно избежать за счет перехода к другой системе начальных векторов.

A-биортогональный алгорифм дает возможность решить полную проблему собственных значений. Так, при нормальном течении алгорифма, полином $q_n(t)$ в последовательности полиномов, построенных по формулам

$$q_{i+1}(t) = tq_i(t) - \gamma_i q_i(t) - \delta_i q_{i-1}(t), \quad q_1(t) = t - \gamma_0, \quad (4)$$

есть характеристический полином. Найдя его корни, соответствующие собственные векторы можно построить по формулам

$$U_i = \frac{q_0(\lambda_i)}{(Aq_0, \tilde{q}_0)} q_0 + \dots + \frac{q_{n-1}(\lambda_i)}{(Aq_{n-1}, \tilde{q}_{n-1})} q_{n-1}. \quad (5)$$

Канонический базис, в случае, если матрица A не приводится к диагональной форме, находится по формулам, аналогичным формулам биортогонального алгорифма.

В табл. VI. 15 приведено решение системы с данными табл. II. 1 по A-биортогональному алгорифму.

Таблица VI.15

Решение системы линейных уравнений A -биортогональным алгоритмом

	q_0	\tilde{q}_0	$A\tilde{q}_0$	q_1	\tilde{q}_1	$A\tilde{q}_1$	$A\tilde{q}_1$
I	0	0	0.17	0.36	0.17	0.36	0.29078666
	1	1	1.00	1.85	-0.1406667	0.2093333	0.03613330
	0	1	0.35	1.67	0.35	0.5293333	-0.03613334
II	0	0	0.43	-1.06	0.43	-1.06	0.58901332
							-1.3242933
III	1.35	1.35	1.95	2.32	$-0.45 \cdot 10^{-7}$	$-0.45 \cdot 10^{-7}$	0.87979994
							-1.3091467
IV	1.1406667						
	-0.39850637						
V	1.2						
	0.88888889						
	q_3	\tilde{q}_3	$A\tilde{q}_3$	$A'\tilde{q}_3$	q_3	\tilde{q}_3	$A\tilde{q}_3$
I	0.02186708	-0.75231677	0.24093636	-0.50208664	0.16530023	0.7785489	-0.09938247
	0.653215750	0.06236619	0.29645727	0.00000002	-0.81490787	-0.14245783	-0.31154287
	-0.58979130	-0.24584873	-0.29645722	0.11091559	0.67245006	0.45252968	0.31154285
	-0.09119504	0.35249539	-0.01166529	0.10821921	0.07771139	-0.33270445	0.06029815
II	$0.452 \cdot 10^{-7}$	$0.213 \cdot 10^{-7}$	0.21987112	-0.28295182	$-0.149 \cdot 10^{-7}$	$-0.242 \cdot 10^{-7}$	-0.03908434
	$-0.284 \cdot 10^{-8}$	$-0.128 \cdot 10^{-7}$			$-0.284 \cdot 10^{-7}$	$-0.734 \cdot 10^{-8}$	
III	-0.08626526	-0.08626526	-0.15002936		0.69097897	0.04097898	
IV	1.7391620						
V	-0.04936660						
	0.57214921						

§ 67. Двучленные формулы метода минимальных итераций и биортогонального алгорифма

Системы векторов p_0, \dots, p_{n-1} и q_0, \dots, q_{n-1} тесно связаны друг с другом, и их одновременное вычисление может быть осуществлено по формулам более простым, чем трехчленные формулы, определяющие каждую из этих систем в отдельности. Именно, справедлива следующая

Теорема 67.1. Если A — положительно-определенная матрица, векторы $X, AX, \dots, A^{n-1}X$ линейно-независимы, векторы p_0, \dots, p_{n-1} получены из них процессом ортогонализации, а векторы q_0, \dots, q_{n-1} процессом A -ортогонализации, то между векторами построенных систем имеются следующие двучленные соотношения

$$p_0 = q_0$$

$$p_{i+1} = Aq_i - p_i p_i$$

$$q_{i+1} = p_{i+1} - \sigma_{i+1} q_i,$$

(1)

где

$$\rho_i = \frac{(p_i, Aq_i)}{(p_i, p_i)} = \frac{(q_i, Aq_i)}{(p_i, p_i)}$$

$$\sigma_{i+1} = \frac{(p_{i+1}, p_{i+1})}{(p_i, Aq_i)} = \frac{(p_{i+1}, p_{i+1})}{(q_i, Aq_i)}. \quad (2)$$

Доказательство. Равенство $p_0 = q_0$ непосредственно следует из построения. Обозначим через P_i подпространство, натянутое на векторы $X, AX, \dots, A^i X$, через \tilde{P}_i множество векторов $A^i X + Y$, где $Y \in P_{i-1}$. Векторы p_0, \dots, p_i , так же как и векторы q_0, \dots, q_i образуют базисы подпространства P_i , причем $p_i \in \tilde{P}_i$, $q_i \in \tilde{P}_i$. Далее, если $Z \in \tilde{P}_i$, то $AZ \in \tilde{P}_{i+1}$, в частности, $Ap_i \in \tilde{P}_{i+1}$ и $Aq_i \in \tilde{P}_{i+1}$. Рассмотрим вектор $p_{i+1} - Aq_i$. Вектор Aq_i ортогонален к векторам q_0, \dots, q_{i-1} и, следовательно, ко всем векторам подпространства P_{i-1} , в частности, к векторам p_0, \dots, p_{i-1} . Вектор p_{i+1} тоже ортогонален к p_0, \dots, p_{i-1} . Следовательно, и вектор $p_{i+1} - Aq_i$ ортогонален к p_0, \dots, p_{i-1} . С другой стороны, вектор $p_{i+1} - Aq_i$, как разность двух векторов из множества \tilde{P}_{i+1} , принадлежит подпространству P_i . Поэтому $p_{i+1} - Aq_i = -\rho_i p_i$, где ρ_i некоторое число. Ясно, что

$$\rho_i = -\frac{(p_{i+1} - Aq_i, p_i)}{(p_i, p_i)} = \frac{(Aq_i, p_i)}{(p_i, p_i)},$$

ибо $(p_{i+1}, p_i) = 0$.

Теперь рассмотрим вектор $q_{i+1} - p_{i+1}$. Этот вектор принадлежит подпространству P_i . Далее, вектор p_{i+1} ортогонален ко всем векторам подпространства P_i , в частности, к векторам Aq_0, \dots, Aq_{i-1} , так что вектор p_{i+1} A -ортогонален к векторам q_0, \dots, q_{i-1} . Вектор же q_{i+1} A -ортогонален к векторам q_0, \dots, q_{i-1} по построению. Следовательно, вектор $q_{i+1} - p_{i+1}$ тоже A -ортогонален к векторам q_0, \dots, q_{i-1} . Поэтому $q_{i+1} - p_{i+1} = -\sigma_{i+1}q_i$, где σ_{i+1} некоторое число. Очевидно, что

$$\sigma_{i+1} = -\frac{(q_{i+1} - p_{i+1}, Aq_i)}{(Aq_i, q_i)} = \frac{(p_{i+1}, Aq_i)}{(Aq_i, q_i)}.$$

Но $Aq_i - p_{i+1} \in P_i$, $q_i - p_i \in P_{i-1}$, так что $(p_{i+1}, Aq_i) = (p_{i+1}, p_i)$, $(Aq_i, q_i) = (Aq_i, p_i)$. Следовательно,

$$\sigma_{i+1} = \frac{(p_{i+1}, p_i)}{(q_i, Aq_i)} = \frac{(p_{i+1}, p_i)}{(p_i, Aq_i)}; \quad \rho_i = \frac{(q_i, Aq_i)}{(p_i, p_i)}.$$

Теорема доказана.

Замечание 1. Если A симметричная, но не положительно-определенная матрица, то теорема остается верной при условии, что процесс A -ортогонализации не имеет тупикового окончания.

Замечание 2. Полиномы $p_i(t)$ и $q_i(t)$ также связаны, очевидно, двучленными соотношениями

$$\begin{aligned} p_{i+1}(t) &= tq_i(t) - \rho_i p_i(t) \\ q_{i+1}(t) &= p_{i+1}(t) - \sigma_{i+1}q_i(t). \end{aligned} \tag{3}$$

Легко устанавливаются связи между коэффициентами двучленных формул ρ_i и σ_i и коэффициентами трехчленных формул α_i и β_i или γ_i и δ_i . Именно,

$$\begin{aligned} \alpha_0 &= \rho_0, \\ \alpha_i &= \rho_i + \sigma_i \quad (i = 1, 2, \dots, n-1) \\ \beta_i &= \rho_{i-1}\sigma_i \quad (i = 1, 2, \dots, n-1) \end{aligned} \tag{4}$$

и

$$\begin{aligned} \gamma_i &= \rho_i + \sigma_{i+1} \quad (i = 0, 1, \dots, n-1) \\ \delta_i &= \rho_i\sigma_i \quad (i = 1, 2, \dots, n-1). \end{aligned} \tag{5}$$

Действительно, в силу двучленных соотношений (1), имеем

$$\begin{aligned} p_{i+1} &= Aq_i - \rho_i p_i \\ Aq_i &= Ap_i - \sigma_i Aq_{i-1} \\ p_i &= Aq_{i-1} - \rho_{i-1} p_{i-1}. \end{aligned} \tag{6}$$

Исключая из этих соотношений Aq_i и Aq_{i-1} , получим

$$\begin{aligned} p_{i+1} &= Ap_i - \sigma_i(p_i + \rho_{i-1}p_{i-1}) - \rho_i p_i = \\ &= Ap_i - (\sigma_i + \rho_i)p_i - \sigma_i\rho_{i-1}p_{i-1} = \\ &= Ap_i - \alpha_i p_i - \beta_i p_{i-1}. \end{aligned}$$

Отсюда, в силу линейной независимости векторов p_i и p_{i-1} , получим

$$\begin{aligned}\alpha_i &= \sigma_i + \rho_i \\ \beta_i &= \rho_{i-1} \sigma_i.\end{aligned}$$

Аналогично, исключая p_{i+1} и p_i из соотношений

$$\begin{aligned}q_{i+1} &= p_{i+1} - \sigma_{i+1} q_i \\ p_{i+1} &= A q_i - \rho_i p_i \\ q_i &= p_i - \sigma_i q_{i-1},\end{aligned}$$

получим

$$\begin{aligned}q_{i+1} &= A q_i - \rho_i q_i - \rho_i \sigma_i q_{i-1} - \sigma_{i+1} q_i = \\ &= A q_i - (\rho_i + \sigma_{i+1}) q_i - \rho_i \sigma_i q_{i-1} = \\ &= A q_i - \gamma_i q_i - \delta_i q_{i-1},\end{aligned}$$

откуда

$$\begin{aligned}\gamma_i &= \rho_i + \sigma_{i+1} \\ \delta_i &= \rho_i \sigma_i.\end{aligned}$$

Выведенные соотношения между коэффициентами показывают, в частности, что

$$\sum_{i=0}^{n-1} \rho_i + \sum_{i=1}^{n-1} \sigma_i = \operatorname{Sp} A.$$

Двучленные формулы существуют и для одновременного построения биортогональных и A -биортогональных систем, в случае, если оба процесса не вырожденные. Именно, если

$$p_0 = q_0 \quad \tilde{p}_0 = \tilde{q}_0,$$

то

$$\begin{aligned}p_{i+1} &= A q_i - \rho_i p_i & \tilde{p}_{i+1} &= A' \tilde{q}_i - \rho_i \tilde{p}_i \\ q_{i+1} &= p_{i+1} - \sigma_{i+1} q_i & \tilde{q}_{i+1} &= \tilde{p}_{i+1} - \sigma_{i+1} \tilde{q}_i.\end{aligned}$$

где

$$\begin{aligned}\rho_i &= \frac{(p_i, A' \tilde{q}_i)}{(p_i, \tilde{p}_i)} = \frac{(\tilde{p}_i, A q_i)}{(p_i, \tilde{p}_i)} \\ \sigma_{i+1} &= \frac{(p_{i+1}, \tilde{p}_{i+1})}{(p_i, A' \tilde{q}_i)} = \frac{(p_{i+1}, \tilde{p}_{i+1})}{(\tilde{p}_i, A q_i)}.\end{aligned}\tag{7}$$

Соотношения (4) и (5) по-прежнему сохраняются. Доказательство принципиально ничем не отличается от доказательства, проведенного выше.

Одновременное вычисление биортого

	p_0	\tilde{p}_0	q_0	\tilde{q}_0	Aq_0	$A'\tilde{q}_0$
I	0	0	0	0	1.870086	0.336964
	1	1	1	1	-11.811654	-7.503621
	0	1	0	1	4.308033	-7.258787
	0	0	0	0	0.269851	0.288043
II					-5.363684	-14.137401
III	1				-7.503621	-7.503621
IV	7.503621		-0.23490436			
V	1		1			
	p_2	\tilde{p}_2	q_2	\tilde{q}_2	Aq_2	$A'\tilde{q}_2$
I	8.6346552	2.6926058	11.594561	3.2259404	-49.43280	-17.688366
	16.298936	0	9.852133	0.3717982	-166.93761	0
	-16.298936	-1.0979091	-9.480335	-0.3385964	166.93761	6.934482
	3.7527932	-1.1323868	4.179904	-0.6764826	-84.06708	11.876143
II	0	0	$2.12 \cdot 10^{-6}$	$0.31 \cdot 10^{-6}$	-133.49987	1.122259
	$0.07 \cdot 10^{-6}$	$0.07 \cdot 10^{-6}$	$-1.24 \cdot 10^{-6}$	$0.66 \cdot 10^{-6}$		
III	36.894859	36.894860			-221.18891	
		36.894860			-221.18892	
IV	-5.9951147		1.3221017			
V	1		1			
	20.963280		22.546044			
	99.233549		111.48181			

нальных и A -биортогональных систем

Таблица VI. 16

p_1	\tilde{p}_1	q_1	\tilde{q}_1	Aq_1	$A'\tilde{q}_1$
1.870086 -4.308033 4.308033 0.269851	0.336964 0 0.244834 0.288043	1.8700860 -4.0731286 4.3080330 0.2698510	0.33696400 0.23490436 0.47973836 0.28804300	-16.096774 73.271617 -73.271617 0.18407987	-1.7636605 0.00000001 -4.3357788 -4.9416849
0	0	$-0.29 \cdot 10^{-6}$	$0.01 \cdot 10^{-6}$	-15.9126943	-11.0411242
1.7626333 1.7626333	1.7626333 1.7626333			-23.310394 -23.310394	-23.310394 -23.310394
-13.224755		-1.5827643			
1 7.503621		1 7.738525			
p_3	\tilde{p}_3	q_3	\tilde{q}_3	Aq_3	p_4
2.332948 -69.223619 69.223619 -61.568654	-1.5458854 0 0.3523910 5.0873542	-12.996241 -82.249141 81.757586 -67.094912	-5.8109067 -0.4915550 0.8000499 5.9817370	-48.220650 1430.8088 -1430.8088 1272.5854	-0.00008 0.00006 -0.00006 -0.00001
0 $-4.78 \cdot 10^{-6}$ $-0.84 \cdot 10^{-6}$	0 $1.03 \cdot 10^{-6}$ $2.48 \cdot 10^{-6}$	$-5.86 \cdot 10^{-6}$ $61.46 \cdot 10^{-6}$ $6.43 \cdot 10^{-6}$	$0.32 \cdot 10^{-6}$ $11.03 \cdot 10^{-6}$ $2.86 \cdot 10^{-6}$		
-292.43424	-292.43426 -292.43424			6044.4321 6044.4321	
-20.669372					
1 28.541159 237.15908 594.91651		1 27.219057 207.35092 447.52622			1 47.888429 797.27875 5349.4555 12296.551

Таблица VI. 17

Одновременное вычисление ортогонального и А-ортогонального базисов

В табл. VI. 16 приведено одновременное вычисление биортогонального и A -биортогонального базисов для матрицы Леверье, в табл. VI. 17 приведено одновременное вычисление ортогонального и A -ортогонального базисов для матрицы (4) § 51.

§ 68. Методы сопряженных направлений и их общие свойства

Методами сопряженных направлений для решения систем линейных уравнений

$$AX = F \quad (1)$$

называются методы, в которых решение X представляется в виде линейной комбинации векторов, ортогональных в некоторой метрике, так или иначе связанной с матрицей системы. Термин „сопряженные направления“ происходит от того, что направления векторов, ортогональных в некоторой метрике R сопряжены по отношению к поверхности второго порядка

$$(RX, X) = \text{const.} \quad (2)$$

Метод A -минимальных итераций, также как и методы $A'A$ и AA' -минимальных итераций являются методами сопряженных направлений.

Каждый отдельный метод характеризуется выбором метрики и выбором исходной системы векторов, подвергающейся ортогонализации. При построении системы R -ортогональных векторов часто используются формулы теоремы 11.20.

Методы сопряженных направлений укладываются в следующую общую схему.

Пусть $R = CAB$ — положительно-определенная матрица. Пусть, далее, построена система R -ортогональных векторов s_1, \dots, s_n (векторов R -сопряженных направлений). Ищем решение системы в виде

$$X = X_0 + B \sum_{i=1}^n a_i s_i. \quad (3)$$

Здесь X_0 — начальное приближение, выбираемое, вообще говоря, произвольно. Обычно берется $X_0 = 0$.

Подстановка (3) в систему дает

$$AB \sum_{i=1}^n a_i s_i = F - AX_0 = r_0.$$

Умножая последнее равенство на C , получим

$$R \sum_{i=1}^n a_i s_i = Cr_0.$$

откуда коэффициенты a_i легко определяются. Именно,

$$a_i = \frac{(Cr_0, s_i)}{(Rs_i, s_i)}. \quad (4)$$

Решение (3) можно представить также в виде

$$X = X_0 + \sum_{i=1}^n a_i Bs_i. \quad (5)$$

Векторы Bs_1, \dots, Bs_n , в свою очередь, ортогональны в метрике, определяемой матрицей $R_1 = B'^{-1}RB^{-1}$, которая, очевидно, положительно определена. Действительно,

$$(R_1Bs_i, Bs_j) = (B'R_1Bs_i, s_j) = (Rs_i, s_j) = 0 \quad (i \neq j).$$

В известных частных методах сопряженных направлений или $R = A$ (A положительно-определенная), или $R = A'A$, или $R = AA'$. В первых двух случаях $B = E$ и системы векторов s_1, \dots, s_n и Bs_1, \dots, Bs_n совпадают. В последнем случае $R_1 = E$ и построение системы Bs_1, \dots, Bs_n осуществляется проще, чем построение системы s_1, \dots, s_n . При этом оказывается возможным исключить векторы s_1, \dots, s_n из формул для решения системы.

Решение в форме (3) можно представить как последний член последовательности векторов X_0, X_1, \dots, X_n , где

$$X_i = X_0 + \sum_{j=1}^i a_j Bs_j = X_{i-1} + a_i Bs_i,$$

а векторы X_0, X_1, \dots, X_n рассматривать как последовательные приближения к решению. При этом последовательные невязки r_0, r_1, \dots будут связаны друг с другом по формуле

$$r_i = r_{i-1} - a_i ABs_i. \quad (6)$$

Действительно,

$$r_i = F - AX_i = F - AX_{i-1} - a_i ABs_i = r_{i-1} - a_i ABs_i.$$

При построении i -го приближения X_i нет необходимости знать всю систему сопряженных направлений, так как оно определяется лишь первыми i векторами системы. Поэтому свойства последовательных приближений X_0, \dots, X_i не будут зависеть от того, каким образом система векторов Bs_1, \dots, Bs_i (или s_1, \dots, s_i) будет продолжена до полной. Построенная система приближений будет обладать следующими экстремальными свойствами.

Теорема 68.1. Среди всех векторов вида $Z = X_0 + V$, где V принадлежит подпространству, натянутому на векторы Bs_1, \dots, Bs_i , вектор X_i наименее по R_1 -длине отличается от точного решения. Иначе говоря, обобщенная функция ошибок $f_{R_1}(Z) = (R_1(X - Z), X - Z)$ будет принимать наименьшее значение при $Z = X_i$.

Доказательство. Пусть $Z = X_0 + \sum_{j=1}^i \gamma_j B s_j$. Тогда

$$f_{R_i}(Z) = (R_i(X - Z), X - Z) =$$

$$= \left(R_i \left(\sum_{j=1}^n a_j B s_j - \sum_{j=1}^i \gamma_j B s_j \right), \sum_{j=1}^n a_j B s_j - \sum_{j=1}^i \gamma_j B s_j \right) = \\ = \sum_{j=1}^i (a_j - \gamma_j)^2 (R s_j, s_j) + \sum_{j=i+1}^n a_j^2 (R s_j, s_j) \geq \sum_{j=i+1}^n a_j^2 (R s_j, s_j),$$

причем равенство достигается при $\gamma_j = a_j$. Тем самым минимум $f_{R_i}(Z)$ достигается при $Z = X_i = X_0 + \sum_{j=1}^i a_j B s_j$.

Теорема 68.2. Обобщенная функция ошибок убывает при возрастании индекса i .

Справедливость теоремы непосредственно вытекает из теоремы 68.1. Легко также проверить ее непосредственно вычислением. Именно,

$$f_{R_i}(X_{i-1}) - f_{R_i}(X_i) = \frac{(Cr_0, s_i)^2}{(s_i, R s_i)}.$$

Отметим еще следующие свойства последовательных невязок r_0, r_1, \dots, r_i и векторов s_1, \dots, s_i .

Теорема 68.3. Справедливы равенства

$$(Cr_i, s_j) = 0 \quad (j = 1, \dots, i) \\ (Cr_i, s_j) = (Cr_0, s_j) \quad (i = 1, \dots, j-1). \quad (7)$$

Доказательство. Имеем

$$r_i = F - AX_i = A(X - X_i) = \sum_{k=i+1}^n \frac{(Cr_0, s_k)}{(s_k, R s_k)} A B s_k.$$

Следовательно,

$$Cr_i = \sum_{k=i+1}^n \frac{(Cr_0, s_k)}{(s_k, R s_k)} R s_k.$$

Отсюда $(Cr_i, s_j) = 0$, если $j < i+1$. Далее,

$$(Cr_i, s_j) = \frac{(Cr_0, s_j)}{(R s_j, s_j)} (R s_j, s_j) = (Cr_0, s_j),$$

если $j \geq i+1$. Тем самым теорема доказана.

Методы сопряженных направлений обладают следующим неприятным свойством. В случае, если система векторов, подвергающихся процессу обобщенной ортогонализации, далека от ортогональной по

отношению к выбранной метрике, то при проведении процесса ортогонализации может произойти значительная потеря точности. Положение может быть исправлено „доортогонализацией“ полученной системы, т. е. по сути дела построением новой ортогональной системы, исходя из уже построенной.

Эта доортогонализация должна быть проведена по общим формулам обобщенной ортогонализации, что требует вычисления некоторых малых поправочных коэффициентов из треугольной системы с сильно преобладающей главной диагональю.

Искомое решение системы затем должно быть найдено по формулам (3) и (4) (или (5) и (6)), примененным к вновь построенной системе векторов.

В случае, если A положительно-определенная матрица, можно принять $R = A$, $B = C = E$. В этом случае процесс решения системы $AX = F$ имеет простой геометрический смысл.

Рассмотрим n -мерное точечное пространство, в котором выбрана декартова система координат. Каждой точке пространства сопоставим вектор, исходящий из начала координат в эту точку. Точку и соответствующий ей вектор будем отождествлять в том смысле, что говоря о результате каких-либо действий над точками мы будем понимать одноименные действия над соответствующими им векторами.

Рассмотрим поверхность, заданную векторным уравнением

$$f(X) = (A(X - X^*), X - X^*) = c.$$

(Здесь X^* точное решение системы). Ясно, что эта поверхность при $c > 0$ есть эллипсоид W_n с центром в точке X^* . При различных значениях c уравнение определяет подобные эллипсоиды с общим центром.

Прямые S_1, S_2, \dots, S_n , проходящие через центр эллипса и в направлениях A -ортогональных векторов s_1, s_2, \dots, s_n , образуют систему сопряженных диаметров эллипса $f(X) = c$.

Пусть X_0 начальное приближение. Возьмем $c > f(X_0)$. В этом предположении точка X_0 будет находиться внутри эллипса $f(X) = c$. Проведем через точку X_0 прямую \tilde{P}_1 , параллельную S_1 , затем плоскость \tilde{P}_2 , параллельную S_1 и S_2 , трехмерную плоскость \tilde{P}_3 , параллельную S_1, S_2, S_3 и т. д. Прямая \tilde{P}_1 пересечется с телом $f(X) \leq c$ эллипса $f(X) = c$ по отрезку W_1 , плоскость \tilde{P}_2 по плоскому эллипсу W_2 , трехмерная плоскость \tilde{P}_3 по трехмерному эллипсу W_3 и т. д. Обозначим через X_1 середину отрезка W_1 , через X_2 центр эллипса W_2 , через X_3 центр эллипса W_3 и т. д. Последней точкой X_n этого ряда будет центр всего n -мерного эллипса W_n , т. е. решение X^* системы $AX = F$.

Геометрически ясно, что из всех векторов, исходящих из центра эллипса W_n и опирающихся на k -мерную плоскость \tilde{P}_k , наименьшую A -длину имеет вектор $X_n - X^*$, направленный в центр X_k эллип-

соида W_k . Но как раз этим же экстремальным свойством обладает k -е приближение в методе сопряженных направлений. Поэтому центры X_1, X_2, \dots, X_n представляют собой не что иное, как последовательные приближения в методе сопряженных направлений. Сам процесс последовательного построения точек X_0, X_1, \dots, X_n может быть описан следующим образом. Через X_0 проводится хорда \bar{W}_1 эллипсоида W_n , параллельная диаметру S_1 . Середина этой хорды есть X_1 . Через X_1 проводится хорда \bar{W}_2 , параллельная S_2 . Эта хорда будет в эллипсе W_2 диаметром, сопряженным с хордой \bar{W}_1 . Ее середина будет центром эллипса W_2 , т. е. точкой X_2 . Далее, через X_2

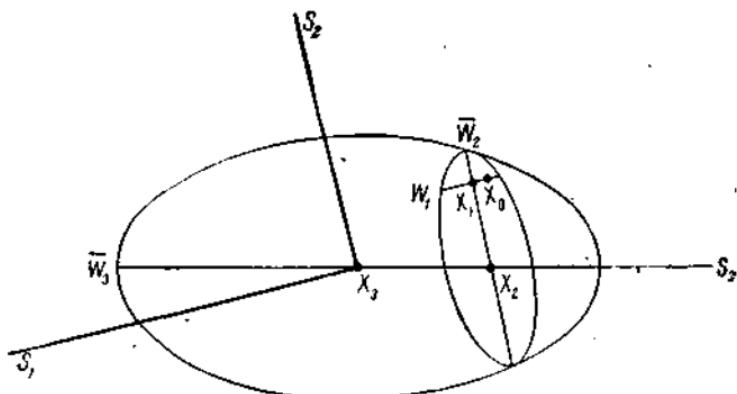


Рис. 5.

проводится хорда \bar{W}_3 , параллельная S_3 . Эта хорда есть диаметр эллипсоида W_3 , сопряженный с его плоским сечением W_2 , и ее середина X_3 есть центр эллипса W_3 и т. д. (см. рис. 5)

В заключение этого параграфа приведем формулы для нахождения обратной матрицы. Условия R -ортогональности векторов s_1, \dots, s_n в матричной форме записываются в виде

$$S'RS = \Lambda,$$

где S — матрица со столбцами s_1, \dots, s_n , $\Lambda = [(s_1, Rs_1), \dots, (s_n, Rs_n)]$. Отсюда следует, что

$$S'CABS = \Lambda$$

и

$$A^{-1} = BSA^{-1}S'C = (BS)\Lambda^{-1}(C'S'). \quad (8)$$

Если $R = A$, $B = C = E$, то

$$A^{-1} = SA^{-1}S'. \quad (9)$$

§ 69. Некоторые методы сопряженных направлений

В современной литературе описаны и изучаются несколько методов сопряженных направлений. В этих методах в качестве системы векторов Z_1, \dots, Z_n , подвергающихся процессу R -ортогонализации

берется или система единичных векторов e_1, \dots, e_n , или система последовательных итераций некоторого произвольного вектора матрицей R .

1. Метод ортогональных векторов. В этом методе¹⁾ A положительно-определенная, $R = A$, $Z_1 = e_1, \dots, Z_n = e_n$. Соответствующие формулы получаются из формул § 68 при $C = B = E$, $R_1 = A$. Векторами направлений будут A -ортогональные векторы

$$s_i = e_i - \gamma_{i1}s_1 - \dots - \gamma_{ii-1}s_{i-1}$$

при

$$\gamma_{ij} = \frac{(Ae_i, s_j)}{(Ae_j, s_j)} \quad \begin{cases} l = 2, \dots, n \\ j = 1, \dots, i-1 \end{cases}. \quad (1)$$

Если $X_0 = 0$, то

$$X = \sum_{i=1}^n a_i s_i, \quad (2)$$

где

$$a_i = \frac{(F, s_i)}{(s_i, As_i)} = \frac{(F, s_i)}{(s_i, Ae_i)}.$$

Отметим, что векторы Ae_i суть не что иное, как столбцы A матрицы A . Таким образом,

$$\begin{aligned} \gamma_{ij} &= \frac{(A_i, s_j)}{(A_j, s_j)} \\ a_i &= \frac{(F, s_i)}{(s_i, A_i)}. \end{aligned} \quad (3)$$

Контроль вычислений осуществляется вычислением величин (A_i, s_j) при $j > i$, которые теоретически должны равняться нулю.

Вычисление решения по методу ортогональных векторов укладывается в схему:

I	A_1	A_2	A_3	A_4	F	s_1	s_2	s_3	s_4	X
II	контроль					(s_i, As_j)				
III							(A_i, s_i)			
IV								γ_{ij}		
V									a_i	

В табл. VI.18 приведено решение системы уравнений (9) § 23 по методу ортогональных векторов.

Установим теперь связь между методом ортогональных векторов и эскалаторным методом (§ 26).

¹⁾ Фокс, Хаски, Уилькинсон [1].

Таблица VI.18

Решение системы уравнений методом ортогональных векторов

	A_1	A_2	A_3	A_4	F	s_1	s_2	s_3	s_4	X
I	1.00	0.42	0.54	0.66	0.3	1	-0.42	-0.4924721	-0.6862368	-1.2577937
	0.42	1.00	0.32	0.44	0.5	0	1	-0.1131617	-0.2227744	0.0434872
	0.54	0.32	1.00	0.22	0.7	0	0	1	0.2218557	1.0391663
	0.66	0.44	0.22	1.00	0.9	0	0	0	1	1.4823929
II						0		-0.14 · 10 ⁻⁷	0.30 · 10 ⁻⁷	
								0.18 · 10 ⁻⁷	-0.32 · 10 ⁻⁷	
									0.20 · 10 ⁻⁷	
III					1.00	0.82360	0.6978533	0.4978712		
IV					0.42					
					0.54	0.1131617				
					0.66	0.1976688	-0.2218557			
V					0.3	0.4541039	0.7102890	1.4823929		

Соотношения

$$s_1 = e_1$$

$$s_2 = e_2 - \gamma_{21}s_1$$

· · · · ·

$$s_n = e_n - \gamma_{n1}s_1 - \cdots - \gamma_{n,n-1}s_{n-1}$$

могут быть записаны в виде

$$e_1 = s_1$$

$$e_1 = \gamma_{21}s_1 + s_2$$

· · · · ·

$$e_n = \gamma_{n1}s_1 + \gamma_{n2}s_2 + \cdots + \gamma_{n,n-1}s_{n-1} + s_n$$

или сокращенно в виде

$$E = \Gamma S', \quad (4)$$

где

$$\Gamma = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ \gamma_{21} & 1 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{n,n-1} & 1 \end{bmatrix},$$

а S матрица, столбцами которой являются векторы направлений. Поэтому

$$S = \Gamma^{-1}$$

и, следовательно, S является правой треугольной матрицей с единичной главной диагональю. Далее, ортогональность векторов s_i к векторам As_j , при $i \neq j$ может быть сокращено записана в виде матричного равенства,

$$S'AS = \Lambda,$$

где Λ — диагональная матрица. Действительно, элемент i -й строки и j -го столбца матрицы $S'AS$ равен, очевидно, (s_i, As_j) и потому равен нулю при $i \neq j$. Таким образом,

$$A = S'^{-1}\Lambda S^{-1} = \Gamma A S^{-1}.$$

Ровно такое же разложение лежит в основе эскалаторного метода при $S = Z$ в обозначениях § 26. При проведении вычислений по обоим методам фактически определяются элементы матриц Γ , Λ и S . Вычислительные схемы методов в основном совпадают.

2. Метод A -минимальных итераций. Метод укладывается в общую схему методов сопряженных направлений при $R = A$, A положительно-определенная, $Z_1 = q_0, \dots, Z_n = A^{n-1}q_0$ (в предположении линейной независимости векторов $q_0, \dots, A^{n-1}q_0$). Как и

в предыдущем методе $C = B = E$, $R_1 = A$. Система векторов сопряженных направлений q_0, \dots, q_{n-1} строится непосредственно по трехчленным рекуррентным формулам (§ 66), минуя вычисления векторов $Aq_0, \dots, A^{n-1}q_0$ и A -ортогонализации системы $q_0, Aq_0, \dots, A^{n-1}q_0$.

3. Метод сопряженных градиентов. В методе сопряженных градиентов¹⁾ $R = A$, A положительно-определенная, $B = C = E$, $R_1 = A$. A -ортогональные векторы s_1, \dots, s_n теоретически строятся процессом A -ортогонализации невязок r_0, r_1, \dots, r_{n-1} последовательных приближений X_0, X_1, \dots, X_{n-1} , определяемых по формулам метода сопряженных направлений, т. е. по формулам

$$X_i = X_0 + \sum_{j=1}^i a_j s_j, \quad (5)$$

где

$$a_j = \frac{(r_0, s_j)}{(s_j, As_j)}.$$

Таким образом, здесь система векторов, подвергающихся процессу ортогонализации, не задается заранее, а строится параллельно с построением векторов сопряженных направлений и соответствующих им последовательных приближений. Отметим, что название метода связано с тем обстоятельством, что невязка r_i является градиентом функции ошибок, вычисленным в точке X_i . Очевидно, что метод сопряженных градиентов естественно связывается с решением лишь одной системы $AX = F$, хотя знание базиса s_1, \dots, s_n позволяет решать так же и системы с отличными от F свободными членами.

Процесс построения последовательных приближений оборвется, как только некоторая невязка r_k окажется равной нулю, т. е. как только приближение совпадает с точным решением системы. А priori мыслима другая причина остановки процесса, не приводящая к точному решению. Именно, процесс может остановиться, если некоторая невязка r_k окажется линейной комбинацией предыдущих. Тогда получится $s_{k+1} = 0$ и $r_{k+1} = r_k$. Однако это невозможно, в силу следующей теоремы.

Теорема 69.1. Ненулевые невязки r_0, r_1, \dots, r_{k-1} взаимно ортогональны.

Доказательство. Рассмотрим (r_i, r_j) при $i > j$. Так как векторы s_1, \dots, s_{j+1} получаются A -ортогонализацией системы векторов r_0, \dots, r_j , заключаем, что вектор r_j есть линейная комбинация векторов s_1, \dots, s_{j+1} и потому (r_i, r_j) есть линейная комбинация чисел (r_i, s_k) при $k = 1, \dots, j+1$. Но, в силу теоремы 68.3, $(r_i, s_k) = 0$ при $k \leq i$, в частности, при $k = 1, \dots, j+1$.

Из доказанной теоремы следует, что если $r_k \neq 0$, то r_k не может быть линейной комбинацией невязок r_0, \dots, r_{k-1} , так что обращение невязки в нуль является единственной причиной остановки процесса.

¹⁾ Штифель [1], Хестинс и Штифель [1].

Выведем теперь расчетные формулы процесса. Заметим, прежде всего, что из формул

$$X_i = X_{i-1} + a_i s_i, \quad a_i = \frac{(r_0, s_i)}{(A s_i, s_i)} \quad (6)$$

следует

$$r_i = r_{i-1} - a_i A s_i, \quad (7)$$

причем $a_i \neq 0$, ибо $r_{i-1} \neq r_i$.

Допустим, что мы уже построили векторы s_1, \dots, s_i и r_0, \dots, r_i . Покажем, что следующий вектор направления s_{i+1} строится по формуле

$$s_{i+1} = r_i + b_i s_i, \quad (8)$$

где

$$b_i = -\frac{(r_i, A s_i)}{(s_i, A s_i)}.$$

Для того чтобы убедиться в этом, достаточно доказать, что так построенный вектор s_{i+1} будет A -ортогонален к векторам s_1, s_2, \dots, s_i . Ясно, что

$$(s_{i+1}, A s_i) = (r_i, A s_i) - \frac{(r_i, A s_i)}{(s_i, A s_i)} (s_i, A s_i) = 0.$$

Рассмотрим теперь $(s_{i+1}, A s_j)$ при $j = 1, 2, \dots, i-1$. Прежде всего отметим, что $(s_{i+1}, A s_j) = (r_i, A s_j)$. Далее, по формуле (7) имеем $A s_j = \frac{1}{a_j} (r_{j-1} - r_j)$, так что

$$(s_{i+1}, A s_j) = \frac{1}{a_j} (r_i, r_{j-1}) - \frac{1}{a_j} (r_i, r_j) = 0$$

в силу того, что $j < i$, $j-1 < i$ и невязки ортогональны. Итак, формула (8) справедлива.

Отметим, что вычисление коэффициентов a_i и b_i можно производить и по формулам

$$\begin{aligned} a_i &= \frac{(s_i, r_{i-1})}{(s_i, A s_i)} = \frac{(r_{i-1}, r_{i-1})}{(s_i, A s_i)} \\ b_i &= \frac{(r_i, r_i)}{(r_{i-1}, r_{i-1})}, \end{aligned} \quad (9)$$

справедливость которых легко устанавливается.

При практическом применении метода коэффициенты следует вычислять по формуле

$$a_i = \frac{(r_{i-1}, r_{i-1})}{(s_i, A s_i)},$$

пользуясь формулой

$$a_i = \frac{(s_i, r_{i-1})}{(s_i, A s_i)}$$

для контроля. Формула же

$$a_i = \frac{(r_0, s_i)}{(s_i, As_i)}$$

чувствительна к ошибкам округления и ее применение нецелесообразно.

Таким образом, расчетные формулы метода сопряженных градиентов таковы. Выбирается произвольное приближение X_0 к решению системы $AX = F$ и вычисляется невязка $r_0 = F - AX_0$. Далее, по рекуррентным соотношениям

$$\begin{aligned} s_1 &= r_0 \\ r_i &= r_{i-1} - a_i As_i \\ a_i &= \frac{(r_{i-1}, r_{i-1})}{(s_i, As_i)} = \frac{(s_i, r_{i-1})}{(s_i, As_i)} \\ s_{i+1} &= r_i + b_i s_i \\ b_i &= \frac{(r_i, r_i)}{(r_{i-1}, r_{i-1})} = -\frac{(r_i, As_i)}{(s_i, As_i)} \end{aligned} \quad (10)$$

строится векторы направлений s_1, s_2, \dots и невязки r_1, r_2, \dots . Процесс заканчивается тем, что при некотором $m \leq n$ окажется $r_m = 0$. Решение системы получается по формуле

$$X = X_0 + \sum_{i=1}^m a_i s_i. \quad (11)$$

Как правило, процесс протекает без вырождения, так что $m = n$.

Контроль вычислений производится при помощи вычисления скалярных произведений (s_i, As_j) или скалярных произведений (r_i, r_j) , которые должны быть нулями.

Из формул (10) видно, что вычислительная схема метода сопряженных градиентов близка к двучленной форме метода минимальных итераций.

В табл. VI.19 приведено решение системы (9) § 23 по методу сопряженных градиентов.

Установим теперь связь между методом сопряженных градиентов и методом A -минимальных итераций.

Из рекуррентных соотношений для построения векторов r_0, r_1, \dots, r_i и s_1, s_2, \dots, s_{i+1} ясно, что как одна, так и другая системы векторов состоят из линейных комбинаций векторов $r_0, Ar_0, \dots, A^i r_0$.

Векторы r_0, r_1, \dots, r_i ортогональны. Следовательно, они лишь скалярными множителями отличаются от векторов p_0, p_1, \dots, p_i метода минимальных итераций, построенных исходя из $p_0 = r_0$. Итак, $p_i = k_i r_i$. Векторы же s_1, \dots, s_{i+1} A -ортогональны, так что они лишь скалярными множителями отличаются от векторов $q_0 = r_0$,

Таблица VI.19

Решение линейной системы методом сопряженных градиентов

	r_0	s_1	As_1	r_1	s_3	As_3	
I	0.3	0.3	1.482	-0.448866929	-0.40302542	-0.12915356	
	0.5	0.5	1.246	-0.12944800	-0.05337488	-0.03272691	
	0.7	0.7	1.220	0.08868655	0.19018892	0.02000433	
	0.9	0.9	1.472	0.15638246	0.29331408	0.04567392	
II			5.420	$1 \cdot 10^{-8}$	$4 \cdot 10^{-8}$	-0.096202222	
III	1.64		3.2464	0.24951983	0.24951983	-0.07100037	
IV		0.15214624	0.50517496		0.001730603	3.5143454	
	r_2	s_3	As_3	r_3	s_4	As_4	X
I	0.00522993	0.00452345	0.00343581	0.00058168	0.000628773	0.000339923	-1.2577936
	-0.01443433	-0.01452670	-0.00983301	-0.00115718	-0.001308415	-0.00079361	0.0434875
	0.01338442	0.01371356	0.01071043	-0.00107748	-0.000934710	-0.00073895	1.0391663
	-0.00413147	-0.00362386	-0.00401315	0.00128733	0.00124960	0.00088325	1.4823930
II	$-11 \cdot 10^{-8}$	$-29 \cdot 10^{-8}$	0.00030008	$28 \cdot 10^{-8}$	$62 \cdot 10^{-8}$	-0.00025008	
	$3 \cdot 10^{-8}$	$1 \cdot 10^{-8}$		$-4 \cdot 10^{-8}$	$1 \cdot 10^{-8}$		
				$2 \cdot 10^{-11}$	$2 \cdot 10^{-11}$		
III	0.0004318197	0.0004318197	0.0003198041	0.0000749560	0.00000449560	0.00000308381	
IV	0.01041083	1.3502631					1.4578070

q_1, \dots, q_i метода A -минимальных итераций, т. е. $s_{i+1} = l_i q_i$. Согласно расчетным формулам метода сопряженных градиентов, получим

$$\begin{aligned}l_i q_i &= k_i p_i + b_i l_{i-1} q_{i-1} \\k_i p_i &= k_{i-1} p_{i-1} - a_i l_{i-1} A q_{i-1}.\end{aligned}$$

Сравнивая эти формулы с двучленными формулами метода минимальных итераций, заключаем, что

$$k_i = l_i \quad \text{и} \quad k_i = -a_i l_{i-1},$$

откуда

$$k_i = l_i = (-1)^i a_1 a_2 \dots a_i.$$

Далее,

$$\begin{aligned}\rho_{i-1} &= \frac{k_{i-1}}{k_i} = -\frac{1}{a_i} \\a_i &= \frac{b_i l_{i-1}}{l_i} = -\frac{b_i}{a_i}.\end{aligned}\tag{12}$$

Коэффициенты γ_i и δ_i метода A -минимальных итераций выражаются через коэффициенты a_i и b_i по формулам

$$\begin{aligned}\gamma_i &= -\frac{1+b_{i+1}}{a_{i+1}} \\ \delta_i &= \frac{b_i}{a_i a_{i+1}}.\end{aligned}\tag{13}$$

Отметим, что последовательные приближения X_i , вычисленные по методу сопряженных градиентов, будут совпадать с последовательными приближениями, вычисленными по методу A -минимальных итераций, исходя из начального приближения X_0 и $q_0 = r_0$. Действительно, соответствующие этим методам векторы направлений отличаются только нормировкой.

Это же обстоятельство следует и из теоремы 68.1. Действительно, приближения X_i обоих методов минимизируют функцию ошибок среди векторов $X_0 + V$, где V принадлежит подпространству, натянутому на $r_0, Ar_0, \dots, A^{i-1}r_0$.

Метод сопряженных градиентов можно применить и к решению полной проблемы собственных значений. Ясно, что для векторов r_i и s_i имеют место представления

$$\begin{aligned}r_i &= r_i(A) r_0 \\s_i &= s_i(A) r_0,\end{aligned}\tag{14}$$

где $r_i(t)$ и $s_i(t)$ полиномы, определяемые по рекуррентным формулам

$$\begin{aligned}r_0(t) &= s_i(t) = 1 \\r_i(t) &= r_{i-1}(t) - a_i t s_i(t) \\s_{i+1}(t) &= r_i(t) + b_i s_i(t).\end{aligned}\tag{15}$$

При невырожденном течении процесса полином $r_n(t)$ лишь скалярным множителем будет отличаться от характеристического полинома матрицы A . Это ясно хотя бы из того, что полиномы $r_i(t)$ лишь постоянными множителями отличаются от полиномов $p_i(t)$ метода минимальных итераций.

Легко проверяется, что если λ_i собственное значение матрицы A , то

$$U_i = \frac{r_0(\lambda_i)}{(r_0, r_0)} r_0 + \dots + \frac{r_{n-1}(\lambda_i)}{(r_{n-1}, r_{n-1})} r_{n-1}$$

есть собственный вектор, принадлежащий собственному значению λ_i . Это следует из аналогичной формулы метода минимальных итераций.

Далее, из рекуррентных соотношений ясно, что $r_i(1) = 1$ при $i = 0, 1, \dots, n$. Поэтому $r_i(t) = \frac{p_i(t)}{p_i(1)}$. Заметим еще, что полиномы $s_i(t)$ при $i = 1, \dots, n$ лишь нормировкой отличаются от полиномов $q_{i-1}(t)$ метода A -минимальных итераций.

4. Метод ортогонализации столбцов. Здесь A — любая неособенная матрица, $R = A'A$, $Z_1 = e_1, \dots, Z_n = e_n$. В этом случае $C = A'$, $B = E$, $R_1 = R$.

Формулами метода будут (мы считаем $X_0 = 0$)

$$\begin{aligned} X &= \sum_{i=1}^n a_i s_i \\ s_i &= e_i - \gamma_{i1}s_1 - \dots - \gamma_{i,i-1}s_{i-1} \\ \gamma_{ij} &= \frac{(s_j, A' A e_i)}{(s_j, A' A e_j)} = \frac{(A s_j, A_i)}{(A s_j, A_j)} \\ a_i &= \frac{(A' F, s_i)}{(s_i, A' A e_i)} = \frac{(F, A s_i)}{(A s_i, A_i)}. \end{aligned} \quad (16)$$

Здесь $A_i = A e_i$ есть i -й столбец матрицы A .

Векторы $A s_1, \dots, A s_n$ образуют ортогональную систему (ибо $(A s_j, A s_i) = (s_j, A' A s_i) = (s_j, R s_j) = 0$ при $j \neq i$) и

$$A s_i = A_i - \gamma_{i1} A s_1 - \dots - \gamma_{i,i-1} A s_{i-1}.$$

Поэтому векторы $A s_1, \dots, A s_n$ фактически находятся процессом ортогонализации столбцов матрицы A .

Векторы направлений s_1, \dots, s_n можно строить затем по формуле

$$s_i = e_i - \gamma_{i1}s_1 - \dots - \gamma_{i,i-1}s_{i-1},$$

используя вычисленные уже коэффициенты ортогонализации γ_{ij} . Однако построения векторов s_1, \dots, s_n можно избежать. Именно, нетрудно проверить, что искомые неизвестные определяются по

рекуррентным формулам

$$\begin{aligned} x_n &= a_n \\ x_{n-1} &= a_{n-1} - \gamma_{n,n-1} x_n \\ &\dots \\ x_1 &= a_1 - \gamma_{21} x_2 - \dots - \gamma_{n1} x_n. \end{aligned} \tag{17}$$

По существу метод ортогонализации столбцов эквивалентен применению метода ортогональных векторов к системе, полученной из данной системы 1-й трансформацией Гаусса.

В табл. VI. 20 дается пример решения системы уравнений методом ортогонализации столбцов по данным табл. II. 1.

Табл. VI. 20 заполняется по схеме.

I	A_1	A_2	A_3	A_4	F	As_1	As_2	As_3	As_4
II	контроль					(As_i, As_j)			
III						(As_i, As_i)			
IV						(A_i, As_i)			
V						γ_{ij}			
VI						a_1	a_2	a_3	a_4
						x_1	x_2	x_3	x_4

5. Метод ортогонализации строк¹⁾). Здесь A любая неособенная матрица, $R = AA'$, $Z_1 = e_1, \dots, Z_n = e_n$. В этом случае $C = E$, $B = A'$, $R_1 = E$.

Формулы метода таковы

$$\begin{aligned} X &= \sum_{i=1}^n a_i A' s_i \\ a_i &= \frac{(F, s_i)}{(AA' s_i, e_i)} = \frac{(F, s_i)}{(A' s_i, A' e_i)}. \end{aligned} \tag{18}$$

Здесь векторы s_i получаются по формулам

$$s_i = e_i - \gamma_{i1} s_1 - \dots - \gamma_{i, i-1} s_{i-1}, \tag{19}$$

где

$$\gamma_{ij} = \frac{(s_j, AA' e_i)}{(s_j, AA' e_j)} = \frac{(A' s_j, A' e_i)}{(A' s_j, A' e_j)} = \frac{(A' s_j, A^i)}{(A' s_j, A^j)}$$

и $A' e_i = A^i$ есть i -я строка матрицы A .

¹⁾ Ю. Шрейдер [1].

Таблица VI. 20

Решение системы уравнений методом ортогонализации столбцов

	A_1	A_2	A_3	A_4	F	AS_1	AS_2	AS_3	AS_4
I	1	0.17	-0.25	0.54	0.3	1	-0.3757506	0.0988025	-0.3617379
	0.47	1	0.67	-0.32	0.5	0.47	0.7434972	-0.2640083	-0.1818692
	-0.11	0.35	1	-0.74	0.7	-0.11	0.4100326	0.5216659	-0.2437552
	0.55	0.43	0.36	1	0.9	0.55	0.1298372	0.1502995	0.7643696
II							-0.420 \cdot 10^{-7}	0.750 \cdot 10^{-7}	-0.720 \cdot 10^{-7}
								-0.389 \cdot 10^{-7}	0.222 \cdot 10^{-7}
									-0.376 \cdot 10^{-7}
III						1.5355	0.8789610	0.3741876	0.8076082
						1.5355	0.8789610	0.3741875	0.8076081
IV						0.5457506	1.1932893		
						0.0995767	-0.6990200	-0.2616262	
						0.6649300			
V						0.6206447	0.7541856	1.0638310	0.3935672
VI						0.4408886	-0.3630312	1.1667985	0.3935672

Таким образом, основную роль в формулах метода играют вспомогательные векторы $A's_i$, которые, очевидно, могут быть получены ортогонализацией строк A^i матрицы A по формулам

$$A's_i = A^i - \gamma_{ii} A's_1 - \dots - \gamma_{i,i-1} A's_{i-1} \quad (20)$$

с прежними значениями коэффициентов γ_{ij} .

Сами векторы s_1, \dots, s_n нужны лишь для вычисления числителей $F_i = (F, s_i)$ коэффициентов a_i . Так же как в предыдущем методе, их можно исключить. Действительно, имеем

$$\begin{aligned} F_i = (F, s_i) &= (F, e_i - \gamma_{ii}s_1 - \dots - \gamma_{i,i-1}s_{i-1}) = \\ &= f_i - \gamma_{ii}F_1 - \dots - \gamma_{i,i-1}F_{i-1}. \end{aligned} \quad (21)$$

Здесь $f_i = (F, e_i)$ есть i -я компонента вектора F . Таким образом, числа F_i находятся параллельно с вычислением ортогонализованных строк $A's_i$ по тем же рекуррентным формулам.

Метод эквивалентен применению метода ортогональных векторов к системе, полученной 2-й трансформацией Гаусса.

В табл. VI. 21 дается решение системы методом ортогонализации строк. Таблица заполняется по схеме

I	A^1	A^2	A^3	A^4	$A's_1$	$A's_2$	$A's_3$	$A's_4$	X
II	f_1	f_2	f_3	f_4	F_1	F_2	F_3	F_4	
III	Контроль				$(A's_i, A's_j)$				
IV					$(A's_i, A's_j)$				
V					$(A's_i, A^i)$				
VI					γ_{ij}				
					a_1	a_2	a_3	a_4	

6. Метод $A'A$ -минимальных итераций. Здесь A неособенная матрица, $R = A'A$, $Z_1 = q_0$, $Z_2 = Rq_0, \dots, Z_n = R^{n-1}q_0$; $C = A'$, $B = E$, $R_1 = R$.

Формулы метода и пример даны в § 65.

7. Метод AA' -минимальных итераций. Здесь A неособенная матрица, $R = AA'$, $Z_1 = q_0$, $Z_2 = Rq_0, \dots, Z_n = R^{n-1}q_0$; $C = E$, $B = A'$, $R_1 = E$.

Формулы метода и пример даны в § 65.

8. Метод сопряженных градиентов после первой трансформации Гаусса. Пусть A неособенная матрица. Применение метода сопряженных градиентов к системе

$$A'AX = A'F,$$

Таблица VI.21

Решение системы уравнений методом ортогонализации строк

	A^1	A^2	A^3	A^4	$A's_1$	$A's_2$	$A's_3$	$A's_4$	X
I	1	0.47	-0.11	0.55	1	0.2532972	0.1949144	-0.2455381	0.4408885
I	0.17	1	0.35	0.43	0.17	0.9631605	-0.3933997	0.0289630	-0.3030308
I	-0.25	0.67	1	0.36	-0.25	0.7241757	0.2979181	0.4495452	1.1667984
I	0.54	-0.32	0.74	1	0.54	-0.4370195	-0.1193276	0.6552979	0.3935672
II	0.3	0.5	0.7	0.9	0.3	0.4349892	0.5061646	0.6654999	
III						1.05 · 10 ⁻⁷	2.20 · 10 ⁻⁷	0.240 · 10 ⁻⁷	
III							2.87 · 10 ⁻⁷	0.614 · 10 ⁻⁷	
III								-0.263 · 10 ⁻⁷	
IV					1.3830	1.7072541	0.2494901	0.6923665	
IV					1.3830	1.7072541	0.2494900	0.6923665	
V					0.2167028				
V					-0.5062184	0.7947344			
V					0.7759219	0.2209139	-0.1864445		
VI					0.2169197	0.2547888	2.0287971	0.9611960	

Таблица VI.22
Решение системы линейных уравнений методом сопряженных градиентов после первой трансформации Гаусса

	r_0	s_1	As_1	r_1	$A'r_1$	s_2	As_2	r_2	$A'r_2$
I	0.3	0.953	1.04047	-0.65773067	-0.09325850	-0.07812450	-0.05997035	0.10480066	0.01012190
	0.5	1.183	2.36831	-0.31426387	-0.12740998	-0.10862350	-0.08061954	-0.09576921	-0.04239773
	0.7	1.284	1.30906	0.24992368	0.14521847	0.16560887	0.02926481	0.17061034	0.06001929
	0.9	0.384	1.87938	0.25394144	0.13838779	0.14448585	0.11442846	-0.05618198	-0.09519513
	II	3.804	6.59692		$0.7 \cdot 10^{-7}$		0.00310338		$-0.51 \cdot 10^{-6}$
III	4.103810	11.936050		0.065170035		0.024046255			$0.65 \cdot 10^{-8}$
IV	0.015880373	0.34381642			0.22348381	2.7101948			0.014564448
	s_3	As_3	r_3	$A'r_3$	s_4	As_4	X		
I	-0.00733766	-0.07689831	0.15056225	0.07946753	0.07448233	0.03405157	0.4408886		
	-0.06667332	0.01501777	-0.10470617	-0.05324986	-0.09853868	-0.0236862	-0.3630310		
	0.09703019	0.12105128	0.09857365	-0.01644605	0.04947622	0.02229369	1.1667984		
	-0.06290488	-0.06067925	-0.02007222	0.02179287	-0.02094468	-0.00453595	0.3935672		
	II	-0.00150851			$0.35 \cdot 10^{-6}$				
III		0.024474267		0.0098956792		0.0022378977			
IV	0.67939954	0.59509231				4.4215959			

полученной из системы $AX = F$ первой трансформацией Гаусса дает

$$X = X_0 + \sum_{i=1}^n a_i s_i$$

$$a_i = \frac{(\bar{r}_{i-1}, \bar{r}_{i-1})}{(s_i, A'As_i)}$$

$$s_1 = \bar{r}_0$$

$$\bar{r}_i = \bar{r}_{i-1} - a_i A'As_i$$

$$s_{i+1} = \bar{r}_i + b_i s_i$$

$$b_i = \frac{(\bar{r}_i, \bar{r}_i)}{(\bar{r}_{i-1}, \bar{r}_{i-1})}.$$

Здесь \bar{r}_i невязка преобразованной системы. Ясно, что

$$\bar{r}_i = A'r_i,$$

где r_i невязка исходной системы. Принимая это во внимание и преобразуя скалярные произведения, придем к следующим расчетным формулам:

$$X = X_0 + \sum_{i=1}^n a_i s_i$$

$$a_i = \frac{(A'r_{i-1}, A'r_{i-1})}{(As_i, As_i)}$$

$$s_1 = A'r_0,$$

$$r_i = r_{i-1} - a_i As_i \quad (22)$$

$$s_{i+1} = A'r_i + b_i s_i$$

$$b_i = \frac{(A'r_i, A'r_i)}{(A'r_{i-1}, A'r_{i-1})}.$$

Метод контролируется выполнением условий ортогональности

$$(A'r_i, A'r_j) = 0.$$

В табл. VI.22 приводится иллюстративный пример с данными табл. II.1.

9. Метод сопряженных градиентов после второй трансформации Гаусса. Пусть A неособенная матрица. Применение метода сопряженных градиентов к вспомогательной системе

$$AA'Y = F,$$

Таблица VI.23

Решение линейной системы методом Крейга

	r_0	g_1	Ag_1	r_1	$A'r_1$	g_2	Ag_2	r_2	$A'r_2$
I	0.3	0.953	1.04047	-0.11580161	-0.26309916	-0.10839731	-0.08891071	0.15303758	0.049890381
	0.5	1.183	2.36831	-0.44644450	-0.34013073	-0.14809265	-0.13742156	-0.03092312	-0.01716715
	0.7	1.284	1.30906	0.17686213	-0.03964159	0.16879197	0.00985241	0.14707141	0.03473364
	0.9	0.384	1.87906	0.14906582	0.09851721	0.16085248	0.09831923	-0.14822190	-0.16451905
II		3.804	6.59692	$0.4 \cdot 10^{-8}$	-0.54435427	0.07315449	-0.11181603	$-0.9 \cdot 10^{-8}$	-0.10614875
III	1.64	4.403810		0.26622354		0.0880456539		$0.2 \cdot 10^{-8}$	
IV		0.16233143	0.39962864			0.25533607	3.0236986	0.067976472	
		g_3	Ag_3	r_3	$A'r_3$	g_4	Ag_4	X	
I	0.01312607	-0.08234041	0.38156664	0.42952349	0.48061945	0.52191943	0.4408885		
	-0.05498055	0.04283899	-0.14982089	-0.10797349	-0.32200171	-0.20492996	-0.3630310		
	0.07783232	0.14849649	-0.26506842	-0.40244471	-0.09946623	-0.36256932	1.1667983		
	-0.12344761	-0.11185027	0.16220935	0.61234865	0.13180318	0.22187530	0.3935672		
II		-0.08746977	-0.00285460	$0.68 \cdot 10^{-7}$	0.53144894	0.19095469	0.17629545		
III				$-0.14 \cdot 10^{-7}$					
IV		0.024492337		0.26461254		0.36194577			0.73108339

полученной из исходной системы $AX=F$ второй трансформацией Гаусса, дает

$$\begin{aligned} Y &= Y_0 + \sum_{i=1}^n a_i s_i \\ a_i &= \frac{(r_{i-1}, r_{i-1})}{(s_i, AA's_i)} \\ s_i &= r_0 \\ r_i &= r_{i-1} - a_i AA's_i \\ s_{i+1} &= r_i + b_i s_i \\ b_i &= \frac{(r_i, r_i)}{(r_{i-1}, r_{i-1})}. \end{aligned}$$

Здесь r_i — невязки преобразованной системы, которые, очевидно, со-впадают с невязками исходной системы.

Обозначим

$$A's_i = g_i.$$

Так как $X = A'Y$, получим после легких преобразований следующие расчетные формулы:

$$\begin{aligned} X &= X_0 + \sum_{i=1}^n a_i g_i \\ a_i &= \frac{(r_{i-1}, r_{i-1})}{(g_i, g_i)} \\ g_i &= A'r_0 \\ r_i &= r_{i-1} - a_i Ag_i \\ g_{i+1} &= A'r_i + b_i g_i \\ b_i &= \frac{(r_i, r_i)}{(r_{i-1}, r_{i-1})}. \end{aligned} \tag{23}$$

Очевидно, что $(r_i, r_j) = 0$ при $i \neq j$.

Последний метод равносителен методу, описанному Крейгом [1], хотя несколько отличается от него по вычислительной схеме.

В табл. VI. 23 дана численная иллюстрация метода для системы табл. II. 1.

ГЛАВА VII

ГРАДИЕНТНЫЕ ИТЕРАЦИОННЫЕ МЕТОДЫ

В настоящей главе будут изложены итерационные методы, пригодные как для решения линейных систем, так и для решения частичной проблемы собственных значений в случае положительно-определенной матрицы и основанные преимущественно на идеи релаксации, которая уже была освещена в гл. III и V. В отличие от методов, изложенных в этих главах, векторами, в направлении которых осуществляется минимизация выбранного функционала, будут не координатные векторы, а векторы, связанные с самим функционалом — именно, это будут градиенты функционала. Именно такой выбор направления минимизации связан с тем, что, как известно (п. 1 § 14 гл. I), направление, противоположное направлению градиента функционала в данной точке, обеспечивает в окрестности этой точки наиболее быстрое убывание функционала. По этой причине некоторые градиентные методы называются также методами наискорейшего спуска.

Идеальным градиентным методом явилось бы построение линии наискорейшего спуска, исходящей из начального приближения и приводящей к точке, дающей минимум функционала, т. е. линии, направление которой в каждой точке противоположно направлению градиента функционала в этой точке. Дифференциальное уравнение линии наискорейшего спуска есть

$$\frac{dX}{dt} = -\rho(t) \operatorname{grad} F(X), \quad (1)$$

где $\rho(t)$ любая положительная функция от параметра t . Выбор функции $\rho(t)$ влияет лишь на параметризацию линии наискорейшего спуска.

Например, если при решении системы $AX = F$ с положительно-определенной матрицей A в качестве функционала взята функция ошибок, то уравнение (1) при $\rho(t) = 1$ будет

$$\frac{dX}{dt} = F - AX. \quad (2)$$

Следовательно, линия наискорейшего спуска будет определяться как решение системы линейных дифференциальных уравнений с постоянными коэффициентами.

При выбранной параметризации искомое решение данной алгебраической системы получается как $\lim_{t \rightarrow \infty} X(t)$, независимо от выбора начального приближения.

Действительно, легко видеть, что общим решением системы (2) является $X = X^* + e^{-At}C$, где X^* точное решение системы $AX = F$, а C произвольный постоянный вектор.

В одношаговых методах наискорейшего спуска траектория наискорейшего спуска, исходящая из данного начального приближения,

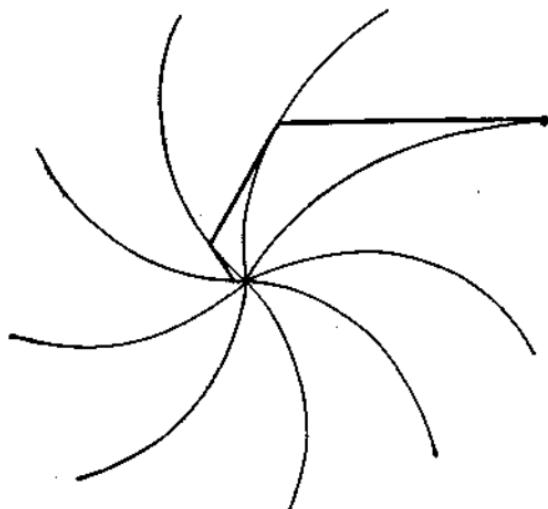


Рис. 6.

заменяется ломаной, составленной из отрезков касательных к траекториям наискорейшего спуска, проходящим через вершины этой ломаной (рис. 6).

Первым среди градиентных методов появился метод наискорейшего спуска, предложенный Л. В. Канторовичем [1] и независимо от него Темплем [1].

Детальное изучение этого метода показало в ряде случаев медленную сходимость процесса, и это обстоятельство привело к развитию других градиентных методов — методов с неполной

релаксацией и s -шаговых процессов. Последние процессы естественным образом (именно при $s = n$) оказались связанными с методом сопряженных градиентов (в случае решения системы) и с методом минимальных итераций (в случае решения проблемы собственных значений). Таким образом, эти последние методы можно рассматривать как предельные случаи многошаговых методов спуска.

В §§ 70—73 мы рассмотрим итерационные градиентные процессы для решения линейных систем, в §§ 74—76 для решения частичной проблемы собственных значений. Как правило, матрица A считается в этой главе положительно-определенной.

§ 70. Метод наискорейшего спуска для решения линейных систем

Пусть A положительно-определенная матрица и пусть

$$AX = F \quad (1)$$

заданная линейная система. В работах Л. В. Канторовича [1]—[3] решение этой системы связывается с задачей отыскания вектора,

дающего минимум функционалу

$$H(X) = (AX, X) - 2(F, X), \quad (2)$$

отличающемуся лишь постоянным, но заранее неизвестным, слагаемым (AX^*, X^*) от функции ошибки $f(X) = (AY, Y)$. Здесь X^* — точное решение системы, совпадающее с вектором, реализующим минимум $H(X)$, $Y = X^* - X$ — вектор ошибки.

Поставленная задача решается следующим образом. Выбирается произвольный вектор X_0 . Вычисляется направление, противоположное градиенту функционала $H(X)$ (или, что то же самое, градиенту функции ошибки) в этой точке, которое совпадает с направлением невязки $r_0 = F - AX_0$ выбранного начального приближения. Из точки X_0 мы двигаемся в выбранном направлении до точки X_1 , в которой функционал $H(X)$ становится минимальным.

Так как

$$\begin{aligned} H(X_0 + \alpha r_0) &= (AX_0 + \alpha Ar_0, X_0 + \alpha r_0) - 2(F, X_0 + \alpha r_0) = \\ &= (AX_0, X_0) + 2\alpha(AX_0, r_0) + \alpha^2(Ar_0, r_0) - 2(F, X_0) - 2\alpha(F, r_0) = \\ &= H(X_0) - 2\alpha(r_0, r_0) + \alpha^2(Ar_0, r_0) = \\ &= H(X_0) - \frac{(r_0, r_0)^2}{(Ar_0, r_0)} + (Ar_0, r_0) \left[\alpha - \frac{(r_0, r_0)}{(Ar_0, r_0)} \right]^2, \end{aligned}$$

то это выражение достигает минимума при

$$\alpha = \alpha_0 = \frac{(r_0, r_0)}{(Ar_0, r_0)}, \quad (3)$$

и этот минимум равен

$$H(X_0) - \frac{(r_0, r_0)^2}{(Ar_0, r_0)}. \quad (4)$$

Итак,

$$X_1 = X_0 + \alpha_0 r_0,$$

где

$$r_0 = F - AX_0$$

$$\alpha_0 = \frac{(r_0, r_0)}{(Ar_0, r_0)}$$

$$H(X_1) = H(X_0) - \frac{(r_0, r_0)^2}{(Ar_0, r_0)}$$

$$f(X_1) = f(X_0) - \frac{(r_0, r_0)^2}{(Ar_0, r_0)}.$$

Далее определяется $X_2 = X_1 + \alpha_1 r_1$, где $r_1 = F - AX_1$ и

$$\alpha_1 = \frac{(r_1, r_1)}{(Ar_1, r_1)}.$$

и процесс продолжается далее по формулам

$$\begin{aligned} X_{k+1} &= X_k + \alpha_k r_k \\ r_k &= F - AX_k = r_{k-1} - \alpha_{k-1} A r_{k-1}, \end{aligned} \quad (5)$$

где

$$\alpha_k = \frac{(r_k, r_k)}{(A r_k, r_k)}. \quad (6)$$

При этом

$$f(X_{k+1}) = f(X_k) - \frac{(r_k, r_k)^2}{(A r_k, r_k)}. \quad (7)$$

Заметим сразу, что при фактическом проведении процесса векторы r_k , особенно при большом порядке матрицы системы, удобнее вычислять по формуле $r_k = r_{k-1} - \alpha_{k-1} A r_{k-1}$. Однако, вследствие ошибок округления, так вычисленные векторы r_k после нескольких шагов процесса могут начать отклоняться от истинных невязок $F - AX_k$. Поэтому следует время от времени вычислять векторы r_k по формуле $r_k = F - AX_k$.

Теорема 70.1 Последовательные приближения X_0, X_1, X_2, \dots сходятся к решению системы $AX = F$ с быстрой геометрической прогрессии¹⁾.

Доказательство. Покажем прежде всего, что последовательность значений функции ошибок стремится к нулю при $k \rightarrow \infty$. Имеем

$$f(X_{k+1}) - f(X_k) = -\frac{(r_k, r_k)^2}{(A r_k, r_k)}.$$

С другой стороны,

$$f(X_k) = (A^{-1} r_k, r_k).$$

Поэтому

$$\frac{f(X_{k+1})}{f(X_k)} - 1 = -\frac{(r_k, r_k)^2}{(A^{-1} r_k, r_k)(A r_k, r_k)}$$

и

$$\frac{f(X_{k+1})}{f(X_k)} = 1 - \frac{(r_k, r_k)^2}{(A^{-1} r_k, r_k)(A r_k, r_k)}.$$

Оценим снизу вычитаемое в правой части последнего равенства.

Пусть $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ собственные значения матрицы A , U_1, U_2, \dots, U_n принадлежащие им собственные векторы, ортогональные друг к другу и нормированные так, что $(U_i, U_j) = 1$ при $i = 1, \dots, n$. Так как A положительно определена, все $\lambda_i > 0$. Пусть $m \leq \lambda_1, \lambda_n \leq M$. Пусть далее

$$r_k = c_1 U_1 + \dots + c_n U_n.$$

1) Л. В. Канторович [2].

причем не все c_i равны нулю. Тогда

$$\begin{aligned} Ar_k &= c_1 \lambda_1 U_1 + \dots + c_n \lambda_n U_n \\ A^{-1} r_k &= c_1 \lambda_1^{-1} U_1 + \dots + c_n \lambda_n^{-1} U_n. \end{aligned}$$

Следовательно,

$$\begin{aligned} (r_k, r_k) &= \sum_{i=1}^n c_i^2 \\ (Ar_k, r_k) &= \sum_{i=1}^n \lambda_i c_i^2 \\ (A^{-1} r_k, r_k) &= \sum_{i=1}^n \frac{1}{\lambda_i} c_i^2 \end{aligned}$$

и потому

$$\frac{(r_k, r_k)^2}{(A^{-1} r_k, r_k)(Ar_k, r_k)} = \frac{\left(\sum_{i=1}^n c_i^2\right)^2}{\sum_{i=1}^n \lambda_i c_i^2 \sum_{i=1}^n \frac{1}{\lambda_i} c_i^2}.$$

Для оценки снизу последнего отношения применим неравенство

$$\frac{\sum_{i=1}^n \gamma_i a_i \sum_{i=1}^n \frac{1}{\gamma_i} a_i}{\left(\sum_{i=1}^n a_i\right)^2} \leq \frac{1}{4} \left[\sqrt{\frac{M}{m}} + \sqrt{\frac{m}{M}} \right]^2, \quad (8)$$

справедливое при условии, что все числа $a_i > 0$, а числа γ_i удовлетворяют неравенствам $0 < m \leq \gamma_i \leq M$.

Это неравенство обосновывается следующим образом. Выражение

$$\Gamma = \sum_{i=1}^n \gamma_i a_i \sum_{i=1}^n \frac{1}{\gamma_i} a_i,$$

рассматриваемое как функция одного из параметров γ_i при фиксированных остальных, имеет вид

$$A\gamma_i + \frac{B}{\gamma_i} + C,$$

причем $A > 0$ и $B > 0$. Последняя функция, очевидно, не имеет максимума при положительных значениях γ_i и потому при изменении γ_i в интервале (m, M) будет иметь максимум на одном из концов. Таким образом, выражение Γ при изменении γ_i в интервале (m, M) принимает максимальное значение, когда некоторые γ_i равны m , а остальные M .

Без нарушения общности можно считать, что $\gamma_i = m$ при $i = 1, \dots, k$ и $\gamma_i = M$ при $i = k+1, \dots, n$.

Введем обозначения

$$S_1 = \sum_{i=1}^k a_i$$

$$S_2 = \sum_{i=k+1}^n a_i.$$

Тогда

$$\begin{aligned}\Gamma &\leq (S_1m + S_2M) \left(\frac{S_1}{m} + \frac{S_2}{M} \right) = S_1^2 + S_2^2 + \left(\frac{m}{M} + \frac{M}{m} \right) S_1 S_2 = \\ &= (S_1 + S_2)^2 + \frac{(m - M)^2}{Mm} S_1 S_2 \leq (S_1 + S_2)^2 \left[1 + \frac{(m - M)^2}{4Mm} \right],\end{aligned}$$

ибо $S_1 S_2 \leq \frac{(S_1 + S_2)^2}{4}$.

Далее

$$1 + \frac{(m - M)^2}{4Mm} = \frac{1}{4} \left[\frac{m}{M} + 2 + \frac{M}{m} \right] = \frac{1}{4} \left[\sqrt{\frac{m}{M}} + \sqrt{\frac{M}{m}} \right]^2.$$

Итак,

$$\Gamma \leq (S_1 + S_2)^2 \frac{1}{4} \left[\sqrt{\frac{m}{M}} + \sqrt{\frac{M}{m}} \right]^2 = \left(\sum_{i=1}^n a_i \right)^2 \frac{1}{4} \left(\sqrt{\frac{m}{M}} + \sqrt{\frac{M}{m}} \right)^2,$$

что и доказывает неравенство (8).

На основании неравенства (8)

$$\frac{(r_k, r_k)^2}{(A^{-1}r_k, r_k)(Ar_k, r_k)} \geq \frac{4}{\left[\sqrt{\frac{m}{M}} + \sqrt{\frac{M}{m}} \right]^2} \quad (9)$$

и, следовательно,

$$\frac{f(X_{k+1})}{f(X_k)} \leq 1 - \frac{4}{\left[\sqrt{\frac{m}{M}} + \sqrt{\frac{M}{m}} \right]^2} = \left[\frac{M-m}{M+m} \right]^2 < 1.$$

Итак,

$$f(X_{k+1}) \leq \left[\frac{M-m}{M+m} \right]^2 f(X_k) \quad (10)$$

и, следовательно,

$$f(X_{k+1}) \leq \left[\frac{M-m}{M+m} \right]^{2(k+1)} f(X_0). \quad (11)$$

Таким образом, $f(X_{k+1}) \rightarrow 0$ при $k \rightarrow \infty$ и потому $X_{k+1} \rightarrow X^*$, где X^* точное решение системы $AX = F$.

Оценим теперь длину вектора ошибки, т. е. вектора $Y_k = X^* - X_k$. Так как

$$f(X_k) = (AY_k, Y_k) \geq m |Y_k|^2,$$

то

$$|Y_k| \leq \sqrt{\frac{f(X_0)}{m}} \left(\frac{M-m}{M+m} \right)^k. \quad (12)$$

Тем самым доказано, что $|Y_k|$ стремится к нулю со скоростью геометрической прогрессии. Теорема доказана.

Оценку для $f(X_k)$ можно придать несколько иную форму, если ввести в рассмотрение P -число обусловленности, т. е. число $\rho = \frac{\lambda_n}{\lambda_1}$. Именно, можно принять, что $m = \lambda_1$, $M = \lambda_n$. Тогда

$$f(X_k) \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^{2k} f(X_0) = \left(\frac{\rho - 1}{\rho + 1} \right)^{2k} f(X_0). \quad (11')$$

Отметим два свойства приближений X_k метода наискорейшего спуска.

1. Невязки двух последовательных приближений ортогональны друг другу.

Действительно, $r_{k+1} = r_k - \alpha_k A r_k$, откуда $(r_{k+1}, r_k) = (r_k, r_k) - \alpha_k (A r_k, r_k) = 0$ на основании определения α_k .

2. Каждое последующее приближение ближе к точному решению, чем предыдущее, т. е.

$$|X^* - X_{k+1}| < |X^* - X_k|.$$

Иначе говоря, длина вектора ошибки при переходе к новому приближению строго убывает.

Действительно,

$$Y_{k+1} = Y_k - \alpha_k r_k$$

и, следовательно,

$$\begin{aligned} (Y_{k+1}, Y_{k+1}) &= (Y_k, Y_k) - 2\alpha_k (Y_k, r_k) + \alpha_k^2 (r_k, r_k) = \\ &= (Y_k, Y_k) - \alpha_k (Y_k, r_k) - \frac{\alpha_k^2}{(r_k, r_k)} \left[\frac{(Y_k, r_k)(r_k, r_k)}{\alpha_k} - (r_k, r_k)^2 \right] = \\ &= (Y_k, Y_k) - \alpha_k (Y_k, r_k) - \frac{\alpha_k^2}{(r_k, r_k)} [(Y_k, r_k)(A r_k, r_k) - (r_k, r_k)^2]. \end{aligned}$$

Покажем, что

$$(Y_k, r_k)(A r_k, r_k) - (r_k, r_k)^2 \geq 0.$$

Положим $A = B^2$, где B положительно-определенная матрица. Тогда

$$\begin{aligned} (Y_k, r_k)(A r_k, r_k) - (r_k, r_k)^2 &= (A^{-1} r_k, r_k)(A r_k, r_k) - (r_k, r_k)^2 = \\ &= (B^{-1} r_k, B^{-1} r_k)(B r_k, B r_k) - (B r_k, B^{-1} r_k)^2 \geq 0 \end{aligned}$$

в силу неравенства Коши—Буняковского. Таким образом,

$$(Y_{k+1}, Y_{k+1}) \leq (Y_k, Y_k) - \alpha_k (Y_k, r_k) < (Y_k, Y_k),$$

ибо $\alpha_k > 0$ и $(Y_k, r_k) = (AY_k, Y_k) > 0$.

Наконец, из сравнения соответствующих формул вытекает, что приближение X_{k+1} совпадает с первым приближением метода сопряженных градиентов, проводимого из начального приближения X_k .

Решение системы линейных уравнений

	X_0	r_0	Ar_0	X_1	$r_1 = r_0 - \alpha_0 Ar_0$
I	0	0.76	0.3616	1.4245790	0.08220033
	0	0.08	0.0496	0.1499557	-0.01297252
	0	1.12	0.6576	2.0993795	-0.11263569
	0	0.68	0.3120	1.2746233	0.09517285
II			1.3808		
III		2.3008	1.227456		
IV		1.8744460			

Приведем примеры применения метода наискорейшего спуска.

Пример 1. Решим систему с матрицей

$$\begin{bmatrix} 0.78 & -0.02 & -0.12 & -0.14 \\ -0.02 & 0.86 & -0.04 & 0.06 \\ -0.12 & -0.04 & 0.72 & -0.08 \\ -0.14 & 0.06 & -0.08 & 0.74 \end{bmatrix}$$

и свободным членом $(0.76, 0.08, 1.12, 0.68)'$. В табл. VII. 1 приведено начало вычислительного процесса (чтобы пояснить вычислительную схему метода) и результаты последующих шагов.

Здесь в первой части таблицы записываются последовательно векторы X_i , r_i , Ar_i , во второй — результат контрольного вычисления по столбцовыми суммам для Ar_i , в третьей — соответствующие скалярные произведения (r_i, r_i) и (Ar_i, r_i) , в четвертой — коэффициенты α_i . Для сравнения вектор r_1 вычислен двумя способами.

Сравнение хода процесса наискорейшего спуска с результатами вычислений по методу последовательных приближений и циклическому одиношаговому процессу показывает, что в данном примере метод наискорейшего спуска сходится быстрее. Именно, восьмой шаг метода наискорейшего спуска дает лучшие результаты, чем десятый шаг в обоих упомянутых методах. Десятый же шаг дает решение уже с точностью до $1 \cdot 10^{-6}$.

Таблица VII. 1

по методу наискорейшего спуска

$r_1 = F - AX_1$	Ar_1	X_2	X_8	X_9	X_{10}
0.08220030	0.06456777	1.5280471	1.5349633	1.5349634	1.5349650
-0.01297254	-0.00258459	0.1336268	0.1220118	0.1220090	0.1220097
-0.11263567	-0.09803664	1.9576015	1.9751502	1.9751560	1.9751560
0.09517284	0.06715236	1.3944203	1.4129515	1.4129545	1.4129552
	0.03107890				
0.02866984	0.02277678				
1.2587313					

В качестве 2-го примера рассмотрим решение системы (9) § 23. Результаты, полученные по методу наискорейшего спуска, помещены в табл. VII. 2.

Таблица VII. 2

Решение системы линейных уравнений по методу наискорейшего спуска

X_0	0	0	0	0
X_1	0.1515525	0.2525875	0.3536225	0.4546575
X_{18}	-1.2546235	0.0434210	1.0365064	1.4786412
X_{26}	-1.2573084	0.0435634	1.0389273	1.4820379
X_{30}	-1.2577348	0.0434859	1.0391170	1.4823232
X_{40}	-1.2577912	0.0434873	1.0391642	1.4823900
X_{52}	-1.2577937	0.0434873	1.0391662	1.4823928

Теоретические оценки быстроты сходимости метода наискорейшего спуска показывают, что с возрастанием числа обусловленности матрицы коэффициентов сходимость метода наискорейшего спуска быстро замедляется. Оказывается, что теоретические оценки почти не являются завышенными, и в действительности процесс очень медленно

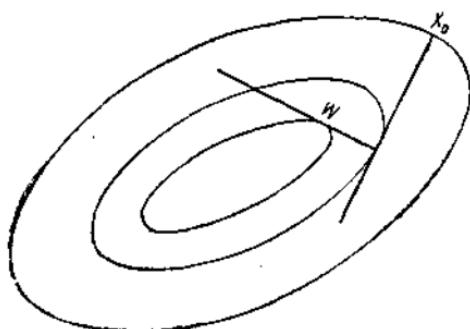


Рис. 7.

где $f(X)$ — функция ошибки. Это семейство есть семейство подобных эллипсоидов. Процесс наискорейшего спуска геометрически может быть объяснен так. Через приближение X_0 проводится эллипсоид

$$f(X) = X_0,$$

в точке X_0 проводится нормаль к этому эллипсоиду и затем в семействе находится эллипсоид, касающийся этой нормали. Точка касания W будет следующим приближением. Для $n = 2$ геометрическая картина показана на рис. 7.

Геометрически очевидно, что если эллипсоиды семейства вытянуты в одном направлении и какое-то приближение попало довольно близко

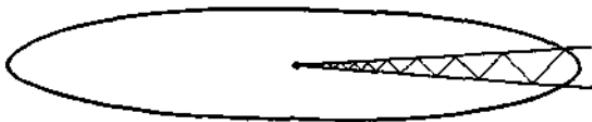


Рис. 8.

к большой оси, то и последующие приближения будут близки к концам больших осей подобных эллипсоидов. Так, для $n = 2$, если начальное приближение находится в точке, нормаль к которой образует угол в 45° с большой осью соответствующего эллипса, то и для всех последующих приближений угол между нормалью и большой осью будет равняться $\pm 45^\circ$ (рис. 8). Нетрудно убедиться (для $n = 2$), что именно при таком выборе начального приближения сходимость процесса будет самой медленной и быстрота сходимости будет совпадать с теоретической оценкой.

§ 71. Градиентный метод с минимальными невязками

Этот метод описан в работе М. А. Красносельского и С. Г. Крейна [1]. Пусть A положительно-определенная матрица, X_0 начальное приближение к решению системы $AX = F$. Следующее приближение X_1 ищется, так же как и в методе наискорейшего спуска, в виде $X_0 + \beta r_0$, но параметр β подбирается так, чтобы минимизировалась длина вектора невязки $|r|$ или, что то же самое, $(r, r) = |r|^2$. Таким образом, здесь минимизируется функционал (r, r) в направлении, противоположном градиенту другого функционала, именно функции ошибок. После выполнения первого шага процесс повторяется.

Выведем формулы, связывающие соседние приближения. Пусть

$$X_{k+1} = X_k + \beta_k r_k, \quad (1)$$

тогда

$$r_{k+1} = r_k - \beta_k A r_k. \quad (2)$$

Следовательно,

$$\begin{aligned} (r_{k+1}, r_{k+1}) &= (r_k, r_k) - 2\beta_k (A r_k, r_k) + \beta_k^2 (A r_k, A r_k) = \\ &= (r_k, r_k) - \frac{(A r_k, r_k)^2}{(A r_k, A r_k)} + (A r_k, A r_k) \left[\beta_k - \frac{(A r_k, r_k)}{(A r_k, A r_k)} \right]^2, \end{aligned}$$

откуда следует, что (r_{k+1}, r_{k+1}) принимает минимальное значение равное $(r_k, r_k) - \frac{(A r_k, r_k)^2}{(A r_k, A r_k)}$ при $\beta_k = \frac{(A r_k, r_k)}{(A r_k, A r_k)}$.

Итак, рабочими формулами метода будут

$$\begin{aligned} X_{k+1} &= X_k + \beta_k r_k \\ r_k &= F - A X_k = r_{k-1} - \beta_{k-1} A r_{k-1} \\ \beta_k &= \frac{(A r_k, r_k)}{(A r_k, A r_k)}. \end{aligned} \quad (3)$$

Теорема 71.1. Последовательные приближения X_0, X_1, \dots сходятся к решению системы $AX = F$ с быстрой геометрической прогрессии.

Справедливость теоремы будет следовать из более общей теоремы, которая будет доказана в следующем параграфе.

В табл. VII.3 приведено решение первого примера § 70.

Быстрота сходимости процесса для данного примера такая же, как и в методе наискорейшего спуска.

Решение системы линейных уравнений градиентным

	X_0	r_0	AX_0	X_1	$r_1 = F - AX_1$
I	0	0.76	0.3616	1.4070460	0.09054233
	0	0.08	0.0496	0.1481101	-0.01182826
	0	1.12	0.6576	2.0735415	-0.09746508
	0	0.68	0.3120	1.2589359	0.10237059
II			1.3808		
III		1.227456	0.66299648		$0.22033655 \cdot 10^{-1}$
IV		1.8513763			1.2620108

§ 72. Градиентные методы с неполной релаксацией

Пусть по-прежнему A положительно-определенная матрица, и ищется решение системы $AX = F$.

Рассмотрим итерационный процесс, в котором каждое последующее приближение получается из предыдущего изменением в направлении, противоположном градиенту функции ошибок, причем так, что на каждом шаге функция ошибок уменьшается. Формулы для получения последовательных приближений, очевидно, должны иметь вид

$$X_{k+1} = X_k + \gamma_k r_k. \quad (1)$$

Для дальнейшего исследования нам удобно положить

$$\gamma_k = q_k \alpha_k, \quad (2)$$

где α_k соответствующий коэффициент в методе наискорейшего спуска. Имеем

$$\begin{aligned}
 f(X_{k+1}) &= f(X_k) - 2\gamma_k(r_k, r_k) + \gamma_k^2(Ar_k, r_k) = \\
 &= f(X_k) - 2q_k \alpha_k(r_k, r_k) + q_k^2 \alpha_k^2(Ar_k, r_k) = \\
 &= f(X_k) - 2q_k \frac{(r_k, r_k)^2}{(Ar_k, r_k)} + q_k^2 \frac{(r_k, r_k)^3}{(Ar_k, r_k)} = \\
 &= f(X_k) - (2q_k - q_k^2) \frac{(r_k, r_k)^2}{(Ar_k, r_k)}. \quad (3)
 \end{aligned}$$

МЕТОДОМ С МИНИМАЛЬНЫМИ НЕВЯЗКАМИ

Таблица VII.3

$r_1 = r_0 - \beta_0 A r_0$	$A r_1$	X_2	X_9	X_{10}
0.09054233	0.06822351	1.5213114	1.5349632	1.5349648
-0.01182826	-0.00194231	0.1331827	0.1220091	0.1220099
-0.09746505	-0.08875643	1.9505396	1.9751557	1.9751558
0.10237059	0.07016581	1.3881287	1.4129543	1.4129551
	0.04769058			
	$0.17459165 \cdot 10^{-1}$			

Из этой формулы ясно, что для того, чтобы $f(X_{k+1})$ было бы меньше, чем $f(X_k)$, необходимо и достаточно выполнение для множителей релаксации q_k неравенств

$$0 < q_k < 2. \quad (4)$$

Будем называть группу методов, в которых не все q_k равны 1, методами неполной градиентной релаксации (одношаговыми). Если все множители релаксации $q_k \leq 1$, но не все равны единице, метод называется методом нижней релаксации, если все $q_k \geq 1$, но не все $q_k = 1$, — методом верхней релаксации.

Так, метод с минимальными невязками является методом нижней релаксации, ибо в нем

$$q_k = \frac{\beta_k}{\alpha_k} = \frac{(A r_k, r_k)^2}{(A r_k, A r_k) (r_k, r_k)} \leq 1 \quad (5)$$

по неравенству Коши—Буняковского. Здесь знак равенства возможен, только если r_k есть собственный вектор матрицы A .

В группу методов неполной релаксации входит и градиентный метод с постоянным множителем $\gamma_k = \gamma$, если этот множитель удовлетворяет неравенству

$$0 < \gamma < \frac{2}{M}, \quad (6)$$

где M наибольшее собственное значение матрицы A .

Действительно, в этом предположении

$$0 < q_k = \frac{\gamma}{\alpha_k} = \frac{\gamma (Ar_k, r_k)}{(r_k, r_k)} \leq \gamma M < 2.$$

Условие $0 < \gamma < \frac{2}{M}$ является так же необходимым для того, чтобы функция ошибки уменьшалась при любом начальном векторе. Действительно,

$$0 < q_k < 2$$

дает

$$0 < \gamma \frac{(Ar_k, r_k)}{(r_k, r_k)} < 2,$$

откуда

$$0 < \gamma < \frac{(2r_k, r_k)}{(Ar_k, r_k)}.$$

Так как последнее неравенство должно выполняться при всех r_k , должно быть выполнено неравенство

$$0 < \gamma < \min \frac{2(z, z)}{(Az, z)} = \frac{2}{M}.$$

Заметим, что градиентный метод с постоянным множителем есть не что иное, как процесс последовательных приближений, примененный к системе, подготовленной следующим образом

$$X = (E - \gamma A) X + \gamma F.$$

Необходимые и достаточные условия сходимости метода совпадают с условием $0 < \gamma < \frac{2}{M}$. Действительно, наибольшее по модулю собственное значение матрицы $E - \gamma A$ будет большее по модулю из чисел $1 - \gamma M$ и $1 - \gamma m$. Таким образом, для сходимости метода последовательных приближений необходимо и достаточно выполнение неравенств

$$-1 < 1 - \gamma m < 1$$

$$-1 < 1 - \gamma M < 1,$$

откуда следует, что $\gamma > 0$, $\gamma < \frac{2}{m}$ и $\gamma < \frac{2}{M}$. Выполнение третьего неравенства обеспечивает выполнение второго.

Наибольшее по модулю собственное значение матрицы $E - \gamma A$ будет наименьшим из возможных, если $1 - \gamma m = -(1 - \gamma M)$, т. е. если $\gamma = \frac{2}{m+M}$. В этом случае $1 - \gamma m = \frac{M-m}{M+m}$. Следовательно, быстрота сходимости при таком выборе множителя γ будет оцениваться неравенством

$$|Y_k| \leq \left(\frac{M-m}{M+m} \right)^k |Y_0|,$$

т. е. имеет тот же порядок, что и в методе наискорейшего спуска¹⁾.

Вся группа методов неполной релаксации (включая и случай полной релаксации, когда все $q_k = 1$, что соответствует методу наискорейшего спуска) естественно укладывается в общую схему итерационных методов, описанных в главе III. Именно, они получаются из общей итерационной формулы

$$X_{k+1} = X_k + H^{(k+1)}(F - AX_k)$$

при

$$H^{(k+1)} = \gamma_k E.$$

Теорема 72.1. Если в процессе неполной градиентной релаксации множители релаксации удовлетворяют условию $\varepsilon < q_k < 2 - \varepsilon$, $0 < \varepsilon < 1$, то процесс сходится к решению со скоростью геометрической прогрессии.

Доказательство. Пусть X_k — k -е приближение метода неполной релаксации, удовлетворяющего условиям теоремы. Обозначим через \bar{X}_{k+1} приближение, получающееся из X_k одним шагом метода наискорейшего спуска, через Y_{k+1} и \bar{Y}_{k+1} — соответствующие векторы ошибок.

Как мы видели выше

$$f(\bar{X}_{k+1}) \leq \tau f(X_k),$$

где

$$\tau = \left(\frac{M-m}{M+m} \right)^2.$$

Следовательно,

$$f(X_k) - f(\bar{X}_{k+1}) \geq (1-\tau)f(X_k).$$

Далее

$$f(X_{k+1}) = f(X_k) - (2q_k - q_k^2) \frac{(r_k, r_k)^2}{(Ar_k, r_k)},$$

откуда

$$f(X_k) - f(X_{k+1}) = (2q_k - q_k^2) \frac{(r_k, r_k)^2}{(Ar_k, r_k)} = (2q_k - q_k^2) [f(X_k) - f(\bar{X}_{k+1})],$$

ибо

$$f(X_k) - f(\bar{X}_{k+1}) = \frac{(r_k, r_k)^2}{(Ar_k, r_k)}.$$

Таким образом,

$$f(X_k) - f(X_{k+1}) \geq (2q_k - q_k^2)(1-\tau)f(X_k),$$

и потому

$$f(X_{k+1}) \leq [1 - (2q_k - q_k^2)(1-\tau)]f(X_k) \leq [1 - (2\varepsilon - \varepsilon^2)(1-\tau)]f(X_k).$$

Положим

$$\tau_1 = 1 - (2\varepsilon - \varepsilon^2)(1-\tau).$$

¹⁾ И. П. Натансон [1].

Ясно, что $0 < \tau_1 < 1$, ибо $0 < 2\epsilon - \epsilon^2 < 1$, $0 < 1 - \tau < 1$. Из последнего неравенства следует, что

$$f(X_{k+1}) \leq \tau_1^k f(X_0). \quad (7)$$

Таким образом, при $k \rightarrow \infty$ $f(X_k) \rightarrow 0$ со скоростью геометрической прогрессии, $Y_k \rightarrow 0$ и $X_k \rightarrow X^*$. Теорема доказана.

Из доказанной теоремы следует, в частности, сходимость метода с минимальными невязками, так как в этом случае множители q_k удовлетворяют неравенству

$$\epsilon \leq q_k \leq 1, \quad \text{где} \quad \epsilon = \frac{4}{\left[\sqrt{\frac{M}{m}} + \sqrt{\frac{m}{M}} \right]^2}.$$

Действительно,

$$q_k = \frac{\beta_k}{\alpha_k} = \frac{(Ar_k, r_k)^2}{(Ar_k, Ar_k)(r_k, r_k)} = \frac{\left(A^{\frac{1}{2}} r_k, A^{\frac{1}{2}} r_k \right)^2}{\left(A^{\frac{1}{2}} r_k, A^{\frac{1}{2}} r_k \right) \left(A^{-\frac{1}{2}} r_k, A^{\frac{1}{2}} r_k \right)} = \frac{(z, z)^2}{(Az, z)(A^{-1}z, z)}.$$

Здесь $z = A^{\frac{1}{2}} r_k$. Поэтому в силу неравенства (9) § 70 имеем

$$q_k \geq \frac{4}{\left[\sqrt{\frac{M}{m}} + \sqrt{\frac{m}{M}} \right]^2}.$$

Неравенство $q_k \leq 1$ было уже отмечено.

Быстрота сходимости метода при таком способе доказательства сильно занижена. В упомянутой выше работе М. А. Красносельского и С. Г. Крейна установлено, что быстрота сходимости имеет тот же порядок, что и в методе наискорейшего спуска.

В оценке (7) никак не учитывается, будет ли на данном шаге q_k больше или меньше 1. Есть основание предполагать, что если применяется верхняя релаксация, приведенные оценки не являются существенно завышенными. Для нижней релаксации они, по всей вероятности, являются завышенными, так как при нижней релаксации ломаная с вершинами в последовательных приближениях будет теснее примыкать к линии наискорейшего спуска, чем аналогичные ломаные для полной и верхней релаксации.

В вычислительном отношении удобно брать множители релаксации q_k не зависящими от k .

В табл. VII. 4 и VII. 5 приводятся результаты вычисления решения системы (9) § 23 градиентным методом с применением неполной релаксации при $q = 0.8$ и $q = 1.2$. Из таблиц видно, что в данном примере нижняя релаксация приводит к решению быстрее, чем полная (табл. VII. 2), а верхняя медленнее.

Таблица VII.4

Решение системы линейных уравнений градиентным методом с неполной релаксацией при $q = 0.8$

X_0	0	0	0	0
X_1	0.1212420	0.2020700	0.2828980	0.3637260
X_{18}	-1.2574919	0.0435581	1.0390689	1.4822318
X_{30}	-1.2577926	0.0434876	1.0391658	1.4823922
X_{35}	-1.2577936	0.0434873	1.0391662	1.4823928

Таблица VII.5

Решение системы линейных уравнений градиентным методом с неполной релаксацией при $q = 1.2$

X_0	0	0	0	0
X_1	0.1818630	0.3031050	0.4243470	0.5455890
X_{18}	-1.2251212	0.0432191	1.0164571	1.4498670
X_{30}	-1.2551442	0.0436275	1.0373248	1.4797554
X_{35}	-1.2566377	0.0437154	1.0386884	1.4816669
X_{52}	-1.2577672	0.0434887	1.0391478	1.4823665
X_{62}	-1.2577904	0.0434875	1.0391640	1.4823896

В заключение этого параграфа рассмотрим метод, являющийся предельным случаем для метода верхней релаксации. Формулы метода¹⁾ следующие

$$X_{k+1} = X_k + 2\alpha_k r_k,$$

где α_k коэффициент метода наискорейшего спуска. В этом случае функция ошибок от шага к шагу не уменьшается и последовательные приближения не стремятся к решению. Однако, если начальное приближение не ортогонально к собственному вектору, принадлежащему наибольшему собственному значению матрицы A , то последовательности X_0, X_1, \dots и X_2, X_3, \dots расходятся. При этом полусумма предлов этих последовательностей дает решение системы, а полуразность есть собственный вектор, принадлежащий наибольшему собственному значению.

Доказательство соответствующей теоремы читатель найдет в упомянутой работе В. Н. Костарчука.¹⁾

¹⁾ В. Н. Костарчук [1].

§ 73. s -шаговые градиентные методы наискорейшего спуска

В предыдущих трех параграфах мы рассматривали одношаговые градиентные методы, связанные с полной или неполной релаксацией функции ошибок и установили, что наилучший результат за один шаг дает метод наискорейшего спуска. Естественно поставить вопрос о том, как определить множители $\gamma_1, \dots, \gamma_s$ с тем, чтобы, исходя из начального приближения $X^{(0)}$, получить наилучший результат после s шагов процесса и как объединить эти s шагов в один шаг нового вычислительного процесса. Иначе говоря, как построить вычислительный процесс, один шаг которого равносителен s шагам одношагового градиентного метода, выбранным согласно наилучшей стратегии.

Пусть

$$\begin{aligned} X_0 &= X^{(0)} \\ X_1 &= X_0 + \gamma_0(F - AX_0) \\ X_2 &= X_1 + \gamma_1(F - AX_1) \\ &\dots \\ X^{(1)} &= X_s = X_{s-1} + \gamma_{s-1}(F - AX_{s-1}). \end{aligned} \quad (1)$$

Тогда, переходя к векторам ошибки, получим

$$\begin{aligned} Y_1 &= Y_0 - \gamma_0 AY_0 = (E - \gamma_0 A) Y_0 \\ Y_2 &= Y_1 - \gamma_1 AY_1 = (E - \gamma_1 A) Y_1 \\ &\dots \\ Y^{(1)} &= Y_s = Y_{s-1} - \gamma_{s-1} AY_{s-1} = (E - \gamma_{s-1} A) Y_{s-1}, \end{aligned} \quad (2)$$

откуда

$$\begin{aligned} Y^{(1)} &= (E - \gamma_0 A)(E - \gamma_1 A) \dots (E - \gamma_{s-1} A) Y_0 = \\ &= (E + c_1 A + c_2 A^2 + \dots + c_s A^s) Y_0 = \\ &= Y_0 + c_1 r_0 + \dots + c_s A^{s-1} r_0, \end{aligned} \quad (3)$$

где

$$r_0 = F - AX^{(0)}$$

и

$$X^{(1)} = X^{(0)} - c_1 r_0 - \dots - c_s A^{s-1} r_0. \quad (4)$$

Таким образом, задача сводится к определению коэффициентов c_1, \dots, c_s так, чтобы значение $f(X^{(1)})$ функции ошибки было бы наименьшим.

Эта задача была уже решена ранее в § 69 при изучении метода сопряженных градиентов. Действительно, на стр. 445 было устано-

влено, что s -е приближение метода сопряженных градиентов минимизирует функцию ошибки среди векторов $X^{(0)} + V$, где вектор V принадлежит подпространству, натянутому на $r_0, Ar_0, \dots, A^{s-1}r_0$.

Итак, результат одного шага s -шагового метода наискорейшего спуска совпадает с результатом s -го приближения метода сопряженных градиентов.

Именно,

$$X^{(1)} = X^{(0)} + \sum_{j=1}^s a_j s_j, \quad (5)$$

где

$$s_1 = r_0,$$

$$r_i = r_{i-1} - a_i As_i, \quad a_i = \frac{(s_i, r_{i-1})}{(s_i, As_i)} = \frac{(r_{i-1}, r_{i-1})}{(s_i, As_i)} \quad (6)$$

$$s_{i+1} = r_i + b_i s_i, \quad b_i = \frac{(r_i, r_i)}{(r_{i-1}, r_{i-1})} = -\frac{(r_i, As_i)}{(s_i, As_i)} \\ (i = 1, 2, \dots, s-1).$$

Поэтому при фактическом проведении s -шагового метода наискорейшего спуска нам не нужны ни коэффициенты c_1, \dots, c_s , ни, тем более, коэффициенты $\gamma_1, \dots, \gamma_s$.

Однако нетрудно отдать себе отчет в том, что представляют собой те и другие коэффициенты. При рассмотрении метода сопряженных градиентов было показано, что

$$r_s = r_s(A) r_0 \quad (7)$$

и, следовательно,

$$Y^{(1)} = r_s(A) Y^{(0)},$$

где $r_s(t)$ полином § 69, п. 3.

Но, с другой стороны,

$$Y^{(1)} = (E + c_1 A + \dots + c_s A^s) Y_0.$$

Следовательно,

$$1 + c_1 t + \dots + c_s t^s = r_s(t),$$

т. е. коэффициенты c_1, \dots, c_s суть не что иное, как коэффициенты полинома $r_s(t)$. Далее, из равенства

$$(1 - \gamma_0 t)(1 - \gamma_1 t) \dots (1 - \gamma_{s-1} t) = 1 + c_1 t + c_2 t^2 + \dots + c_s t^s = r_s(t)$$

следует, что числа $\gamma_0, \dots, \gamma_{s-1}$ суть числа, обратные к корням полинома $r_s(t)$.

Отметим, что если вместо формулы (5) вектор $X^{(1)}$ вычислять рекуррентно по формулам (1), то на некоторых шагах этого процесса может произойти даже увеличение функции ошибки, так как не все числа $\gamma_0, \dots, \gamma_{s-1}$ обеспечивают релаксацию.

Таблица VII.6

Двухшаговый метод наискорейшего спуска. Схема сопряженных градиентов

$X^{(0)}$	$r_0 = s_1$	$A s_1$	r_1	s_2	$A s_2$	$X^{(1)}$	$X^{(2)}$	$X^{(3)}$	$X^{(4)}$
0	0.3	1.482	-0.44866829	-0.40302542	-0.12915358	-1.2648181	-1.2572773	-1.2577994	-1.2577932
0	0.5	1.246	-0.12944800	-0.05337489	-0.03272692	0.0650097	0.0434885	0.0435032	0.0434873
0	0.7	1.220	0.08368655	0.19018891	0.02000431	1.0220120	1.0387884	1.0391674	1.0391658
0	0.9	1.472	0.15638246	0.29331407	0.04567390	1.4854645	1.4818603	1.4823840	1.4823923
		5.420			-0.09620229				
	1.64	3.2464	-0.49392752	0.24951983	0.071000367				
	-0.15214623	0.50517496			3.5143456				

Таблица VII.7

Двухшаговый метод наискорейшего спуска. Схема минимальных итераций

$X^{(0)}$	$r_0 = q_0$	$A q_0$	q_1	$A q_1$	$X^{(1)}$	$X^{(2)}$	$X^{(3)}$	$X^{(4)}$
0	0.3	1.482	0.79779375	0.25566113	-1.2648180	-1.2572775	-1.2577994	-1.2577932
0	0.5	1.246	0.10565625	0.06478338	0.0650097	0.0434887	0.0435032	0.0434873
0	0.7	1.220	-0.37648125	-0.03959875	1.0220119	1.0387863	1.0391673	1.0391659
0	0.9	1.472	-0.58061875	-0.09041200	1.4854643	1.4818599	1.4823841	1.4823923
		5.420		0.19043375				
	3.2464	7.404024	0.27821271					
	2.2806875							
	1.64		-0.49392750					
	0.50517496		-1.7753592					

Действительно, если $s = n$ и за γ_0 принять число $\frac{1}{\lambda_1}$, а за r_0 вектор U_n , то $\alpha_0 = \frac{(r_0, r_0)}{(Ar_0, r_0)} = \frac{1}{\lambda_n}$, $\gamma_0 = q_0 \alpha_0 = q_0 \lambda_n^{-1}$, так что $q_0 = \frac{\lambda_n}{\lambda_1} = \rho$ и потому $q_0 > 2$, если $\rho > 2$.

Вычисление приближения $X^{(1)}$ можно проводить и пользуясь методом A -минимальных итераций. Действительно, как мы видели (стр. 445), последовательные приближения, полученные по этому методу, при $q_0 = r_0$ совпадают с соответствующими приближениями метода сопряженных градиентов.

Многошаговый метод наискорейшего спуска впервые был описан в работах Л. В. Канторовича [2], [3]. В этих работах для составления последовательных приближений фактически находились коэффициенты c_1, \dots, c_s посредством решения системы s линейных уравнений. Эта система для первого шага имеет вид

$$\begin{aligned} (r_0, r_0) - (r_0, Ar_0)c_1 - \dots - (r_0, A^s r_0)c_s &= 0 \\ (r_0, Ar_0) - (r_0, A^2 r_0)c_1 - \dots - (r_0, A^{s+1} r_0)c_s &= 0 \\ \cdot &\quad \cdot \\ (r_0, A^{s-1} r_0) - (r_0, A^s r_0)c_1 - \dots - (r_0, A^{2s-1} r_0)c_s &= 0. \end{aligned} \quad (8)$$

Для каждого последующего шага начальная невязка r_0 должна быть заменена невязкой, полученной в результате предыдущего шага.

Рассмотрим применение двухшагового метода для нахождения решения системы (9) § 23.

Вычисления по разным схемам приведены в табл. VII. 6, VII. 7 и VII. 8.

Таблица VII. 8

Двухшаговый метод наискорейшего спуска. Схема с решением систем

$X^{(0)}$	r_0	Ar_0	$A^2 r_0$	$X^{(1)}$	$X^{(2)}$	$X^{(3)}$	$X^{(4)}$
0	0.3	1.482	3.63564	-1.2648177	-1.2572769	-1.2577994	-1.2577931
0	0.5	1.246	2.90652	0.0650098	0.0434889	0.0435034	0.0434873
0	0.7	1.220	2.74284	1.0220119	1.0387865	1.0391674	1.0391657
0	0.9	1.472	3.26676	1.4854642	1.4818605	1.4823838	1.4823928
		5.420	12.55176				

Для нахождения $X^{(1)}$ в табл. VII. 8 вычисляем коэффициенты c_1 и c_2 из системы уравнений

$$\begin{aligned} 3.2464c_1 + 7.404024c_2 &= 1.64 \\ 7.404024c_1 + 17.164478c_2 &= 3.2464. \end{aligned}$$

Это дает

$$c_1 = 4.5542139 \quad c_2 = 1.7753589.$$

При $s > 2$ 1-я и 2-я схемы предпочтительнее, так как в последней схеме на каждом шаге процесса нужно решать вспомогательную систему линейных уравнений.

Так же как и одношаговые градиентные процессы, s -шаговый метод наискорейшего спуска можно включить в общую схему итерационных процессов

$$X^{(k)} = X^{(k-1)} + H^{(k)}(F - AX^{(k-1)}),$$

полагая

$$H^{(k)} = H_s^{(k)}(A)$$

при

$$H_s^{(k)}(t) = \frac{1 - r_s^{(k)}(t)}{t}. \quad (9)$$

Действительно,

$$Y^{(k)} = r_s^{(k)}(A) Y^{(k-1)} = Y^{(k-1)} + (r_s^{(k)}(A) - E) Y^{(k-1)}.$$

Отсюда

$$\begin{aligned} X^{(k)} &= X^{(k-1)} + (E - r_s^{(k)}(A)) Y^{(k-1)} = X^{(k-1)} + H_s^{(k)}(A) A Y^{(k-1)} = \\ &= X^{(k-1)} + H_s^{(k)}(A)(F - AX^{(k-1)}). \end{aligned}$$

Ясно, что полиномы $H_s^{(k)}(t)$ здесь меняются от шага к шагу.

Установим теперь сходимость s -шагового процесса наискорейшего спуска и оценим быстроту сходимости.

Прежде всего заметим, что один шаг s -шагового процесса наискорейшего спуска дает не худший результат в смысле уменьшения функции ошибок, чем s шагов одношагового метода наискорейшего спуска. Отсюда непосредственно следует, что s -шаговый метод наискорейшего спуска сходится, и для функции ошибок имеет место оценка

$$f(X^{(k)}) \leq \left(\frac{M-m}{M+m}\right)^{2s} f(X^{(k-1)}), \quad (10)$$

и, следовательно,

$$f(X^{(k)}) \leq \left(\frac{M-m}{M+m}\right)^{2sk} f(X_0). \quad (11)$$

Однако для *s*-шагового метода наискорейшего спуска можно дать и лучшие оценки¹⁾, сравнивая его со стационарным итерационным процессом

$$X^{(k)} = X^{(k-1)} + H_s(A)(F - AX^{(k-1)}), \quad (12)$$

где $H_s(t)$ некоторый полином степени $s-1$.

Ясно, что каков бы ни был полином $H_s(t)$, каждый шаг *s*-шагового метода наискорейшего спуска будет обеспечивать не худшее уменьшение функции ошибок, чем один шаг, проведенный по формуле (12), исходя из предыдущего приближения метода наискорейшего спуска. Пусть $X^{(k-1)}$ это приближение, $X^{(k)}$ следующее приближение *s*-шагового метода наискорейшего спуска,

$$\bar{X}^{(k)} = X^{(k-1)} + H_s(A)(F - AX^{(k-1)}). \quad (13)$$

Пусть далее $Y^{(k-1)}$, $Y^{(k)}$ и $\bar{Y}^{(k)}$ соответствующие векторы ошибки. Из формулы (13) следует, что

$$\bar{Y}^{(k)} = \Phi_s(A) Y^{(k-1)},$$

где

$$\Phi_s(t) = 1 - tH_s(t).$$

Для оценки функции ошибки положим, как это делалось неоднократно, $A = B^2$, где B положительно-определенная матрица. Тогда

$$f(\bar{X}^{(k)}) = (A\bar{Y}^{(k)}, \bar{Y}^{(k)}) = (B\bar{Y}^{(k)}, B\bar{Y}^{(k)}) = |B(\bar{Y}^{(k)})|^2.$$

Но

$$B\bar{Y}^{(k)} = B\Phi_s(A)Y^{(k-1)} = \Phi_s(A)BY^{(k-1)},$$

и потому

$$|B\bar{Y}^{(k)}| = |\Phi_s(A)BY^{(k-1)}| \leq \|\Phi_s(A)\| |BY^{(k-1)}|.$$

Пусть

$$Q_s = \|\Phi_s(A)\|. \quad (14)$$

Тогда

$$f(\bar{X}^{(k)}) \leq Q_s^2 f(X^{(k-1)}) \quad (15)$$

и подавно

$$f(X^{(k)}) \leq Q_s^2 f(X^{(k-1)}). \quad (16)$$

Матрица $\Phi_s(A)$ — симметричная, так что ее норма Q_s равна наибольшему из модулей ее собственных значений, которые равны, очевидно, $\Phi_s(\lambda_1), \dots, \Phi_s(\lambda_n)$.

Полином $\Phi_s(t)$, очевидно, обладает свойством $\Phi_s(0) = 1$, его степень равна s , в остальном он произволен. Мы получим наилучшую

1) М. Ш. Бирман [1].

оценку (16), если среди полиномов указанного вида выберем полином $\Phi_s(t)$ таким, что $Q_s = \max_i \Phi_s(\lambda_i)$ будет наименьшим. Эта наилучшая оценка, конечно, зависит от $\lambda_1, \dots, \lambda_n$.

Для класса матриц, собственные значения которых заключены в промежутке (m, M) можно дать оценку, зависящую от чисел m и M , наилучшую для этого класса в целом. Для этого в качестве полинома $\Phi_s(t)$ следует взять полином, наименее отклоняющийся от нуля в промежутке (m, M) и нормированный так, что $\Phi_s(0) = 1$.

Линейной заменой $t = \frac{M-m}{2}\tau - \frac{M+m}{2}$ мы сведем задачу к построению полинома, наименее отклоняющегося от нуля в промежутке $-1 \leq \tau \leq 1$ и принимающего в точке $\tau_0 = \frac{M+m}{M-m}$ значение 1. Решение последней задачи¹⁾ дается полиномом

$$\tilde{T}_s(\tau) = \frac{\cos s \arccos \tau}{\cos s \arccos \tau_0}, \quad (17)$$

причем максимум отклонения равен

$$L_s = \max_{-1 \leq \tau \leq 1} |\tilde{T}_s(\tau)| = \frac{1}{|\cos s \arccos \tau_0|} = \frac{1}{T_s(\tau_0)},$$

где $T_s(\tau) = \cos s \arccos \tau$ — полином Чебышева.

Известно²⁾, что

$$T_s(\tau) = \frac{(\tau + \sqrt{\tau^2 - 1})^s + (\tau - \sqrt{\tau^2 - 1})^s}{2}.$$

Поэтому

$$L_s = \frac{2}{\left(\frac{M+m}{M-m} + \sqrt{\left(\frac{M+m}{M-m} \right)^2 - 1} \right)^s + \left(\frac{M+m}{M-m} - \sqrt{\left(\frac{M+m}{M-m} \right)^2 - 1} \right)^s}. \quad (18)$$

В частности,

$$\begin{aligned} L_1 &= \frac{M-m}{M+m}, & L_2 &= \frac{(M-m)^3}{(M+m)^3 + 4mM}, \\ L_3 &= \frac{(M-m)^5}{(M+m)[(M+m)^3 + 2mM]}, \dots \end{aligned} \quad (19)$$

Легко видеть, что

$$1 > L_1 > \sqrt{L_2} > \sqrt[3]{L_3} > \dots \quad (20)$$

¹⁾ В. Л. Гончаров. Теория интерполяции и приближения функций, ГТТИ, 1934, стр. 281.

²⁾ В. Л. Гончаров. Там же, стр. 27.

Последние неравенства означают, что при достаточно большом N результат применения $\left[\frac{N}{s}\right]$ шагов s -шагового процесса дает лучшее приближение, чем $\left[\frac{N}{s-1}\right]$ шагов $s-1$ -шагового процесса.

Заметим, что в приведенных рассуждениях мы не использовали результатов Л. В. Канторовича об оценке функции ошибок в одноступенчатом методе наискорейшего спуска, и потому равенство

$$L_1 = \frac{M-m}{M+m}$$

дает другой вывод оценки (10) § 70.

Можно дать и другие, более точные оценки для быстроты сходимости s -шагового метода наискорейшего спуска, если иметь более точную информацию о расположении собственных значений матрицы A .

Так, в работе Б. А. Самокиша [1] рассматривается ситуация, когда известно наибольшее собственное значение λ_1 и известен промежуток (m, M_1) , в котором расположены все остальные собственные значения.

В этом случае в качестве $\Phi_s(t)$ можно взять полином наименее уклоняющийся от нуля на множестве, состоящем из точки λ_1 и промежутка (m, M_1) . Обозначим отклонение от нуля выбранного полинома через \bar{L}_s . Тогда

$$\bar{L}_s < L_s, \text{ если } a = \frac{2\lambda_1 - M_1 - m}{M_1 - m} > \frac{3 - \cos \frac{\pi}{s}}{1 + \cos \frac{\pi}{s}}.$$

Далее, \bar{L}_s , монотонно возрастающая, стремится при $a \rightarrow \infty$ к L_{s-1} , построенному для промежутка (m, M_1) .

Можно подсчитать, что для двухшагового метода наискорейшего спуска

$$\bar{L}_2 = \frac{M_1 - m}{M_1 + m + \frac{2mM_1}{\lambda_1 - M_1}}, \text{ если } M_1 < \frac{\lambda_1 + m}{2}.$$

Для произвольного s соответствующие оценки получаются при помощи полиномов, изучавшихся Е. И. Золотаревым¹⁾, и оказываются громоздкими. Б. А. Самокишием [1] предложена приближенная формула

$$\begin{aligned} \bar{L}_s &\approx \frac{1}{R_s(\tau, a)}, \quad \tau = \frac{M_1 + m}{M_1 - m} \\ R_s(\tau, a) &= \frac{1}{2} \left[v^s \frac{v-a}{1+av} + v^{-s} \frac{1-av}{v-a} \right] \\ v &= \tau + \sqrt{\tau^2 - 1}, \quad a = a + \sqrt{a^2 - 1}. \end{aligned}$$

¹⁾ Е. И. Золотарев. Приложение эллиптических функций к вопросу о функциях, наименее и наиболее отклоняющихся от нуля. [Полн. собр. соч., вып. 2].

§ 74. Определение алгебраически наибольшего собственного значения симметричной матрицы и принадлежащего ему собственного вектора градиентными методами

Экстремальная теория собственных значений позволяет применить релаксационные градиентные методы и к определению крайних собственных значений (т. е. алгебраически наибольшего λ_1 и алгебраически наименьшего λ_n) симметричной матрицы A , так же как и принадлежащих им собственных векторов. Действительно,

$$\lambda_1 = \max \frac{(AX, X)}{(X, X)}$$

$$\lambda_n = \min \frac{(AX, X)}{(X, X)},$$

причем собственными векторами, принадлежащими этим собственным значениям, будут векторы, реализующие экстремум. Таким образом задача отыскания λ_1 или λ_n связывается с задачей максимизации или минимизации функционала

$$\mu(X) = \frac{(AX, X)}{(X, X)}.$$

Ввиду полной аналогии теории мы рассмотрим вопрос об отыскании алгебраически наибольшего собственного значения и принадлежащего ему собственного вектора.

Как было выяснено в § 14, градиентом функционала $\mu(X)$ будет вектор $\frac{2}{(X, X)} [AX - \mu(X)X] = \frac{2}{(X, X)} \xi$, где

$$\xi = AX - \mu(X)X. \quad (1)$$

Направление градиента совпадает с направлением вектора ξ , так как $(X, \xi) > 0$. Если X не является каким-либо собственным вектором матрицы A , то $\xi \neq 0$. Во всем дальнейшем, говоря о градиенте функционала $\mu(X)$, мы будем подразумевать под этим вектор ξ .

Пусть X_0 — произвольный вектор, не являющийся собственным вектором матрицы A . Пусть

$$X_1 = X_0 + \gamma \xi_0. \quad (2)$$

Из свойств градиента следует, что при достаточно малом по модулю γ справедливы неравенства $\mu(X_1) > \mu(X_0)$ при $\gamma > 0$ и $\mu(X_1) < \mu(X_0)$ при $\gamma < 0$.

Выясним подробнее, как ведет себя разность $\mu(X_1) - \mu(X_0)$ при изменении γ по всей вещественной оси.

Имеем

$$(AX_1, X_1) = (AX_0, X_0) + 2\gamma(AX_0, \xi_0) + \gamma^2(A\xi_0, \xi_0) = \\ = (AX_0, X_0) + 2\gamma(\xi_0, \xi_0) + \gamma^2(A\xi_0, \xi_0),$$

ибо

$$(AX_0, \xi_0) = (\xi_0, \xi_0).$$

Далее

$$(X_1, X_1) = (X_0, X_0) + \gamma^2(\xi_0, \xi_0),$$

ибо $(X_0, \xi_0) = 0$.

Таким образом,

$$\mu(X_1) = \mu(X_0 + \gamma\xi_0) = \frac{(AX_0, X_0) + 2\gamma(\xi_0, \xi_0) + \gamma^2(A\xi_0, \xi_0)}{(X_0, X_0) + \gamma^2(\xi_0, \xi_0)} = \\ = \frac{\mu(X_0) + 2\gamma t_0^2 + \gamma^2 t_0^2 \mu(\xi_0)}{1 + \gamma^2 t_0^2},$$

где

$$t_0^2 = \frac{(\xi_0, \xi_0)}{(X_0, X_0)}. \quad (3)$$

Поэтому

$$\mu(X_1) - \mu(X_0) = \frac{\mu(X_0) + 2\gamma t_0^2 + \gamma^2 t_0^2 \mu(\xi_0) - \mu(X_0) - \gamma^2 t_0^2 \mu(X_0)}{1 + \gamma^2 t_0^2} = \\ = \frac{2\gamma t_0^2 - \gamma^2 t_0^2 [\mu(X_0) - \mu(\xi_0)]}{1 + \gamma^2 t_0^2} = \frac{2\gamma - \gamma^2 s_0}{1 + \gamma^2 t_0^2} t_0^2, \quad (4)$$

где

$$s_0 = \mu(X_0) - \mu(\xi_0). \quad (5)$$

Из равенства (4) ясно, что $\mu(X_1) = \mu(X_0)$ при $\gamma = 0$ и при $\gamma = \frac{2}{s_0}$.

Кроме того, ясно, что $\mu(X_1) - \mu(X_0)$ есть непрерывная функция от γ при всех вещественных γ , включая $\gamma = \infty$, так как $\lim_{\gamma \rightarrow +\infty} [\mu(X_1) - \mu(X_0)] = \lim_{\gamma \rightarrow -\infty} [\mu(X_1) - \mu(X_0)] = -s_0$. Отметим, что $\mu(X_1) - \mu(X_0) = -s_0$ еще в одной точке, именно при $\gamma = -\frac{s_0}{2t_0^2}$. Таким образом, график $\mu(X_1) - \mu(X_0)$, в зависимости от знака s_0 , имеет вид, изображенный на рис. 9, 10, 11.

Из графиков видно, что неравенство $\mu(X_1) - \mu(X_0) > 0$ выполняется в области $0 < \gamma < \frac{2}{s_0}$, если $s_0 > 0$, в области $\gamma > 0$, если $s_0 = 0$, и в области $\gamma > 0$ или $\gamma < \frac{2}{s_0}$, если $s_0 < 0$.

Покажем, что для данной матрицы A можно указать такой промежуток изменения γ , в котором $\mu(X_1) - \mu(X_0) > 0$ независимо от выбора X_0 , т. е. независимо от величины s_0 . Именно таким промежутком является $0 < \gamma < \frac{2}{M-m}$. Действительно, при $s_0 \leq 0$, $\mu(X_1) - \mu(X_0) > 0$ при любом положительном γ , если же $s_0 > 0$, то $\mu(X_1) - \mu(X_0) > 0$, при $0 < \gamma < \frac{2}{M-m}$, ибо $\frac{2}{s_0} = \frac{2}{\mu(X_0) - \mu(\xi_0)} \geq \frac{2}{M-m}$.

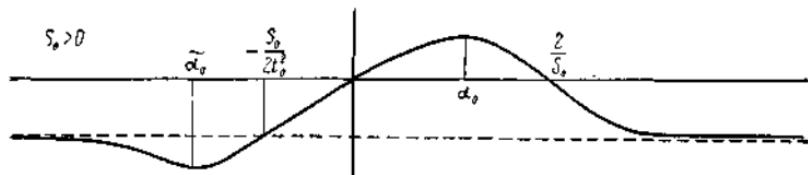


Рис. 9.

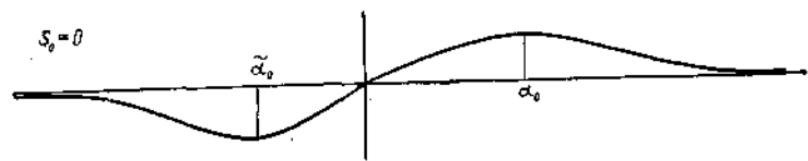


Рис. 10.

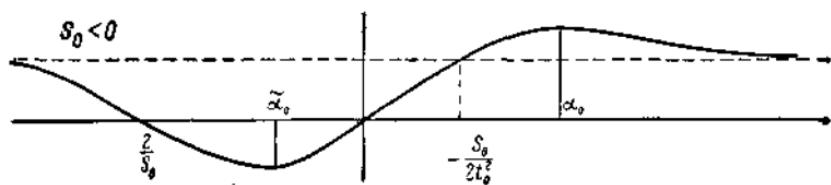


Рис. 11.

Найдем теперь значения параметра γ , при которых $\mu(X_1) - \mu(X_0)$ будет принимать экстремальные значения. Из приведенных графиков ясно, что максимум лежит направо, минимум — налево. Производная от $\mu(X_1) - \mu(X_0)$ по γ (с точностью до положительного множителя) будет

$$(2 - 2\gamma s_0)(1 + \gamma^2 t_0^2) - 2\gamma t_0^2(2\gamma - \gamma^2 s_0) = -2\gamma^2 t_0^2 - 2\gamma s_0 + 2,$$

и потому критические значения α и $\tilde{\alpha}$ параметра γ определяются из квадратного уравнения

$$\gamma^2 t_0^2 + \gamma s_0 - 1 = 0.$$

Положительный корень этого уравнения

$$\alpha_0 = \frac{-s_0 + \sqrt{s_0^2 + 4t_0^2}}{2t_0^2} = \frac{2}{\sqrt{s_0^2 + 4t_0^2} + s_0} \quad (6)$$

реализует максимум, отрицательный корень

$$\tilde{\alpha}_0 = \frac{-s_0 - \sqrt{s_0^2 + 4t_0^2}}{2t_0^2} = -\frac{2}{\sqrt{s_0^2 + 4t_0^2} - s_0} \quad (7)$$

реализует минимум. При этом

$$\mu(X_0 + \alpha_0 \xi_0) - \mu(X_0) = \frac{2\alpha_0 - \alpha_0^2 s_0}{1 + \alpha_0^2 t_0^2} t_0^2 = \frac{(2 - \alpha_0 s_0) \alpha_0}{1 + 1 - \alpha_0 s_0} t_0^2 = \alpha_0 t_0^2, \quad (8)$$

$$\mu(X_0 + \tilde{\alpha}_0 \xi_0) - \mu(X_0) = \tilde{\alpha}_0 t_0^2. \quad (9)$$

Отметим два свойства корня α_0 .

$$1. \quad \alpha_0 = \frac{1}{\mu(X_1) - \mu(\xi_0)} \geq \frac{1}{M - m}. \quad (10)$$

Действительно,

$$\mu(X_1) - \mu(\xi_0) = \mu(X_0 + \alpha_0 \xi_0) - \mu(X_0) + \mu(X_0) - \mu(\xi_0) = \alpha_0 t_0^2 + s_0 = \frac{1}{\alpha_0},$$

на основании уравнения для α_0 .

$$2. \quad \alpha_0 t_0^2 = \mu(X_0 + \alpha_0 \xi_0) - \mu(X_0) \leq M - m. \quad (11)$$

Коэффициент α_0 будем называть оптимальным коэффициентом.

Рассмотрим теперь следующую группу итерационных процессов, которые естественно назвать градиентными. Пусть X_0 некоторый начальный вектор, отличный от нулевого. Построим последовательность векторов

$$X_k = X_{k-1} + \gamma_{k-1} \xi_{k-1} \quad (k = 1, 2, \dots), \quad (12)$$

где

$$\xi_{k-1} = A X_{k-1} - \varphi_{k-1} X_{k-1}, \quad (13)$$

$\varphi_{k-1} = \mu(X_{k-1})$, а γ_{k-1} некоторое положительное число, выбираемое так, чтобы во всяком случае μ_k было бы меньше, чем μ_{k-1} .

Отметим, что если $X_{k-1} \neq 0$, то и $X_k \neq 0$. Действительно,

$$(X_k, X_k) = (X_{k-1} + \gamma_{k-1} \xi_{k-1}, X_{k-1} + \gamma_{k-1} \xi_{k-1}) = \\ = (X_{k-1}, X_{k-1}) + \gamma_{k-1}^2 (\xi_{k-1}, \xi_{k-1}) > 0.$$

Таким образом, описанный процесс продолжим неограниченно.

Две следующие теоремы¹⁾ дают достаточные условия сходимости градиентных методов.

¹⁾ Хестинс и Каруш [1].

Теорема 74.1. Если на всех шагах градиентного процесса

$$\mu(X_{k+1}) - \mu(X_k) \geq \delta \frac{(\xi_k, \xi_k)}{(X_k, X_k)} \quad (\delta > 0),$$

то последовательность $\mu(X_k)$ сходится к наибольшему собственному значению матрицы A в инвариантном подпространстве, порожденном вектором X_0 , а последовательность векторов X_k сходится по направлению к соответствующему собственному вектору.

Доказательство. Пусть U_1, \dots, U_r — нормированные собственные векторы, образующие базис циклического подпространства P_0 , порожденного вектором X_0 , $\lambda_1 > \lambda_2 > \dots > \lambda_r$ — соответствующие им собственные значения. Пусть

$$X_0 = a_1^{(0)} U_1 + \dots + a_r^{(0)} U_r.$$

Тогда (теорема 11.8) $a_1^{(0)} \neq 0, \dots, a_r^{(0)} \neq 0$. Без нарушения общности можно считать, что $a_1^{(0)} > 0$.

Ясно, что все векторы X_1, \dots, X_k, \dots содержатся в подпространстве P_0 и потому

$$X_k = a_1^{(k)} U_1 + \dots + a_r^{(k)} U_r.$$

Покажем, что $a_1^{(k)} > 0$. Имеем

$$\begin{aligned} a_1^{(k)} &= (X_k, U_1) = (X_{k-1}, U_1) + \gamma_{k-1} (\xi_{k-1}, U_1) = \\ &= (X_{k-1}, U_1) + \gamma_{k-1} (AX_{k-1} - \mu_{k-1} X_{k-1}, U_1) = \\ &= (X_{k-1}, U_1) + \gamma_{k-1} \lambda_1 (X_{k-1}, U_1) - \gamma_{k-1} \mu_{k-1} (X_{k-1}, U_1) = \\ &= [1 + \gamma_{k-1} (\lambda_1 - \mu_{k-1})] (X_{k-1}, U_1) = \\ &= [1 + \gamma_{k-1} (\lambda_1 - \mu_{k-1})] a_1^{(k-1)} = \dots = \\ &= [1 + \gamma_{k-1} (\lambda_1 - \mu_{k-1})] \dots [1 + \gamma_0 (\lambda_1 - \mu_0)] a_1^{(0)} > 0 \end{aligned}$$

ибо все множители последнего произведения положительны.

Отметим прежде всего, что процесс может стабилизироваться, если на каком-либо шагу окажется, что $\xi_k = 0$. В этом случае вектор X_k будет собственным вектором матрицы A и так как $(X_k, U_1) = a_1^{(k)} > 0$, то этот собственный вектор будет пропорционален U_1 . Таким образом, в случае стабилизации процесса теорема доказана.

Обратимся к рассмотрению общего случая, когда $\xi_k \neq 0$. Введем обозначение

$$b_1^{(k)} = \frac{a_1^{(k)}}{\|X_k\|}.$$

Тогда

$$X_k = \|X_k\| (b_1^{(k)} U_1 + \dots + b_r^{(k)} U_r),$$

причем $b_1^{(k)} > 0$.

Далее

$$\xi_k = a_1^{(k)}(\lambda_1 - \mu_k)U_1 + \dots + a_r^{(k)}(\lambda_r - \mu_k)U_r.$$

Следовательно,

$$t_k^2 = \frac{(\xi_k, \xi_k)}{(X_k, X_k)} = \frac{\sum_{i=1}^r a_i^{(k)2}(\lambda_i - \mu_k)^2}{|X_k|^2} = \sum_{i=1}^r b_i^{(k)2}(\lambda_i - \mu_k)^2.$$

Так как $\mu_1 < \mu_2 < \dots < \mu_k < \dots < M = \lambda_1$, то $\lim_{k \rightarrow \infty} \mu_k$ существует.

Обозначим его через μ . По условию теоремы

$$t_k^2 \leq \frac{1}{b} (\mu_{k+1} - \mu_k) \rightarrow 0 \quad (k \rightarrow \infty).$$

Поэтому $\sum_{i=1}^r b_i^{(k)2}(\lambda_i - \mu_k)^2 \rightarrow 0$ при $k \rightarrow \infty$, т. е. $b_i^{(k)2}(\lambda_i - \mu_k)^2 \rightarrow 0$ при всех $i = 1, 2, \dots, r$. Если $\lambda_i - \mu \neq 0$, то $b_i^{(k)} \rightarrow 0$ при $k \rightarrow \infty$. Этого однако не может быть при всех $i = 1, 2, \dots, r$, ибо $\sum_{i=1}^r b_i^{(k)2} = 1$. Итак, найдется такое j , что $\mu = \lambda_j$. Тогда $\lim_{k \rightarrow \infty} b_j^{(k)2} = 0$ при $i \neq j$, а $\lim_{k \rightarrow \infty} b_j^{(k)2} = 1$.

Покажем, что $j = 1$. Если допустить, что $j > 1$, то

$$\left| \frac{b_1^{(k)}}{b_j^{(k)}} \right| = \left| \frac{a_1^{(k)}}{a_j^{(k)}} \right| \rightarrow 0.$$

С другой стороны, имеем

$$\begin{aligned} \left| \frac{a_1^{(k)}}{a_j^{(k)}} \right| &= \left| \frac{1 + \gamma_{k-1}(\lambda_1 - \mu_{k-1})}{1 + \gamma_{k-1}(\lambda_j - \mu_{k-1})} \right| \left| \frac{a_j^{(k-1)}}{a_j^{(k-1)}} \right| = \\ &= \left| 1 + \frac{\gamma_{k-1}(\lambda_1 - \lambda_j)}{1 + \gamma_{k-1}(\lambda_j - \mu_{k-1})} \right| \left| \frac{a_j^{(k-1)}}{a_j^{(k-1)}} \right| > \left| \frac{a_j^{(k-1)}}{a_j^{(k-1)}} \right| \end{aligned}$$

ибо $\frac{\gamma_{k-1}(\lambda_1 - \lambda_j)}{1 + \gamma_{k-1}(\lambda_j - \mu_{k-1})} > 0$, так как $\gamma_{k-1} > 0$ по построению и $\mu_{k-1} < \mu = \lambda_j < \lambda_1$. Полученное противоречие показывает, что $j = 1$.

Итак, мы доказали, что $\mu_k \rightarrow \lambda_1$ при $k \rightarrow \infty$. Тем самым установлено, что $b_1^{(k)2} \rightarrow 1$ и так как $b_1^{(k)} > 0$, то $b_1^{(k)} \rightarrow 1$. При этом $b_i^{(k)} \rightarrow 0$ при $i = 2, \dots, r$. Но

$$\frac{X_k}{|X_k|} = b_1^{(k)}U_1 + \dots + b_r^{(k)}U_r.$$

Следовательно,

$$\frac{X_k}{\|X_k\|} \rightarrow U_1,$$

т. е. последовательность X_k сходится к U_1 по направлению.

Теорема 74.2. Если сверх условия теоремы 74.1 все числа γ_k ограничены сверху в совокупности, то $\lim_{k \rightarrow \infty} X_k = LU_1$, где L некоторое положительное число.

Доказательство. В силу теоремы 74.1 нам надо доказать, что $\|X_k\| \rightarrow L$ при $k \rightarrow \infty$. Но

$$(X_k, X_k) = (X_{k-1}, X_{k-1}) + \gamma_{k-1}^2 (\xi_{k-1}, \xi_{k-1}) = (1 + \gamma_{k-1}^2 t_{k-1}^2) (X_{k-1}, X_{k-1}).$$

Следовательно,

$$(X_k, X_k) = (1 + \gamma_0^2 t_0^2)(1 + \gamma_1^2 t_1^2) \cdots (1 + \gamma_{k-1}^2 t_{k-1}^2) (X_0, X_0).$$

Бесконечное произведение

$$\prod_{k=0}^{\infty} (1 + \gamma_k^2 t_k^2)$$

сходится, ибо $\sum_{k=0}^{\infty} t_k^2$ сходится (так как он мажорируется сходящимся

рядом $\sum_{k=0}^{\infty} \frac{1}{\delta} (\mu_{k+1} - \mu_k)$), а γ_k ограничены сверху по условию теоремы.

Рассмотрим теперь несколько частных градиентных методов.

1. $\gamma_k = \gamma = \text{const}$ — метод постоянного множителя. В этом случае для возрастания μ_k на каждом шаге процесса при любом начальном векторе необходимо, как мы видели, выполнение неравенства

$$0 < \gamma < \frac{2}{M-m}.$$

Покажем, что при выполнении этого неравенства выполняются и условия теорем 74.1 и 74.2. Действительно, пусть $\gamma = \frac{\beta}{M-m}$, $0 < \beta < 2$. Тогда

$$\begin{aligned} \frac{\mu(X_k) - \mu(X_{k-1})}{t_{k-1}^2} &= \frac{2\gamma - \gamma^2 s_{k-1}}{1 + \gamma^2 t_{k-1}^2} = \\ &= \frac{\gamma \left(2 - \frac{\beta}{M-m} s_{k-1}\right)}{1 + \gamma^2 t_{k-1}^2} \geq \frac{\gamma(2-\beta)}{1+\gamma^2 t^2} = \delta > 0, \end{aligned}$$

ибо $t_{k-1}^2 \leq l^2$, где l сферическая норма матрицы A . Действительно,

$$\begin{aligned} t_k^2 &= \frac{(\xi_k, \xi_k)}{(X_k, X_k)} = \frac{(AX_k, \xi_k)}{(X_k, X_k)} = \frac{(AX_k, AX_k)}{(X_k, X_k)} - \\ &- \frac{(AX_k, X_k)^2}{(X_k, X_k)^2} \leq \frac{(AX_k, AX_k)}{(X_k, X_k)} = \frac{\|AX_k\|^2}{\|X_k\|^2} \leq \|A\|^2. \end{aligned}$$

Остальные условия теоремы, очевидно, выполняются.

2. $\gamma_k = \alpha_k$, где α_k — оптимальный коэффициент k -го шага — метод наискорейшего спуска. В этом случае $\frac{\mu(X_k) - \mu(X_{k-1})}{t_{k-1}^2} = \alpha_k$,

и потому для проверки условий выполнения теоремы 74.1 надо убедиться в ограниченности снизу чисел α_k . Но мы установили ранее, что $\alpha_k \geq \frac{1}{M-m}$, так что ограниченность снизу действительно имеет место.

Чтобы убедиться, что условие теоремы 74.2 тоже выполнено, нужно доказать ограниченность чисел α_k сверху. Здесь, в отличие от предыдущих оценок, оказывается, что верхняя граница существует, но зависит от начального приближения. Именно, нетрудно видеть, что все значения α_k при достаточно больших k удовлетворяют условию

$$\alpha_k < \frac{2}{\lambda_1 - \lambda_2}.$$

Действительно,

$$\alpha_k = \frac{1}{\mu(X_k) - \mu(\xi_{k-1})}.$$

При достаточно большом k , $\mu(X_k)$ становится сколь угодно близко к λ_1 , т. е. $\mu(X_k) > \lambda_1 - \varepsilon$, при $\varepsilon > 0$. С другой стороны, ξ_{k-1} ортогонален X_{k-1} и, следовательно, ξ_{k-1} после нормировки (к единичной длине) подходит сколь угодно близко к подпространству, ортогональному к U_1 . В этом подпространстве $\mu(X)$ не превосходит λ_2 . Поэтому $\mu(\xi_{k-1}) \leq \lambda_2 + \varepsilon$, при достаточно большом k . Следовательно, при достаточно большом k

$$\alpha_k \leq \frac{1}{\lambda_1 - \varepsilon - \lambda_2 - \varepsilon} = \frac{1}{\lambda_1 - \lambda_2 - 2\varepsilon} < \frac{2}{\lambda_1 - \lambda_2},$$

если взять $\varepsilon < \frac{\lambda_1 - \lambda_2}{4}$. Таким образом, теорема 74.2 оказывается справедливой и для метода наискорейшего спуска.

3. $\gamma_k = \beta \alpha_k$, где α_k — оптимальный коэффициент k -го шага, $0 < \beta < 1$ — метод неполного наискорейшего спуска. Для справедливости теоремы 74.1 надо доказать, что последовательность

$\frac{\mu(X_{k+1}) - \mu(X_k)}{t_k^2}$ ограничена снизу. Имеем

$$\begin{aligned} \frac{\mu(X_{k+1}) - \mu(X_k)}{t_k^2} &= \frac{2\alpha_k\beta - \alpha_k^2\beta^2 s_k}{1 + \alpha_k^2\beta^2 t_k^2} = \\ &= \frac{2\alpha_k\beta - \alpha_k\beta^2(1 - t_k^2\alpha_k^2)}{1 + \alpha_k^2\beta^2 t_k^2} = \alpha_k\beta \frac{2 - \beta + \beta t_k^2\alpha_k^2}{1 + \alpha_k^2\beta^2 t_k^2} = \\ &= \alpha_k\beta \left(1 + \frac{1 - \beta + \beta(1 - \beta)\alpha_k^2 t_k^2}{1 + \alpha_k^2\beta^2 t_k^2}\right) = \\ &= \alpha_k\beta \left(1 + (1 - \beta) \frac{1 + \beta\alpha_k^2 t_k^2}{1 + \alpha_k^2\beta^2 t_k^2}\right) \geq \alpha_k\beta. \end{aligned}$$

Следовательно,

$$\frac{\mu(X_{k+1}) - \mu(X_k)}{t_k^2} \geq \frac{\beta}{M - m}.$$

Справедливость теоремы 74.2 доказывается так же, как и в методе наискорейшего спуска.

4. $\gamma_k = \frac{1}{\mu(X_k)} = \frac{1}{\mu_k}$ — степенной метод. В этом случае

$$X_{k+1} = \frac{1}{\mu_k} A X_k.$$

Имеем

$$\frac{\mu(X_{k+1}) - \mu(X_k)}{t_k^2} = \frac{\frac{2}{\mu_k} - \frac{s_k}{\mu_k^2}}{1 + \frac{t_k^2}{\mu_k^2}} = \frac{2\mu_k - s_k}{\mu_k^2 + t_k^2} = \frac{\mu(X_k) + \mu(E_k)}{\mu^2(X_k) + t_k^2}.$$

Если A положительно-определенная матрица, то

$$\frac{\mu(X_{k+1}) - \mu(X_k)}{t_k^2} \geq \frac{2m}{M^2 + l^2} > 0,$$

где l — сферическая норма матрицы A . Очевидно также, что $\gamma_k = \frac{1}{\mu_k} \leq \frac{1}{m}$. Таким образом, в случае положительно-определенной матрицы степенной метод является не только сходящимся, но и релаксационным, ибо на каждом шаге происходит увеличение $\mu(X_k)$.

В случае произвольной симметричной матрицы этого может не быть на первых шагах процесса. Однако, если $\lambda_1 > -\lambda_n$, т. е. алгебраически наибольшим является собственное значение, наибольшее по

Таблица VII. 9
Определение наибольшего собственного значения градиентным методом с постоянным множителем $\gamma = 0.6$

	X_0	$A X_0$	ξ_0	X_1	X_2	X_3	X_4	\tilde{X}_4	\tilde{X}_7
I	0.34	1.4050	0.78278286	0.809686972	0.81117064	0.80557137	0.80682530	1.0000	1.00000
	0.20	0.9608	0.59478992	0.556687395	0.63814979	0.63972334	0.64009468	0.7934	0.79359
	0.90	1.3126	-0.33444536	0.69933278	0.69957215	0.60444577	0.60307521	0.7475	0.74781
	0.75	1.2604	-0.11213780	0.68271732	0.70658793	0.71601739	0.71543204	0.8867	0.88732
II		4.9388							
III	1.5281	2.7965							
IV	1.8300504			2.3087048	2.3226218	2.3227458	2.3227486		2.3227488

Таблица VII. 10
Определение наибольшего собственного значения методом наискорейшего спуска

X_0	$A X_0$	ξ_0	$A \xi_0$	X_1	X_2	X_3	X_4	X_5	\tilde{X}_4	\tilde{X}_5
0.34	1.4050	0.78278286	0.777998318	0.87034015	0.82112664	0.82844787	0.82774590	0.82787672	1.0000	1.00000
	0.9608	0.59478992	0.76719557	0.60297379	0.65447856	0.65670073	0.65693195	0.65697730	0.7936	0.79357
	1.3126	-0.33444536	0.25391984	0.67341123	0.62913176	0.62014585	0.61926655	0.61910096	0.7481	0.74783
	0.75	1.2604	-0.11213780	0.59262847	0.67402596	0.73573673	0.73327297	0.73460252	0.8875	0.88727
	4.9388	-0.18 · 10 ⁻⁷	2.39172706							
1.5281	2.7965	1.09095264	0.91393373							
1.8300504				2.3137407	2.3226083	2.3227463	2.3227487	2.3227488		

модулю, то из факта сходимости по направлению векторов X_k и U , следует, что $\mu(X_k) \rightarrow \lambda_1$ а $\mu(\xi_k) \geq m$ и, следовательно,

$$\frac{\mu(X_{k+1}) - \mu(X_k)}{t_k^2} \geq \frac{(\lambda_1 - \epsilon) + m}{\lambda_1^2 + t^2},$$

начиная с некоторого места. Таким образом, и в этом случае степенной метод сохраняет релаксационный характер, начиная с некоторого шага процесса.

В табл. VII.9 определяется наибольшее собственное значение матрицы (4) § 51 градиентным методом с постоянным γ , в табл. VII.10 методом наискорейшего спуска. В табл. VII.11 для той же матрицы определяется наименьшее собственное значение.

В последней строке таблиц записываются значения $\mu(X)$.

На каждом шаге процесса в табл. VII.10 оптимальный коэффициент находится по формуле (6). Для первого шага имеем

$$s_0 = 0.99231126, \quad t_0^2 = 0.71392752,$$

так что

$$\alpha_0 = \frac{2}{\sqrt{3.8403917} + 0.99231126} = 0.67750609.$$

Приведем также значения

$$\begin{aligned} \alpha_1 &= 0.60271425, & \alpha_2 &= 0.51669127, & \alpha_3 &= 0.59837767, \\ \alpha_4 &= 0.51670397. \end{aligned}$$

Таблица VII. 11

Определение наименьшего собственного значения методом наискорейшего спуска

X_0	AX_0	ξ_0	$A\xi_0$	X_1	X_2	X_{13}	\tilde{X}_{13}
0.34	1.4050	0.78278286	0.77798318	-1.2783558	-2.0012022	-2.0088782	1.0000
0.20	0.9608	0.59478992	0.76719557	-1.0296919	-0.2776540	-0.2691368	0.3140
0.90	1.3126	-0.33444536	0.25391984	1.5914454	1.0838913	1.0828515	-0.5390
0.75	1.2604	-0.11213780	0.59262847	0.9818381	1.5991528	1.5917026	-0.7923
	4.9388	$-0.18 \cdot 10^{-7}$	2.39172706				
1.5281	2.7965	1.09095264	0.91303373				
1.8300504		0.83773917			0.24227728	0.24226089	

На каждом шаге процесса в табл. VII. 11 оптимальный коэффициент $\tilde{\alpha}_k$ находится по формуле (7). Для первого шага имеем

$$s_0 = 0.99231126, \quad t_0^2 = 0.71392752$$

$$\tilde{\alpha}_0 = \frac{-2}{\sqrt{3.8403917 - 0.99231126}} = -2.0674390.$$

Сравнение табл. VII. 9 и VII. 10 показывает, что в данном примере метод постоянного множителя с $\gamma = 0.6$ сходится лишь немного медленнее, чем метод наискорейшего спуска. Объем же вычислений для одного шага метода постоянного множителя вдвое меньше, чем в методе наискорейшего спуска.

Вопрос о выборе значения постоянного множителя пока не исследован. Для положительно-определенной матрицы во всяком случае в качестве γ может быть взято число не превосходящее $\frac{2}{\|A\|_1}$.

Вычисления, проведенные для матрицы (4) § 51 при $\gamma = 0.5$, $\gamma = 0.4$, $\gamma = 0.3$ и $\gamma = 0.25$, показывают, что с уменьшением γ сходимость процесса замедляется.

Применение неполного наискорейшего спуска при $\beta = 0.8$ и $\beta = 0.9$ в рассматриваемом примере оказалось не целесообразным, так как сходимость процесса не улучшилась.

Возратимся теперь снова к методу постоянного множителя, причем будем считать на этот раз, что

$$0 < \gamma < \frac{1}{M-m}.$$

Отметим следующее важное свойство последовательных приближений.

Лемма. Если в методе постоянного множителя $\gamma = \frac{\beta}{M-m}$ и $0 < \beta < 1$, то $\lim_{k \rightarrow \infty} \frac{a_j^{(k)}}{a_i^{(k)}} = 0$ при $i < j$.

Здесь $a_1^{(k)}, a_2^{(k)}, \dots, a_r^{(k)}$ коэффициенты в разложении приближения X_k по собственным векторам U_1, \dots, U_r , содержащимся в инвариантном подпространстве P_0 , порожденном начальным приближением.

Доказательство. Пусть

$$X_0 = a_1^{(0)}U_1 + a_2^{(0)}U_2 + \dots + a_r^{(0)}U_r.$$

Тогда

$$X_k = a_1^{(k)}U_1 + a_2^{(k)}U_2 + \dots + a_r^{(k)}U_r,$$

причем

$$a_i^{(k)} = [1 + \gamma(\lambda_i - \mu_{k-1})] a_i^{(k-1)}.$$

Отсюда следует, что $a_i^{(k)} > 0$, ибо $1 + \gamma(\lambda_i - \mu_{k-1}) = 1 + \beta \frac{\lambda_i - \mu_{k-1}}{M - m} > 0$.
Обозначим

$$\delta_i = 1 + \gamma(\lambda_i - \lambda_1) = 1 + \beta \frac{\lambda_i - \lambda_1}{M - m}.$$

Очевидно, в силу выбора β

$$0 < \delta_r < \delta_{r-1} < \dots < \delta_2 < \delta_1 = 1.$$

Так как $\mu_k \rightarrow \lambda_1$, то $\frac{a_i^{(k)}}{a_i^{(k-1)}} \rightarrow \delta_i$. Поэтому

$$\frac{a_j^{(k)}}{a_i^{(k)}} : \frac{a_j^{(k-1)}}{a_i^{(k-1)}} \rightarrow \frac{\delta_j}{\delta_i}$$

и $0 < \frac{\delta_j}{\delta_i} < 1$, если $i < j$. Следовательно, начиная с некоторого места,

$$\frac{a_j^{(k)}}{a_i^{(k)}} \leq L \left(\frac{\delta_j}{\delta_i} + \varepsilon \right)^k \quad (k > k_0).$$

Выбирая ε так, что $\frac{\delta_j}{\delta_i} + \varepsilon < 1$, получим, что $\frac{a_j^{(k)}}{a_i^{(k)}} \rightarrow 0$ при $i < j$.

В этом „суженном“ методе постоянного множителя интересным является и поведение последовательности ξ_k .

Теорема 74.3.¹⁾ *Если в методе постоянного множителя*

$$\gamma = \frac{\beta}{M - m} \text{ и } 0 < \beta < 1, \text{ то } \lim_{k \rightarrow \infty} \frac{\xi_k}{|\xi_k|} = U_2, \lim_{k \rightarrow \infty} \mu(\xi_k) = \lambda_2.$$

Доказательство. Имеем

$$X_k = a_1^{(k)} U_1 + a_2^{(k)} U_2 + \dots + a_r^{(k)} U_r, \quad a_i^{(k)} > 0,$$

$$\xi_k = (\lambda_1 - \mu_k) a_1^{(k)} U_1 + (\lambda_2 - \mu_k) a_2^{(k)} U_2 + \dots + (\lambda_r - \mu_k) a_r^{(k)} U_r,$$

причем

$$(\xi_k, \xi_k) = (\lambda_1 - \mu_k)^2 a_1^{(k)2} + (\lambda_2 - \mu_k)^2 a_2^{(k)2} + \dots + (\lambda_r - \mu_k)^2 a_r^{(k)2} \neq 0,$$

так как $r \geq 2$. Имеем далее

$$(X_k, \xi_k) = 0 = (\lambda_1 - \mu_k) a_1^{(k)2} + (\lambda_2 - \mu_k) a_2^{(k)2} + \dots + (\lambda_r - \mu_k) a_r^{(k)2}.$$

Отсюда, поделив на $a_2^{(k)2}$, после перехода к пределу получим, на основании предыдущей леммы,

$$\frac{\lambda_1 - \mu_k}{a_2^{(k)2}} a_1^{(k)2} \rightarrow \lambda_1 - \lambda_2.$$

¹⁾ Хестинс и Каруш [1].

Но

$$\frac{1}{a_2^{(k)}} \xi_k = \frac{(\lambda_1 - \mu_k) a_1^{(k)2}}{a_2^{(k)2}} \frac{a_2^{(k)}}{a_1^{(k)}} U_1 + (\lambda_2 - \mu_k) U_2 + \dots + \frac{(\lambda_r - \mu_k) a_r^{(k)}}{a_2^{(k)}} U_r$$

и, следовательно,

$$\frac{1}{a_2^{(k)}} \xi_k \rightarrow (\lambda_2 - \lambda_1) U_2.$$

На основании § 13 (стр. 118) заключаем, что последовательность ξ_k стремится по направлению к U_2 .

Далее,

$$\mu(\xi_k) = \mu\left(\frac{\xi_k}{|\xi_k|}\right) \rightarrow \mu(U_2) = \lambda_2.$$

В качестве примера проведем описанный процесс для матрицы

$$\begin{bmatrix} 0.22 & 0.02 & 0.12 & 0.14 \\ 0.02 & 0.14 & 0.04 & -0.06 \\ 0.12 & 0.04 & 0.28 & 0.08 \\ -0.14 & -0.06 & 0.08 & 0.26 \end{bmatrix}.$$

Здесь точными значениями являются $\lambda_1 = 0.48$, $\lambda_2 = 0.24$, $\lambda_3 = 0.12$, $\lambda_4 = 0.06$, $U_1 = (1, 0, 1, 1)^T$, $U_2 = (0, -1, -1, 1)^T$.

Для применения метода постоянного множителя можно взять $\gamma = 1$. Имеем

X_0	X_{25}	ξ_{25}	X_{43}	ξ_{43}	$\tilde{\xi}_{43}$
0.34	0.72458928	$0.008256 \cdot 10^{-4}$	0.724591532513	$0.000237 \cdot 10^{-6}$	0.001
0.20	0.00013557	$-0.323536 \cdot 10^{-4}$	0.000000981434	$-0.235470 \cdot 10^{-6}$	-0.998
0.90	0.72473074	$-0.336452 \cdot 10^{-4}$	0.724592515860	$-0.235923 \cdot 10^{-6}$	-1.000
0.75	0.72445458	$0.328216 \cdot 10^{-4}$	0.724590551128	$0.235687 \cdot 10^{-6}$	0.999
μ	0.48000000	0.239942	0.48000000	0.23999989	

Мы видим, что при $k = 43$ второе собственное значение определяется с точностью до $1 \cdot 10^{-7}$. Собственный вектор, ему принадлежащий, определяется значительно хуже.

Совсем иную картину мы имеем в методе наискорейшего спуска. Здесь последовательность ξ_k не является сходящейся. Более того, справедлива следующая теорема.

Теорема 74.4. Градиенты соседних приближений метода наискорейшего спуска ортогональны.

Доказательство. Имеем

$$\xi_k = AX_k - \mu_k X_k$$

$$\xi_{k+1} = AX_{k+1} - \mu_{k+1} X_{k+1} = A(X_k + \alpha_k \xi_k) - \mu_{k+1}(X_k + \alpha_k \xi_k).$$

Следовательно,

$$\begin{aligned} (\xi_{k+1}, \xi_k) &= (AX_k, \xi_k) + \alpha_k (A\xi_k, \xi_k) - \mu_{k+1}(X_k, \xi_k) - \alpha_k \mu_{k+1}(\xi_k, \xi_k) = \\ &= [1 + \alpha_k \mu(\xi_k) - \alpha_k \mu_{k+1}] (\xi_k, \xi_k) = 0, \end{aligned}$$

так как

$$\alpha_k = \frac{1}{\mu(X_{k+1}) - \mu(\xi_k)} = \frac{1}{\mu_{k+1} - \mu(\xi_k)}.$$

§ 75. Решение частичной проблемы собственных значений с помощью полиномов Ланцоша

Установленное в теореме 74.3 свойство последовательности градиентов приближений суженного метода постоянного множителя допускает обобщение, позволяющее находить несколько собственных значений и принадлежащих им собственных векторов с заранее фиксированным числом m собственных значений, подлежащих определению.

Введем следующее обозначение. Через $p_0^{(k)}, p_1^{(k)}, \dots, p_{m-1}^{(k)}$ обозначим первые m векторов Ланцоша, построенные из начального вектора X_k , через $p_0^{(k)}(t), p_1^{(k)}(t), \dots, p_{m-1}^{(k)}(t)$ соответствующие им полиномы. Ясно, что $p_0^{(k)} = X_k, p_1^{(k)} = \xi_k$.

Теорема 75.1. Пусть $X_k = p_0^{(k)} = a_1^{(k)}U_1 + a_2^{(k)}U_2 + \dots + a_r^{(k)}U_r$ последовательность векторов в подпространстве, натянутом на нормированные собственные векторы U_1, U_2, \dots, U_r симметричной матрицы A , соответствующие собственным значениям

$\lambda_1 > \lambda_2 > \dots > \lambda_r$. Пусть $\frac{a_2^{(k)}}{a_1^{(k)}} \rightarrow 0, \frac{a_3^{(k)}}{a_2^{(k)}} \rightarrow 0, \dots, \frac{a_r^{(k)}}{a_{r-1}^{(k)}} \rightarrow 0$ при $k \rightarrow \infty$.

Тогда, если $p_i^{(k)} = p_i^{(k)}(A)p_0^{(k)}$ i -й вектор Ланцоша, то соответствующие полиномы $p_i^{(k)}(t)$ сходятся к полиному $\bar{p}_i(t) = (t - \lambda_1) \dots (t - \lambda_i)$, а векторы $p_i^{(k)}$ сходятся по направлению к вектору U_{i+1} .

Доказательство. Будем доказывать теорему по индукции. Для $i = 0$ утверждение справедливо, ибо

$$\frac{1}{a_1^{(k)}} p_0^{(k)} = U_1 + \frac{a_2^{(k)}}{a_1^{(k)}} U_2 + \dots + \frac{a_r^{(k)}}{a_1^{(k)}} U_r \rightarrow U_1.$$

Допустим, что утверждение теоремы справедливо для индексов $0, 1, \dots, i-1$ и в этом предположении докажем его для индекса i . Имеем

$$\begin{aligned} p_j^{(k)} &= a_1^{(k)} p_j^{(k)}(\lambda_1) U_1 + \dots + a_{j+1}^{(k)} p_j^{(k)}(\lambda_{j+1}) U_{j+1} + \dots + a_r^{(k)} p_j^{(k)}(\lambda_r) U_r = \\ &= a_{j+1}^{(k)} p_j^{(k)}(\lambda_{j+1}) [b_{1j}^{(k)} U_1 + \dots + b_{jj}^{(k)} U_j + U_{j+1} + \\ &\quad + b_{j+2,j}^{(k)} U_{j+2} + \dots + b_{rj}^{(k)} U_r] = a_{j+1}^{(k)} p_j^{(k)}(\lambda_{j+1}) V_j^{(k)}, \end{aligned}$$

где

$$b_{sj}^{(k)} = \frac{a_s^{(k)}}{a_{j+1}^{(k)}} \frac{p_j^{(k)}(\lambda_s)}{p_j^{(k)}(\lambda_{j+1})}$$

$$V_j^{(k)} = b_{1j}^{(k)} U_1 + \dots + b_{jj}^{(k)} U_j + U_{j+1} + b_{j+2,j}^{(k)} U_{j+2} + \dots + b_{rj}^{(k)} U_r.$$

В силу индукционного предположения

- 1) $p_j^{(k)}(t) \rightarrow \bar{p}_j(t)$ при $j < i$,
- 2) $V_j^{(k)} \rightarrow U_{j+1}$ при $j < i$.

Докажем то же самое для $j = i$. Имеем

$$p_i^{(k)}(t) = (t - \alpha_{i-1}^{(k)}) p_{i-1}^{(k)}(t) - \beta_{i-1}^{(k)} p_{i-2}^{(k)}(t),$$

где

$$\alpha_{i-1}^{(k)} = \frac{(A p_{i-1}^{(k)}, p_{i-1}^{(k)})}{(p_{i-1}^{(k)}, p_{i-1}^{(k)})} = \frac{(AV_{i-1}^{(k)}, V_{i-1}^{(k)})}{(V_{i-1}^{(k)}, V_{i-1}^{(k)})}$$

$$\beta_{i-1}^{(k)} = \frac{(p_{i-1}^{(k)}, p_{i-1}^{(k)})}{(p_{i-2}^{(k)}, p_{i-2}^{(k)})} = \left[\frac{a_i^{(k)}}{a_{i-1}^{(k)}} \right]^2 \left[\frac{p_{i-1}^{(k)}(\lambda_i)}{p_{i-2}^{(k)}(\lambda_i)} \right]^2 \frac{(V_{i-1}^{(k)}, V_{i-1}^{(k)})}{(V_{i-2}^{(k)}, V_{i-2}^{(k)})}.$$

В силу индукционного предположения

$$\alpha_{i-1}^{(k)} \rightarrow \frac{(AU_i, U_i)}{(U_i, U_i)} = \lambda_i.$$

Далее,

$$\frac{(V_{i-1}^{(k)}, V_{i-1}^{(k)})}{(V_{i-2}^{(k)}, V_{i-2}^{(k)})} \rightarrow \frac{(U_i, U_i)}{(U_{i-1}, U_{i-1})} = 1$$

$$\frac{p_{i-1}^{(k)}(\lambda_i)}{p_{i-2}^{(k)}(\lambda_i)} \rightarrow \frac{\bar{p}_{i-1}(\lambda_i)}{\bar{p}_{i-2}(\lambda_i)},$$

а $\frac{a_i^{(k)}}{a_{i-1}^{(k)}} \rightarrow 0$ в силу условия теоремы. Следовательно, $\beta_{i-1}^{(k)} \rightarrow 0$ при $k \rightarrow \infty$ и потому

$$p_i^{(k)}(t) \rightarrow (t - \lambda_i) \bar{p}_{i-1}(t) = \bar{p}_i(t).$$

Для доказательства второго утверждения нам надо установить, что коэффициенты $b_{si}^{(k)} \rightarrow 0$ при $k \rightarrow \infty$, если $s \neq i+1$.

Для $s \geq i+2$ это устанавливается очень просто. Именно,

$$b_{si}^{(k)} = \frac{a_s^{(k)}}{a_{i+1}^{(k)}} \frac{p_i^{(k)}(\lambda_s)}{p_i^{(k)}(\lambda_{i+1})}.$$

В силу условия теоремы $\frac{a_s^{(k)}}{a_{i+1}^{(k)}} \rightarrow 0$ при $s \geq i+2$ и, в силу уже доказанного, $p_i^{(k)}(\lambda_s) \rightarrow \bar{p}_i(\lambda_s)$, $p_i^{(k)}(\lambda_{i+1}) \rightarrow \bar{p}_i(\lambda_{i+1}) = (\lambda_{i+1} - \lambda_i) \dots (\lambda_{i+1} - \lambda_i) \neq 0$.

Труднее устанавливаются предельные соотношения $b_{si}^{(k)} \rightarrow 0$, если $s \leq i$. Для доказательства положим

$$b_{si}^{(k)} = c_{si}^{(k)} \frac{a_{i+1}^{(k)}}{a_s^{(k)}} \quad (s \leq i), \quad (1)$$

где

$$c_{si}^{(k)} = \left[\frac{a_s^{(k)}}{a_{i+1}^{(k)}} \right]^2 \frac{p_i^{(k)}(\lambda_s)}{p_i^{(k)}(\lambda_{i+1})}. \quad (2)$$

В силу условия теоремы $\frac{a_{i+1}^{(k)}}{a_s^{(k)}} \rightarrow 0$ при $s \leq i$. Докажем, что $c_{si}^{(k)}$ стремится при $k \rightarrow \infty$ к конечным пределам. Тем самым будет установлено, что $b_{si}^{(k)} \rightarrow 0$ при $s \leq i$.

Для доказательства используем ортогональность вектора $p_i^{(k)}$ к векторам $p_0^{(k)}, \dots, p_{i-1}^{(k)}$. Имеем

$$\begin{aligned} 0 &= (p_i^{(k)}, p_s^{(k)}) = a_1^{(k)2} p_i^{(k)}(\lambda_1) p_s^{(k)}(\lambda_1) + \dots \\ &\quad \dots + a_i^{(k)2} p_i^{(k)}(\lambda_i) p_s^{(k)}(\lambda_i) + a_{i+1}^{(k)2} p_i^{(k)}(\lambda_{i+1}) p_s^{(k)}(\lambda_{i+1}) + \\ &\quad \dots + a_{i+2}^{(k)2} p_i^{(k)}(\lambda_{i+2}) p_s^{(k)}(\lambda_{i+2}) + \dots + a_r^{(k)2} p_i^{(k)}(\lambda_r) p_s^{(k)}(\lambda_r). \end{aligned}$$

Поделим это равенство на $a_{i+1}^{(k)2} p_i^{(k)}(\lambda_{i+1})$ и перенесем члены, начиная с $i+1$ -го, в другую часть равенства. Получим

$$c_{1i}^{(k)} p_s^{(k)}(\lambda_1) + c_{2i}^{(k)} p_s^{(k)}(\lambda_2) + \dots + c_{ii}^{(k)} p_s^{(k)}(\lambda_i) = d_{is}^{(k)}, \quad (3)$$

где

$$\begin{aligned} d_{is}^{(k)} &= -p_s^{(k)}(\lambda_{i+1}) - \left[\frac{a_{i+2}^{(k)}}{a_{i+1}^{(k)}} \right]^2 \frac{p_i^{(k)}(\lambda_{i+2}) p_s^{(k)}(\lambda_{i+2})}{p_i^{(k)}(\lambda_{i+1})} - \dots \\ &\quad \dots - \left[\frac{a_r^{(k)}}{a_{i+1}^{(k)}} \right]^2 \frac{p_i^{(k)}(\lambda_r) p_s^{(k)}(\lambda_r)}{p_i^{(k)}(\lambda_{i+1})}. \end{aligned}$$

Система равенств (3) при $s = 0, \dots, t-1$ может рассматриваться как система линейных уравнений относительно коэффициентов $c_{11}^{(k)}, \dots, c_{ii}^{(k)}$.

При стремлении k к бесконечности коэффициенты этой системы $p_s^{(k)}(\lambda_j)$ стремятся к конечным пределам $p_s(\lambda_j)$. В свою очередь, к конечным пределам стремятся и свободные члены системы. Именно,

$$d_{is}^{(k)} \rightarrow -\bar{p}_s(\lambda_{i+1}),$$

ибо

$$\frac{p_i^{(k)}(\lambda_j)p_s^{(k)}(\lambda_j)}{p_i^{(k)}(\lambda_{i+1})} \rightarrow \frac{\bar{p}_i(\lambda_j)\bar{p}_s(\lambda_j)}{\bar{p}_i(\lambda_{i+1})},$$

2

$$\frac{a_j^{(k)}}{a_{j+1}^{(k)}} \rightarrow 0 \quad \text{при} \quad j > i + 1.$$

Рассмотрим предельную систему

$$c_{1i}\bar{p}_0(\lambda_1) + c_{2i}\bar{p}_0(\lambda_2) + \dots + c_{ii}\bar{p}_0(\lambda_i) = -\bar{p}_0(\lambda_{i+1})$$

$$c_{1i}\bar{p}_1(\lambda_1) + c_{2i}\bar{p}_1(\lambda_2) + \dots + c_{ii}\bar{p}_1(\lambda_i) = -\bar{p}_1(\lambda_{i+1})$$

• •

$$c_{1i}p_{i-1}(k_1) + c_{2i}p_{i-1}(k_2) + \dots + c_{ii}p_{i-1}(k_i) = -p_{i-1}(k_{i+1})$$

Эта система имеет треугольную матрицу, ибо $\bar{p}_1(\lambda_1) = \bar{p}_2(\lambda_1) = \dots = \bar{p}_{i-1}(\lambda_2) = \dots = \bar{p}_{i-1}(\lambda_{i-1}) = 0$, и ее определитель $\Delta = p_0(\lambda_1)p_1(\lambda_2) \dots \dots p_{i-1}(\lambda_i)$ не равен нулю.

В силу теоремы 13.1 ее решение c_{1i}, \dots, c_{4i} будет предельным для $c_{1i}^{(k)}, \dots, c_{4i}^{(k)}$.

Итак, мы установили, что последовательность $c_{1i}^{(k)}, \dots, c_{ii}^{(k)}$ имеет конечные пределы. Следовательно,

$$b_{si}^{(k)} = c_{si}^{(k)} \frac{a_{i+1}^{(k)}}{a_i^{(k)}} \rightarrow 0 \quad \text{при} \quad s \leq i.$$

Таким образом,

$$V_i^{(k)} \rightarrow U_{i+1},$$

а потому векторы $p_i^{(k)}$ сходятся к U_{i+1} по направлению. Теорема полностью доказана.

Заметим, что в процессе доказательства теоремы мы получили оценку быстроты сходимости. Именно

$$b_{si}^{(k)} = O\left(\frac{a_{i+1}^{(k)}}{a_s^{(k)}}\right) \quad \text{при } s \leq i$$

$$b_{si}^{(k)} = O\left(\frac{a_s^{(k)}}{a_i^{(k)}}\right) \quad \text{при } s \geq i+2.$$

В лемме § 74 было доказано, что последовательные приближения X_k , вычисленные по методу постоянного множителя при $0 < \gamma < \frac{1}{M-m}$, удовлетворяют условиям теоремы 75.1.

Тем самым доказано, что если

$$\begin{aligned} X_{k+1} &= X_k - \gamma \xi_k \\ \xi_k &= AX_k - \mu(X_k)X_k, \quad 0 < \gamma < \frac{1}{M-m}. \end{aligned}$$

$p_1^{(k)}, \dots, p_{m-1}^{(k)}$ векторы Ланцюша, построенные исходя из $p_0^{(k)} = X_k$, то $p_0^{(k)}, \dots, p_{m-1}^{(k)}$ сходятся по направлению к первым (в порядке убывания собственных значений) собственным векторам, лежащим в подпространстве, порожденном начальным вектором X_0 .

Отметим, что при вычислении векторов $p_0^{(k)}, \dots, p_{m-1}^{(k)}$, исходя из вектора X_k , приходится производить вычитание близких величин, так что указанный процесс мало пригоден для практических вычислений.

Мы это уже видели на последнем примере § 74.

В заключение заметим, что условиям теоремы 75.1 удовлетворяет и степенной метод. Поэтому, если

$$X_{k+1} = AX_k \quad (k = 0, 1, \dots),$$

$p_1^{(k)}, \dots, p_{m-1}^{(k)}$ векторы Ланцюша, построенные исходя из $p_0^{(k)} = X_k$, то $p_0^{(k)}, \dots, p_{m-1}^{(k)}$ сходятся по направлению к соответствующим собственным векторам, лежащим в подпространстве, порожденном начальным вектором X_0 .

§ 76. s -шаговый метод наискорейшего спуска

По сути дела, первое приближение X_1 любого градиентного метода лежит в подпространстве $P^{(2)}$, натянутом на векторы X_0 и AX_0 , и в методе наискорейшего спуска оно осуществляет максимизацию функционала $\mu(X)$ в этом подпространстве. Второе приближение X_2 любого градиентного метода лежит в подпространстве $P^{(3)}$, натянутом на X_0, AX_0, A^2X_0 , третье в подпространстве $P^{(4)}$, натянутом на $X_0, AX_0, A^2X_0, A^3X_0$ и т. д. Однако уже второе приближение, даже по методу наискорейшего спуска, не будет максимизировать $\mu(X)$ в подпространстве $P^{(3)}$. Естественно поставить вопрос об отыскании вектора, осуществляющего максимизацию функционала $\mu(X)$ в подпространстве $P^{(s+1)}$, натянутом на векторы X_0, AX_0, \dots, A^sX_0 , при заранее фиксированном числе s . Решение поставленной задачи позволяет построить итерационный процесс для определения алгебраически наибольшего собственного значения симметричной матрицы и принадлежащего ему собственного вектора — так называемый s -шаговый метод наискорейшего спуска, который заключается в следующем.

Берется начальное приближение X_0 , строится вектор X_1 , осуществляющий максимизацию $\mu(X)$ в подпространстве $P_0^{(s+1)}$, натянутом на векторы $X_0, AX_0, \dots, A^s X_0$, затем ищется вектор X_2 , осуществляющий максимизацию $\mu(X)$ в подпространстве $P_1^{(s+1)}$, натянутом на векторы $X_1, AX_1, \dots, A^s X_1$ и т. д.

Прежде чем выяснить сходимость процесса, покажем, как осуществляется решение задачи о максимизации $\mu(X)$ в подпространстве $P_0^{(s+1)}$. В этом подпространстве выбирается какой-либо базис V_0, V_1, \dots, V_s . Пусть X любой вектор подпространства и пусть

$$X = b_0 V_0 + b_1 V_1 + \dots + b_s V_s, \quad (1)$$

причем не все b_i равны нулю.

Сопоставим вектору X вектор \bar{X} арифметического пространства $R^{(s+1)}$ с компонентами b_0, b_1, \dots, b_s . Тогда

$$\begin{aligned} (X, X) &= \sum_{i=0, j=0}^s c_{ij} b_i b_j = (C\bar{X}, \bar{X}) \\ (AX, X) &= \sum_{i=0, j=0}^s d_{ij} b_i b_j = (D\bar{X}, \bar{X}), \end{aligned} \quad (2)$$

где

$$\begin{aligned} c_{ij} &= (V_i, V_j), \quad C = (c_{ij}) \\ d_{ij} &= (AV_i, V_j), \quad D = (d_{ij}). \end{aligned} \quad (3)$$

Очевидно, что матрица C положительно определена. Таким образом, наша задача свелась к максимизации функционала

$$\frac{(D\bar{X}, \bar{X})}{(C\bar{X}, \bar{X})} \quad (4)$$

в пространстве $R^{(s+1)}$.

Искомый максимум $\mu^{(0)}$ получается как наибольший корень уравнения $|D - Ct| = 0$, а реализующий его вектор определяется из системы линейных однородных уравнений $(D - \mu^{(0)}C)\bar{X} = 0$. Соответствующий ему вектор X_1 получается затем из разложения (1). Таким образом, наша задача сводится к решению частичной обобщенной проблемы собственных значений для матрицы $s+1$ -го порядка. Если за базис взять векторы $X_0, AX_0, \dots, A^s X_0$, то

$$\begin{aligned} c_{ij} &= (A^i X_0, A^j X_0) = (A^{i+j} X_0, X_0) \\ d_{ij} &= (A^{i+1} X_0, A^j X_0) = (A^{i+j+1} X_0, X_0). \end{aligned} \quad (5)$$

Решение задачи сильно упрощается, если в подпространстве $P_0^{(s+1)}$ взять какой-либо ортогональный базис. Очень удобным для этой

цели оказывается базис p_0, p_1, \dots, p_s , состоящий из первых $(s+1)$ векторов Ланцюша¹⁾. В этом случае

$$c_{ij} = (p_i, p_j)$$

и потому

$$c_{ij} = 0 \quad i \neq j; \quad c_{ii} = (p_i, p_i). \quad (6)$$

Далее $d_{ij} = (Ap_i, p_j)$, так что

$$\begin{aligned} d_{ij} &= 0 \quad \text{при } |i-j| > 1; \quad d_{ii} = (Ap_i, p_i) = \alpha_i(p_i, p_i); \\ d_{ii-1} &= d_{i-ii} = \beta_i(p_{i-1}, p_{i-1}). \end{aligned} \quad (7)$$

Таким образом,

$$|D - Ct| =$$

$$= \begin{vmatrix} \alpha_0(p_0, p_0) - t(p_0, p_0) & \beta_1(p_0, p_0) & 0 & \dots & 0 \\ \beta_1(p_0, p_0) & \alpha_1(p_1, p_1) - t(p_1, p_1) & \beta_2(p_1, p_1) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \alpha_s(p_s, p_s) - t(p_s, p_s) \end{vmatrix} =$$

$$= \begin{vmatrix} \alpha_0(p_0, p_0) - t(p_0, p_0) & \beta_1(p_0, p_0) & 0 & \dots & 0 \\ (p_1, p_1) & \alpha_1(p_1, p_1) - t(p_1, p_1) & \beta_2(p_1, p_1) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \alpha_s(p_s, p_s) - t(p_s, p_s) \end{vmatrix} =$$

$$[\text{ибо } \beta_i(p_{i-1}, p_{i-1}) = (p_i, p_i)]$$

$$= (p_0, p_0)(p_1, p_1) \dots (p_s, p_s) \begin{vmatrix} \alpha_0 - t & \beta_1 & 0 & \dots & 0 & 0 \\ 1 & \alpha_1 - t & \beta_2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & \alpha_s - t \end{vmatrix} = \\ = l_s p_{s+1}(t), \quad (8)$$

где $l_s = (-1)^{s+1} (p_0, p_0) \dots (p_s, p_s)$, а $p_{s+1}(t)$ есть $(s+1)$ -й полином Ланцюша.

Итак, значение $\mu^{(0)}$ максимума $\mu(X)$ в подпространстве $P_0^{(s+1)}$ есть наибольший корень $(s+1)$ -го полинома Ланцюша $p_{s+1}(t)$.

Обратимся теперь к определению вектора, реализующего этот максимум. Координаты b_0, b_1, \dots, b_s этого вектора (относительно базиса p_0, p_1, \dots, p_s) определяются из системы линейных однородных уравнений

$$\begin{aligned} [\alpha_0(p_0, p_0) - \mu^{(0)}(p_0, p_0)] b_0 + \beta_1(p_0, p_0) b_1 &= 0 \\ \beta_1(p_0, p_0) b_0 + [\alpha_1(p_1, p_1) - \mu^{(0)}(p_1, p_1)] b_1 + \beta_2(p_1, p_1) b_2 &= 0 \\ \dots & \dots \\ \beta_s(p_{s-1}, p_{s-1}) b_{s-1} + (\alpha_s - \mu^{(0)})(p_s, p_s) b_s &= 0. \end{aligned} \quad (9)$$

¹⁾ Каруш [1].

Сделав подстановку

$$b_i = \frac{b'_i}{(p_i, p_i)},$$

и принимая во внимание, что $\beta_i = \frac{(p_i, p_i)}{(p_{i-1}, p_{i-1})}$, получим для определения b'_i систему

$$\begin{aligned} (\alpha_0 - \mu^{(0)}) b'_0 + b'_1 &= 0 \\ \beta_1 b'_0 + (\alpha_1 - \mu^{(0)}) b'_1 + b'_2 &= 0 \\ \vdots &\quad \vdots \\ \beta_s b'_{s-1} + (\alpha_s - \mu^{(0)}) b'_s &= 0. \end{aligned}$$

Положим

$$b'_0 = 1 = p_0(\mu^{(0)}).$$

Тогда

$$b'_1 = \mu^{(0)} - \alpha_0 = p_1(\mu^{(0)})$$

$$b'_2 = (\mu^{(0)} - \alpha_1) b'_1 - \beta_1 b'_0 = (\mu^{(0)} - \alpha_1) p_1(\mu^{(0)}) - \beta_1 p_0(\mu^{(0)}) = p_2(\mu^{(0)})$$

и т. д. Из предпоследнего уравнения получим

$$b'_s = p_s(\mu^{(0)}).$$

Последнее уравнение окажется удовлетворенным, так как

$$p_{s+1}(\mu^{(0)}) = 0.$$

Итак,

$$b_i = \frac{p_i(\mu^{(0)})}{(p_i, p_i)} \quad (i = 0, \dots, s)$$

и вектор X_1 , реализующий максимум, дается формулой

$$X_1 = \sum_{i=1}^s \frac{p_i(\mu^{(0)})}{(p_i, p_i)} p_i. \quad (10)$$

Выведенные формулы позволяют придать следующую форму s-шаговому методу наискорейшего спуска. За начальное приближение берется произвольный вектор X_0 . После того как построен вектор X_k , строится вектор X_{k+1} по формуле

$$X_{k+1} = \sum_{i=0}^s \frac{p_i^{(k)}(\mu^{(k)})}{(p_i^{(k)}, p_i^{(k)})} p_i^{(k)}, \quad (11)$$

где векторы $p_0^{(k)}, p_1^{(k)}, \dots, p_s^{(k)}$ суть векторы Ланцюша, построенные исходя из вектора $p_0^{(k)} = X_k$ по рекуррентным соотношениям метода минимальных итераций; $\mu^{(k)}$ — наибольший корень полинома Ланцюша $p_{s+1}^{(k)}(t)$.

Теорема 76.1. Пусть последовательность векторов $X_0, X_1, \dots, X_k, \dots$ строится так, что X_k реализует максимум $\frac{(AX, X)}{(X, X)}$ в подпространстве, натянутом на векторы $X_{k-1}, \dots, AX_{k-1}, \dots, A^s X_{k-1}$.

Тогда отношения $\frac{(AX_k, X_k)}{(X_k, X_k)}$ сходятся к наибольшему собственному значению матрицы A в инвариантном подпространстве, порожденном вектором X_0 , а векторы X_k сходятся по направлению к принадлежащему этому собственному значению собственному вектору.

Доказательство. Пусть $r = r_0$ размерность инвариантного подпространства, порожденного вектором X_0 , $\lambda_1 > \lambda_2 > \dots > \lambda_r$ — собственные значения матрицы A на этом подпространстве, U_1, U_2, \dots, U_r соответствующие нормированные собственные векторы. Тогда

$$X_0 = a_1^{(0)} U_1 + a_2^{(0)} U_2 + \dots + a_r^{(0)} U_r.$$

Все коэффициенты $a_1^{(0)}, a_2^{(0)}, \dots, a_r^{(0)}$ отличны от нуля, и без нарушения общности их можно считать положительными.

Далее, через r_k обозначим размерность инвариантного подпространства, порожденного вектором X_k . Так как каждый последующий вектор X_k содержится в инвариантном подпространстве, порожденном предшествующим вектором X_{k-1} , то $r_0 \geq r_1 \geq \dots \geq r_k \geq \dots$

Рассмотрим прежде всего вырожденный случай, когда при некотором k окажется $r_k \leq s+1$. Тогда подпространство, натянутое на векторы $X_k, AX_k, \dots, A^s X_k$, само будет инвариантным (если $r_k < s+1$, то среди перечисленных векторов будут линейно-зависимые) и вектор X_{k+1} , на котором реализуется максимум $\frac{(AX, X)}{(X, X)}$ в этом подпространстве, окажется собственным вектором матрицы A . На следующем шаге подпространство, натянутое на векторы $X_{k+1}, AX_{k+1}, \dots, A^s X_{k+1}$, будет одномерным, так что процесс стабилизируется. В дальнейшем мы покажем, что так полученный собственный вектор будет пропорционален U_1 , и тем самым для вырожденного случая теорема будет доказана.

Обратимся теперь к рассмотрению одного невырожденного шага процесса. Пусть

$$X_k = a_1^{(k)} U_1 + a_2^{(k)} U_2 + \dots + a_r^{(k)} U_r,$$

$p_i^{(k)}, i = 0, \dots, s, s+1$ — векторы Ланцюша, построенные исходя из вектора X_k , $p_i^{(k)}(t)$ — соответствующие полиномы, $\mu_i^{(k)}$ их наибольшие корни. Как мы видели выше, $\mu_{s+1}^{(k)}$ есть максимум $\frac{(AX, X)}{(X, X)}$ в подпространстве, натянутом на векторы $X_k, AX_k, \dots, A^s X_k$, который мы раньше обозначали через $\mu^{(k)}$, так что

$$\mu_{s+1}^{(k)} = \mu(X_{k+1}). \quad (12)$$

Согласно формуле (11)

$$X_{k+1} = \sum_{i=0}^s \frac{p_i^{(k)}(\mu_{s+1}^{(k)})}{(p_i^{(k)}, p_i^{(k)})} p_i^{(k)}.$$

В свою очередь,

$$p_i^{(k)} = a_1^{(k)} p_i^{(k)}(\lambda_1) U_1 + \dots + a_r^{(k)} p_i^{(k)}(\lambda_r) U_r. \quad (13)$$

Следовательно,

$$X_{k+1} = a_1^{(k+1)} U_1 + \dots + a_r^{(k+1)} U_r,$$

где

$$a_1^{(k+1)} = a_1^{(k)} \sum_{i=0}^s \frac{p_i^{(k)}(\mu_{s+1}^{(k)}) p_i^{(k)}(\lambda_1)}{(p_i^{(k)}, p_i^{(k)})} = a_1^{(k)} M_1^{(k)}$$

.....

$$a_r^{(k+1)} = a_r^{(k)} \sum_{i=0}^s \frac{p_i^{(k)}(\mu_{s+1}^{(k)}) p_i^{(k)}(\lambda_r)}{(p_i^{(k)}, p_i^{(k)})} = a_r^{(k)} M_r^{(k)}.$$

Сделаем некоторые выводы из построенных формул. Так как корни полиномов Ланцоша разделяются, мы имеем $\mu_1^{(k)} < \mu_2^{(k)} < \dots < \mu_s^{(k)} < \mu_{s+1}^{(k)} \leq \lambda_1$, так что $\mu_{s+1}^{(k)}$ и λ_1 строго больше всех корней полиномов $p_i^{(k)}(t)$. Следовательно, $M_1^{(k)} > 0$. Мы предположили, что $a_1^{(0)} > 0$. Поэтому $a_1^{(1)} > 0, \dots, a_1^{(k+1)} > 0$, так что инвариантное подпространство, порожденное вектором X_{k+1} , включает собственный вектор U_1 , принадлежащий собственному значению λ_1 . Следовательно, если для X_{k+1} имеет место вырожденный случай, то на следующем шаге в подпространстве, натянутом на векторы $X_{k+1}, AX_{k+1}, \dots, A^s X_{k+1}$ (оно будет инвариантным!), максимум отношения $\frac{(AX, X)}{(X, X)}$ будет достигаться именно на собственном векторе U_1 .

Теперь остается рассмотреть процесс, протекающий без вырождения. Прежде всего заметим, что

$$\mu(X_0) < \mu(X_1) < \dots < \lambda_1,$$

так что существует

$$\lim_{k \rightarrow \infty} \mu(X_k) = \mu.$$

Положим

$$X_k = |X_k|(b_1^{(k)} U_1 + \dots + b_r^{(k)} U_r). \quad (14)$$

Так же как при доказательстве теоремы 74.1, прежде всего докажем, что один из коэффициентов $b_j^{(k)} \rightarrow 1$, а остальные коэффициенты $b_i^{(k)} \rightarrow 0$, $i \neq j$. С этой целью снова рассмотрим отношение

$$t_k^2 = \frac{(\xi_k, \xi_k)}{(X_k, X_k)} = b_1^{(k)2} [\lambda_1 - \mu(X_k)]^2 + \dots + b_r^{(k)2} [\lambda_r - \mu(X_k)]^2. \quad (15)$$

В наших обозначениях

$$t_k^2 = \frac{(p_1^{(k)}, p_1^{(k)})}{(p_0^{(k)}, p_0^{(k)})} = \beta_1^{(k)}.$$

Но

$$\beta_1^{(k)} = (t - \alpha_0^{(k)})(t - \alpha_1^{(k)}) - p_2^{(k)}(t). \quad (16)$$

Здесь $\alpha_0^{(k)} = \mu(X_k)$, $\alpha_1^{(k)} = \mu(p_1^{(k)})$, $p_2^{(k)}(t)$ — второй полином Ланцша.

Подставим в (16) $t = \mu(X_{k+1}) = \mu_{s+1}^{(k)}$. Ясно, что $p_2^{(k)}(\mu_{s+1}^{(k)}) > 0$, ибо $\mu_{s+1}^{(k)}$ при $s > 1$ больше обоих корней полинома $p_2^{(k)}(t)$. Поэтому

$$\begin{aligned} \beta_1^{(k)} &< (\mu_{s+1}^{(k)} - \alpha_0^{(k)})(\mu_{s+1}^{(k)} - \alpha_1^{(k)}) = \\ &= [\mu(X_{k+1}) - \mu(X_k)][\mu(X_{k+1}) - \mu(p_1^{(k)})]. \end{aligned} \quad (17)$$

Отсюда следует, что $\beta_1^{(k)} \rightarrow 0$, так как первый множитель правой части неравенства (17) стремится к нулю, а второй ограничен.

Из равенства (15) заключаем, что $b_i^{(k)2} [\lambda_i - \mu(X_k)]^2 \rightarrow 0$ при всех $i = 1, 2, \dots, r$. Из множителей $(\lambda_i - \mu(X_k))$ стремиться к нулю может не более чем один, все же $b_i^{(k)}$ не могут стремиться к нулю,

ибо $\sum_{i=1}^r b_i^{(k)2} = 1$. Поэтому $\mu(X_k) \rightarrow \lambda_j$ при некотором определенном j , а все $b_i^{(k)}$ при $i \neq j$ стремятся к нулю и $b_j^{(k)2} \rightarrow 1$.

Таким образом, векторы X_k сходятся по направлению к U_j . Стается доказать, что $j = 1$.

Если допустить, что $j > 1$, мы получим, что

$$\frac{b_1^{(k)}}{b_j^{(k)}} = \frac{\alpha_1^{(k)}}{\alpha_j^{(k)}} \rightarrow 0 \quad \text{при } k \rightarrow \infty.$$

Но, с другой стороны,

$$\frac{\alpha_1^{(k+1)}}{\alpha_j^{(k+1)}} = \frac{\alpha_1^{(k)}}{\alpha_j^{(k)}} \frac{M_1^{(k)}}{M_j^{(k)}}.$$

Покажем, что $\frac{M_1^{(k)}}{M_j^{(k)}} > 1$. Действительно,

$$\frac{M_1^{(k)}}{M_j^{(k)}} = \frac{\sum_{i=0}^s \frac{p_i^{(k)}(\mu_{s+1}^{(k)}) p_i^{(k)}(\lambda_1)}{(p_i^{(k)}, p_i^{(k)})}}{\sum_{i=0}^s \frac{p_i^{(k)}(\mu_{s+1}^{(k)}) p_i^{(k)}(\lambda_j)}{(p_i^{(k)}, p_i^{(k)})}}}.$$

Но $\mu_{s+1}^{(k)} = \mu(X_{k+1}) < \lambda_j$, так что λ_j больше всех корней $(s+1)$ -го полинома Ланцша при любом k и подавно больше всех корней полиномов $p_i^{(k)}(t)$. Так как, кроме того, $\lambda_1 > \lambda_j$, то $p_i^{(k)}(\lambda_1) > p_i^{(k)}(\lambda_j)$ для $i = 1, 2, \dots, s$ и, следовательно,

$$M_1^{(k)} > M_j^{(k)} > 0.$$

Таким образом,

$$\left| \frac{a_1^{(k+1)}}{a_j^{(k+1)}} \right| > \left| \frac{a_1^{(k)}}{a_j^{(k)}} \right|.$$

Это неравенство выполняется при всех k и находится в противоречии с предельным соотношением

$$\frac{a_1^{(k)}}{a_j^{(k)}} \rightarrow 0.$$

Поэтому неравенство $j > 1$ невозможно. Теорема доказана.

Замечание. Если процесс наискорейшего спуска вести по формулам

$$X_{k+1} = X_k + \sum_{i=1}^s \frac{(p_0^{(k)}, p_0^{(k)}) p_i^{(k)} (\mu_{s+1}^{(k)})}{(p_i^{(k)}, p_i^{(k)})} p_i^{(k)},$$

то вектор $X_{k+1} - X_k$ будет ортогонален к вектору X_k . В этом случае последовательность векторов X_k будет сходиться к U_1 (а не только сходиться по направлению). Этот факт, доказанный Карушем [1], легко вытекает из оценок быстроты сходимости процесса.

Теорема 76.2. В s -шаговом методе наискорейшего спуска, при достаточно больших k , имеет место неравенство

$$\lambda_1 - \mu_{k+1} \leq (1 + \varepsilon) Q_s^2 (\lambda_1 - \mu_k).$$

Здесь Q_s есть наименьшее отклонение от нуля в точках совокупности $\lambda_2, \dots, \lambda_r$ полиномов $F_s(t)$ s -й степени, нормированных условием $F_s(\lambda_1) = 1$.

Доказательство. Пусть $\Phi_s(t)$ — полином, реализующий наименьшее уклонение от нуля на совокупности точек $\lambda_2, \dots, \lambda_r$ при нормировке $\Phi_s(\lambda_1) = 1$.

Допустим, что построено приближение X_k . Наряду со следующим приближением X_{k+1} рассмотрим вектор $\tilde{X}_{k+1} = \Phi_s(A) X_k$. Через μ_k , μ_{k+1} , $\tilde{\mu}_{k+1}$ обозначим соответствующие значения функционала $\mu(X)$. Ясно, что

$$\mu_{k+1} \geq \tilde{\mu}_{k+1},$$

так как μ_{k+1} есть максимум $\mu(X)$ в подпространстве $P_k^{(s+1)}$, а $\tilde{\mu}_{k+1}$ есть одно из значений $\mu(X)$ в том же подпространстве.

Сравним теперь $\lambda_1 - \mu_{k+1}$ и $\lambda_1 - \mu_k$. Пусть

$$X_k = a_1^{(k)} U_1 + a_2^{(k)} U_2 + \dots + a_r^{(k)} U_r$$

разложение вектора X_k по собственным векторам, содержащимся в инвариантном подпространстве, порожденным вектором X_0 . Тогда

$$\tilde{X}_{k+1} = a_1^{(k)} U_1 + \Phi_s(\lambda_2) a_2^{(k)} U_2 + \dots + \Phi_s(\lambda_r) a_r^{(k)} U_r.$$

Далее

$$\begin{aligned} \lambda_1 - \mu_k &= \lambda_1 - \frac{(AX_k, X_k)}{(X_k, X_k)} = \\ &= \lambda_1 - \frac{\lambda_1 a_1^{(k)2} + \lambda_2 a_2^{(k)2} + \dots + \lambda_r a_r^{(k)2}}{a_1^{(k)2} + a_2^{(k)2} + \dots + a_r^{(k)2}} = \\ &= \frac{(\lambda_1 - \lambda_2) a_2^{(k)2} + \dots + (\lambda_1 - \lambda_r) a_r^{(k)2}}{a_1^{(k)2} + a_2^{(k)2} + \dots + a_r^{(k)2}}. \end{aligned}$$

Таким же образом

$$\lambda_1 - \tilde{\mu}_{k+1} = \frac{(\lambda_1 - \lambda_2) \Phi_s^2(\lambda_2) a_2^{(k)2} + \dots + (\lambda_1 - \lambda_r) \Phi_s^2(\lambda_r) a_r^{(k)2}}{a_1^{(k)2} + \Phi_s^2(\lambda_2) a_2^{(k)2} + \dots + \Phi_s^2(\lambda_r) a_r^{(k)2}}.$$

Так как

$$|\Phi_s(\lambda_i)| \leq Q_s \quad \text{при } i = 2, \dots, r,$$

то

$$\lambda_1 - \tilde{\mu}_{k+1} \leq Q_s^2 \frac{(\lambda_1 - \lambda_2) a_2^{(k)2} + \dots + (\lambda_1 - \lambda_r) a_r^{(k)2}}{a_1^{(k)2}}.$$

Поэтому

$$\frac{\lambda_1 - \tilde{\mu}_{k+1}}{\lambda_1 - \mu_k} \leq Q_s^2 \frac{a_1^{(k)2} + \dots + a_r^{(k)2}}{a_1^{(k)2}} = Q_s^2 \frac{1}{b_1^{(k)2}},$$

где

$$b_1^{(k)2} = \frac{a_1^{(k)2}}{a_1^{(k)2} + \dots + a_r^{(k)2}}.$$

Таким образом,

$$\frac{\lambda_1 - \mu_{k+1}}{\lambda_1 - \mu_k} \leq \frac{\lambda_1 - \tilde{\mu}_{k+1}}{\lambda_1 - \mu_k} \leq Q_s^2 \frac{1}{b_1^{(k)2}} = Q_s^2 (1 + \sigma_k),$$

где $\sigma_k = \frac{1 - b_1^{(k)2}}{b_1^{(k)2}}$. В силу теоремы 75.1, $b_1^{(k)2} \rightarrow 1$ при $k \rightarrow \infty$, так что σ_k станет меньше ε при достаточно больших k , и потому

$$\lambda_1 - \mu_{k+1} \leq Q_s^2 (1 + \varepsilon) (\lambda_1 - \mu_k) \quad \text{при } k > k_0.$$

Полученные оценки показывают, что быстрота сходимости последовательности μ_k к λ_1 не меньше, чем быстрота сходимости геометрической прогрессии со знаменателем $Q_s^2(1+\epsilon)$.

Ясно, что Q_s не превосходит уклонения от нуля полинома, наименее уклоняющегося от нуля на промежутке (λ_2, λ_r) при прежней нормировке в точке $t = \lambda_1$. Это наименьшее уклонение равно как известно

$$E_s = \frac{1}{T_s(t)} = \frac{2}{(t + \sqrt{t^2 - 1})^s + (t - \sqrt{t^2 - 1})^s}$$

при

$$t = \frac{2\lambda_1 - \lambda_2 - \lambda_r}{\lambda_2 - \lambda_r}.$$

Применение двухшагового метода наискорейшего спуска к матрице (4) § 51 при $X = (0.34, 0.20, 0.90, 0.75)'$ дает

$$\mu^{(0)} = 1.8300504$$

$$\mu^{(1)} = 2.3227312$$

$$\mu^{(2)} = 2.3227487.$$

При этом

$$\tilde{X}_1 = (1.000, 0.799, 0.752, 0.887)'$$

$$\tilde{X}_2 = (1.0000, 0.7931, 0.7479, 0.8867)'.$$

Таким образом, применение двух шагов оказывается здесь достаточным для получения первого собственного значения с высокой точностью. Соответствующий же ему собственный вектор определяется со значительно меньшей точностью.

ГЛАВА VIII

ИТЕРАЦИОННЫЕ МЕТОДЫ ДЛЯ РЕШЕНИЯ ПОЛНОЙ ПРОБЛЕМЫ СОБСТВЕННЫХ ЗНАЧЕНИЙ

Итерационные методы решения полной проблемы собственных значений появились в самое последнее время. Естественно, что они значительно более трудоемки, чем итерационные методы решения частичной проблемы. Как правило, их практическое осуществление даже для матриц не очень высокого порядка, требует применения быстродействующих вычислительных машин.

Однако несомненным их преимуществом перед точными методами (гл. IV) является возможность вычисления всех собственных значений, минуя вычисления характеристического полинома. Как уже отмечалось выше, ошибки округления в коэффициентах характеристического полинома могут сильно влиять на точность вычисления его корней.

§ 77. Алгорифм деления и вычитания

Алгорифм (Рутисхаузер [2]), к описанию которого мы переходим, является надстройкой над биортогональным алгорифмом. Этот алгорифм позволяет определять собственные значения матрицы A как пределы последовательностей чисел, которые строятся рекуррентно по несложным формулам. Начальными данными алгорифма служат коэффициенты p_k и σ_k двучленной формы биортогонального алгорифма.

Прежде всего мы изложим теорию метода в простейшем случае, предполагая матрицу A , для которой строится алгорифм, положительно-определенной.

Алгорифм основан на свойствах нескольких бесконечных последовательностей векторов $p_i^{(k)}$, лежащих в циклическом инвариантном подпространстве, порожденном начальным вектором X_0 .

Это дает право при изложении теории метода считать все собственные значения матрицы A попарно различными и все компоненты в разложении начального вектора X_0 по собственным векторам, отличными от нуля.

При фиксированном k векторы $p_i^{(k)}$, $i = 0, 1, \dots, n - 1$, строятся посредством A^k -ортогонализации последовательности векторов X_0 .

$AX_0, \dots, A^{n-1}X_0$. При всех k будет $p_0^{(k)} = X_0$, $p_i^{(k)} = p_i^{(k)}(A)X_0$, где $p_i^{(k)}(t) = t^i + \dots$ полиномы степени i .

Покажем, что векторы $p_i^{(k)}$ связаны простыми рекуррентными соотношениями

$$\begin{aligned} Ap_{i-1}^{(k+1)} - p_i^{(k)} &= \rho_{i-1}^{(k+1)} p_{i-1}^{(k)} \quad (i = 1, 2, \dots, n) \\ p_i^{(k)} - p_i^{(k+1)} &= \sigma_i^{(k+1)} p_{i-1}^{(k+1)} \quad (i = 1, 2, \dots, n-1). \end{aligned} \quad (1)$$

(При $i = n$ считаем $p_n^{(k)} = 0$).

Действительно, вектор $Ap_{i-1}^{(k+1)} - p_i^{(k)}$ принадлежит подпространству $P^{(i)}$, натянутому на векторы $X_0, AX_0, \dots, A^{i-1}X_0$, и A^k -ортогонален к векторам $X_0, AX_0, \dots, A^{i-2}X_0$. В силу единственности нормированного вектора, удовлетворяющего этим условиям, вектор $Ap_{i-1}^{(k+1)} - p_i^{(k)}$ лишь численным множителем отличается от вектора $p_{i-1}^{(k)}$.

Аналогично, вектор $p_i^{(k)} - p_i^{(k+1)}$ принадлежит подпространству $P^{(i)}$, A^{k+1} -ортогонален к $X_0, AX_0, \dots, A^{i-2}X_0$ и, следовательно, лишь численным множителем отличается от вектора $p_{i-1}^{(k+1)}$.

Между полиномами $p_i^{(k)}(t)$, очевидно, тоже выполняются соотношения

$$\begin{aligned} tp_{i-1}^{(k+1)}(t) - p_i^{(k)}(t) &= \rho_{i-1}^{(k+1)} p_{i-1}^{(k)}(t) \\ p_i^{(k)}(t) - p_i^{(k+1)}(t) &= \sigma_i^{(k+1)} p_{i-1}^{(k+1)}(t). \end{aligned} \quad (2)$$

Выведем теперь зависимость между числами $\rho_i^{(k)}$ и $\sigma_i^{(k)}$. С этой целью перейдем к трехчленным соотношениям, связывающим векторы $p_{i+1}^{(k)}, p_i^{(k)}, p_{i-1}^{(k)}$. Такие соотношения можно построить двумя способами. Во-первых, исключая из соотношений

$$\begin{aligned} Ap_i^{(k+1)} - p_{i+1}^{(k)} &= p_i^{(k+1)} p_i^{(k)} \\ Ap_{i-1}^{(k+1)} - p_i^{(k)} &= \rho_{i-1}^{(k+1)} p_{i-1}^{(k)}, \\ Ap_i^{(k)} - Ap_i^{(k+1)} &= \sigma_i^{(k+1)} Ap_{i-1}^{(k+1)} \end{aligned}$$

векторы $Ap_i^{(k+1)}$ и $Ap_{i-1}^{(k+1)}$, получим

$$p_{i+1}^{(k)} + (\sigma_i^{(k+1)} + \rho_i^{(k+1)}) p_i^{(k)} - Ap_i^{(k)} + \rho_{i-1}^{(k+1)} \sigma_i^{(k+1)} p_{i-1}^{(k)} = 0. \quad (3)$$

Во-вторых, исключая из соотношений

$$\begin{aligned} Ap_i^{(k)} - p_{i+1}^{(k-1)} &= \rho_i^{(k)} p_i^{(k-1)} \\ p_{i+1}^{(k-1)} - p_{i+1}^{(k)} &= \sigma_{i+1}^{(k)} p_i^{(k)} \\ p_i^{(k-1)} - p_i^{(k)} &= \sigma_i^{(k)} p_{i-1}^{(k)} \end{aligned}$$

векторы $p_{i+1}^{(k-1)}$ и $p_i^{(k-1)}$, получим

$$p_{i+1}^{(k)} + (\sigma_{i+1}^{(k)} + \rho_i^{(k)}) p_i^{(k)} - Ap_i^{(k)} + \sigma_i^{(k)} p_i^{(k)} p_{i-1}^{(k)} = 0. \quad (4)$$

Сравнивая трехчленные соотношения (3) и (4), заключаем, что

$$\begin{aligned}\sigma_i^{(k+1)} + p_i^{(k+1)} &= \sigma_{i+1}^{(k)} + p_i^{(k)} \\ \sigma_i^{(k+1)} p_{i-1}^{(k+1)} &= \sigma_i^{(k)} p_i^{(k)}.\end{aligned}\quad (5)$$

Последние соотношения позволяют последовательно строить числа $p_i^{(k+1)}$, $\sigma_i^{(k+1)}$ в порядке $p_0^{(k+1)}, \sigma_1^{(k+1)}, p_1^{(k+1)}, \dots, p_{n-2}^{(k+1)}, \sigma_{n-1}^{(k+1)}, p_{n-1}^{(k+1)}$, как только числа $p_i^{(k)}$, $\sigma_i^{(k)}$ уже построены.

Вычислительными формулами алгорифма (схема деления и вычитания — QD) являются, при $k \geq 1$,

$$\begin{aligned}p_i^{(k+1)} &= \sigma_{i+1}^{(k)} + p_i^{(k)} - \sigma_i^{(k+1)} \quad (i = 0, 1, \dots, n-1) \\ \sigma_i^{(k+1)} &= \frac{\sigma_i^{(k)} p_i^{(k)}}{p_{i-1}^{(k+1)}} \quad (i = 1, 2, \dots, n-1).\end{aligned}\quad (6)$$

При этом надлежит считать $\sigma_0^{(k)} = \sigma_n^{(k)} = 0$. Начальная строка алгорифма

$$p_0^{(1)}, \sigma_1^{(1)}, \dots, \sigma_{n-1}^{(1)}, p_{n-1}^{(1)}$$

определяется коэффициентами двучленных формул метода минимальных итераций или получается из коэффициентов α_i и β_i трехчленных формул этого метода в виде

$$p_0^{(1)} = \alpha_0, \quad \sigma_i^{(1)} = \frac{\beta_i}{p_{i-1}^{(1)}}, \quad p_i^{(1)} = \alpha_i - \sigma_i^{(1)}. \quad (7)$$

Для вычисления собственных значений матрицы (на инвариантном подпространстве, порожденном вектором X_0) нужно составить только последовательность чисел $p_i^{(k)}$, $\sigma_i^{(k)}$. Векторы же $p_i^{(k)}$ строить нет необходимости, так что их введение нужно только для пояснения теории метода. Именно, верна следующая теорема.

Теорема 77.1. Пусть A положительно-определенная матрица, X_0 данный вектор, $\lambda_1 > \lambda_2 > \dots > \lambda_r$ собственные значения матрицы A на инвариантном подпространстве, порожденном вектором X_0 . Пусть $p_i^{(k)}$ ($i = 0, 1, \dots, r-1$) и $\sigma_i^{(k)}$ ($i = 1, \dots, r-1$) — числа, построенные согласно схеме деления и вычитания. Тогда

$$\lim_{k \rightarrow \infty} p_i^{(k)} = \lambda_{i+1}, \quad \lim_{k \rightarrow \infty} \sigma_i^{(k)} = 0. \quad (8)$$

Доказательство. Полиномы $p_i^{(k)}(t) = t^i + \dots$ характеризуются условиями ортогональности $(A^k p_i^{(k)}(A) X_0, p_j^{(k)}(A) X_0) = 0$ при $i \neq j$. Эти условия можно переписать в форме $(p_i^{(k)}(A) A^{\frac{k}{2}} X_0, p_j^{(k)}(A) A^{\frac{k}{2}} X_0) = 0$, что означает, что $p_i^{(k)}(t)$ суть полиномы Ланцша, построенные для начального вектора $Y_0^{(k)} = A^{\frac{k}{2}} X_0$.

Последовательность векторов $A^{\frac{k}{2}}X_0$, очевидно, удовлетворяет условиям теоремы 75.1. Следовательно, имеет место предельное соотношение

$$\lim_{k \rightarrow \infty} p_i^{(k)}(t) = \bar{p}_i(t) = (t - \lambda_1) \dots (t - \lambda_i).$$

Переходя к пределу в вытекающем из (2) равенстве

$$p_i^{(k)} = \frac{tp_i^{(k)}(t) - p_{i+1}^{(k-1)}(t)}{p_i^{(k-1)}(t)},$$

получим

$$\lim_{k \rightarrow \infty} p_i^{(k)} = \frac{t(t - \lambda_1) \dots (t - \lambda_i) - (t - \lambda_1) \dots (t - \lambda_i)(t - \lambda_{i+1})}{(t - \lambda_1) \dots (t - \lambda_i)} = \lambda_{i+1}.$$

Для $\sigma_i^{(k)}$ получим аналогично

$$\lim_{k \rightarrow \infty} \sigma_i^{(k)} = \frac{\bar{p}_i(t) - \bar{p}_{i-1}(t)}{p_{i-1}(t)} = 0.$$

Теорема доказана.

Теперь перейдем к обобщению алгорифма в двух направлениях. Во-первых, распространим его на любые матрицы с вещественными различными собственными значениями и, во-вторых, введем в рассмотрение более широкий класс весовых функций, управляющих ортогонализацией. Как мы увидим ниже, рациональный выбор последовательности весовых функций может значительно ускорить сходимость процесса.

Пусть

$\varphi_1(t) = 1$, $\varphi_1(t) = t$, $\varphi_k(t) = t(t - t_2) \dots (t - t_k)$ ($k = 2, 3, \dots$) (9)

последовательность полиномов. Каждый последующий полином получается из предшествующего умножением на линейный двучлен.

Пусть, далее, A данная матрица, собственные значения которой вещественны и различны, X_0 и Y_0 некоторые начальные векторы.

Исходя из системы векторов X_0 , $AX_0, \dots, A^{n-1}X_0$ и системы векторов Y_0 , $A'Y_0, \dots, A'^{n-1}Y_0$, построим системы векторов $p_0^{(k)}, p_1^{(k)}, \dots, p_{n-1}^{(k)}$ и $\tilde{p}_0^{(k)}, \tilde{p}_1^{(k)}, \dots, \tilde{p}_{n-1}^{(k)}$ биортогональных по весу $\varphi_k(A)$. Мы будем предполагать, что начальные векторы X_0 и Y_0 выбраны так, что при всех $k = 0, 1, \dots$ построение векторов $p_0^{(k)}, \dots, p_{n-1}^{(k)}$ и $\tilde{p}_0^{(k)}, \dots, \tilde{p}_{n-1}^{(k)}$ возможно, т. е. что процесс ортогонализации проходит при любом k без вырождения.

Векторы $p_i^{(k)}$ вполне характеризуются выполнением следующих двух требований:

- 1) $p_i^{(k)} = p_i^{(k)}(A)X_0$, где $p_i^{(k)}(t) = t^i + c_1^{(k)}t^{i-1} + \dots$ (10)
- 2) $(\varphi_k(A)p_i^{(k)}, A^jY_0) = 0$ при $j = 0, 1, \dots, i-1$.

Нетрудно дать явные формулы для векторов $p_i^{(k)}$ и полиномов $p_i^{(k)}(t)$. Именно, положив

$$\tilde{p}_i^{(k)}(t) = \begin{vmatrix} m_0^{(k)} \dots m_{i-1}^{(k)} & 1 \\ m_1^{(k)} \dots m_i^{(k)} & t \\ \vdots & \ddots & \ddots \\ m_i^{(k)} \dots m_{2i-1}^{(k)} & t^i \end{vmatrix}, \quad (11)$$

где

$$m_j^{(k)} = (\varphi_k(A) A^j X_0, Y_0),$$

мы будем иметь

$$(\varphi_k(A) \tilde{p}_i^{(k)}(A) X_0, A^j Y_0) = (\varphi_k(A) A^j \tilde{p}_i^{(k)}(A) X_0, Y_0) =$$

$$= \begin{vmatrix} m_0^{(k)} \dots m_{i-1}^{(k)} & m_j^{(k)} \\ m_1^{(k)} \dots m_i^{(k)} & m_{j+1}^{(k)} \\ \vdots & \ddots & \ddots \\ m_i^{(k)} \dots m_{2i-1}^{(k)} & m_{j+i}^{(k)} \end{vmatrix} = 0 \quad (j = 0, 1, \dots, i-1).$$

Поэтому вектор $\tilde{p}_i^{(k)}(A) X_0$ лишь численным множителем отличается от вектора $p_i^{(k)}$. Этот множитель равен $\frac{1}{\Delta_i^{(k)}}$, где

$$\Delta_i^{(k)} = \begin{vmatrix} m_0^{(k)} \dots m_{i-1}^{(k)} \\ \vdots & \ddots & \ddots & \ddots \\ m_{i-1}^{(k)} \dots m_{2i-2}^{(k)} \end{vmatrix}. \quad (12)$$

Таким образом, для обеспечения невырожденного течения процесса ортогонализации на каждом шагу нужно предположить, что все определители $\Delta_i^{(k)}$ отличны от нуля.

Тогда

$$p_i^{(k)} = p_i^{(k)}(A) X_0,$$

где

$$p_i^{(k)}(t) = \frac{1}{\Delta_i^{(k)}} \tilde{p}_i^{(k)}(t). \quad (13)$$

Отметим одно, важное для дальнейшего, свойство полиномов $\tilde{p}_i^{(k)}(t)$. Именно, докажем справедливость тождества

$$\tilde{p}_i^{(k)}(t_{k+1}) = (-1)^i \Delta_i^{(k+1)}. \quad (14)$$

Действительно, в силу равенства $\varphi_{k+1}(t) = \varphi_k(t)(t - t_{k+1})$, имеем

$$m_j^{(k+1)} = (\varphi_k(A)(A - t_{k+1}E) A^j X_0, Y_0) =$$

$$= (\varphi_k(A) A^{j+1} X_0, Y_0) - t_{k+1} (\varphi_k(A) A^j X_0, Y_0) = m_{j+1}^{(k)} - t_{k+1} m_j^{(k)}.$$

Отнимая в определителе

$$\tilde{p}_i^{(k)}(t_{k+1}) = \begin{vmatrix} m_0^{(k)} & \dots & m_{i-1}^{(k)} & 1 \\ m_1^{(k)} & \dots & m_i^{(k)} & t_{k+1} \\ \vdots & \ddots & \vdots & \vdots \\ m_i^{(k)} & \dots & m_{2i-1}^{(k)} & t_{k+1}^i \end{vmatrix}$$

из каждой строки предыдущую, умноженную на t_{k+1} , получим

$$\tilde{p}_i^{(k)}(t_{k+1}) = \begin{vmatrix} m_0^{(k)} & \dots & m_{i-1}^{(k)} & 1 \\ m_0^{(k+1)} & \dots & m_{i-1}^{(k+1)} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ m_{i-1}^{(k+1)} & \dots & m_{2i-2}^{(k+1)} & 0 \end{vmatrix} = (-1)^i \Delta_i^{(k+1)}.$$

Легко устанавливается, что векторы $p_i^{(k)}$ связаны рекуррентными соотношениями:

$$\begin{aligned} (A - t_{k+1}E) p_{i-1}^{(k+1)} - p_i^{(k)} &= p_{i-1}^{(k+1)} p_{i-1}^{(k)} \\ p_i^{(k)} - p_i^{(k+1)} &= \sigma_i^{(k+1)} p_{i-1}^{(k+1)}, \end{aligned} \quad (15)$$

а полиномы $p_i^{(k)}(t)$ — соотношениями:

$$\begin{aligned} (t - t_{k+1}) p_{i-1}^{(k+1)}(t) - p_i^{(k)}(t) &= \rho_{i-1}^{(k+1)} p_{i-1}^{(k)}(t) \\ p_i^{(k)}(t) - p_i^{(k+1)}(t) &= \sigma_i^{(k+1)} p_{i-1}^{(k+1)}(t). \end{aligned} \quad (16)$$

Коэффициенты $\rho_i^{(k)}$ и $\sigma_i^{(k)}$ в свою очередь удовлетворяют соотношениям:

$$\begin{aligned} \sigma_i^{(k+1)} + \rho_i^{(k+1)} + t_{k+1} &= \sigma_{i+1}^{(k)} + \rho_i^{(k)} + t_k \\ \sigma_i^{(k+1)} \rho_{i-1}^{(k+1)} &= \sigma_i^{(k)} \rho_i^{(k)}. \end{aligned} \quad (17)$$

Эти соотношения выводятся посредством сравнения трехчленных соотношений, связывающих векторы $p_{i+1}^{(k)}$, $p_i^{(k)}$, $p_{i-1}^{(k)}$, которые получаются двумя способами из двухчленных соотношений, точно так же, как это делалось выше.

Соотношения (17) позволяют последовательно вычислять коэффициенты $\rho_i^{(k+1)}$ и $\sigma_i^{(k+1)}$ в порядке $\rho_0^{(k+1)}, \sigma_1^{(k+1)}, \rho_1^{(k+1)}, \sigma_2^{(k+1)}, \dots, \rho_{n-2}^{(k+1)}, \sigma_{n-1}^{(k+1)}, \rho_{n-1}^{(k+1)}$, как только числа предыдущей строки уже вычислены.

Вычислительными формулами (схема деления и вычитания со сдвигом) являются

$$\begin{aligned} \rho_i^{(k+1)} &= \sigma_{i+1}^{(k)} + \rho_i^{(k)} - \sigma_i^{(k+1)} - t_{k+1} + t_k \quad (i = 0, 1, \dots, n-1) \\ \sigma_i^{(k+1)} &= \frac{\sigma_i^{(k)} \rho_i^{(k)}}{\rho_{i-1}^{(k+1)}} \quad (i = 1, \dots, n-1). \end{aligned} \quad (18)$$

При этом нужно считать, что $\sigma_0^{(k)} = \sigma_n^{(k)} = 0$.

Начальная строка составляется из коэффициентов двучленных формул биортогонального алгоритма, которые могут быть вычислены непосредственно или найдены при помощи коэффициентов трехчленных формул.

Если считать, что все $t_k = 0$, то формулы (18) в точности совпадут с формулами схемы QD , выведенными выше для положительно-определенной матрицы A .

Для дальнейшего окажется полезной следующая явная формула для коэффициента $p_{i-1}^{(k)}$. Именно

$$p_{i-1}^{(k+1)} = \frac{\Delta_{i-1}^{(k)} \Delta_i^{(k+1)}}{\Delta_i^{(k)} \Delta_{i-1}^{(k+1)}}. \quad (19)$$

Для вывода этой формулы положим $t = t_{k+1}$ в первом из соотношений (16). Тогда

$$p_{i-1}^{(k+1)} = -\frac{p_i^{(k)}(t_{k+1})}{p_{i-1}^{(k)}(t_{k+1})} = -\frac{\Delta_{i-1}^{(k)} \tilde{p}_i^{(k)}(t_{k+1})}{\Delta_i^{(k)} \tilde{p}_{i-1}^{(k)}(t_{k+1})} = \frac{\Delta_{i-1}^{(k)} \Delta_i^{(k+1)}}{\Delta_i^{(k)} \Delta_{i-1}^{(k+1)}}.$$

Теорема 77.2. Если весовые полиномы $\varphi_k(t)$ и начальные векторы X_0 и Y_0 выбраны так, что

- 1) все $\Delta_i^{(k)} \neq 0$,
- 2) последовательность t_k сходится к конечному пределу τ ,
- 3) при некоторой нумерации собственных значений

$$|\lambda_1 - \tau| > |\lambda_2 - \tau| > \dots > |\lambda_n - \tau|,$$

то последовательности $p_i^{(k)}$ сходятся к пределам $\lambda_{i+1} - \tau$, а последовательности $\varphi_i^{(k)}$ сходятся к нулю.

Доказательство. Заметим прежде всего, что, при выполнении условий теоремы, $\tau \neq \lambda_i$ при $i = 1, \dots, n-1$. Равенство же $\tau = \lambda_n$ не исключено.

Выведем асимптотические формулы для $p_i^{(k+1)}$. Для этого оценим прежде всего определитель $\Delta_i^{(k)}$. Пусть

$$X_0 = c_1 U_1 + \dots + c_n U_n$$

$$Y_0 = d_1 V_1 + \dots + d_n V_n,$$

где U_1, \dots, U_n — собственные векторы матрицы A , принадлежащие собственным значениям $\lambda_1, \dots, \lambda_n$, занумерованным согласно условию 3) теоремы. Соответственно, V_1, \dots, V_n собственные векторы матрицы A' . Тогда

$$(31) \quad m_j^{(k)} = b_1 \varphi_k(\lambda_1) \lambda_1^j + \dots + b_n \varphi_k(\lambda_n) \lambda_n^j,$$

где

$$b_i = c_i d_i (U_i, V_i).$$

Следовательно,

$$\Delta_i^{(k)} = \begin{vmatrix} \Sigma b_s \varphi_k(\lambda_s) & \Sigma b_s \varphi_k(\lambda_s) \lambda_s & \dots & \Sigma b_s \varphi_k(\lambda_s) \lambda_s^{i-1} \\ \Sigma b_s \varphi_k(\lambda_s) \lambda_s & \Sigma b_s \varphi_k(\lambda_s) \lambda_s^2 & \dots & \Sigma b_s \varphi_k(\lambda_s) \lambda_s^i \\ \dots & \dots & \dots & \dots \\ \Sigma b_s \varphi_k(\lambda_s) \lambda_s^{i-1} & \Sigma b_s \varphi_k(\lambda_s) \lambda_s^i & \dots & \Sigma b_s \varphi_k(\lambda_s) \lambda_s^{2i-2} \end{vmatrix}.$$

Здесь все суммы распространены на $s = 1, 2, \dots, n$.

Матрицу, находящуюся под знаком определителя $\Delta_i^{(k)}$, можно представить как произведение следующих двух прямоугольных матриц:

$$\begin{bmatrix} b_1 \varphi_k(\lambda_1) & b_2 \varphi_k(\lambda_2) & \dots & b_n \varphi_k(\lambda_n) \\ \lambda_1 b_1 \varphi_k(\lambda_1) & \lambda_2 b_2 \varphi_k(\lambda_2) & \dots & \lambda_n b_n \varphi_k(\lambda_n) \\ \dots & \dots & \dots & \dots \\ \lambda_1^{i-1} b_1 \varphi_k(\lambda_1) & \lambda_2^{i-1} b_2 \varphi_k(\lambda_2) & \dots & \lambda_n^{i-1} b_n \varphi_k(\lambda_n) \end{bmatrix}$$

и

$$\begin{bmatrix} 1 & \lambda_1 & \dots & \lambda_1^{i-1} \\ 1 & \lambda_2 & \dots & \lambda_2^{i-1} \\ \dots & \dots & \dots & \dots \\ 1 & \lambda_n & \dots & \lambda_n^{i-1} \end{bmatrix}.$$

Воспользовавшись известной теоремой об определителе произведения двух прямоугольных матриц, получим

$$\Delta_i^{(k)} = \sum_{s_1 < s_2 < \dots < s_i} b_{s_1} \varphi_k(\lambda_{s_1}) \dots b_{s_i} \varphi_k(\lambda_{s_i}) \begin{vmatrix} 1 & \lambda_{s_1} & \dots & \lambda_{s_1}^{i-1} \\ \dots & \dots & \dots & \dots \\ 1 & \lambda_{s_i} & \dots & \lambda_{s_i}^{i-1} \end{vmatrix}.$$

Нетрудно видеть, что при достаточно большом k

$$|\varphi_k(\lambda_1)| > |\varphi_k(\lambda_2)| > \dots > |\varphi_k(\lambda_n)|$$

и более того,

$$\lim_{k \rightarrow \infty} \frac{\varphi_k(\lambda_i)}{\varphi_k(\lambda_{i-1})} = 0.$$

Действительно,

$$\frac{\varphi_k(\lambda_i)}{\varphi_k(\lambda_{i-1})} = \frac{(\lambda_i - t_1) \dots (\lambda_i - t_k)}{(\lambda_{i-1} - t_1) \dots (\lambda_{i-1} - t_k)} = \frac{\varphi_{k_0}(\lambda_i)}{\varphi_{k_0}(\lambda_{i-1})} \prod_{s=k_0+1}^k \frac{\lambda_i - t_s}{\lambda_{i-1} - t_s}.$$

Выберем k_0 настолько большим, чтобы при $s > k_0$

$$\left| \frac{\lambda_i - t_s}{\lambda_{i-1} - t_s} \right| < \left| \frac{\lambda_i - \tau}{\lambda_{i-1} - \tau} \right| + \varepsilon,$$

где ε малое число, такое, что $\left| \frac{\lambda_i - \tau}{\lambda_{i-1} - \tau} \right| + \varepsilon = q < 1$.

Тогда

$$\left| \frac{\varphi_k(\lambda_i)}{\varphi_k(\lambda_{i-1})} \right| < \left| \frac{\varphi_{k_0}(\lambda_i)}{\varphi_{k_0}(\lambda_{i-1})} \right| q^{k-k_0} \rightarrow 0$$

при $k \rightarrow \infty$.

Поэтому преобладающим слагаемым в $\Delta_i^{(k)}$ будет то, в котором $s_1 = 1, s_2 = 2, \dots, s_i = i$. Следующим по величине будет то, в котором $s_1 = 1, s_2 = 2, \dots, s_{i-1} = i-1, s_i = i+1$. Следовательно,

$$\Delta_i^{(k)} = b_1 \dots b_i \varphi_k(\lambda_1) \dots \varphi_k(\lambda_i) \begin{vmatrix} 1 & \lambda_1 & \dots & \lambda_1^{i-1} \\ \cdot & \cdot & \cdot & \cdot \\ 1 & \lambda_1 & \dots & \lambda_1^{i-1} \end{vmatrix} \left[1 + O\left(\frac{\varphi_k(\lambda_{i+1})}{\varphi_k(\lambda_i)}\right) \right]. \quad (20)$$

Эта формула верна для $i = 1, 2, \dots, n-1$. При $i=0$ мы должны считать $\Delta_0^{(k)} = 1$, при $i=n$ верна точная формула

$$\Delta_i^{(k)} = b_1 \dots b_n \varphi_k(\lambda_1) \dots \varphi_k(\lambda_n) \begin{vmatrix} 1 & \lambda_1 & \dots & \lambda_1^{n-1} \\ \cdot & \cdot & \cdot & \cdot \\ 1 & \lambda_n & \dots & \lambda_n^{n-1} \end{vmatrix}. \quad (21)$$

Далее,

$$\begin{aligned} \frac{\Delta_i^{(k+1)}}{\Delta_i^{(k)}} &= \frac{b_1 \dots b_i \varphi_{k+1}(\lambda_1) \dots \varphi_{k+1}(\lambda_i) \begin{vmatrix} 1 & \lambda_1 & \dots & \lambda_1^{i-1} \\ \cdot & \cdot & \cdot & \cdot \\ 1 & \lambda_i & \dots & \lambda_i^{i-1} \end{vmatrix} \left[1 + O\left(\frac{\varphi_{k+1}(\lambda_{i+1})}{\varphi_{k+1}(\lambda_i)}\right) \right]}{b_1 \dots b_i \varphi_k(\lambda_1) \dots \varphi_k(\lambda_i) \begin{vmatrix} 1 & \lambda_1 & \dots & \lambda_1^{i-1} \\ \cdot & \cdot & \cdot & \cdot \\ 1 & \lambda_i & \dots & \lambda_i^{i-1} \end{vmatrix} \left[1 + O\left(\frac{\varphi_k(\lambda_{i+1})}{\varphi_k(\lambda_i)}\right) \right]} = \\ &= (\lambda_1 - t_{k+1}) \dots (\lambda_i - t_{k+1}) \frac{1 + O\left(\frac{\varphi_{k+1}(\lambda_{i+1})}{\varphi_{k+1}(\lambda_i)}\right)}{1 + O\left(\frac{\varphi_k(\lambda_{i+1})}{\varphi_k(\lambda_i)}\right)}. \end{aligned}$$

Для $i < n-1$ величины $\frac{\varphi_k(\lambda_{i+1})}{\varphi_k(\lambda_i)}$ и $\frac{\varphi_{k+1}(\lambda_{i+1})}{\varphi_{k+1}(\lambda_i)}$ имеют одинаковые порядки малости, ибо

$$\frac{\varphi_{k+1}(\lambda_{i+1})}{\varphi_{k+1}(\lambda_i)} = \frac{\varphi_k(\lambda_{i+1})}{\varphi_k(\lambda_i)} \frac{(\lambda_{i+1} - t_{k+1})}{(\lambda_i - t_k)} \approx \frac{\varphi_k(\lambda_{i+1})}{\varphi_k(\lambda_i)} \frac{\lambda_{i+1} - \tau}{\lambda_i - \tau}.$$

При $i=n-1$ порядок малости $\frac{\varphi_{k+1}(\lambda_n)}{\varphi_{k+1}(\lambda_{n-1})}$ может быть выше порядка $\frac{\varphi_k(\lambda_n)}{\varphi_k(\lambda_{n-1})}$, если $\tau = \lambda_n$.

Поэтому при $i \leq n-1$ верна формула

$$\frac{\Delta_i^{(k+1)}}{\Delta_i^{(k)}} = (\lambda_1 - t_{k+1}) \dots (\lambda_i - t_{k+1}) \left[1 + O\left(\frac{\varphi_k(\lambda_{i+1})}{\varphi_k(\lambda_i)}\right) \right]. \quad (22)$$

Для $i = n$ верна точная формула

$$\frac{\Delta_n^{(k+1)}}{\Delta_n^{(k)}} = (\lambda_1 - t_{k+1}) \dots (\lambda_n - t_{k+1}). \quad (23)$$

Отсюда получаем асимптотические формулы для $\rho_i^{(k+1)}$. Именно, при $i = 0, 1, \dots, n-2$

$$\begin{aligned} \rho_i^{(k+1)} &= \frac{\Delta_{i+1}^{(k+1)}}{\Delta_{i+1}^{(k)}} : \frac{\Delta_i^{(k+1)}}{\Delta_i^{(k)}} = \\ &= \frac{(\lambda_1 - t_{k+1}) \dots (\lambda_{i+1} - t_{k+1}) \left[1 + O\left(\frac{\varphi_k(\lambda_{i+2})}{\varphi_k(\lambda_{i+1})}\right) \right]}{(\lambda_1 - t_{k+1}) \dots (\lambda_i - t_{k+1}) \left[1 + O\left(\frac{\varphi_k(\lambda_{i+1})}{\varphi_k(\lambda_i)}\right) \right]} = \\ &= (\lambda_{i+1} - t_{k+1}) \left[1 + O\left(\frac{\varphi_k(\lambda_{i+2})}{\varphi_k(\lambda_{i+1})}\right) + O\left(\frac{\varphi_k(\lambda_{i+1})}{\varphi_k(\lambda_i)}\right) \right]. \end{aligned} \quad (24)$$

Для $\rho_{n-1}^{(k+1)}$ асимптотическая формула имеет более простой вид

$$\rho_{n-1}^{(k+1)} = (\lambda_n - t_{k+1}) \left[1 + O\left(\frac{\varphi_k(\lambda_n)}{\varphi_k(\lambda_{n-1})}\right) \right]. \quad (25)$$

Переходя к пределу, получим

$$\lim_{k \rightarrow \infty} \rho_i^{(k+1)} = \lambda_{i+1} - \tau \quad (i = 0, 1, \dots, n-1),$$

что и требовалось доказать.

При доказательстве теоремы мы получили и оценку быстроты сходимости последовательностей $\rho_i^{(k)}$.

Остается доказать, что последовательности $\sigma_i^{(k)}$ стремятся к нулю. Имеем

$$\frac{\sigma_i^{(k+1)}}{\sigma_i^{(k)}} = \frac{\rho_{i+1}^{(k)}}{\rho_i^{(k+1)}}.$$

Мы уже установили, что $\frac{\rho_{i+1}^{(k)}}{\rho_i^{(k+1)}} \rightarrow \frac{\lambda_{i+1} - \tau}{\lambda_i - \tau}$, и, следовательно, при $k > k_0$

$$\left| \frac{\sigma_i^{(k+1)}}{\sigma_i^{(k)}} \right| < \left| \frac{\lambda_{i+1} - \tau}{\lambda_i - \tau} \right| + \varepsilon = q < 1.$$

Поэтому $|\sigma_i^{(k)}| \leq |\sigma_i^{(k_0)}| q^{k-k_0} \rightarrow 0$.

Из доказанной теоремы следует, в частности, что если взять $t_2 = t_3 = \dots = 0$ (схема QD), то при $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$, получим

$$\lim_{k \rightarrow \infty} p_i^{(k)} = \lambda_{i+1}.$$

Точнее

$$p_i^{(k)} = \lambda_{i+1} + O\left(\frac{\lambda_{i+2}}{\lambda_{i+1}}\right)^k + O\left(\frac{\lambda_{i+1}}{\lambda_i}\right)^k. \quad (26)$$

Таким образом, результат теоремы 77.1 распространяется на любые матрицы с вещественными собственными значениями, удовлетворяющими неравенствам $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$, если только предположить, что все $\Delta_i^{(k)}$ не равны нулю.

В работе Рутисхаузера [2] рассматривается и случай комплексных собственных значений, а также возможные вырождения процесса. Задача вычисления собственных векторов рассмотрена в работе [4].

Сходимость алгорифма деления и вычитания, особенно при наличии близких собственных значений, довольно медленна. Так как этот процесс к тому же не самоисправляющийся, при длительном его проведении возникает опасность некоторого накопления ошибок округления. Тем более интересна возможность использования сдвигов для ускорения сходимости процесса. Именно, при некотором определенном выборе сдвигов получается процесс с квадратичной сходимостью поочередно для каждого собственного значения.

Теорема 77.3. Пусть на некотором шаге процесса QD (с сдвигом или без сдвига) в условиях предыдущей теоремы уже получено, что

$$|\lambda_n - p_{n-1}^{(k)} - t_k| < \epsilon.$$

Тогда, взяв

$$t_{k+1} = p_{n-1}^{(k)} + t_k,$$

получим

$$|\lambda_n - p_{n-1}^{(k+1)} - t_{k+1}| < \mu \epsilon^2,$$

где μ — некоторая константа.

Доказательство. Можно доказать, что в асимптотической формуле (25) порядок остаточного члена, в условиях теоремы 77.2, оказывается точным. Именно,

$$p_{n-1}^{(k)} = (\lambda_n - t_k) \left[1 + M \frac{\varphi_{k-1}(\lambda_n)}{\varphi_{k-1}(\lambda_{n-1})} (1 + \epsilon_k) \right],$$

где M — некоторая константа, $\epsilon_k \rightarrow 0$. Поэтому

$$\frac{p_{n-1}^{(k+1)} - \lambda_n + t_{k+1}}{p_{n-1}^{(k)} - \lambda_n + t_k} = \frac{\varphi_{k+1}(\lambda_n) \varphi_{k-1}(\lambda_{n-1})}{\varphi_k(\lambda_{n-1}) \varphi_k(\lambda_n)} (1 + \epsilon'_k) = \frac{\lambda_n - t_{k+1}}{\lambda_{n-1} - t_k} (1 + \epsilon'_k),$$

где $\varepsilon'_k \rightarrow 0$. При

$$t_{k+1} = p_{n-1}^{(k)} + t_k$$

имеем

$$\frac{\lambda_n - p_{n-1}^{(k+1)} - t_{k+1}}{\lambda_n - p_{n-1}^{(k)} - t_k} = \frac{\lambda_n - p_{n-1}^{(k)} - t_k}{\lambda_{n-1} - t_k} (1 + \varepsilon'_k),$$

откуда

$$|\lambda_n - p_{n-1}^{(k+1)} - t_{k+1}| = \frac{(\lambda_n - p_{n-1}^{(k)} - t_k)^2}{|\lambda_{n-1} - t_k|} (1 + \varepsilon'_k), \quad (27)$$

т. е.

$$|\lambda_n - p_{n-1}^{(k+1)} - t_{k+1}| < \nu \varepsilon^2,$$

при

$$\nu < \frac{2}{|\lambda_{n-1} - t_k|}.$$

Опишем теперь упомянутый процесс с ускорением. Допустим для простоты, что все собственные значения положительны и $\lambda_1 > \lambda_2 > \dots > \lambda_n$.

Пусть сделано несколько шагов алгорифма *QD* без сдвига, так что в грубом приближении последний столбец начал стабилизоваться. Пусть это произошло на k_1 шаге. Тогда берем $t_{k_1+1} = p_{n-1}^{(k_1)}$, $t_{k_1+2} = p_{n-1}^{(k_1+1)} + t_{k_1+1}, \dots$. В новом процессе мы получим квадратичную сходимость последовательности t_k к λ_n ; последовательности же $\sigma_{n-1}^{(k)}$ и $p_{n-1}^{(k)}$ сходятся с той же быстротой к нулю.

Пусть при $k = k_2$ числа $\sigma_{n-1}^{(k)}$ и $p_{n-1}^{(k)}$ практически станут равными нулю. Тогда можно принять, с той же степенью точности, $\lambda_n \approx t_{k_2}$ и перейти к процессу с ускорением для определения λ_{n-1} . Именно, полагаем $t_{k_2+1} = p_{n-2}^{(k_2)}$, $t_{k_2+2} = p_{n-2}^{(k_2+1)} + t_{k_2+1}, \dots$. При этом мы вычеркиваем из схемы столбец, состоящий из значений $p_{n-1}^{(k)}$, так как последние более не нужны ни для определения λ_n , ни для продолжения схемы, ибо $\sigma_{n-1}^{(k)}$ стало и остается равным нулю. После определения λ_{n-1} переходим, поочередно, к определению $\lambda_{n-2}, \lambda_{n-3}, \dots, \lambda_1$.

То, что в этом процессе для каждого собственного значения будет иметь место квадратичная сходимость, следует из того, что по ходу процесса векторы $p_0^{(k)}, p_1^{(k)}, \dots, p_{n-1}^{(k)}$ будут попадать в инвариантные подпространства убывающих размерностей и $\lambda_{n-1}, \lambda_{n-2}, \dots, \lambda_1$ будут играть поочередно роль наименьших собственных значений.

Указанный процесс можно применять уже при переходе ко второй строке, полагая $t_2 = p_{n-1}^{(1)}$, $t_3 - t_2 = p_{n-1}^{(2)} \dots$. При этом, однако, несколько первых сдвигов будут иметь случайный характер и процесс начнет быстро сходиться, лишь как только один из сдвигов окажется близким к какому-либо собственному значению. Именно

Таблица VIII.1

Схема QD без сдвигов

k	$\rho_0^{(k)}$	$\sigma_1^{(k)}$	$\rho_1^{(k)}$	$\sigma_2^{(k)}$	$\rho_2^{(k)}$	$\sigma_3^{(k)}$	$\rho_3^{(k)}$	Σ
1	0	2.3	0.018	0.45668599	0.11169690	0.39432359	0.02841890	0.69087470
2	0	2.318	0.00354631	0.56483656	0.07797792	0.34476467	0.05694870	0.63392600
3	0	2.32154631	0.00086282	0.64195158	0.04187854	0.358983483	0.10032731	0.53359869
13	0	2.32274877	0.00000001	0.77144752	0.00436073	0.65910897	0.00004616	0.24228792
14	0	2.32274878	0	0.77580825	0.00370478	0.65545035	0.00001706	0.24227066
21	0			0.79176600	0.00095776	0.64226676	0.00000002	0.24226076
22				0.79272376	0.00077598	0.64149080	0.00000001	0.24226075
69				0.79570669	0.00000002	0.638828384		

Схема QD со сдвигами

k	$\rho_0^{(k)}$	$\sigma_1^{(k)}$	$\rho_1^{(k)}$	$\sigma_2^{(k)}$	$\rho_2^{(k)}$	$\sigma_3^{(k)}$	$\rho_3^{(k)}$	λ_1
1	0	2.3	0.018	0.45668599	0.11169690	0.39432359	0.02841890	0.69087470
2	0	1.6712530	0.00505207	-0.12754388	-0.34532995	0.07719774	0.25433256	-0.25433256
3	0	1.88650993	-0.00034156	-0.21819971	0.12217565	0.46366871	-0.13850148	0.43654214
4	0	1.74665689	0.0004267	-0.23556821	-0.24048782	0.42517207	-0.04577126	0.57604362
5	0	1.70093830	-0.00000591	-0.52182138	0.19594579	0.13768376	-0.01521609	0.62181488
6	0	1.88571630	0.000001183	-0.34109351	-0.07909430	0.16634588	-0.00124247	0.63703097
7	0	1.88447566	-0.000000337	-0.42142991	0.03497354	0.14888740	-0.00001037	0.63827344
8	0	1.88446492	0.00000039	-0.38646683	-0.01347365	0.16234031	0	0.63828381
9	0	1.52212470	-0.00000092	-0.56228077	0.003889008	-0.003889008	0.80062412	0.79673404
10	0	1.53601476	0.00000001	-0.55450062	0.000002729	-0.000002729	0	0.79670675
11				-0.56444604	0	0		0.24226071
12				0				0.24226071

это собственное значение будет определено первым. Следующим, вообще говоря, определится собственное значение, ближайшее к полученному.

Применение сдвига возможно, и на первом шагу процесса, т. е. можно взять весовой полином $\varphi_k(t)$ равным $(t - t_1)(t - t_2) \dots (t - t_k)$ при $t_1 \neq 0$. Это влечет за собой изменение начальной строки процесса. Легко видеть, что она в этом случае должна быть построена по формулам

$$\tilde{p}_0^{(1)} = \alpha_0 - t_1$$

$$\tilde{p}_i^{(1)} = \frac{\beta_i}{\tilde{p}_{i-1}^{(1)}} \quad (i = 1, 2, \dots, n-1)$$

$$\tilde{p}_i^{(1)} = -\tilde{p}_i^{(1)} + \alpha_i - t_1 \quad (i = 1, 2, \dots, n-1).$$

В этой форме QD процесс полезен для уточнения полученного каким-либо другим способом грубого приближения к одному из собственных значений. За t_1 следует взять это известное грубое приближение и далее применять процесс QD со сдвигами так, как он описан выше.

В табл. VIII. 1 — VIII. 6 приводится числовой материал, иллюстрирующий ход QD процесса.

В табл. VIII. 1 дается ход QD процесса без ускорения для положительно-определенной матрицы (4) § 51. Первая строка таблицы заполняется по данным табл. VI. 16.

В табл. VIII. 2 приведен для той же матрицы QD процесс со сдвигами

$$t_1 = 0, \quad t_{k+1} = t_k + p_3^{(k)}.$$

В этом случае все четыре собственных значения определяются после двенадцати шагов процесса.

В табл. VIII. 3 дается ход QD процесса для не положительно-определенной (и даже несимметричной) матрицы Леверье. Первая строка схемы взята из табл. VI. 17.

В табл. VIII. 4 приводится для матрицы Леверье QD процесс со сдвигами

$$t_1 = t_2 = t_3 = t_4 = t_5 = t_6 = 0$$

$$t_{k+1} = t_k + p_3^{(k)} \quad \text{при } k \geq 6.$$

Наибольшие корни получились с невысокой точностью из-за некоторой потери точности на 7-м и 8-м шагах.

В табл. VIII. 5 дается уточнение наибольших корней матрицы Леверье при помощи схемы QD с постоянными сдвигами.

Наконец, в табл. VIII. 6 дается уточнение первого собственного значения с изменением начальной строки процесса на $t_1 = -17.863248$ и с последующими постоянными сдвигами.

Таблица VIII. 3

Схема QD без сдвигов

k	$\rho_0^{(k)}$	$\sigma_1^{(k)}$	$\rho_1^{(k)}$	$\sigma_2^{(k)}$	$\rho_2^{(k)}$	$\sigma_3^{(k)}$	$\rho_3^{(k)}$	Σ
1	-7.5036210	-0.23490436	-13.224755	-1.5827643	-5.9951147	1.3221017	-20.669372	-47.888430
2	-7.7388254	-0.40143988	-14.406079	-0.65867010	-4.014349	-6.8073437	-27.476717	-47.888431
3	-8.1398653	-0.71046673	-1.354282	-0.18420479	-2.972056	-62.825173	35.348456	-47.888430
4	-8.8504320	-1.15922072	-13.386200	0.04966872	-59.888936	-37.081521	1.73806560	-47.888430
5	-10.002719	-1.5420554	-11.203176	0.20187398	-23.015289	2.7922606	4.52932256	-47.888430
6	-11.544774	-1.5765705	-10.018732	0.47753345	-20.700562	0.61041324	5.1357388	-47.888431
35	-18.6535638	0.06358643	-16.425749	-0.00000061	-7.5740424	-0.00000061	-5.29696996	
137	-17.869245	0.00024064	-17.146587					

Схема QD со сдвигами

k	$\rho_0^{(k)}$	$\sigma_1^{(k)}$	$\rho_1^{(k)}$	$\sigma_2^{(k)}$	$\rho_2^{(k)}$	$\sigma_3^{(k)}$	$\rho_3^{(k)}$	t_k
6	-11.544774	-1.5765705	-10.018732	0.47753345	-20.700562	0.61041324	-5.1357388	0
7	-7.9856057	-1.9779636	-2.4274962	4.0721838	-19.026594	0.16476533	-0.16476533	5.1357388
8	-9.7988040	-0.49000869	-2.2994616	-33.694752	14.997589	-0.00181012	0.00181012	5.3005041
9	-10.290623	-1.0949348	-31.506394	-16.903290	-1.0452212	-0.00000313	-0.00000313	5.2988940
10	-10.181126	-15.806140	0.33883939	1.06063588	-2.10585337	0	0	5.2988971
11	-7.7366329	-0.69227548	-13.331923	0.16753398	-0.16753398	-0.00195743	-0.00195743	7.4045508
12	-5.7662234	-1.3388991	-14.3388991	0.00195743	-0.00195743	0	0	7.5720848
13	-5.5325301	3.4784943	-17.813547	0.00000022	-0.00000022	0	0	7.5740422
14	-2.0504356	30.167151	-47.980721					
15	76.093836	-19.021799	-19.021799					-55.554753
16	38.050238	9.50092398	9.50092398					-36.332961
17	19.031758	-4.7513026	4.7513026					-27.023724
18	9.5291528	-2.36903029	2.36903029					-22.272421
19	4.7910870	-1.1714078	1.1714078					-19.903388
20	2.4482714	0.560447554	-0.560447554					-18.731960
21	1.3273203	0.238666894	-0.238666894					-18.171504
22	0.853098642	0.065558798	-0.065558798					-17.984587
23	0.72281046	0.00595147	-0.00595147					-17.869249
24	0.71080752	0.00004982	-0.00004982					-17.863248
25	0.71080758	0	0					-17.152440

Таблица VIII. 5

Уточнение собственных значений по схеме QD с постоянными сдвигами

k	$p_0^{(k)}$	$\sigma_1^{(k)}$	$p_1^{(k)}$	$\sigma_2^{(k)}$	$p_2^{(k)}$	$\sigma_3^{(k)}$	$p_3^{(k)}$	t_k	λ_k
1	-7.5036210	-0.23490436	-13.224755	-1.5827643	-5.9951147	1.3221017	-20.669372	0	
2	10.124723	0.306682839	2.7489003	3.4518725	9.7383625	-2.8061198	-0.000004420	-17.863248	
3	10.431551	0.08085477	6.1199180	5.4928163	1.4394264	0.0300319	-0.00001239	-17.863248	
4	10.512406	0.04707054	11.565664	0.68361875	0.75581584	0	-0.00001239	-17.863260	
1	-7.5036210	-0.23490436	-13.224755	-1.5827643	-5.9951147	1.3221017	-20.669372	0	
2	9.4139146	0.32999583	2.0149249	4.7092839	7.7701431	-3.5169251	-0.00000690	-17.152440	
3	9.7439104	0.06823921	6.6559696	5.4975926	-1.2443746	-0.00001950	0.00001260	-17.152440	
4	9.8121496	0.04628936	12.10723	-0.56503756	-0.67935654	0	0.00001260	-17.152427	

Таблица VIII. 6

Уточнение собственного значения по схеме QD с измененной первой строкой

k	$\bar{p}_0^{(k)}$	$\bar{\sigma}_1^{(k)}$	$\bar{p}_1^{(k)}$	$\bar{\sigma}_2^{(k)}$	$\bar{p}_2^{(k)}$	$\bar{\sigma}_3^{(k)}$	$\bar{p}_3^{(k)}$	t_k	λ_k
1	10.359627	0.17014447	4.2334445	4.9443593	5.3410097	-1.4840175	-0.00000550	-17.863248	
2	10.529771	0.68405768	8.4937461	3.1090958	0.74789640	0.00001091	-0.00001641	-17.863248	
3	11.213829	0.51812919	11.084713	0.20977372	0.53813359	0	-0.00001641	-17.863264	

§ 78. Треугольный степенной метод

Треугольный степенной метод (Бауэр [7]) является обобщением ступенчатого степенного метода, приспособленным для нахождения всех собственных значений матрицы A . Мы будем предполагать, что все собственные значения матрицы A вещественны и различны по абсолютной величине. Обозначим их $\lambda_1, \dots, \lambda_n$, занумеровав в порядке убывания абсолютных величин.

Для обоснования треугольного степенного метода проведем следующие рассуждения. Пусть K и M произвольные матрицы. Рассмотрим последовательность матриц $A^{(k)} = KA^kM$. Каждую из этих матриц представим в виде произведения левой треугольной матрицы C_k с единичными диагональными элементами, диагональной D_k и правой треугольной B_k с единичными диагональными элементами:

$$A^{(k)} = C_k D_k B_k.$$

Предполагается, что такое разложение возможно при всех k . Тогда, при некоторых дополнительных ограничениях, матрицы $D_k = [d_1^{(k)}, \dots, d_n^{(k)}]$ таковы, что

$$\lim_{k \rightarrow \infty} \frac{d_i^{(k)}}{d_i^{(k-1)}} = \lambda_i, \quad (1)$$

а матрицы C_k и B_k стремятся к предельным матрицам.

Докажем это. Прежде всего отметим (§ 1, п. 12), что

$$d_i^{(k)} = \frac{\Delta_i^{(k)}}{\Delta_{i-1}^{(k)}}, \quad (2)$$

где $\Delta_i^{(k)}$ верхний главный минор i -го порядка матрицы $A^{(k)}$. Оценим $\Delta_i^{(k)}$. Пусть $A = P^{-1}\Lambda P$, где $\Lambda = [\lambda_1, \dots, \lambda_n]$. Тогда

$$A^{(k)} = KA^kM = KP^{-1}\Lambda^kPM.$$

Положим

$$KP^{-1} = \begin{bmatrix} q_{11} & \dots & q_{1n} \\ \vdots & \ddots & \vdots \\ q_{n1} & \dots & q_{nn} \end{bmatrix}, \quad PM = \begin{bmatrix} p_{11} & \dots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \dots & p_{nn} \end{bmatrix}, \quad (3)$$

Тогда

$$\Lambda^k PM = \begin{bmatrix} \lambda_1^k p_{11} & \lambda_1^k p_{12} & \dots & \lambda_1^k p_{1n} \\ \lambda_2^k p_{21} & \lambda_2^k p_{22} & \dots & \lambda_2^k p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_n^k p_{n1} & \lambda_n^k p_{n2} & \dots & \lambda_n^k p_{nn} \end{bmatrix}.$$

Матрица минора $\Delta_i^{(k)}$ есть произведение двух прямоугольных матриц, из которых первая составлена из первых i строк матрицы KP^{-1} , вторая из первых i столбцов матрицы $\Lambda^k PM$. Поэтому

$$\Delta_i^{(k)} = \sum_{j_1 < \dots < j_i} \begin{vmatrix} q_{1j_1} & \dots & q_{1j_i} \\ \dots & \dots & \dots \\ q_{ij_1} & \dots & q_{ij_i} \end{vmatrix} \cdot \begin{vmatrix} p_{j_1 1} & \dots & p_{j_1 i} \\ \dots & \dots & \dots \\ p_{j_i 1} & \dots & p_{j_i i} \end{vmatrix} \lambda_{j_1}^k \dots \lambda_{j_i}^k. \quad (4)$$

При достаточно большом k , преобладающим слагаемым в $\Delta_i^{(k)}$ будет, вообще говоря, слагаемое, соответствующее $j_1 = 1, \dots, j_i = i$. Вторым по величине слагаемым будет слагаемое, соответствующее $j_1 = 1, \dots, j_{i-1} = i-1, j_i = i+1$. Это верно, если

$$Q_{ii} = \begin{vmatrix} q_{11} & \dots & q_{ii} \\ \dots & \dots & \dots \\ q_{ii} & \dots & q_{ii} \end{vmatrix} \neq 0 \text{ и } P_{ii} = \begin{vmatrix} p_{11} & \dots & p_{ii} \\ \dots & \dots & \dots \\ p_{ii} & \dots & p_{ii} \end{vmatrix} \neq 0. \quad (5)$$

Последние условия мы будем предполагать выполненными при всех i . В этом предположении

$$\Delta_i^{(k)} = Q_{ii} P_{ii} \lambda_1^k \lambda_2^k \dots \lambda_i^k \left[1 + O\left(\frac{\lambda_{i+1}}{\lambda_i}\right)^k \right]. \quad (6)$$

Следовательно,

$$d_i^{(k)} = \frac{\Delta_i^{(k)}}{\Delta_{i-1}^{(k)}} = \frac{Q_{ii} P_{ii}}{Q_{i-1} P_{i-1}} \lambda_i^k \left[1 + O\left(\frac{\lambda_{i+1}}{\lambda_i}\right)^k + O\left(\frac{\lambda_i}{\lambda_{i-1}}\right)^k \right]. \quad (7)$$

Подобным же образом

$$d_i^{(k-1)} = \frac{Q_{ii} P_{ii}}{Q_{i-1} P_{i-1}} \lambda_i^{k-1} \left[1 + O\left(\frac{\lambda_{i+1}}{\lambda_i}\right)^k + O\left(\frac{\lambda_i}{\lambda_{i-1}}\right)^k \right].$$

Поэтому

$$\frac{d_i^{(k)}}{d_i^{(k-1)}} = \lambda_i + O\left(\frac{\lambda_{i+1}}{\lambda_i}\right)^k + O\left(\frac{\lambda_i}{\lambda_{i-1}}\right)^k \quad (i = 1, \dots, n-1). \quad (8)$$

Для наименьшего по модулю собственного значения мы получим

$$\frac{d_n^{(k)}}{d_n^{(k-1)}} = \lambda_n + O\left(\frac{\lambda_n}{\lambda_{n-1}}\right)^k. \quad (9)$$

Обратимся теперь к исследованию матриц B_k и C_k . Напомним явные формулы для элементов матриц B_k и C_k (§ 1, п. 12). Именно,

$$b_{ij}^{(k)} = \frac{\beta_{ij}^{(k)}}{\beta_{ii}^{(k)}}$$

$$c_{ji}^{(k)} = \frac{\gamma_{ji}^{(k)}}{\gamma_{ii}^{(k)}},$$

где $\beta_{ii}^{(k)} = \gamma_{ii}^{(k)} = \Delta_i^{(k)}$, а $\beta_{ij}^{(k)}$ и $\gamma_{ji}^{(k)}$ суть некоторые миноры i -го порядка матрицы $A^{(k)}$.

Очевидно, что рассуждения, которые мы применяли для оценки главных миноров $\Delta_i^{(k)}$, остаются в силе для оценки любых миноров, так что

$$\begin{aligned}\beta_{ij}^{(k)} &= Q_{ij}P_{ij}\lambda_1^k \dots \lambda_i^k \left[1 + O\left(\frac{\lambda_{i+1}}{\lambda_i}\right)^k \right] \\ \gamma_{ji}^{(k)} &= Q_{ji}P_{ji}\lambda_1^k \dots \lambda_i^k \left[1 + O\left(\frac{\lambda_{i+1}}{\lambda_i}\right)^k \right].\end{aligned}\quad (10)$$

Здесь Q_{ij} , P_{ij} , Q_{ji} , P_{ji} некоторые миноры i -го порядка, составленные из элементов матриц KP^{-1} и PM . Поэтому

$$\begin{aligned}b_{ij}^{(k)} &= \frac{Q_{ij}P_{ij}}{Q_{ii}P_{ii}} \left[1 + O\left(\frac{\lambda_{i+1}}{\lambda_i}\right)^k \right] \\ c_{ji}^{(k)} &= \frac{Q_{ji}P_{ji}}{Q_{ii}P_{ii}} \left[1 + O\left(\frac{\lambda_{i+1}}{\lambda_i}\right)^k \right].\end{aligned}\quad (11)$$

Таким образом, при сделанном выше предположении о неравенстве нулю всех определителей Q_{ii} и P_{ii} , все элементы $b_{ij}^{(k)}$ и $c_{ji}^{(k)}$ имеют пределы при $k \rightarrow \infty$, причем равенства (11) дают оценку быстроты сходимости.

Перейдем теперь к описанию вычислительной схемы метода.

Пусть C_0 произвольная матрица, A матрица с различными по абсолютной величине вещественными собственными значениями, для которой требуется решить полную проблему собственных значений.

Строим последовательность левых треугольных матриц $C_1, C_2, \dots, C_k, \dots$ с единичными диагональными элементами посредством рекуррентных соотношений:

$$\begin{aligned}AC_0 &= C_1R_1 \\ AC_1 &= C_2R_2 \\ &\vdots \\ AC_{k-1} &= C_kR_k \\ &\vdots\end{aligned}\quad (12)$$

Здесь $R_1, R_2, \dots, R_k, \dots$ — правые треугольные матрицы.

Процесс следует вести до тех пор, пока матрицы C_k не стабилизируются с достаточной точностью. То, что $\lim_{k \rightarrow \infty} C_k$ существует, легко доказывается. Именно, исключая из соотношений (12) матрицы C_1, C_2, \dots, C_{k-1} , получим

$$A^kC_0 = C_kR_kR_{k-1} \dots R_1. \quad (13)$$

Матрица $R_k R_{k-1} \dots R_1$ — правая треугольная. Поэтому матрицы C_k и $R_k R_{k-1} \dots R_1$ совпадают с матрицами C_k и $D_k B_k$ в предыдущих обозначениях, построенных для матрицы

$$A^{(k)} = A^k C_0.$$

Следовательно, $\lim_{k \rightarrow \infty} C_k = C$ существует.

Одновременно со стабилизацией матриц C_k стабилизируются и матрицы R_k , ибо при $k \rightarrow \infty$

$$R_k = C_k^{-1} A C_{k-1} \rightarrow C^{-1} A C = R. \quad (14)$$

Знание предельных матриц C и R дает возможность решить полную проблему собственных значений для матрицы A .

Действительно, из равенства (14) следует, что собственные значения матриц A и R совпадают, а собственные векторы U_1, \dots, U_n матрицы A равны соответственно CV_1, \dots, CV_n , где V_1, \dots, V_n собственные векторы матрицы R . Матрица R треугольная, так что ее собственные значения равны ее диагональным элементам, а собственные векторы V_1, \dots, V_n легко определяются из решения треугольной системы.

Быстрота сходимости диагональных элементов матриц к искомым собственным значениям может быть оценена из следующих соображений.

Прежде всего из равенства

$$R_k R_{k-1} \dots R_1 = D_k B_k$$

следует

$$R_k = D_k B_k B_{k-1}^{-1} D_{k-1}^{-1}.$$

Поэтому диагональные элементы матрицы R_k совпадают с диагональными элементами матрицы $D_k D_{k-1}^{-1}$, быстрота сходимости которых к собственным значениям оценивается формулами (8) и (9).

Матрицы C_k и R_k можно вычислять на каждом шагу, пользуясь компактной схемой метода Гаусса. Однако матрицы R_k до стабилизации процесса вычислять нет необходимости, так как для решения поставленной задачи нужна лишь предельная матрица R .

Это позволяет ограничиться лишь вычислением матриц C_k , для чего можно использовать, например, схему единственного деления (с исключением по столбцам), применявшуюся в § 17 для вычисления определителя. Та же схема доставит нам диагональные элементы матрицы R_k , так что приближенное определение собственных значений матрицы A не потребует дополнительных вычислений. Если же нужно определить и собственные векторы, то на последнем шаге процесса необходимо вычислить всю матрицу R_k , которая дает приближенное значение для предельной матрицы R .

Треугольный степенной метод является самонаправляющимся процессом, ибо каждое приближение C_k можно считать начальным приближением.

В § 57 мы упомянули о связи ступенчатого степенного метода с треугольным степенным методом. Теперь можно сказать об этом несколько подробнее. Именно, r -ступенчатый степенный метод можно излагать так. Исходя из прямоугольной матрицы с r столбцами $C_0^{(r)}$, надо образовать прямоугольную матрицу $A C_0^{(r)}$, которую затем представлять в виде произведения левой ступенчатой прямоугольной матрицы $C_1^{(r)}$ с единичной главной диагональю и правой треугольной матрицы $R_1^{(r)}$ r -го порядка. Далее процесс повторяется. Легко видеть, что если матрицу $C_0^{(r)}$ каким-либо образом дополнить до квадратной матрицы C_0 и применить к ней треугольный степенной метод, то матрицы $C_k^{(r)}$ будут совпадать с матрицами, составленными из первых r столбцов матриц C_k , а матрицы $R_k^{(r)}$ будут левыми верхними клетками r -го порядка матриц R_k .

Поэтому суждения о сходимости треугольного степенного метода переносятся без изменения на r -ступенчатый метод.

Треугольный степенной метод допускает модификацию со сдвигами. При надлежащем выборе сдвигов можно добиться квадратичной сходимости поочередно к каждому собственному значению.

Обоснование такой модификации заключается в следующем. Пусть

$$\varphi_k(t) = (t - t_1) \dots (t - t_k) \quad (15)$$

последовательность полиномов, таких, что каждый последующий полином получается из предыдущего умножением на линейный двучлен. Предполагается, что $t_k \rightarrow \tau$ и собственные значения матрицы A в некоторой нумерации удовлетворяют условиям

$$|\lambda_1 - \tau| > |\lambda_2 - \tau| > \dots > |\lambda_n - \tau|.$$

Тогда, как мы видели в § 77,

$$\frac{\varphi_k(\lambda_i)}{\varphi_k(\lambda_{i-1})} \rightarrow 0 \quad \text{при } k \rightarrow \infty.$$

Рассмотрим последовательность матриц

$$A^{(k)} = K \varphi_k(A) M, \quad (16)$$

где K и M некоторые фиксированные матрицы. Нетрудно провести оценки миноров любого порядка, составленных из элементов матриц $A^{(k)}$. Именно, любой минор порядка i равен

$$\hat{Q} \hat{P} \varphi_k(\lambda_1) \dots \varphi_k(\lambda_i) \left[1 + O\left(\frac{\varphi_k(\lambda_{i+1})}{\varphi_k(\lambda_i)}\right)\right],$$

где \hat{Q} и \hat{P} некоторые числа, зависящие от расположения минора внутри матрицы A^k . В частности, для верхних главных миноров имеем

$$\Delta_i^{(k)} = Q_{ii} P_{ii} \varphi_k(\lambda_1) \dots \varphi_k(\lambda_i) \left[1 + O\left(\frac{\varphi_k(\lambda_{i+1})}{\varphi_k(\lambda_i)}\right) \right], \quad (17)$$

причем Q_{ii} и P_{ii} имеют тот же смысл, что и в предыдущем параграфе. Вывод этих асимптотических формул ничем не отличается от только что сделанного расчета для случая $\varphi_k(t) = t^k$.

Представляя матрицы $A^{(k)}$ в виде

$$A^{(k)} = C_k D_k B_k,$$

где матрицы C_k , D_k и B_k имеют прежнее строение, получим, очевидно, что

$$\lim_{k \rightarrow \infty} \frac{d_i^{(k)}}{d_i^{(k-1)}} = \lambda_i - \tau, \quad (18)$$

а матрицы C_k и B_k стремятся при $k \rightarrow \infty$ к некоторым предельным матрицам.

Вычислительная схема треугольного степенного метода со сдвигом такова. Берется произвольная матрица C_0 , составляется матрица $(A - t_1 E) C_0$, которая затем раскладывается в произведение левой треугольной матрицы C_1 с единичной главной диагональю и правой треугольной матрицей R_1 . Далее, составляется матрица $(A - t_2 E) C_1$ и раскладывается в произведение треугольных матриц C_2 и R_2 прежнего строения. Процесс продолжается. Ясно, что

$$\varphi_k(A) C_0 = C_k R_k R_{k-1} \dots R_1. \quad (19)$$

Отсюда следует, что существуют предельные матрицы

$$C = \lim_{k \rightarrow \infty} C_k \text{ и } R = \lim_{k \rightarrow \infty} R_k.$$

и что

$$R_k = D_k B_k B_{k-1}^{-1} D_{k-1}^{-1}, \quad (20)$$

так что диагональные элементы $r_i^{(k)}$ матриц R_k совпадают с диагональными элементами матриц $D_k D_{k-1}^{-1}$, которые, как мы видели, стремятся к числам $\lambda_i - \tau$. Точнее,

$$r_i^{(k)} = (\lambda_i - t_k) \left[1 + O\left(\frac{\varphi_k(\lambda_i)}{\varphi_k(\lambda_{i-1})}\right) + O\left(\frac{\varphi_k(\lambda_{i+1})}{\varphi_k(\lambda_i)}\right) \right] \quad (21)$$

при $i = 1, 2, \dots, n-1$.

Для последнего диагонального элемента $r_n^{(k)}$ асимптотическая формула будет проще. Именно,

$$r_n^{(k)} = (\lambda_n - t_k) \left[1 + O\left(\frac{\varphi_k(\lambda_n)}{\varphi_k(\lambda_{n-1})}\right) \right]. \quad (22)$$

Для обеспечения квадратичной сходимости следует за t_{k+1} принимать $t_k + r_n^{(k)}$ до тех пор, пока, при некотором $k = k_1$, число $r_n^{(k)}$ не станет равным нулю с требуемой степенью точности. Затем нужно вычеркнуть из матрицы C_{k_1} последний столбец и перейти к ступенчатому алгорифму при числе столбцов, равном $n - 1$, принимая за сдвиги числа $t_{k+1} = t_k + r_{n-1}^{(k)}$, $k > k_1$. После того как $r_{n-1}^{(k)}$ окажется практически равным нулю, из матрицы C_{k_1} вычеркивается последний столбец и процесс продолжается как ступенчатый с $n - 2$ столбцами и т. д.

Приближенными значениями для собственных значений будут числа $t_{k_1}, t_{k_2}, \dots, t_{k_n}$.

§ 79. LR-алгорифм

LR-алгорифм (Рутисхаузер [5]) заключается в следующем. Матрица A раскладывается в произведение двух треугольных матриц (левой и правой)

$$A = L_1 R_1, \quad (1)$$

причем левая матрица берется с единичными диагональными элементами. Далее составляется матрица $R_1 L_1$ и для нее строится аналогичное разложение

$$R_1 L_1 = L_2 R_2. \quad (2)$$

Затем процесс повторяется. Таким образом, в результате процесса строятся две последовательности матриц L_1, L_2, \dots и R_1, R_2, \dots связанных соотношениями

$$\begin{aligned} A &= L_1 R_1 \\ R_1 L_1 &= L_2 R_2 \\ &\cdot \cdot \cdot \cdot \cdot \\ R_{k-1} L_{k-1} &= L_k R_k \\ &\cdot \cdot \cdot \cdot \cdot \end{aligned} \quad (3)$$

Установим связь между *LR*-алгорифмом и треугольным степенным методом при $C_0 = E$. С этой целью положим

$$L_1 L_2 \cdots L_k = C_k. \quad (4)$$

Ясно, что C_k есть левая треугольная матрица с единичной главной диагональю. Докажем теперь, что

$$A C_{k-1} = C_k R_k. \quad (5)$$

Для $k = 1$ имеем

$$A C_0 = A = L_1 R_1 = C_1 R_1.$$

Допустим, что равенство (5) справедливо для индексов, меньших k . И докажем, что оно верно и для индекса k . Имеем

$$AC_{k-1} = AC_{k-2}L_{k-1} = C_{k-1}R_{k-1}L_{k-1} = C_{k-1}L_kR_k = C_kR_k.$$

Тем самым равенство (5) доказано.

Из равенства (5) следует, что матрицы C_k и L_k LR-алгорифма совпадают с одноименными матрицами треугольного степенного метода при $C_0 = E$. Тем самым доказано, что в условиях сходимости треугольного степенного метода диагональные элементы матриц R_k сходятся к собственным значениям матрицы A .

Алгорифм LR по своей вычислительной схеме несколько проще, чем степенной треугольный метод. Однако он не является самоисправляющимся алгорифмом. Кроме того, он менее приспособлен к определению собственных векторов матрицы, ибо для решения этой задачи требуется восстановить матрицу C_k , что сводится к перемножению большого числа треугольных матриц и может сопровождаться нарастанием ошибок округления.

Следует отметить связь алгорифма LR с алгорифмом QD. Именно, соотношения

$$\sigma_i^{(k+1)} + \rho_i^{(k+1)} = \sigma_{i+1}^{(k)} + \rho_i^{(k)}$$

$$\sigma_i^{(k+1)} \rho_{i-1}^{(k+1)} = \sigma_i^{(k)} \rho_i^{(k)},$$

лежащие в основе QD алгорифма, равносильны следующим матричным равенствам

$$\begin{bmatrix} \rho_0^{(k)} & 1 & & 0 \\ \rho_1^{(k)} & 1 & & \\ \vdots & \ddots & & \\ & & 1 & \\ 0 & & \rho_{n-1}^{(k)} & \end{bmatrix} \begin{bmatrix} 1 & & & 0 \\ \sigma_1^{(k)} & 1 & & \\ \vdots & \ddots & & \\ & & 1 & \\ 0 & & \sigma_{n-1}^{(k)} & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & & & 0 \\ \sigma_1^{(k+1)} & 1 & & \\ \vdots & \ddots & & \\ & & 1 & \\ 0 & & \sigma_{n-1}^{(k+1)} & 1 \end{bmatrix} \begin{bmatrix} \rho_0^{(k+1)} & 1 & & 0 \\ \rho_1^{(k+1)} & 1 & & \\ \vdots & \ddots & & \\ & & 1 & \\ 0 & & \rho_{n-1}^{(k+1)} & \end{bmatrix}.$$

так что алгорифм *QD* можно рассматривать как *LR*-алгорифм, примененный к некоторой начальной матрице

$$J = L_1 R_1 = \left[\begin{array}{ccc|cc} 1 & & 0 & p_0^{(1)} & 1 & 0 \\ \sigma_1^{(1)} & 1 & & p_1^{(1)} & 1 & \\ \vdots & \ddots & & \vdots & \ddots & \\ 0 & & \sigma_{n-1}^{(1)} & 1 & 0 & p_{n-1}^{(1)} \end{array} \right] = \\ = \left[\begin{array}{ccc|cc} \alpha_0 & 1 & & & \\ \beta_1 & \alpha_1 & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ & & & & 1 \\ & & & & \beta_{n-1} \alpha_{n-1} \end{array} \right].$$

где $\alpha_i = p_i^{(1)} + \sigma_i^{(1)}$, $\beta_i = p_{i-1}^{(1)} \sigma_i^{(1)}$.

Напомним, что в алгорифме *QD* в качестве исходных значений $p_i^{(1)}$ и $\sigma_i^{(1)}$ берутся коэффициенты двучленных соотношений биортогонального алгорифма. Тем самым числа α_i и β_i , определяющие матрицу *J*, являются коэффициентами трехчленных соотношений биортогонального алгорифма. Как мы видели, матрица *J'* получается из матрицы *A* преобразованием подобия, вызываемым переходом к базису, состоящему из векторов p_0, \dots, p_{n-1} . Поэтому собственные значения матрицы *J'*, а следовательно, и матрицы *J* совпадают с собственными значениями матрицы *A*.

Применение алгорифма *LR* к матрицам общего вида требует очень большого числа вычислительных операций. Число операций значительно сокращается, если исходная матрица является ленточной, т. е. такой, для элементов a_{ij} которой выполняются равенства $a_{ij} = 0$ при $|i - j| > m$, где m некоторое число, значительно меньшее, чем порядок n матрицы. Иными словами, ленточная матрица имеет вид

$$\left[\begin{array}{cccc|cc} & & & 0 & & & \\ & & & & & & \\ & & & & & & \\ 0 & & & & & & \end{array} \right].$$

Сокращение числа операций в этом случае происходит за счет того, что все последующие матрицы $L_k R_k$ будут оставаться ленточными того же строения.

LR-алгоритм допускает модификацию со сдвигами, равносильную соответствующей модификации степенного треугольного метода при $C_0 = E$. В этой модификации процесс происходит по следующему предписанию:

$$\begin{aligned} (A - t_1 E) &= L_1 R_1 \\ R_1 L_1 - (t_2 - t_1) E &= L_2 R_2 \\ \vdots &\quad \vdots \\ R_{k-1} L_{k-1} - (t_k - t_{k-1}) E &= L_k R_k \end{aligned} \tag{6}$$

где L_k и R_k матрицы прежнего строения.

Обозначим

$$L_1 L_2 \dots L_k = c_k$$

и убедимся в том, что

$$(A - t_k E) C_{k+1} = C_k R_k.$$

Для $k = 1$ это верно при $C_0 = E$. Пусть утверждение верно для индексов, меньших k . Докажем, что оно верно и для индекса k . Действительно,

$$\begin{aligned}
 (A - t_k E) C_{k-1} &= (A - t_{k-1} E) C_{k-1} - (t_k - t_{k-1}) C_{k-1} = \\
 &= (A - t_{k-1} E) C_{k-2} L_{k-1} - (t_k - t_{k-1}) C_{k-1} = \\
 &= C_{k-1} R_{k-1} L_{k-1} - (t_k - t_{k-1}) C_{k-1} = \\
 &= C_{k-1} (R_{k-1} L_{k-1} - (t_k - t_{k-1}) E) = \\
 &= C_{k-1} L_k R_k = C_k R_k.
 \end{aligned}$$

Тем самым мы убедились в том, что матрицы C_k и R_k совпадают с одноименными матрицами треугольного степенного метода со сдвигами t_1, t_2, \dots

Так же как и в § 78, можно подобрать сдвиги так, чтобы соответствующий метод LR со сдвигами имел квадратичную сходимость. Именно, при этом следует брать

$$t_{k+1} - t_k = r_n^{(k)},$$

где $r_n^{(k)}$ — последний диагональный элемент матрицы R_k .

§ 80. АР-алгорифм

Вычислительная схема АР-алгорифма¹⁾ близка к вычислительной схеме LR-алгорифма, но несколько более трудоемка на каждом шагу. Сходимость же процесса, в условиях сходимости треугольного ступенчатого метода, является квадратичной.

¹⁾ Рутисхайзер и Бауэр [1].

Алгорифм позволяет последовательно вычислять матрицы C_k треугольного степенного метода с номерами k , являющимися степенями двойки (при $C_0 = E$).

Положим

$$A^{2^m} = \Lambda_m \Sigma_m, \quad (1)$$

где Λ_m — левая треугольная матрица с единичной главной диагональю, Σ_m — правая треугольная матрица. Ясно, что $\Lambda_m = C_{2^m}$ в обозначениях треугольного степенного метода при $C_0 = E$.

Возводя равенство (1) в квадрат, получим

$$A^{2^{m+1}} = \Lambda_m \Sigma_m \Lambda_m \Sigma_m. \quad (2)$$

Разложим теперь матрицу $\Sigma_m \Lambda_m$ в произведение левой треугольной матрицы \tilde{L}_m с единичной главной диагональю и правой треугольной матрицы \tilde{R}_m

$$\Sigma_m \Lambda_m = \tilde{L}_m \tilde{R}_m. \quad (3)$$

Тогда получим

$$A^{2^{m+1}} = \Lambda_m \tilde{L}_m \tilde{R}_m \Sigma_m,$$

откуда

$$\begin{aligned} \Lambda_{m+1} &= \Lambda_m \tilde{L}_m \\ \Sigma_{m+1} &= \tilde{R}_m \Sigma_m. \end{aligned} \quad (4)$$

Эти формулы позволяют последовательно вычислять матрицы Λ_m и Σ_m при $m = 0, 1, 2, \dots$, начиная с матриц Λ_0 и Σ_0 , которые находятся разложением исходной матрицы A в произведение двух треугольных.

Матрицы Λ_m , в условиях сходимости треугольного степенного метода, будут сходиться к предельной матрице C , причем сходимость будет порядка $O\left(\left|\frac{\lambda_i}{\lambda_{i-1}}\right|^{2^m}\right)$, т. е. сходимость будет квадратичной.

Найдя матрицу C , строим затем матрицу $R = C^{-1}AC$, что не представляет труда, ибо C треугольная матрица с единичной главной диагональю. Матрица R будет правой треугольной матрицей, той самой, которая появлялась как предельная для матриц R_k в треугольном степенном методе. Собственные значения матрицы A равны диагональным элементам матрицы R , собственные же векторы определяются при помощи матриц C и R , как в треугольном степенном методе.

Недостатком только что описанной вычислительной схемы является стремительный рост (или стремительное исчезновение) элементов матриц Σ_m с возрастанием m . Это явление частично устраняется посредством нормировки на каждом шагу правых треугольных матриц Σ_m к единичной главной диагонали. Под АР-алгорифмом и

подразумевают процесс с такой нормировкой. Выведем расчетные формулы алгорифма АР.

Положим

$$\Sigma_m = \Delta_m P_m, \quad (5)$$

где Δ_m диагональная матрица, P_m правая треугольная матрица с единичной главной диагональю. Тогда

$$A^{2^m} = \Lambda_m \Delta_m P_m.$$

Возводя в квадрат это равенство, получим

$$A^{2^{m+1}} = \Lambda_m \Delta_m P_m \Lambda_m \Delta_m P_m.$$

Разложим теперь матрицу $P_m \Lambda_m$ в произведение левой треугольной с единичной главной диагональю L_m , диагональной D_m и правой треугольной с единичной главной диагональю R_m

$$P_m \Lambda_m = L_m D_m R_m. \quad (6)$$

Тогда

$$A^{2^{m+1}} = \Lambda_m \Delta_m L_m D_m R_m \Delta_m P_m = \Lambda_m (\Delta_m L_m \Delta_m^{-1}) \Delta_m D_m \Delta_m (\Delta_m^{-1} R_m \Delta_m) P_m.$$

Ясно, что $\Delta_m L_m \Delta_m^{-1}$ есть левая треугольная матрица с единичной главной диагональю, матрица $\Delta_m^{-1} R_m \Delta_m$ есть правая треугольная матрица с единичной главной диагональю, а матрица $\Delta_m D_m \Delta_m = \Delta_m^2 D_m$ диагональная.

Поэтому

$$\begin{aligned} \Lambda_{m+1} &= \Lambda_m (\Delta_m L_m \Delta_m^{-1}) \\ P_{m+1} &= (\Delta_m^{-1} R_m \Delta_m) P_m \\ \Delta_{m+1} &= \Delta_m^2 D_m, \end{aligned} \quad (7)$$

где

$$P_m \Lambda_m = L_m D_m P_m.$$

Последние формулы и являются предписанием для алгорифма АР. Начало процесса определяется разложением данной матрицы A в произведение $\Lambda_0 \Delta_0 P_0$.

Заметим, что каждый из множителей $\Delta_m L_m \Delta_m^{-1}$ и $\Delta_m^{-1} R_m \Delta_m$ стремится к единичной матрице.

Алгорифм АР заканчивается стабилизацией матриц Λ_m . По предельной матрице C определяется матрица $R = C^{-1} A C$. Ее диагональные элементы дают собственные значения матрицы A , собственные же векторы матрицы A суть

$$U_1 = CV_1, \dots, U_n = CV_n,$$

где V_1, \dots, V_n собственные векторы матрицы R .

§ 81. Итерационные процессы, основанные на применении вращений

Вращения, уже рассматривавшиеся нами в § 51 в связи с преобразованием симметричной матрицы к трехдиагональной, можно использовать для построения итерационных процессов, решающих полную проблему собственных значений.

Эти процессы для симметрических матриц состоят в цепочке преобразований подобия, в результате которых в пределе получается диагональная матрица, так что ее собственные значения определяются непосредственно. Впервые такой процесс был предложен Якоби [1] в 1846 г. Однако практическое применение его стало возможным лишь с развитием быстродействующих счетных устройств. В настоящее время имеется целый ряд модификаций метода Якоби.

Элементарный шаг каждого якобиева процесса заключается в преобразовании подобия посредством матрицы

$$T_{ij} = \begin{vmatrix} 1 & & & & & \\ & c & \dots & -s & & \dots & i \\ & \vdots & & \ddots & & & \\ & & 1 & & & & \\ & & \vdots & & & & \\ & s & \dots & c & & \dots & j \\ & & & & & & \\ & & & & & & 1 \end{vmatrix} \quad (i < j)$$

при $c^2 + s^2 = 1$. Как мы видели в § 51, матрица T_{ij} есть матрица вращения плоскости, натянутой на i -й и j -й координатные векторы, на угол θ такой, что $\cos \theta = c$, $\sin \theta = s$. Матрица T_{ij} ортогональна, так что $T_{ij}' = T_{ij}^{-1}$.

Процесс в целом состоит в построении последовательности матриц $A = A^{(0)}, A^{(1)}, A^{(2)}, \dots$, каждой из которых получается из предыдущей при помощи элементарного шага. Эти элементарные шаги должны быть подобраны так, чтобы матрицы $A^{(k)}$ безгранично приближались к диагональной матрице при $k \rightarrow \infty$.

Близость симметричной матрицы A к диагональной мы будем характеризовать числом $f^2(A)$, равным сумме квадратов всех недиагональ-

ных элементов матрицы A . Эта близость может быть также охарактеризована любой нормой матрицы $A - D$, где D диагональная матрица, составленная из диагональных элементов матрицы A .

Якобиев процесс будем называть релаксационным или монотонным, если $t^2(A^{(k)})$ уменьшается на каждом шагу.

Выясним, как надо выбирать матрицу T_{ij} при фиксированных индексах i и j , чтобы $t^2(T'_{ij}AT_{ij})$ было меньше, чем $t^2(A)$.

Обозначим

$$T'_{ij}AT_{ij} = C = (c_{kl}).$$

Напомним (§ 51), что элементы матрицы C совпадают с элементами матрицы A за исключением элементов, находящихся в строках с номерами i и j или в столбцах с номерами i и j . В частности, $c_{kk} = a_{kk}$ при $k \neq i, k \neq j$.

Пусть

$$n^2(A) = \sum_{i,j=1}^n a_{ij}^2 = \operatorname{Sp} A'A = \operatorname{Sp} A^2.$$

Легко видеть, что $n^2(A) = n^2(C)$. Действительно,

$$n^2(C) = \operatorname{Sp} C^2 = \operatorname{Sp} (T_{ij}^{-1}AT_{ij})^2 = \operatorname{Sp} A^2 = n^2(A).$$

Далее пусть

$$\tilde{C} = \begin{bmatrix} c_{ii} & c_{ij} \\ c_{ji} & c_{jj} \end{bmatrix}, \quad \tilde{A} = \begin{bmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{bmatrix}.$$

Тогда $\tilde{C} = \tilde{T}'\tilde{A}\tilde{T}$, где $\tilde{T} = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$, и следовательно, $n^2(\tilde{C}) = n^2(\tilde{A})$.

Ясно, что

$$\begin{aligned} t^2(C) - t^2(A) &= n^2(C) - \sum_{k=1}^n c_{kk}^2 - n^2(A) + \\ &\quad + \sum_{k=1}^n a_{kk}^2 = a_{ii}^2 + a_{jj}^2 - c_{ii}^2 - c_{jj}^2, \end{aligned}$$

ибо $n^2(C) = n^2(A)$ и $c_{kk} = a_{kk}$ при $k \neq i, k \neq j$. Следовательно,

$$t^2(C) - t^2(A) = n^2(\tilde{A}) - 2a_{ij}^2 - n^2(\tilde{C}) + 2c_{ij}^2 = 2(c_{ij}^2 - a_{ij}^2),$$

так как $n^2(\tilde{A}) = n^2(\tilde{C})$.

Таким образом, для релаксационности процесса на данном шаге нужно, чтобы $|c_{ij}| < |a_{ij}|$. Этого можно добиться, только если $a_{ij} \neq 0$.

Легко проверить, что

$$\begin{aligned} c_{ij} &= a_{ij}(c^2 - s^2) + (a_{jj} - a_{ii})cs = a_{ij} \cos 2\theta + \frac{1}{2}(a_{jj} - a_{ii}) \sin 2\theta = \\ &= a_{ij} \sin(2\theta - 2\theta_0), \end{aligned}$$

где $a_{ij} = \pm \sqrt{a_{ij}^2 + \frac{1}{4}(a_{jj} - a_{ii})^2}$, $\operatorname{tg} 2\theta_0 = \frac{2a_{ij}}{a_{ii} - a_{jj}}$. Угол θ_0 определен с точностью до целого кратного $\frac{\pi}{2}$. Мы будем считать, что $-\frac{\pi}{4} < \theta_0 \leq \frac{\pi}{4}$. При таком условии выбора θ_0

$$a_{ij} = \operatorname{sign}(a_{jj} - a_{ii}) \sqrt{a_{ij}^2 + \frac{1}{4}(a_{jj} - a_{ii})^2}.$$

На рис. 12 приведен график c_{ij} как функции от θ , в предположении $a_{ij} > 0$ и $\theta_0 > 0$. Ясно, что при других знаках a_{ij} и θ_0 соответствующие графики для c_{ij} будут иметь такой же вид с точностью

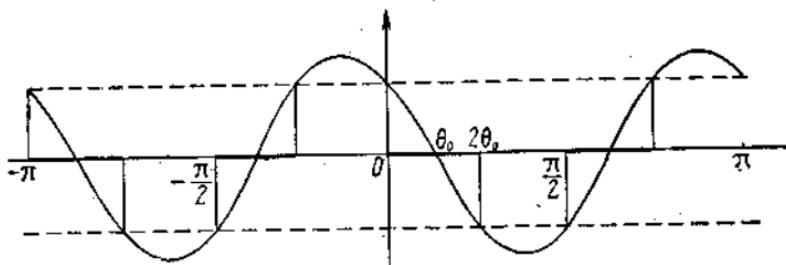


Рис. 12.

до симметрии относительно оси абсцисс (при изменении знака a_{ij}) или относительно оси ординат (при изменении знака θ_0).

Из графиков можно заключить, что значения θ , обеспечивающие неравенство $t^2(c) < t^2(A)$, заполняют четыре интервала на промежутке $(-\pi, \pi)$, каждый из которых примыкает одним из концов к точкам $0, \frac{\pi}{2}, -\frac{\pi}{2}, -\pi$ (или $+\pi$). В дальнейшем мы будем считать, что угол вращения θ на каждом шагу якобиева процесса берется из промежутка $(0, 2\theta_0)$ или $(2\theta_0, 0)$, примыкающего к точке 0, так что $\theta = q\theta_0$, где $0 < q < 2$ и $|\theta_0| \leq \frac{\pi}{4}$.

Наибольшее понижение величины $t^2(A)$ за один шаг получится при $c_{ij} = 0$, что будет обеспечено при $\theta = \theta_0$. По аналогии с релаксационными процессами для решения систем линейных уравнений мы будем называть якобиевы процессы, в которых на каждом шагу $\theta = \theta_0$, процессами с полной релаксацией. Аналогично процессы, в которых $0 < q < 1$, мы будем называть процессами с нижней релаксацией, процессы, в которых $1 < q < 2$, — процессами с верхней релаксацией.

Выбор пар индексов $(i_1, j_1), (i_2, j_2), \dots, (i_k, j_k)$ для последовательных шагов процесса можно осуществить априорно или с управлением по ходу процесса. Наиболее естественным для априорного выбора является циклическое чередование пар, занумерованных тем или другим способом.

Перейдем к рассмотрению нескольких исследованных в настоящее время якобиевых процессов.

1. Классический метод Якоби. Матрица вращения T_{ij} выбирается на каждом шаге так, чтобы элемент i -й строки и j -го столбца преобразованной матрицы стал нулем. При этом пары индексов (i, j) выбираются на каждом шаге так, чтобы аннулировался наибольший по абсолютной величине недиагональный элемент матрицы, полученной в результате предыдущего шага процесса. Таким образом, метод Якоби является методом полной релаксации с управлением.

Легко доказывается сходимость метода. Действительно, пусть $A^{(k)}$ — матрица, полученная на k -м шагу процесса, $a_{i_k j_k}^{(k)}$ ее наибольший по модулю недиагональный элемент. Тогда

$$t^2(A^{(k+1)}) = t^2(A^{(k)}) - 2(a_{i_k j_k}^{(k)})^2.$$

С другой стороны, $t^2(A^{(k)}) \leq n(n-1)(a_{i_k j_k}^{(k)})^2$, откуда $(a_{i_k j_k}^{(k)})^2 \geq \frac{1}{n(n-1)} t^2(A^{(k)})$. Следовательно,

$$t^2(A^{(k+1)}) \leq t^2(A^{(k)}) \left(1 - \frac{2}{n(n-1)}\right),$$

и потому

$$t^2(A^{(k+1)}) \leq t^2(A) \left(1 - \frac{2}{n(n-1)}\right)^{k+1}.$$

Таким образом, $t(A^{(k)}) \rightarrow 0$ при $k \rightarrow \infty$, что доказывает сходимость процесса Якоби.

Дадим расчетные формулы одного шага метода, обозначая снова исходную матрицу шага через A и полагая $C = T'_{ij} A T_{ij}$.

Как уже сказано выше, пара (i, j) выбирается так, чтобы a_{ij} был наибольшим по модулю элементом матрицы A . Далее

$$\begin{aligned} c_{kl} &= a_{kl} && \text{при } k \neq i, k \neq j, l \neq i, l \neq j \\ c_{ki} &= c_{ik} = ca_{ki} + sa_{kj} && \\ c_{kj} &= c_{jk} = -sa_{ki} + ca_{kj} && \text{при } k \neq i, k \neq j. \end{aligned} \tag{1}$$

Наконец,

$$\begin{aligned} c_{ii} &= c^2 a_{ii} + 2csa_{ij} + s^2 a_{jj} \\ c_{jj} &= s^2 a_{ii} - 2csa_{ij} + c^2 a_{jj} \\ c_{ij} &= (c^2 - s^2) a_{ij} + cs(a_{jj} - a_{ii}) = 0. \end{aligned} \tag{2}$$

Числа c и s определяются по формулам

$$c = \cos \theta, \quad s = \sin \theta, \tag{3}$$

где

$$\operatorname{tg} 2\theta = \frac{2a_{ij}}{a_{ii} - a_{jj}}, \quad |\theta| \leq \frac{\pi}{4}. \tag{4}$$

Формулы (3) и (4) можно заменить на

$$c = \sqrt{\frac{1}{2} \left(1 + \frac{|a_{ii} - a_{jj}|}{d} \right)} \quad (5)$$

$$s = \operatorname{sign}(a_{ij}(a_{ii} - a_{jj})) \sqrt{\frac{1}{2} \left(1 - \frac{|a_{ii} - a_{jj}|}{d} \right)},$$

где

$$d = \sqrt{(a_{ii} - a_{jj})^2 + 4a_{ij}^2}.$$

Формулы (2) можно заменить на

$$c_{ii} = \frac{a_{ii} - a_{jj}}{2} + \operatorname{sign}(a_{ii} - a_{jj}) \frac{d}{2}$$

$$c_{jj} = \frac{a_{ii} + a_{jj}}{2} - \operatorname{sign}(a_{ii} - a_{jj}) \frac{d}{2}$$

$$c_{ij} = c_{ji} = 0.$$

Построение матрицы C , после того как числа c, s вычислены, можно осуществлять и непосредственно по формулам § 51.

Если матрица не имеет кратных собственных значений, то, аннулируя внедиагональные элементы с точностью до ϵ , мы получим, что диагональные элементы приближаются к собственным значениям уже с точностью до ϵ^2 . Действительно, если недиагональные элементы симметричной матрицы

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

являются малыми числами, то для ее собственных значений справедлива приближенная формула

$$\lambda_i \approx a_{ii} + \sum_{\substack{j=1 \\ j \neq i}}^n \frac{a_{ij}^2}{a_{ii} - a_{jj}}, \quad (6)$$

верная с точностью до малых 3-го порядка,

В самом деле, пусть $a_{ij} = \epsilon a_{ij}$, где ϵ — малое число. Положим

$$\lambda_1 = a_{11} + k\epsilon^2.$$

Тогда для определения k получим уравнение

$$(1) \quad \begin{vmatrix} -k\epsilon^2 & \epsilon a_{12} & \dots & \epsilon a_{1n} \\ \epsilon a_{21} & a_{22} - a_{11} - k\epsilon^2 & \dots & \epsilon a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon a_{n1} & \epsilon a_{n2} & \dots & a_{nn} - a_{11} - k\epsilon^2 \end{vmatrix} = 0.$$

Вынесем ϵ из первой строки и первого столбца определителя и заменим ϵ нулем в получившемся уравнении. Получим

$$\left| \begin{array}{ccccc} -k & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - a_{11} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ a_{n1} & 0 & \dots & a_{nn} - a_{11} \end{array} \right| \approx 0,$$

откуда

$$k \approx \frac{a_{12}a_{21}}{a_{11} - a_{22}} + \frac{a_{13}a_{31}}{a_{11} - a_{33}} + \dots + \frac{a_{1n}a_{n1}}{a_{11} - a_{nn}}$$

с точностью до малых порядка ϵ . Следовательно,

$$\begin{aligned} \lambda_1 = a_{11} + \frac{a_{12}a_{21}}{a_{11} - a_{22}} + \frac{a_{13}a_{31}}{a_{11} - a_{33}} + \dots + \frac{a_{1n}a_{n1}}{a_{11} - a_{nn}} + O(\epsilon^3) = \\ = a_{11} + \sum_{j=2}^n \frac{a_{1j}^2}{a_{11} - a_{jj}} + O(\epsilon^3). \end{aligned}$$

Аналогично

$$\lambda_i = a_{ii} + \sum_{\substack{j=1 \\ j \neq i}}^n \frac{a_{ij}^2}{a_{ii} - a_{jj}} + O(\epsilon^3).$$

Огрубляя последнюю формулу, получим

$$\lambda_i = a_{ii} + O(\epsilon^2).$$

Применим метод Якоби для определения собственных значений матрицы (4) § 51.

Первый шаг процесса заключается в преобразовании вращением при помощи матрицы T_{14} .

Вычисляя по формулам (4), получим

$$c = s = 0.70710678.$$

Далее

$$T'_{14} A T_{14} = \begin{bmatrix} 1.66 & 0.60811183 & 0.53740115 & 0 \\ 0.60811183 & 1 & 0.32 & 0.01414214 \\ 0.53740115 & 0.32 & 1 & -0.22627417 \\ 0 & 0.01414214 & -0.22627417 & 0.34 \end{bmatrix}.$$

Следующие шаги требуют преобразований матрицами T_{12} , T_{13} , T_{34} , T_{14} , T_{13} , T_{24} , T_{23} . В результате получилась матрица

$$\begin{bmatrix} 2.3227487 & -0.00048637 & 0.00001483 & 0.00004994 \\ -0.00048637 & 0.63828393 & 0 & 0.00004930 \\ 0.00001483 & 0 & 0.79670201 & 0.00161630 \\ 0.00004994 & 0.00004930 & 0.00161630 & 0.24226544 \end{bmatrix}.$$

Ее недиагональные элементы уже настолько малы, что для определения собственных значений можно применить формулу (6). Вычисляя, получим

$$\lambda_1 = 2.3227487 + 0.00000014 = 2.32274884$$

$$\lambda_2 = 0.63828393 - 0.00000014 = 0.63828379$$

$$\lambda_3 = 0.79670201 + 0.00000471 = 0.79670672$$

$$\lambda_4 = 0.24226544 - 0.00000471 = 0.24226073.$$

Результаты получились верными уже с точностью до $4 \cdot 10^{-8}$ для всех собственных значений.

Нетрудно показать, что в случае отсутствия кратных собственных значений у исходной матрицы, метод Якоби обладает квадратичной сходимостью.

Допустим, что процесс проведен настолько далеко, что все недиагональные элементы матрицы стали по модулю меньше малого числа ϵ . Тогда, как мы видели выше, диагональные элементы приближаются к собственным значениям с точностью до ϵ^2 . Поэтому на каждом шагу процесса Якоби угол поворота θ будет иметь порядок ϵ , ибо $\operatorname{tg} 2\theta = \frac{2a_{ij}}{a_{ii} - a_{jj}}$, числитель имеет порядок ϵ , а знаменатель, равный $\lambda_i - \lambda_j + O(\epsilon^2)$, ограничен снизу.

Все меняющиеся на этом шагу недиагональные элементы, кроме a_{ij} , изменятся на величину порядка ϵ^2 . Действительно,

$$c_{ki} = c_{ik} = a_{ki} \cos \theta + a_{kj} \sin \theta,$$

откуда

$$c_{ki} - a_{ki} = a_{ki} (\cos \theta - 1) + a_{kj} \sin \theta = O(\epsilon^2).$$

Аналогично

$$c_{kj} - a_{kj} = O(\epsilon^2).$$

В частности, элемент, аннулированный на предыдущем шаге, станет, самое большое, величиной порядка ϵ^2 . Поэтому после, самое большое, $\frac{n(n-1)}{2}$ шагов процесса, все внедиагональные элементы станут величинами порядка ϵ^2 . Это и доказывает квадратичную сходимость процесса.

Подсчет констант в приведенных оценках осуществлен в работе Хенриси [1].

Скажем несколько слов о нахождении собственных векторов. Пусть процесс доведен до того, что матрица

$$A^{(k)} = \left[\prod_m T'_{i_m j_m} \right] A \left[\prod_m T_{i_m j_m} \right]$$

оказалась практически диагональной. Тогда столбцы матрицы $\prod T_{i_m j_m}$ и будут собственными векторами исходной матрицы A .

В случае, если собственные значения матрицы A попарно различны и внедиагональные элементы матрицы анулированы с точностью до малого числа ϵ , легко дать прием, позволяющий вычислять компоненты собственных векторов U_i с точностью до величины ϵ^2 .

Именно,

$$U_i = \prod T_{i_m j_m} V_i,$$

где

$$V_i = \left(\frac{a_{1i}}{a_{ii} - a_{11}}, \dots, \frac{a_{i-1i}}{a_{ii} - a_{i-1i-1}}, 1, \frac{a_{i+1i}}{a_{ii} - a_{i+1i+1}}, \dots, \frac{a_{ni}}{a_{ii} - a_{nn}} \right)'.$$

Здесь a_{ij} — суть элементы матрицы $A^{(k)}$, на которой мы остановили процесс.

В качестве примера определим собственный вектор матрицы, рассмотренной выше в связи с определением собственных значений, принадлежащий наименьшему собственному значению.

Имеем

$$\alpha_{44} - \alpha_{11} = -2.08048326, \quad \alpha_{44} - \alpha_{22} = -0.39601849,$$

$$\alpha_{44} - \alpha_{33} = -0.55445657,$$

так что

$$V_4 = (-0.00002400, -0.00012449, -0.00291521, 1)'.$$

Далее находим по формулам (8) § 51 последовательно векторы

$$T_{23}V_4, \quad T_{24}T_{23}V_4, \quad T_{13}T_{24}T_{23}V_4, \dots,$$

$$T_{14}T_{12}T_{13}T_{34}T_{14}T_{13}T_{24}T_{23}V_4 = U_4,$$

располагая их компоненты последовательно в графах I—IV табл. VIII. 7.

В графах V и VI приведены соответствующие коэффициенты c и s . В последней строке приведен собственный вектор U_4 , нормированный к единичной первой компоненте.

Классический метод Якоби на каждом шагу процесса требует выбора наибольшего внедиагонального элемента. Эта операция при выполнении на быстродействующих машинах требует значительной затраты машинного времени. Более удобными оказываются циклические якобиевы процессы и, в частности, циклические процессы с препятствиями.

Таблица VIII. 7

Определение собственного вектора методом Якоби

<i>i</i>	<i>j</i>	I	II	III	IV	V	VI
2	3	-0.00002400	-0.00003556	-0.00291765	1	0.99953524	0.03048475
2	4	-0.00002400	-0.02812808	-0.00291765	0.99960433	0.99960533	0.02809253
1	3	0.00004740	-0.02812808	-0.00291736	0.99960433	0.99970055	0.02447059
1	4	0.03967421	-0.02812808	-0.00291736	0.99881670	0.99921394	-0.03964253
3	4	0.03967421	-0.02812808	0.40999603	0.91079448	0.91066679	-0.41314164
1	3	-0.13986866	-0.02812808	0.38743715	0.91079448	0.90350564	0.42857618
1	2	-0.10581153	-0.09569928	0.38743715	0.91079448	0.85934868	0.51139013
1	4	-0.71884900	-0.09569928	0.38743715	0.56920890	0.70710678	0.70710678
	1		0.133128	-0.538969	-0.791834		

2. Циклические якобиевы процессы.¹⁾ При проведении циклического процесса выбирается определенная нумерация пар (*i*, *j*) и аннулирование внедиагональных элементов происходит по циклам, в течение каждого из которых аннулируются по очереди элементы a_{ij} в порядке выбранной нумерации пар индексов.

Элементарный шаг процесса ничем не отличается от элементарного шага классического метода Якоби, так что рабочими формулами остаются формулы, выведенные выше. Как указано в статье Грегори [3], сходимость метода установлена в неопубликованной еще работе Форсайта и Хенричи.

После того, как внедиагональные элементы станут достаточно малыми, сходимость процесса становится квадратичной, в чем легко убедиться теми же рассуждениями, которые были приведены для метода Якоби.

Наиболее естественной нумерацией пар является нумерация по строкам слева направо и сверху вниз, именно

$$(1, 2), (1, 3), \dots, (1, n), (2, 3), \dots, (2, n), \dots, (n-1, n)$$

или по столбцам сверху вниз и слева направо

$$(1, 2), (1, 3), (2, 3), (1, 4), (2, 4), (3, 4), \dots$$

$$\dots, (1, n), (2, n), \dots, (n-1, n).$$

3. Циклические процессы с преградами.²⁾ Промежуточное положение между классическим методом Якоби и циклическим занимает циклический метод с преградами. Недостатком циклического метода является то обстоятельство, что по ходу процесса приходится анну-

¹⁾ Грегори [3].

²⁾ Пол, Томпкинс [1].

лировать малые внедиагональные элементы, хотя в матрице еще присутствуют большие. Этот недостаток устраняется введением „преград“. Именно, вводится монотонно убывающая к нулю последовательность чисел $\alpha_1, \alpha_2, \dots$, и при последовательном аннулировании внедиагональных элементов пропускаются те шаги, при которых приходилось бы аннулировать элементы, меньшие чем α_1 . После того, как все внедиагональные элементы станут по модулю не больше α_1 , „преграда“ сдвигается — число α_1 заменяется на число α_2 и т. д.

Легко устанавливается, что процесс с преградами сходится.

Действительно, „преграда“ α_1 будет перейдена не более чем через $t^2(A)/\alpha_1^2$ элементарных шагов, так как за каждый шаг до преодоления преграды α_1 число $t^2(A)$ уменьшается не менее чем на α_1^2 . Аналогично, через конечное число шагов будут преодолеваться последующие преграды α_2, α_3 и т. д. Так как $\alpha_k \rightarrow 0$, то через достаточно большое число элементарных шагов все внедиагональные элементы матрицы станут сколь угодно малыми. Это и доказывает сходимость процесса.

Как и прежде, при попарно различных собственных значениях, сходимость процесса будет, начиная с некоторого места, квадратичной.

4. Процессы Якоби с рациональными формулами. Элементарный шаг процессов этой группы отличается от элементарного шага процессов с полной релаксацией выбором угла поворота вращения. Именно, вместо формулы $\operatorname{tg} 2\theta_0 = \frac{2a_{ij}}{a_{ii} - a_{jj}}$ берется формула

$$\operatorname{tg} \frac{\theta}{2} = \frac{a_{ij}}{2(a_{ii} - a_{jj})} = \alpha, \quad \text{если } \left| \frac{a_{ij}}{2(a_{ii} - a_{jj})} \right| < \sqrt{2} - 1 \approx 0.414,$$

$$\theta = \frac{\pi}{4}, \quad \text{если } \left| \frac{a_{ij}}{2(a_{ii} - a_{jj})} \right| \geq \sqrt{2} - 1.$$

При малых углах $\theta \approx \theta_0$, только θ всегда немного больше θ_0 . Нетрудно проверить, что $|\theta_0| < |\theta| < |2\theta_0|$, так что указанный выбор угла вращения обеспечивает верхнюю релаксацию.

Для вычисления чисел c и s получаем рациональные формулы

$$c = \frac{1 - \alpha^2}{1 + \alpha^2}, \quad s = \frac{2\alpha}{1 + \alpha^2} \quad (\alpha < \sqrt{2} - 1)$$

$$c = \frac{\sqrt{2}}{2}, \quad s = \pm \frac{\sqrt{2}}{2} \quad (\alpha \geq \sqrt{2} - 1).$$

Само преобразование вращения следует осуществлять по формулам § 51.

Выбор пар может осуществляться либо циклически, либо по индексам наибольшего по модулю внедиагонального элемента.

В литературе описаны и некоторые другие якобиевы процессы.

Описанные методы почти без изменения переносятся на диагонализацию эрмитовых матриц. Вместо элементарных вращений приходится производить трансформации унитарными матрицами вида

$$T_{ij} = \begin{bmatrix} -1 & & & & & \\ & c_1 & & & -s_1 & \\ & & 1 & & & \\ & & & \ddots & & \\ & & & & c_2 & \\ s_2 & & & & & \dots j \\ & & & & & \\ & & & & & 1 \end{bmatrix}$$

(унитарные вращения).

Условия унитарности

$$|c_1|^2 + |s_2|^2 = 1$$

$$|s_1|^2 + |c_2|^2 = 1$$

$$-c_1s_1 + s_2c_2 = 0$$

позволяют выбирать числа c_1, s_1, c_2 и s_2 в виде

$$c_1 = \cos \theta e^{i\alpha}, \quad s_1 = \sin \theta e^{i\beta},$$

$$s_2 = \sin \theta e^{i\gamma}, \quad c_2 = \cos \theta e^{i\delta}.$$

Здесь $\theta, \alpha, \beta, \gamma, \delta$ — вещественные числа, причем

$$\alpha - \beta - \gamma + \delta \equiv 0 \pmod{2\pi}.$$

Мы не будем здесь приводить соответствующих рабочих формул.

В случае произвольной матрицы преобразованиями подобия посредством вращений привести матрицу к диагональной форме уже невозможно.

Однако верна следующая теорема И. Шура. Для любой комплексной матрицы A существует унитарная матрица U такая, что $U^{-1}AU$ есть верхняя треугольная матрица.

Для доказательства рассмотрим матрицу A как матрицу оператора в некотором ортонормальном базисе и перейдем к новому ортонормальному базису, включающему один из собственных векторов

матрицы. В этом новом базисе оператору будет соответствовать матрица

$$U_1^{-1}AU_1 = \begin{bmatrix} \lambda_1 & b_{12} & \dots & b_{1n} \\ 0 & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & b_{n2} & \dots & b_{nn} \end{bmatrix}.$$

Применяя то же рассуждение к матрице $(n - 1)$ -го порядка

$$\begin{bmatrix} b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots \\ b_{n2} & \dots & b_{nn} \end{bmatrix},$$

получим

$$U_2^{-1}U_1^{-1}AU_1U_2 = \begin{bmatrix} -\lambda_1 & c_{12} & c_{13} & \dots & c_{1n} \\ 0 & \lambda_2 & c_{23} & \dots & c_{2n} \\ 0 & 0 & c_{33} & \dots & c_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & c_{n3} & \dots & c_{nn} \end{bmatrix}.$$

Продолжая процесс далее, принимая во внимание, что произведение унитарных матриц есть унитарная матрица, придем к требуемому результату.

Теорема Шура дает основание предполагать, что, подбирая подходящим образом последовательность унитарных матриц вращений, можно добиться преобразованиями подобия приведения данной матрицы к треугольному виду.

Два таких процесса описаны в работах Гринштадта [1] и Лоткина [4]. В первой из них предлагается аннулировать поочередно за счет подходящих унитарных вращений поддиагональные элементы. Доказательство сходимости процесса отсутствует, дается только указание на экспериментальную проверку метода. Во второй работе вращения подбираются так, чтобы сумма квадратов модулей поддиагональных элементов уменьшалась на каждом шагу. Процесс оказывается довольно сложным — на каждом шагу приходится решать вспомогательное кубическое уравнение.

Мы не будем описывать подробности этих процессов.

§ 82. Решение полной проблемы собственных значений при помощи спектрального анализа последовательных итераций

Этот метод, который принадлежит Ланцошу [7], заключается в следующем. Пусть матрица A симметрична¹⁾, и ее собственные

1) Рассуждения остаются в силе и в случае, если матрица A не симметрична, но имеет попарно различные вещественные собственные значения.

значения заключены в интервале (m, M) . Вводится матрица

$$B = \frac{M+m}{M-m} \left(E - \frac{2}{M+m} A \right), \quad (1)$$

собственные значения μ_1, \dots, μ_n , которой принадлежат уже промежутку $(-1, 1)$. В случае положительно-определенной матрицы A в качестве m можно взять, например, нуль.

Вычислим последовательность векторов $X_0, X_1 = BX_0$.

$$X_k = 2BX_{k-1} - X_{k-2} \quad (k = 2, 3, \dots, N).$$

Пусть

$$X_0 = U_1 + \dots + U_r, \quad (2)$$

где U_1, \dots, U_r собственные векторы матрицы B , а следовательно, и матрицы A .

Тогда

$$X_k = T_k(\mu_1)U_1 + \dots + T_k(\mu_r)U_r, \quad (3)$$

где $T_k(t) = \cos k \arccos t$.

Положим $\mu_i = \cos \theta_i$ при $0 < \theta_i < \pi$. Ясно, что $T_k(\mu_i) = \cos k \arccos \mu_i = \cos k \theta_i$, так что

$$X_k = \cos k \theta_1 U_1 + \dots + \cos k \theta_r U_r. \quad (4)$$

Отсюда и любая компонента вектора X_k , которую мы обозначим через x_k , будет иметь вид

$$x_k = a_1 \cos k \theta_1 + \dots + a_r \cos k \theta_r, \quad (5)$$

где a_1, \dots, a_r некоторые определенные, но заранее неизвестные нам числа (равные выбранным компонентам векторов U_1, \dots, U_r).

Ланцош предлагает определять углы $\theta_1, \dots, \theta_r$ (а следовательно, и собственные значения μ_1, \dots, μ_r), подвергая гармоническому анализу последовательность x_0, x_1, \dots, x_N (для контроля следует взять также и вторую последовательность, состоящую из других компонент итераций X_k). При этом попутно определяются и амплитуды a_1, \dots, a_r , которые являются выбранными компонентами собственных векторов U_1, \dots, U_r . После того как собственные значения определены, остальные компоненты собственных векторов определяются, как мы увидим ниже, без проведения полного гармонического анализа последовательности соответствующих компонент векторов x_0, x_1, \dots, x_N .

Рассмотрим сначала случай, когда

$$x_k = a \cos k \theta, \quad k = 0, \dots, N. \quad (6)$$

Пусть

$$\begin{aligned} y_m &= \frac{x_0}{2} + x_1 \cos \frac{\pi m}{N} + x_2 \cos \frac{2\pi m}{N} + \dots \\ &\quad \dots + x_{N-1} \cos \frac{(N-1)\pi m}{N} + x_N \cos \frac{N\pi m}{N}, \quad (m = 0, \dots, N). \end{aligned} \quad (7)$$

Последовательность y_0, y_1, \dots, y_N есть „трансформация Фурье“ последовательности x_0, x_1, \dots, x_N . Нетрудно проверить, что

$$y_m = (-1)^m \frac{a}{2} \sin N\theta \sin \theta \frac{1}{\cos \frac{\pi m}{N} - \cos \theta}, \quad (8)$$

Исследуем поведение y_m в зависимости от изменения m при фиксированном θ . Допустим сначала, что

$$\theta = \frac{\pi}{N} p, \quad (9)$$

где p — целое число.

В этом случае знаменатель выражения (8) обращается в нуль при $m = p$ и $\sin N\theta = 0$. Следовательно, $y_m = 0$ при $m \neq p$ и, как нетрудно подсчитать, $y_p = \frac{aN}{2}$. Таким образом, среди значений y_m встретится единственное отличное от нуля значение y_p .

Аналогично, если

$$x_k = a_1 \cos \theta_1 + \dots + a_r \cos \theta_r, \quad (10)$$

и

$$\theta_i = \frac{\pi}{N} p_i, \quad (11)$$

где p_i — целые числа, то среди значений y_m встретится r различных от нуля значений $y_{p_1} = \frac{a_1 N}{2}, \dots, y_{p_r} = \frac{a_r N}{2}$, а все остальные значения будут нулевыми. Это позволит определить как углы $\theta_1, \dots, \theta_r$ (по номерам p_1, \dots, p_r), так и амплитуды a_1, \dots, a_r (по значениям y_{p_1}, \dots, y_{p_r}). y_m

Вернемся теперь к исследованию случая

$$x_k = a \cos k\theta$$

в предположении, что $\theta \neq \frac{\pi}{N} p$, при целом p . Пусть

$$\theta = \frac{\pi}{N} (p + \tau) \quad 0 < \tau < 1. \quad (12)$$

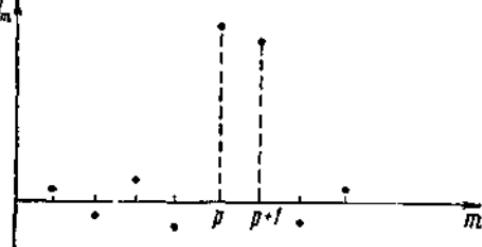


Рис. 13.

В этом случае $\cos \frac{\pi m}{N} - \cos \theta > 0$ при $0 \leq m \leq p$ и $\cos \frac{\pi m}{N} - \cos \theta < 0$ при $p+1 \leq m \leq N$. Поэтому значения y_m будут иметь чередующиеся знаки при m изменяющемся от 0 до p . Знаки y_p и y_{p+1} совпадут, а затем снова будут чередоваться. Наибольшие абсолютные значения y_m будут при $m = p$ и $m = p + 1$ и при удалении m от p будут убывать (см. рис. 13).

Такое поведение последовательности y_m дает возможность определить число p , а следовательно, в силу (13), и значение θ с точностью до $\frac{\pi}{N}$.

После определения p можно получить уточненное значение для θ из легко проверяемой формулы

$$\cos \theta = \frac{y_{p+1} \cos \frac{\pi}{N}(p+1) + y_p \cos \frac{\pi}{N} p}{y_{p+1} + y_p}. \quad (13)$$

При переходе к общему случаю

$$x_k = a_1 \cos k\theta_1 + \dots + a_r \cos k\theta_r \quad (14)$$

картина будет несколько более сложной. В этом случае

$$y_m = y_m^{(1)} + y_m^{(2)} + \dots + y_m^{(r)},$$

где

$$y_m^{(i)} = (-1)^m \frac{a_i}{2} \sin N\theta_i \sin \theta_i \frac{1}{\cos \frac{\pi m}{N} - \cos \theta_i}, \quad (15)$$

$$\theta_i = \frac{\pi}{N}(p_i + \tau_i), \quad 0 \leq \tau_i < 1.$$

При m , близком к p_i , слагаемое $y_m^{(i)}$ будет, вообще говоря, преобладать над остальными. В частности, $y_{p_i} \approx y_{p_i}^{(i)}$ и $y_{p_i+1} \approx y_{p_i+1}^{(i)}$. Поэтому y_{p_i} и y_{p_i+1} будут одного знака, и это позволит определить p_i , а следовательно, и θ_i с точностью до $\frac{\pi}{N}$.

Однако при уточнении значения θ_i указанным выше приемом, влияние других слагаемых $y_m^{(j)}$, $j \neq i$, может быть еще слишком значительным и им нельзя пренебречь.

Ланцош предлагает, наряду с числами y_m , ввести в рассмотрение числа

$$z_m = y_{m-1} - 2y_m + y_{m+1}. \quad (16)$$

Ясно, что

$$z_m = z_m^{(1)} + \dots + z_m^{(i)} + \dots + z_m^{(r)}. \quad (17)$$

Величина $z_m^{(i)}$, при удалении m от p_i , убывает значительно быстрее, чем $y_m^{(i)}$, так что пики для z_m выражены отчетливее, чем для y_m .

Подсчитаем приближенно значение $z_m^{(i)}$ при m , близких к p_i . Для этого предварительно заменим точную формулу для $y_m^{(i)}$ более удобной приближенной формулой, справедливой при m , близких к p_i .

Положим $m = p_i + q$, где q небольшое целое число. Имеем

$$\begin{aligned} y_m^{(i)} &= (-1)^{p_i+q} \frac{a_i}{2} \sin N \frac{\pi(p_i + \tau_i)}{N} \sin \theta_i \frac{1}{\cos \frac{(p_i + q)\pi}{N} - \cos \theta_i} \approx \\ &\approx (-1)^{p_i+q} \frac{a_i}{2} (-1)^{p_i} \sin \pi \tau_i \sin \theta_i \frac{1}{\sin \theta_i \left(\theta_i - \frac{(p_i + q)\pi}{N} \right)} = \\ &= (-1)^q \frac{a_i N}{2\pi} \frac{\sin \pi \tau_i}{\tau_i - q}. \end{aligned} \quad (18)$$

Соответственно

$$\begin{aligned} z_m^{(i)} &= \frac{(-1)^q a_i N \sin \pi \tau_i}{2\pi} \left(\frac{1}{\tau_i - q + 1} - \frac{2}{\tau_i - q} + \frac{1}{\tau_i - q - 1} \right) = \\ &= \frac{(-1)^q a_i N \sin \pi \tau_i}{\pi} \frac{1}{(\tau_i - q)(\tau_i - q + 1)(\tau_i - q - 1)}, \end{aligned} \quad (19)$$

откуда следует, что $z_m^{(i)}$ убывает, как $\frac{1}{|q|^3}$ при возрастании $|q|$.

Это дает основание считать

$$z_{p_i} \approx z_{p_i}^{(i)} \approx \frac{a_i N \sin \pi \tau_i}{\pi} \frac{1}{\tau_i (\tau_i + 1) (\tau_i - 1)}$$

и

$$z_{p_i+1} \approx z_{p_i+1}^{(i)} \approx -\frac{a_i N \sin \pi \tau_i}{\pi} \frac{1}{(\tau_i - 1) \tau_i (\tau_i - 2)}.$$

Отсюда

$$\frac{z_{p_i}}{z_{p_i+1}} \approx -\frac{\tau_i - 2}{\tau_i + 1},$$

и, следовательно,

$$\tau_i \approx \frac{2 - \frac{z_{p_i}}{z_{p_i+1}}}{1 + \frac{z_{p_i}}{z_{p_i+1}}}. \quad (20)$$

Точность этого приближенного равенства будет удовлетворительной, если углы θ_i не очень близки друг к другу.

Определив τ_i , находим амплитуды a_i по формуле

$$a_i \approx \frac{\pi}{N \sin \pi \tau_i} \tau_i (\tau_i + 1) (\tau_i - 1) z_{p_i}. \quad (21)$$

Однако амплитуды a_i можно вычислять так же и по формулам

$$a_i \approx \frac{2\pi}{N \sin \pi \tau_i} \tau_i (1 - \tau_i) [y_{p_i} + y_{p_i+1}]. \quad (22)$$

Последняя формула, несколько менее точная, чем предыдущая, требует знания лишь двух соседних значений y_m . Ею разумно пользоваться при нахождении всех компонент собственного вектора U_i после того, как соответствующее собственное значение μ_i вычислено.

Именно, нужно вычислить векторы y_{p_i}, y_{p_i+1} при $i=1, \dots, r$ и применять формулу (22) ко всем их компонентам.

При практическом проведении метода в качестве N следует брать достаточно большие числа, во всяком случае во много раз (10,20)

превосходящие порядок матрицы. Для удобства вычислений следует также брать число N кратным 180.

Метод требует очень большого числа операций (несколько миллионов умножений при $N = 1080$), и его осуществление возможно лишь на быстродействующих счетных устройствах. В настоящее время он был применен лишь для матриц невысокого порядка.

В работе Ланцоша рассматривается вопрос о влиянии на ход процесса пары близких собственных значений и собственных значений, близких по модулю.

ГЛАВА IX

УНИВЕРСАЛЬНЫЕ АЛГОРИФМЫ

Настоящая глава посвящена изложению теории и описанию вычислительных схем, так называемых универсальных алгорифмов, в применении к задаче решения системы линейных уравнений $AX = F$ или, в подготовленном виде, $X = BX + G$.

Под универсальным алгорифмом мы подразумеваем итерационный процесс, вообще говоря, не стационарный, осуществляемый по формулам вида

$$X^{(k+1)} = X^{(k)} + h^{(k)}(A)[F - AX^{(k)}]$$

(или в случае подготовленной системы по формулам

$$X^{(k+1)} = X^{(k)} + f^{(k)}(B)[BX^{(k)} + G - X^{(k)}],$$

в котором последовательность полиномов $h^{(k)}(t)$ (или $f^{(k)}(t)$) строится раз навсегда для обширного класса матриц с заданным расположением собственных значений, например, для класса симметричных матриц, все собственные значения которых расположены в известном промежутке.

Характерной особенностью универсальных алгорифмов является то, что быстрота их сходимости не зависит от порядка матрицы, а определяется лишь ее обусловленностью.

Простейшим универсальным алгорифмом является алгорифм, в котором

$$h^{(k)}(t) = h = \text{const.}$$

Такой алгорифм есть не что иное, как процесс последовательных приближений, примененный к системе $AX = F$, подготовленной к виду

$$X = (E - hA)^{-1}F.$$

Основным принципом построения универсальных алгорифмов является идея „подавления компонент“. В следующем параграфе мы ее поясним в простейшем случае.

§ 83. Общая идея подавления компонент

Пусть $AX = F$ система линейных уравнений, причем известно, что все собственные значения матрицы A вещественны, положительны и различны. Пусть далее X некоторый исходный вектор и

$$X' = X + h(A)[F - AX], \quad (1)$$

где $h(t)$ — некоторый полином (может быть, нулевой степени).

Как мы видели в гл. VII, последовательные приближения при применении градиентных итерационных процессов в случае положительно-определенной матрицы A определяются по формуле (1), причем полином $h(t)$ может меняться от шага к шагу. При этом в рассмотренных выше градиентных методах коэффициенты полиномов существенно зависят как от матрицы A , так и от начального приближения, так что описанные в гл. VII методы не являются универсальными в смысле данного выше определения.

Для построения универсальных алгорифмов исследуем формулу (1) со следующей точки зрения. Обозначим через Y и Y' соответствующие векторы ошибок для приближений X и X' . Тогда

$$Y' = (E - Ah(A))Y = g(A)Y,$$

где

$$g(t) = 1 - th(t). \quad (2)$$

Пусть $\lambda_1, \dots, \lambda_n$ попарно различные собственные значения матрицы A и U_1, \dots, U_n соответствующие им собственные векторы. Пусть

$$Y = a_1U_1 + \dots + a_nU_n.$$

Тогда

$$Y' = a_1g(\lambda_1)U_1 + \dots + a_ng(\lambda_n)U_n. \quad (3)$$

Таким образом, при переходе от приближения X к приближению X' компоненты вектора ошибки в разложении по собственным векторам матрицы A умножаются, соответственно, на значения $g(\lambda_1), \dots, g(\lambda_n)$, которые естественно называть множителями затухания. Переход от вектора X к вектору X' можно считать удачным, если один или несколько множителей $g(\lambda_1), \dots, g(\lambda_n)$ очень малы, а остальные не слишком велики. „Подавив“ таким образом одну или несколько компонент вектора ошибки, можно перейти к следующему шагу процесса, распорядившись им так, чтобы в нем подавлялись другие компоненты.

В выборе полинома $g(t)$ для одного шага процесса имеется широкий произвол. Единственным условием, связывающим выбор $g(t)$, является требование $g(0) = 1$.

Идеальным выбором полинома $g(t)$ для данной матрицы A является такой, при котором $g(\lambda_1) = \dots = g(\lambda_n) = 0$, ибо при таком выборе

уже один шаг процесса приводит к точному решению. По существу к этому мы приходили в результате применения метода сопряженных градиентов.

Выясним теперь критерии, которыми следует руководствоваться при выборе универсального полинома $g(t)$, исходя из естественного предположения о том, что мы ничего не знаем о расположении собственных значений матрицы A , кроме промежутка $(0, M)$, в котором они содержатся.

Пусть на рис. 14 изображен график $g(t)$.

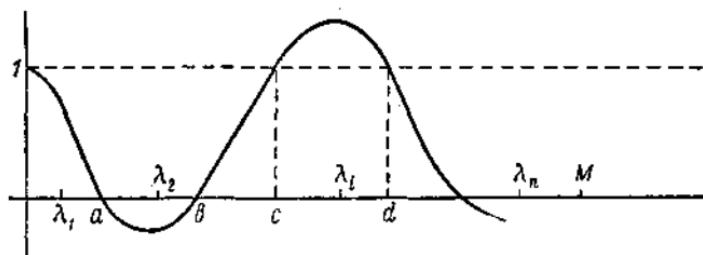


Рис. 14.

Тогда множителями затухания $g(\lambda_1), \dots, g(\lambda_n)$ будут ординаты графика с абсциссами $\lambda_1, \dots, \lambda_n$.

Ясно, что затухание будет наиболее эффективным в окрестности корней полинома $g(t)$, мало эффективным в окрестности точек, где $g(t)$ близко к ± 1 , и вместо затухания будет возрастание компонент в точках, где $|g(t)|$ больше 1. Так, на рис. 14 затухание сильно для собственных значений, близких к точкам a и b , а для собственных значений из промежутка (c, d) затухание совсем не имеет места. Таким образом, на каждом шагу процесса желательно подобрать полином $g(t)$ так, чтобы он возможно теснее примыкал к оси абсцисс и чтобы $|g(t)|$ нигде в промежутке $(0, M)$ не превышал единицы. Ясно, что при этом наибольшую трудность будет представлять подавление компонент, отвечающих близким к нулю собственным значениям, так как $g(0) = 1$. Это совершенно естественно, так как наличие близких к нулю собственных значений матрицы A свидетельствует о плохой обусловленности системы.

Выбор полинома $h(t)$ (а вместе с ним и $g(t)$) можно осуществлять, применяя другой критерий, основанный на сравнении компонент вектора ошибки Y' с компонентами невязки $r = F - AX = AY$ приближения X . Пусть

$$r = b_1 U_1 + \dots + b_n U_n.$$

Тогда

$$Y = \frac{b_1}{\lambda_1} U_1 + \dots + \frac{b_n}{\lambda_n} U_n$$

и

$$\begin{aligned} Y' &= \frac{b_1}{\lambda_1} g(\lambda_1) U_1 + \dots + \frac{b_n}{\lambda_n} g(\lambda_n) U_n = \\ &= b_1 \left(\frac{1}{\lambda_1} - h(\lambda_1) \right) U_1 + \dots + b_n \left(\frac{1}{\lambda_n} - h(\lambda_n) \right) U_n. \end{aligned} \quad (4)$$

При этом подходе естественным становится такой выбор полинома $h(t)$, чтобы его значения на промежутке $(0, M)$ возможно менее отклонялись от значений функции $\frac{1}{t}$.

Оба указанных критерия по существу близки друг другу. Именно, выполнение одного из них влечет выполнение, в большей или меньшей степени, второго. Однако второй критерий предъявляет более сильные требования к подавлению компонент, отвечающих малым собственным значениям. Действительно, если

$$\left| \frac{1}{\lambda_i} - h(\lambda_i) \right| < \delta (\delta > 0),$$

то

$$|g(\lambda_i)| = |1 - \lambda_i h(\lambda_i)| < \delta \lambda_i.$$

Первым критерием целесообразно руководствоваться в тех случаях, когда пользование формулой (1) является элементарным шагом в итерационном процессе, ибо множители затухания при проведении нескольких шагов итерационного процесса просто перемножаются. Если же формула (1) применяется всего лишь один раз, то имеет смысл руководствоваться вторым критерием, если есть основания предполагать, что компоненты исходной невязки (например, свободного члена F) имеют коэффициенты одного порядка в своем разложении по собственным векторам.

В случае, если о распределении собственных значений матрицы A имеются какие-либо дополнительные сведения, можно ставить вопрос о наилучшем выборе полинома $g(t)$ данной степени в смысле первого или второго критерия. Так, если известно, что все собственные значения заключены в промежутке (m, M) , то наилучшим в смысле первого критерия будет полином $g(t)$, наименее отклоняющийся от нуля на промежутке (m, M) , нормированный условием $g(0) = 1$, наилучшим в смысле второго критерия будет полином $h(t)$, наименее отклоняющийся от $\frac{1}{t}$ в том же промежутке. Ниже мы рассмотрим универсальные алгоритмы, основанные на осуществлении такого выбора.

Для системы, подготовленной к виду

$$X = BX + G, \quad (5)$$

при условии, что все собственные значения матрицы B лежат в промежутке $(-1, 1)$, один шаг универсального алгорифма производится по формуле

$$X' = X + f(B)[BX + G - X]. \quad (6)$$

Векторы ошибки двух соседних приближений в этом случае связаны соотношением

$$Y' = Y + f(B)(BY - Y) = [E - (E - B)f(B)]Y = e(B)Y,$$

где $e(t) = 1 - (1-t)f(t)$. Таким образом, здесь полином $e(t)$ должен удовлетворять требованию $e(1) = 1$. Множителями затухания будут значения $e(\mu_1), \dots, e(\mu_n)$, где μ_1, \dots, μ_n собственные значения матрицы B . Критериями для выбора полинома $e(t)$ или полинома $f(t)$ являются, во-первых, возможно малое отклонение от нуля значений полинома $e(t)$ на промежутке $(-1, 1)$, нормированного условием $e(1) = 1$, или, во-вторых, возможно малое отклонение полинома $f(t)$ от функции $\frac{1}{1-t}$ на том же промежутке.

§ 84. Прием Л. А. Люстерника для ускорения сходимости метода последовательных приближений при решении системы линейных уравнений¹⁾

При решении системы $X = BX + G$ методом последовательных приближений в самом ходе процесса получается некоторая информация о расположении собственных значений матрицы B . Именно, если наибольшее по модулю собственное значение μ_1 матрицы B оторвано от остальных собственных значений, то оно может быть практически определено из отношений одноименных компонент векторов $X^{(k+1)} - X^{(k)}$ и $X^{(k)} - X^{(k-1)}$. Действительно, $X^{(k+1)} - X^{(k)} = B^k(X_1 - X_0)$, $X^{(k)} - X^{(k-1)} = B^{k-1}(X_1 - X_0)$, где X_0 начальное приближение.

Зная μ_1 , можно уничтожить коэффициент при собственном векторе U_1 в векторе ошибки, исходя от приближения $X^{(k)}$, уже построенного методом последовательных приближений. Для этого достаточно взять, при переходе от $X^{(k)}$ к следующему приближению $\bar{X}^{(k+1)}$, в качестве полинома $f^{(k)}(t)$ константу $\frac{1}{1-\mu_1}$. Действительно, соответствующий полином $e^{(k)}(t)$ равен $1 - \frac{1-t}{1-\mu_1}$, так что $e^{(k)}(\mu_1) = 0$. Последний шаг тогда будет осуществлен по формуле

$$\begin{aligned} \bar{X}^{(k+1)} &= X^{(k)} + \frac{1}{1-\mu_1}(BX^{(k)} + G - X^{(k)}) = \\ &= X^{(k)} + \frac{1}{1-\mu_1}(X^{(k+1)} - X^{(k)}), \end{aligned} \quad (1)$$

¹⁾ Л. А. Люстерник [1]

где $X^{(k+1)} = BX^{(k)} + G$ есть следующее за $X^{(k)}$ приближение метода последовательных приближений.

Если метод последовательных приближений проводится по формуле

$$X^{(k)} = \sum_{i=0}^{k-1} B^i G,$$

то формула (1) приобретает еще более простой вид

$$\bar{X}^{(k+1)} = X^{(k)} + \frac{1}{1-\mu_1} B^k G. \quad (2)$$

Компоненты вектора ошибки для приближений, найденных по формулам (1) и (2), будут, очевидно, иметь порядок $O(|\mu_2|^k)$, где μ_2 следующее за μ_1 по величине модуля собственное значение матрицы B .

Проиллюстрируем описанный прием на примере § 30. На основе табл. III. 1 в § 53 (пример 3) было определено $\mu_1 = 0.4800$ из отношений компонент 14-й и 13-й итераций вектора $G = (0.76, 0.08, 1.12, 0.68)'$ матрицей B . Из той же таблицы

$$X^{(13)} = (1.53490847, 0.12200958, 1.97509985, 1.41289889)'.$$

Вычислим

$$\frac{B^{14}G}{1 - 0.4800} = (0.00005656, 0, 0.00005656, 0.00005656)'.$$

Таким образом,

$$\bar{X}^{(14)} = (1.534965, 0.122010, 1.975156, 1.412955)'.$$

Указанное решение совпадает с точностью до $1 \cdot 10^{-6}$ с найденным по методу Гаусса.

Отношения 7-й и 6-й итераций дают (§ 53) для μ_1 значение $\mu_1 = 0.4792$.

Так как

$$X^{(6)} = (1.52533, 0.12201, 1.96551, 1.40333)',$$

и

$$\frac{B^7G}{0.5208} = (0.00962, 0.00002, 0.00963, 0.00960)',$$

то

$$\bar{X}^{(7)} = (1.53495, 0.12203, 1.97514, 1.41293)'.$$

Из табл. III. 1 видно, что $\bar{X}^{(7)}$ ближе к точному, чем $X^{(14)}$.

Прием Люстерника может быть применен и к циклическому одношаговому процессу, ибо этот метод, применяемый к системе $X = BX + G$, равносителен методу последовательных приближений для некоторой эквивалентной системы.

§ 85. Подавление компонент при помощи полиномов низших степеней

Рассмотрим в свете идеи подавления компонент приемы подготовки системы $AX = F$ с положительно-определенной матрицей A к применению метода последовательных приближений.

Для системы, подготовленной к виду $X = X + h(F - AX) = = (E - hA)X + hF$, формула метода последовательных приближений будет

$$X^{(k)} = X^{(k-1)} + h(F - AX^{(k-1)}). \quad (1)$$

В этом случае коэффициентами затухания будут значения $1 - h\lambda_i$ при $i = 1, \dots, n$.

Все коэффициенты затухания, очевидно, будут меньше единицы, и, следовательно, процесс будет сходящимся, если $\frac{1}{h} > \frac{M}{2}$, т. е. $h < \frac{2}{M}$ (рис. 15). Быстрота затухания компонент на разных частях

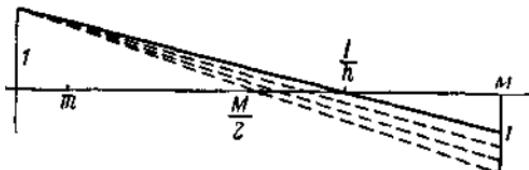


Рис. 15.

промежутка (m, M) будет различной и наибольшая быстрота, очевидно, будет в зоне, примыкающей к точке $\frac{1}{h}$ (если $h > \frac{1}{M}$), или на правом конце промежутка (если $h \leq \frac{1}{M}$). Наиболее медленное затухание будет на правом конце промежутка, если $h < \frac{1}{M+m}$, и на левом, если $h > \frac{1}{M+m}$. Ясно, что для ускорения процесса последовательных приближений целесообразно время от времени брать $h = \frac{1}{M}$ или $h = \frac{1}{m}$, соответственно, если одно из этих чисел известно.

Именно в этом и состоит описанный выше прием Люстерника в применении к подготовленной системе $X = BX + G$ при $B = E - hA$, $G = hF$.

Действительно, собственными значениями μ_i матрицы B являются как раз коэффициенты затухания $\mu_i = 1 - h\lambda_i$. Наибольшим из них по модулю будет $\mu_1 = 1 - hM$ или $\mu_2 = 1 - hm$. Один шаг при $h = \frac{1}{M}$ дает

$$\bar{X}^{(k+1)} = X^{(k)} + \frac{1}{M} (F - AX^{(k)})$$

или, переходя к матрице B ,

$$\begin{aligned}\bar{X}^{(k+1)} &= X^{(k)} + \frac{h}{1-\mu_1} \left[F - \frac{1}{h} (E - B) X^{(k)} \right] = \\ &= X^{(k)} + \frac{1}{1-\mu_1} [G - (E - B) X^{(k)}] = \\ &= X^{(k)} + \frac{BX^{(k)} + G - X^{(k)}}{1-\mu_1} = X^{(k)} + \frac{1}{1-\mu_1} (X^{(k+1)} - X^{(k)}).\end{aligned}$$

К такому же результату мы придем, если наибольшим собственным значением матрицы B окажется $1 - hm$, и мы положим на $k+1$ -м шагу $h = \frac{1}{M}$.

Довольно эффективный прием подавления компонент мы получим, если будем изменять константу h от шага к шагу так, чтобы корень $\frac{1}{hk}$ полинома $g^{(k)}(t) = 1 - ht$ двигался по промежутку $(\frac{M}{2}, M)$ справа налево. При этом зона эффективного подавления компонент вектора ошибки будет передвигаться, накрывая, в конце концов, весь промежуток.

Если мы захотим быстро подавлять компоненты собственных векторов, принадлежащих собственным значениям, лежащим в промежутке $(0, \frac{M}{2})$, то нам придется брать $h > \frac{2}{M}$, что, однако, будет приводить к возрастанию компонент, отвечающих собственным значениям, близким к M .

Если же стремиться строить процесс так, чтобы ни на одном шагу не происходило увеличения каких-либо компонент, то для дальнейшего сдвига зоны наиболее эффективного подавления компонент влево нужно обратиться к полиномам высших степеней.

Рассмотрим грубую схему выбора полиномов $g_s^{(k)}(t)$ невысоких степеней s , обслуживающих значительную часть промежутка $(0, M)$, с постепенным сдвигом зоны наиболее эффективного подавления компонент справа налево (метод „утюга“, см. рис. 16). Как мы уже видели, промежуток $(\frac{M}{2}, M)$ хорошо обслуживается полиномами первой степени.

Известно, что из всех полиномов данной степени s , удовлетворяющих требованиям $g_s(0) = 1$ и $|g_s(t)| \leq 1$ при $0 \leq t \leq M$, наиболее

близким к нулю наименьшим корнем обладает полином Чебышева для промежутка $(0, M)$, т. е. полином

$$\cos s \arccos \frac{M - 2t}{M}.$$

Наименьший корень этого полинома равен $t_s = M \sin^2 \frac{\pi}{4s}$, так что

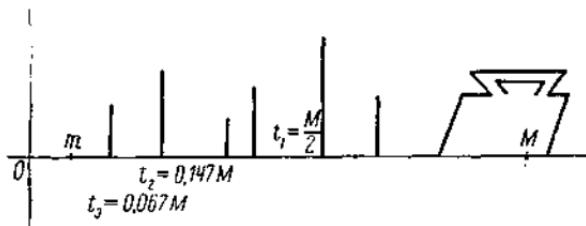


Рис. 16.

для последовательных полиномов Чебышева наименьшими корнями будут

$$t_1 = \frac{M}{2}, \quad t_2 = \frac{2 - \sqrt{2}}{4} M \approx 0.147M, \quad t_3 = \frac{2 - \sqrt{3}}{4} M \approx 0.067M, \dots$$

$$\dots, \quad t_s \approx \frac{\pi^2}{16s^2} M.$$

Поэтому за счет полиномов второй степени мы можем расширить зону действия метода от $\frac{M}{2}$ до $0.147M$, за счет полиномов третьей степени расширить ее до $0.067M$ и т. д. Таким образом, для матрицы с обусловленностью, меньше чем 15, можно ограничиться лишь полиномами до третьей степени включительно.

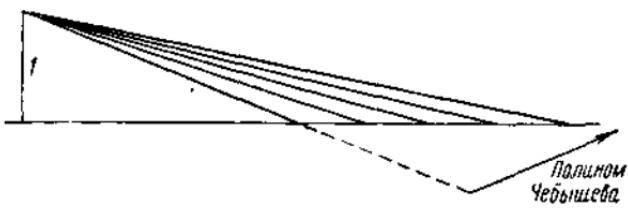


Рис. 17.

Выбор полиномов первой степени можно осуществить, например, так. На промежутке $(\frac{M}{2}, M)$ возьмем точки $a_0 = M > a_1 > \dots > a_{p_1} \geq \frac{M}{2}$ и положим $g_1^{(k)}(t) = 1 - \frac{t}{a_k}$, $k = 0, \dots, p_1$. (рис. 17)

Затем, для построения полиномов второй степени выберем последовательность точек $b_0 = \frac{M}{2} > b_1 > \dots > b_{p_1} = 0.147M$ и положим

$$g_2^{(k)}(t) = 1 - \frac{t^2 - Mt}{b_{k-p_1}^2 - Mb_{k-p_1}} \quad (k = p_1 + 1, \dots, p_1 + p_2)$$

(см. рис. 18). Эти полиномы, очевидно, удовлетворяют первым двум требованиям, наложенным на полиномы $g_2^{(k)}(t)$, и обращаются в нуль при $t = b_{k-p_1}$.

Для обслуживания зоны от $0.147M$ до $0.067M$ служат полиномы 3-й степени

$$g_3^{(k)}(t) = \frac{(t^2 - Mt)(2t - M)}{(c_j^2 - Mc_j)(2c_j - M)},$$

где c_j точки между t_2 и t_3 .

Мы не будем входить в подробности относительно выбора полиномов более высокой степени, так как ниже мы рассмотрим другие,

более просто выполняемые способы подавления компонент, отвечающих собственным значениям, близким к нулю. Отметим только, что во всем изложенном важную роль играет значение числа M . Ошибка в оценке этого числа в сторону уменьшения имеет некоторую опасность. Ошибка же в сторону увеличения может лишь несколько увеличить объем работы.

Выбор точек деления обусловливается требованием точности с одной стороны и желанием иметь их как можно меньше (обслуживать зону подлиннее) с другой стороны.

Приведем результат использования полиномов низших степеней к нахождению решения системы (9) § 23.

Здесь $M = 2.62$. Взяв

$$\begin{aligned} a_1 &= 2.62, & a_2 &= 2.30, & a_3 &= 2.00, & a_4 &= 1.70, & a_5 &= 1.40 \\ b_1 &= 1.20, & b_2 &= 0.9, & b_3 &= 0.6, & b_4 &= 0.3, \end{aligned}$$

получим

$$X' = (-1.2284428, \quad 0.0473912, \quad 1.0233490, \quad 1.4591532)^T$$

Приближение X' получено в результате применения 13 итераций. Длина вектора ошибки построенного приближения составляет 1.85% длины вектора ошибки начального приближения $X_0 = (0, 0, 0, 0)^T$.

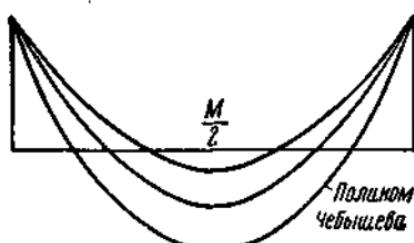


Рис. 18.

Описанный прием выбора подавляющих полиномов является довольно грубым, и как мы увидим ниже, при более тщательном выборе полиномов может быть получен лучший результат при использовании того же числа итераций.

§ 86. Различные формы проведения универсальных алгорифмов

Прежде чем перейти к описанию конкретных универсальных алгорифмов, коснемся вопроса об их численном осуществлении, ограничиваясь рассмотрением одного шага процесса при различных способах задания полинома, определяющего этот шаг.

1. Известны коэффициенты полинома $h(t)$, или, что то же самое, коэффициенты $g(t)$. Пусть

$$h(t) = c_0 t^{s-1} + c_1 t^{s-2} + \dots + c_{s-1}. \quad (1)$$

В этом случае вычисление X' можно производить двумя следующими способами.

a) Прежде всего вычисляется невязка $r = F - AX$. Затем последовательно вычисляются векторы $Ar, A^2r, \dots, A^{s-1}r$ и, наконец, X' находится как известная линейная комбинация уже известных векторов

$$X' = X + c_{s-1}r + c_{s-2}Ar + \dots + c_0A^{s-1}r. \quad (2)$$

При таком построении вектора X' возможен рост ошибок округления с одной стороны, а также уничтожение значащих цифр с другой стороны, ибо, как правило, коэффициенты c_0, c_1, \dots, c_{s-1} имеют разные знаки.

b) Вычислив $r = F - AX$, находим последовательно векторы

$$\begin{aligned} Z_0 &= c_0r \\ Z_1 &= AZ_0 + c_1r \\ &\vdots \\ Z_{s-1} &= AZ_{s-2} + c_{s-1}r. \end{aligned} \quad (3)$$

Ясно, что $h(A)r = Z_{s-1}$ и, следовательно,

$$X' = X + Z_{s-1}.$$

Этот способ есть по существу не что иное, как применение известной схемы Хорнера. При таком построении вектора X' ошибки округления сказываются несколько менее, чем в предыдущем способе.

2. Известны корни $\epsilon_1, \dots, \epsilon_s$ полинома $g(t)$. Тогда

$$g(t) = \left(1 - \frac{t}{\epsilon_1}\right) \cdots \left(1 - \frac{t}{\epsilon_s}\right). \quad (4)$$

В этом случае вектор X' может быть найден как последний член Z_s , следующей последовательности векторов:

$$\begin{aligned} Z_1 &= X + \frac{1}{\varepsilon_1} (F - AX) \\ Z_i &= Z_{i-1} + \frac{1}{\varepsilon_i} (F - AZ_{i-1}) \quad (i = 2, \dots, s). \end{aligned} \tag{5}$$

Действительно, при всех i

$$X^* = X^* + \frac{1}{\varepsilon_i} (F - AX^*),$$

где X^* — точное решение системы, и потому

$$X^* - Z_i = (X^* - Z_{i-1}) - \frac{1}{\varepsilon_i} A(X^* - Z_{i-1}) = \left(E - \frac{1}{\varepsilon_i} A\right)(X^* - Z_{i-1}).$$

Следовательно,

$$\begin{aligned} X^* - Z_s &= \left(E - \frac{1}{\varepsilon_s} A\right) \dots \left(E - \frac{1}{\varepsilon_1} A\right)(X^* - X) = \\ &= g(A)Y = Y - h(A)AY = Y - h(A)(F - AX) \end{aligned}$$

$$Z_s = X^* - Y + h(A)(F - AX) = X + h(A)(F - AX) = X'.$$

Эта схема впервые описана в работе Ричардсона [1].

При применении этой схемы наиболее опасными в смысле влияния ошибок округления являются шаги, соответствующие малым корням ε_i . Целесообразно располагать корни в порядке их убывания.

3. Известны рекуррентные соотношения для определения полинома $g(t)$. Мы рассмотрим случай наиболее часто встречающихся трехчленных соотношений. Пусть $g(t) = g_s(t)$, где

$$\begin{aligned} g_0(t) &= 1, \quad g_1(t) = \alpha_1 t + 1 \\ g_i(t) &= [\alpha_i t + (\beta_i + 1)] g_{i-1}(t) - \beta_i g_{i-2}(t) \quad (i = 2, \dots, s). \end{aligned} \tag{6}$$

Вид рекуррентных соотношений выбран так, чтобы все полиномы $g_i(t)$ были нормированы условием $g_i(0) = 1$.

Строим последовательность векторов

$$X_i = X_{i-1} - \alpha_i(F - AX_{i-1}) + \beta_i(X_{i-1} - X_{i-2}) \tag{7}$$

при $X_0 = X$ и $X_1 = X_0 - \alpha_1(F - AX)$.

Тогда

$$X' = X_s.$$

Действительно, из (7) следует, что

$$Y_i = Y_{i-1} + \alpha_i AY_{i-1} + \beta_i(Y_{i-1} - Y_{i-2}).$$

Отсюда заключаем по индукции, что

$$Y_i = g_i(A)Y_0.$$

В частности,

$$Y_s = g_s(A) Y_0$$

и потому

$$X_s = X'.$$

Следующие приемы относятся к системе, подготовленной к виду

$$X = BX + G.$$

4. Известны коэффициенты полинома $f(t)$, именно

$$f(t) = b_0 t^{s-1} + b_1 t^{s-2} + \dots + b_{s-1}. \quad (8)$$

a) Вычислив невязку начального приближения $r = BX + G - X$, вычисляем векторы $B^s r, \dots, B^{s-1} r$ и находим X' в виде известной линейной комбинации уже известных векторов

$$X' = X + b_{s-1} r + \dots + b_0 B^{s-1} r. \quad (9)$$

b) Вычислив невязку r , строим последовательность векторов

$$\begin{aligned} Z_0 &= b_0 r \\ Z_i &= BZ_{i-1} + b_i r \quad (i = 1, 2, \dots, s-1). \end{aligned} \quad (10)$$

Тогда

$$X' = X + Z_{s-1}.$$

5. Известны коэффициенты полинома $e(t)$, именно

$$e(t) = a_0 t^s + a_1 t^{s-1} + \dots + a_s. \quad (11)$$

a) Вычисляем последовательные приближения $X_0 = X, X_1 = BX_0 + G, \dots, X_s = BX_{s-1} + G$ и составляем их линейную комбинацию

$$a_0 X_s + a_1 X_{s-1} + \dots + a_s X_0. \quad (12)$$

Она и будет искомым приближением X' .

Действительно, пусть

$$Y_0 = X^* - X_0, \quad Y_1 = X^* - X_1, \dots, \quad Y_s = X^* - X_s.$$

Тогда

$$\begin{aligned} Y_1 &= BX_0, \dots, Y_s = B^s Y_0 \quad \text{и} \quad a_0 X_s + a_1 X_{s-1} + \dots + a_s X_0 = \\ &= (a_0 + a_1 + \dots + a_s) X^* - a_0 B^s Y_0 - a_1 B^{s-1} Y_0 - \dots - a_s Y_0 = \\ &= e(1) X^* - e(B) Y_0 = X^* - Y_0 + f(B)(E - B) Y_0 = \\ &= X_0 + f(B)(BX_0 + G - X_0) = X'. \end{aligned}$$

b) Вычисляем последовательность векторов по формулам

$$\begin{aligned} Z_0 &= a_0 X \\ Z_{i+1} &= BZ_i + a_{i-1} X + (a_0 + \dots + a_i) G \quad (i = 0, \dots, s-1). \end{aligned} \quad (13)$$

Тогда последний вектор Z_s будет равен X' .

6. Известны корни $\epsilon_1, \dots, \epsilon_s$ полинома $e(t)$. Тогда

$$e(t) = \frac{(t - \epsilon_1) \dots (t - \epsilon_s)}{(1 - \epsilon_1) \dots (1 - \epsilon_s)}, \quad (14)$$

ибо $e(1) = 1$.

Переход от вектора X к вектору X' можно осуществить в s шагов, полагая на i -м шагу $e_i(t) = \frac{t - \epsilon_i}{1 - \epsilon_i}$, ибо при применении нескольких шагов полиномы $e_i(t)$ перемножаются. При этом $f_i(t) = \frac{1}{1 - \epsilon_i}$, так как

$$e_i(t) = 1 - \frac{1}{1 - \epsilon_i} (1 - t) = 1 - f_i(t)(1 - t).$$

Следовательно, вектор X' находится как s -й член Z_s последовательности

$$Z_i = Z_{i-1} + \frac{1}{1 - \epsilon_i} (BZ_{i-1} + G - Z_{i-1}), \quad (15)$$

начиная с $Z_0 = X$.

7. Известны рекуррентные соотношения для определения полинома $f(t)$. Пусть $f(t) = f_{s-1}(t)$, где

$$\begin{aligned} f_0(t) &= \beta_0, \quad f_1(t) = \alpha_1 t + \beta_1 \\ f_i(t) &= (\alpha_i t + \beta_i) f_{i-1}(t) - \gamma_i f_{i-2}(t). \end{aligned} \quad (16)$$

Вычисляем последовательность векторов

$$\begin{aligned} r &= BX + G - X, \quad Z_0 = \beta_0 r, \quad Z_1 = \alpha_1 Br + \beta_1 r, \\ Z_i &= \alpha_i BZ_{i-1} + \beta_i Z_{i-1} - \gamma_i Z_{i-2} \quad (i = 1, 2, \dots, s-1). \end{aligned} \quad (17)$$

Ясно, что

$$\begin{aligned} Z_{s-1} &= f(B)(BX + G - X) \\ X' &= X + Z_{s-1}. \end{aligned}$$

8. Известны рекуррентные соотношения для определения полинома $e(t)$. Пусть $e(t) = e_s(t)$, где

$$\begin{aligned} e_0(t) &= 1, \quad e_1(t) = \alpha_1 t + \beta_1 \\ e_i(t) &= (\alpha_i t + \beta_i) e_{i-1}(t) - \gamma_i e_{i-2}(t). \end{aligned} \quad (18)$$

Предполагается, кроме того, что все полиномы $e_i(t)$ удовлетворяют условию $e_i(1) = 1$, так что $\alpha_1 + \beta_1 = 1$, $\alpha_i + \beta_i - \gamma_i = 1$, $i = 2, \dots, s$. В этом случае X' будет равен вектору X_s в последовательности

$$\begin{aligned} X_0 &= X, \quad X_1 = \alpha_1(BX_0 + G) + \beta_1 X_0 \\ X_i &= \alpha_i(BX_{i-1} + G) + \beta_i X_{i-1} - \gamma_i X_{i-2}. \end{aligned} \quad (19)$$

Действительно, в силу условий $\alpha_1 + \beta_1 = 1$, $\alpha_i + \beta_i - \gamma_i = 1$ имеем

$$X^* = \alpha_1(BX^* + G) + \beta_1 X^*$$

$$X^* = \alpha_i(BX^* + G) + \beta_i X^* - \gamma_i X^*.$$

Вычитая, получим

$$\begin{aligned} Y_1 &= \alpha_1 B Y_0 + \beta_1 Y_0 = e_1(B) Y_0 \\ Y_i &= (\alpha_i B + \beta_i E) Y_{i-1} - \gamma_i Y_{i-2}, \end{aligned}$$

откуда по индукции

$$Y_i = e_i(B) Y_0,$$

в частности, $Y_s = e(B) Y_0$ и, следовательно,

$$X' = X_s.$$

Как правило, наиболее удобными оказываются схемы, использующие рекуррентные соотношения.

§ 87. Универсальный алгорифм, наилучший в смысле первого критерия¹⁾

Пусть известно, что все собственные значения матрицы A расположены в промежутке (m, M) , $0 < m < M$, и различны. Подготовим систему $AX = F$ к виду $X = BX + G$ так, чтобы собственные значения матрицы B расположились в симметричном интервале с центром в начале координат. Очевидно, что для этого нужно взять $h = \frac{2}{M+m}$, $B = E - hA$, $G = hF$. Тогда все собственные значения матрицы B попадут в интервал $\left(-\frac{1}{\gamma}, \frac{1}{\gamma}\right)$, где $\gamma = \frac{M+m}{M-m} > 1$.

Построим универсальный алгорифм

$$X_s = X_0 + f_{s-1}(B)(BX_0 + G - X_0). \quad (1)$$

подобрав полином $f_{s-1}(t)$ так, чтобы при данной его степени $s-1$ было обеспечено максимально возможное подавление компонент для всего класса матриц B с собственными значениями, заключенными в промежуток $\left(-\frac{1}{\gamma}, \frac{1}{\gamma}\right)$.

Для этого в качестве полинома $e_s(t) = 1 - (1-t)f_{s-1}(t)$ нужно, очевидно, взять полином, наименее уклоняющийся от нуля на промежутке $\left(-\frac{1}{\gamma}, \frac{1}{\gamma}\right)$, нормированный условием $e_s(1) = 1$.

Таким полиномом является

$$\tilde{T}_s(t) = \frac{T_s(\gamma t)}{T_s(1)}, \quad (2)$$

¹⁾ М. Ш. Бирман [1].

где

$$T_s(t) = \cos s \arccos t.$$

На рис. 19 дан график $\tilde{T}_s(t)$ при $\gamma = \frac{5}{4}$.

Полиномы $\tilde{T}_i(t)$ связаны простыми рекуррентными соотношениями. Действительно, для полиномов Чебышева такими являются соотношения

$$T_i(t) = 2tT_{i-1}(t) - T_{i-2}(t)$$

при $T_0(t) = 1$, $T_1(t) = t$. Поэтому

$$\tilde{T}_i(t) = \left[1 + \frac{T_{i-2}(\gamma)}{T_i(\gamma)} \right] t \tilde{T}_{i-1}(t) - \frac{T_{i-2}(\gamma)}{T_i(\gamma)} \tilde{T}_{i-2}(t), \quad \tilde{T}_0(t) = 1, \quad \tilde{T}_1(t) = t. \quad (3)$$

В соответствии с рекуррентными соотношениями (3) универсальный алгорифм (§ 86 п. 8) строится по формулам

$$\begin{aligned} X_1 &= BX_0 + G \\ X_i &= \left[1 + \frac{T_{i-2}(\gamma)}{T_i(\gamma)} \right] (BX_{i-1} + G) - \\ &\quad - \frac{T_{i-2}(\gamma)}{T_i(\gamma)} X_{i-2} \quad (i = 1, 2, \dots, s). \end{aligned} \quad (4)$$

Для вычисления по формулам (4) нужно предварительно заготовить по рекуррентным соотношениям значения $\frac{T_{i-2}(\gamma)}{T_i(\gamma)}$. При возрастающем i эти значения довольно быстро стремятся к пределу

$$\alpha = (\gamma - \sqrt{\gamma^2 - 1})^2.$$

Быстрота сходимости описанного процесса при $s \rightarrow \infty$ будет иметь порядок

$$\frac{1}{T_s(\gamma)} = \frac{2}{(\gamma + \sqrt{\gamma^2 - 1})^s + (\gamma - \sqrt{\gamma^2 - 1})^s},$$

так что процесс сходится значительно быстрее метода последовательных приближений, для которого быстрота сходимости будет γ^{-s} .

Так, например, при $\gamma = \frac{25}{24}$ ($\gamma^{-1} = 0.96$) будет

$$\frac{1}{T_s(\gamma)} = \frac{2}{\left(\frac{4}{3}\right)^s + \left(\frac{3}{4}\right)^s} \approx 2 \left(\frac{3}{4}\right)^s,$$

в то время как $\gamma^{-s} = \left(\frac{24}{25}\right)^s$.

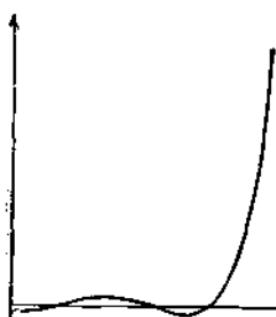


Рис. 19.

При пользовании формулами (4) вместо того, чтобы брать все увеличивающиеся значения для s , можно брать не очень большие s , но повторять процесс несколько раз, принимая приближение, полученное в конце цикла, за начальное приближение нового процесса.

Для данной матрицы A осуществить подготовку к виду, позволяющему использовать описанный процесс, не просто. Действительно, если для оценки числа M можно воспользоваться хотя бы оценкой Гершгорина, то определение числа m оказывается гораздо более трудоемким. Процесс же определяется выбором числа $\gamma = \frac{M+m}{M-m}$, которое быстро приближается к единице при $m \rightarrow 0$, и слишком грубый выбор γ может сильно замедлить быстроту сходимости процесса (если истинное значение γ будет значительно больше взятого на основании грубых оценок M и m).

Описанный процесс по существу дела идентичен тому, с которым сравнивался метод наискорейшего спуска при выводе оценок для быстроты сходимости.

Для полинома $T_s(t)$ легко вычислить коэффициенты, что дает возможность использовать схему п. 5 § 86. Однако при пользовании этой схемой происходит значительное уничтожение значащих цифр.

Несколько лучшей оказывается схема п. 6 § 86, использующая корни полинома $T_s(t)$. Очевидно, что эти корни равны

$$\varepsilon_i = \frac{1}{\gamma} \cos \frac{(2i-1)\pi}{2s} \quad (i=1, \dots, s).$$

При пользовании этой схемой приближение X_s получится как последний член последовательности

$$Z_i = Z_{i-1} + \frac{1}{1-\varepsilon_i} (BZ_{i-1} + G - Z_{i-2}) \quad (i=1, \dots, s),$$

начинаяющейся с $Z_0 = X_0$.

Ясно, что при таком способе построения X_s нужно заранее фиксировать значение s . Сходимость процесса обеспечивается циклическим повторением.

Для уменьшения влияния ошибок округления следует располагать корни ε в порядке их убывания, так как процесс наиболее чувствителен к ошибкам округления на тех шагах, где ε близко к единице.

Покажем ход процесса на примере системы (9) § 23, принимая за M и m трехзначные приближения к наибольшему и наименьшему собственным значениям матрицы системы, именно $M = 2.322$ и $m = 0.242$. Это дает $\gamma = 0.78003120$, так что

$$B = \begin{bmatrix} 0.21996880 & -0.32761310 & -0.42121685 & -0.51482059 \\ -0.32761310 & 0.21996880 & -0.24960998 & -0.34321373 \\ -0.42121685 & -0.24960998 & 0.21996880 & -0.17160686 \\ -0.51482059 & -0.34321373 & -0.17160686 & 0.21996880 \end{bmatrix}$$

$$G = (0.23400936, \quad 0.39001560, \quad 0.54602184, \quad 0.70202808)',$$

Далее $\gamma = 1.2326923$, $\alpha = 0.26204959$.

Имеем

X_0	0	0	0	0
X_1	-0.23400936	0.39001560	0.54602184	0.70202808
X_2	-0.64639927	0.03264644	0.52125464	0.75775989
X_3	-1.2463278	0.0430531	1.0292117	1.4685821
X_{13}	-1.2571934	0.0436358	1.0389835	1.4821027
X_{18}	-1.2577816	0.0434862	1.0391543	1.4823759
X_{22}	-1.2577930	0.0434873	1.0391655	1.4823918
X_{24}	-1.2577936	0.0434873	1.0391661	1.4823926

§ 88. Универсальный алгорифм, наилучший в смысле второго критерия

Пусть для системы $X = BX + G$ известно, что все собственные значения матрицы B лежат в промежутке $(-\frac{1}{\gamma}, \frac{1}{\gamma})$ при $\gamma > 1$. Универсальный алгорифм будет наилучшим в смысле второго критерия, если определяющий его полином $f(t)$ будет полиномом, наименее отклоняющимся от $\frac{1}{1-t}$ на промежутке $(-\frac{1}{\gamma}, \frac{1}{\gamma})$. Очевидно, что $f(t) = \gamma F(\gamma t)$, где $F(t)$ есть полином, наименее уклоняющийся от функции $\frac{1}{1-t}$ на промежутке $(-1, 1)$.

Полином степени $s = 1$, удовлетворяющий последнему требованию, был известен еще Чебышеву. Именно,

$$F_{s=1}(t) = -\frac{2\alpha^{\frac{s}{2}}}{(1-\alpha)^2} \frac{1}{\gamma-t} [T_s(t) - 2\sqrt{\alpha} T_{s-1}(t) + \alpha T_{s-2}(t)] + \frac{1}{\gamma-t},$$

где

$$\alpha = (\gamma - \sqrt{\gamma^2 - 1})^2, \quad T_s(t) = \cos s \arccos t.$$

Таким образом,

$$f_{s=1}(t) = \frac{1}{1-t} - \frac{2\alpha^{\frac{s}{2}}}{(1-\alpha)^2} \frac{1}{1-t} [T_s(\gamma t) - 2\sqrt{\alpha} T_{s-1}(\gamma t) + \alpha T_{s-2}(\gamma t)].$$

В работе М. К. Гавурина [1], впервые предложившего такой выбор полинома $f(t)$, рекомендуется вычислять коэффициенты полинома $f(t)$ и затем искать решение как линейную комбинацию векторов $B^i r_0$ согласно схеме п. 4, а) § 86. При больших степенях s происходит сильное уничтожение значащих цифр.

Значительно более удобная вычислительная схема получается, если перейти к полиному $e_s(t)$. Имеем

$$e_s(t) = 1 - (1-t)f_{s-1}(t) =$$

$$= \frac{2\alpha^{\frac{s}{2}}}{(1-\alpha)^2} [T_s(\gamma t) - 2\sqrt{\alpha} T_{s-1}(\gamma t) + \alpha T_{s-2}(\gamma t)].$$

Положим

$$\tilde{e}_s(t) = \alpha^{-\frac{s}{2}} e_s(t) = \frac{2}{(1-\alpha)^2} [T_s(\gamma t) - 2\sqrt{\alpha} T_{s-1}(\gamma t) + \alpha T_{s-2}(\gamma t)].$$

Полиномы $T_i(\gamma t)$ связаны рекуррентным соотношением

$$T_i(\gamma t) = 2\gamma t T_{i-1}(\gamma t) - T_{i-2}(\gamma t)$$

с коэффициентами, не зависящими от номера полинома. Поэтому любая комбинация нескольких соседних полиномов Чебышева связана соотношениями такого же вида. В частности,

$$\tilde{e}_i(t) = 2\gamma t \tilde{e}_{i-1}(t) - \tilde{e}_{i-2}(t).$$

Умножив на $\alpha^{\frac{i}{2}}$ и переходя к полиномам $e_i(t)$, получим

$$e_i(t) = 2\gamma \sqrt{\alpha} e_{i-1}(t) - \alpha e_{i-2}(t) = (1+\alpha)t e_{i-1}(t) - \alpha e_{i-2}(t),$$

ибо $2\gamma \sqrt{\alpha} = 2\sqrt{\alpha} \frac{\sqrt{\alpha} + \sqrt{\alpha-1}}{2} = 1 + \alpha$. При этом

$$e_1(t) = \left(\frac{1+\alpha}{1-\alpha}\right)^2 (t-1) + 1, \quad e_2(t) = t e_1(t) + \frac{2\alpha}{1-\alpha} (t-1).$$

Таким образом, последовательные приближения можно вычислять по рекуррентным соотношениям

$$X_i = (1+\alpha)(BX_{i-1} + G) - \alpha X_{i-2}$$

(согласно п. 8 § 86), начиная с начальных приближений X_1 и X_2 , которые вычисляются по формулам

$$X_1 = X + \left(\frac{1+\alpha}{1-\alpha}\right)^2 (BX + G - X)$$

$$X_2 = BX_1 + G + \frac{2\alpha}{1-\alpha} (BX + G - X).$$

При численном осуществлении алгорифма, наилучшего в смысле 2-го критерия, нужно располагать такой же информацией о расположении собственных значений матрицы коэффициентов, как и при пользовании алгорифмом, наилучшим в смысле 1-го критерия. Мы покажем ход процесса на примере системы (9) § 23, подготовленной так же, как на стр. 569.

Здесь $\alpha = 0.26204959$. Имеем

X_0	0	0	0	0
X_1	0.6844342	1.1407237	1.5970131	2.0533026
X_2	-1.5527282	-0.4096498	0.3597126	0.6343461
X_3	-0.7490792	0.3334810	1.1875589	1.6324174
X_8	-1.2636344	0.0348866	1.0266917	1.4668835
X_{13}	-1.2571287	0.0438702	1.0393853	1.4826318
X_{18}	-1.2578043	0.0434744	1.0391495	1.4823719
X_{24}	-1.2577941	0.0434870	1.0391660	1.4823925
X_{25}	-1.2577935	0.0434875	1.0391665	1.4823930

§ 89. Прием А. А. Абрамова для ускорения сходимости метода последовательных приближений при решении систем линейных уравнений

А. А. Абрамовым [1] предложен следующий прием ускорения сходимости процесса последовательных приближений для системы, записанной в виде $X = BX + G$, где B — матрица, все собственные значения которой вещественны и лежат в промежутке $(-1, 1)$.

Время от времени нормальное течение процесса последовательных приближений, происходящего по формуле

$$X^{(k)} = BX^{(k-1)} + G, \quad (1)$$

применение которой мы для краткости будем называть B -шагом, прерывается одним или несколькими более сложными B_2 -шагами, B_4 -шагами, к описанию которых мы переходим.

B_2 -шаг заключается в построении приближения $\bar{X}^{(k+2)}$ по формуле

$$\bar{X}^{(k+2)} = 2X^{(k+2)} - X^{(k)}, \quad (2)$$

где $X^{(k+2)}$ второе последовательное приближение, построенное из $X^{(k)}$, т. е.

$$\begin{aligned} X^{(k+2)} &= BX^{(k+1)} + G \\ X^{(k+1)} &= BX^{(k)} + G. \end{aligned} \quad (3)$$

Таким образом, B_2 -шаг требует применения двух B -шагов и составления одной линейной комбинации.

Далее, B_4 -шаг заключается в построении приближения $\bar{X}^{(k+4)}$ по формуле

$$\bar{X}^{(k+4)} = 2\bar{X}^{(k+4)} - X^{(k)}, \quad (4)$$

где $\bar{X}^{(k+4)}$ получается из $\bar{X}^{(k)}$ двукратным применением B_2 -шага.

Аналогично определяются B_8 -шаг, B_{16} -шаг и т. д.

Как правило, при практических вычислениях следует ограничиться употреблением лишь B , B_2 , B_4 , B_8 -шагов, чередуя их друг с другом в некоторой последовательности.

Из описания следует, что вычислительная схема процесса Абрамова почти не сложнее вычислительной схемы классического метода последовательных приближений. Объем же вычислений для процесса $B^{k_0}B^{k_1}B^{k_2}B^{k_3}$ (т. е. процесса, состоящего из k_0 B -шагов, k_1 B_2 -шагов и т. д.) лишь немного больше, чем объем вычислений при $k_0 + 2k_1 + 4k_2 + 8k_3$ шагах процесса последовательных приближений.

Поясним теперь, почему применение процесса Абрамова дает лучший результат по сравнению с эквивалентным по объему вычислений результатом, полученным по методу последовательных приближений.

Для этой цели вычислим множители затухания в компонентах векторов ошибки для отдельных шагов процесса Абрамова.

Ясно, что

$$Y^{(k+1)} = BY^{(k)}$$

$$\bar{Y}^{(k+2)} = 2Y^{(k+2)} - Y^{(k)} = (2B^2 - E)Y^{(k)} = B_2Y^{(k)} = T_2(B)Y^{(k)}$$

$$\bar{Y}^{(k+4)} = (2B_2^2 - E)Y^{(k)} = B_4Y^{(k)} = T_2(B_2)Y^{(k)} = T_4(B)Y^{(k)}.$$

Здесь полиномы $T_2(t)$, $T_4(t)$, ... суть не что иное, как полиномы Чебышева

$$T_s(t) = \cos s \arccos t.$$

Это очевидно для $s=2$, для $s=4, 8, \dots$ это верно в силу известного соотношения

$$T_{s_1 s_2}(t) = T_{s_1}(T_{s_2}(t)).$$

На рис. 20 даны графики функций t^2 и $|T_2(t)| = |2t^2 - 1|$, на рис. 21 графики функций t^4 и $|T_4(t)|$, на рис. 22 графики функций

$$|T_2^2(t)| \text{ и } |T_4(t)|.$$

Рассмотрение этих графиков позволяет сравнить множители затухания компонент в разложении вектора ошибки для почти эквивалентных по объему вычислений двух B -шагов и одного B_2 -шага (рис. 20), четырех B -шагов и одного B_4 -шага (рис. 21) и, наконец, двух B_2 -шагов и одного B_4 -шага (рис. 22).

Именно из рис. 20 мы видим, что два B -шага выгоднее одного B_2 -шага при $0 < t < \frac{1}{\sqrt{3}} \approx 0.58$. Но один B_2 -шаг значительно выгоднее двух B -шагов при $t > \frac{1}{\sqrt{3}}$. Поэтому „подавив“ достаточно компоненты вектора ошибки для собственных значений из промежутка

$0 \leq t \leq \frac{1}{\sqrt{3}}$ методом последовательных приближений, следует перейти к B_2 -шагам.

Далее, B_4 -шаг оказывается более выгодным, чем два B_2 -шага для промежутка $0.89 < t < 1$ (рис. 22). Из рис. 21 следует также, что один B_4 -шаг выгоднее четырех B -шагов для промежутка $0.86 < t < 1$. Поэтому применение B_4 -шагов следует начинать после того, как B и B_2 -шагами уже подавлены компоненты вектора ошибки для промежутка $0 < t < 0.89$.

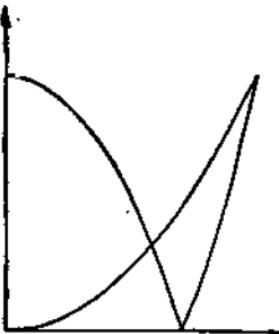


Рис. 20.



Рис. 21.

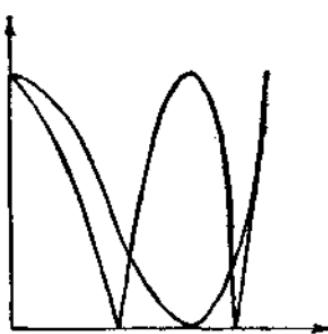


Рис. 22.

Время от времени следует возвращаться к низшим шагам, для подавления накопившихся ошибок округления на компонентах, отвечающих малым собственным значениям.

Мы не приводим здесь численного примера, так как метод Абрамова укладывается как частный случай в группу BT -процессов, имеющих более единообразную вычислительную схему.

§ 90. BT -процессы

Метод Абрамова может быть несколько обобщен без усложнения вычислительной схемы. Именно, процесс Абрамова можно рассматривать как частный случай „ BT -процессов“, которые заключаются в следующем. Пусть дана система

$$X = BX + G \quad (1)$$

с матрицей B , имеющей вещественные собственные значения, расположенные в интервале $(-1, 1)$. Пусть далее дана последовательность букв B и T , начинающаяся с буквы B . Например,

$$BB\ TTT\ BT\ BT\ BB\ TT\dots \quad (2)$$

В соответствии с этой последовательностью строим последовательные приближения $X^{(0)}, X^{(1)}, X^{(2)}, \dots$, задавшись начальным приближением $X^{(0)}$ произвольно. Приближение $X^{(k)}$ строим следующим

образом: если в последовательности (2) на k -м месте находится буква B , то полагаем

$$X^{(k)} = BX^{(k-1)} + G. \quad (3)$$

Если же на k -м месте находится буква T , то полагаем

$$X^{(k)} = 2[BX^{(k-1)} + G] - X^{(k-2)}. \quad (4)$$

Выясним, как изменяются компоненты вектора ошибки в *BT*-процессе. С этой целью разобьем последовательность (2), определяющую процесс на „слова“ так, чтобы каждое слово содержало букву *B* один раз и с нее начиналось. Например,

$$BB\,TTT\,BT\,BT\,BB\,TT \dots = B(BTTT)(BT)(BT)B(BTT) \dots = \\ = B_1B_4B_2B_5B_3B_1B_3 \dots$$

Здесь

$$B_s = BT \dots T = BT^{s-1}, \quad B_1 = B.$$

Ясно, что

$$Y^{(k)} = BY^{k-1},$$

если на k -м шагу находится буква B и

$$Y^{(k)} = 2BY^{(k-1)} - Y^{(k-2)},$$

если на k -м месте находится буква T .

Рассмотрим, как влияет на вектор ошибки применение слова B_8 .
Имеем

$$Y^{(k+1)} = BY^{(k)} = T_1(B)Y^{(k)}$$

$$Y^{(k+2)} = 2BY^{(k+1)} - Y^{(k)} = (2B^2 - E)Y^{(k)} = T_2(B)Y^{(k)}$$

$$Y^{(k+3)} = 2BY^{(k+2)} - Y^{(k+1)} = (2BT_s(B) - T_1(B))Y^{(k)} = T_s(B)Y^{(k)}$$

$$Y^{(k+8)} = [2BT_{s-1}(B) - T_{s-2}(B)] Y^{(k)} = T_s(B) Y^{(k)}.$$

Здесь полиномы $T_i(t)$ определяются рекуррентными соотношениями

$$T_i(t) = 2tT_{i-1}(t) - T_{i-2}(t)$$

при начальных условиях $T_0(t) = 1$, $T_1(t) = t$ и, следовательно, совпадают с полиномами Чебышева $\cos t \arccos t$. Таким образом, применение слова B_g влечет за собой умножение компонент вектора ошибок на множители $T_s(\mu_1), \dots, T_s(\mu_i)$. Применение же последовательности из букв B и T влечет за собой умножение каждой компоненты вектора ошибок на произведение значений в соответствующей точке полиномов Чебышева, отвечающих отдельным словам B_g , из которых состоит определяющая последовательность (2).

Процессы Абрамова являются частными случаями BT -процессов, в которых определяющие последовательности состоят из слов B, B_2, B_4, B_8, \dots . Метод последовательных приближений записывается как $BBBB\dots$ -процесс.

Так же как в процессах Абрамова, применение слов с высокими номерами ускоряет убывание компонент, отвечающих лишь собственным значениям матрицы B , близким по модулю к единице. Поэтому целесообразно вначале работать только со словами B , вводя более длинные слова на последующих стадиях и возвращаясь к более коротким словам для погашения ошибок от округлений.

Применим BT -процесс для нахождения решения системы (9) § 23, подготовленной с $h = \frac{2}{2.62}$ ((3) § 31).

Применение слова

$$B^{10}(BT)^8 B^2 (BTT)^2 (BT)^2$$

дает

$$X = (-1.2577936, \quad 0.0434873, \quad 1.0391661, \quad 1.4823926)'.$$

Взятое слово состоит из 38 букв, так что для получения приближения потребовалось вычисление 38 итераций.

Как мы видели (§ 31), метод последовательных приближений дал такую же точность при 75 итерациях.

Слово

$$B^{10}(BT)^4 (BTT)^3 (BTTT)^2 B^2 (BT)^3,$$

состоящее из 43 букв, дает

$$X = (-1.2577945, \quad 0.0434870, \quad 1.0391667, \quad 1.4823935)'.$$

Это приближение хуже предыдущего.

Дополнив взятое слово еще буквами $T B B T$, получим

$$X = (-1.2577936, \quad 0.0434874, \quad 1.0391662, \quad 1.4823927)'.$$

Более удачный выбор первого слова по сравнению со вторым объясняется тем, что обусловленность взятой системы такова, что нет необходимости прибегать к словам, более длинным, чем BTT .

Разумный выбор слова, управляющего BT -процессом, может быть осуществлен посредством следующего „ BT -процесса с управлением“. По ходу этого процесса параллельно с последовательными приближениями X_k должны вычисляться три системы вспомогательных векторов — невязки r_k , векторы $s_k = Br_k$ и векторы $\bar{s}_k = 2s_k - r_{k-1}$. Положим

$$Z_{k+1} = BX_k + O$$

$$\bar{Z}_{k+1} = 2(BX_k + O) - X_{k-1}.$$

Эти векторы получаются из вектора X_k применением B -шага и T -шага. Вычисление векторов Z_{k+1} и \bar{Z}_{k+1} не требует проведения итераций, ибо

$$\begin{aligned} Z_{k+1} &= X_k + r_k \\ \bar{Z}_{k+1} &= 2(X_k + r_k) - X_{k-1}. \end{aligned}$$

Вычислим их невязки w_{k+1} и \bar{w}_{k+1} . Легко видеть, что

$$\begin{aligned} w_{k+1} &= Br_k = s_k \\ \bar{w}_{k+1} &= 2Br_k - r_{k-1} = 2s_k - r_{k-1} = \bar{s}_k. \end{aligned}$$

Сравним нормы векторов s_k и \bar{s}_k и положим

$$\begin{aligned} X_{k+1} &= Z_{k+1} = X_k + r_k \\ r_{k+1} &= s_k, \end{aligned} \tag{5}$$

если $\|s_k\| \leq \|\bar{s}_k\|$, и

$$\begin{aligned} X_{k+1} &= \bar{Z}_{k+1} = 2(X_k + r_k) - X_{k-1} \\ r_{k+1} &= \bar{s}_k, \end{aligned} \tag{6}$$

если $\|s_k\| > \|\bar{s}_k\|$.

Затем вычисляются векторы $s_{k+1} = Br_{k+1}$ и $\bar{s}_{k+1} = 2s_{k+1} - r_k$, и процесс продолжается дальше. Время от времени, для уменьшения ошибок округления, следует непосредственно вычислять невязку r_{k+1} по формуле $r_{k+1} = BX_{k+1} + G - X_{k+1}$, а не по формулам (5) или (6).

Применение BT -процесса с управлением к прежнему примеру привело к слову $BTBTBTBTBTBBBTBTBBBTBTBBBTBTBBBTBTBBBTBT = = (BT)^6 B(BT)B^2(BT)^2 B(BT)B(BT)B(BT)B^2(BT)^2$ (34 буквы), в результате получено приближение

$$(-1.2577936, \quad 0.0434874, \quad 1.0391661, \quad 1.4823928)',$$

отличающееся от точного решения не более чем на $2 \cdot 10^{-7}$ в каждой компоненте. Невязки вычислялись непосредственно на 11, 16, 21, 26 и 31-м шагах.

§ 91. Общие трехчленные итерационные процессы

Изложенные выше BT -процессы и наилучший в смысле 1-го критерия алгоритм укладываются в следующую общую схему трехчленных универсальных процессов¹⁾.

Для решения системы

$$X = BX + G \tag{1}$$

¹⁾ Д. К. Фаддеев [2].

с матрицей B , собственные значения которой заключены в промежуток $(-1, 1)$, строится, исходя из некоторого начального приближения $X^{(0)}$, последовательность приближений по формулам

$$X^{(1)} = BX^{(0)} + G \quad (2)$$

$$X^{(k)} = (1 + \alpha_k)(BX^{(k-1)} + G) - \alpha_k X^{(k-2)}. \quad (3)$$

Первый шаг можно формально рассматривать протекающим по общей формуле (3), считая $\alpha_1 = 0$.

Ясно, что если $\alpha_{s+1} = 0$, то процесс, начиная с $s+1$ -го шага, протекает так, как будто s -е приближение принято за начальное, и далее применяется процесс при $\alpha'_2 = \alpha_{s+2}$, $\alpha'_3 = \alpha_{s+3}$, ... В частности, циклическое повторение процесса с некоторыми $\alpha_2, \dots, \alpha_s$ равносильно применению единого процесса, в котором последовательность $\alpha_1, \alpha_2, \dots, \alpha_s, \alpha_{s+1}, \dots$ состоит из циклически повторяющихся отрезков $0, \alpha_2, \dots, \alpha_s$.

В методе последовательных приближений все $\alpha_k = 0$, в BT -процессах последовательность α_k состоит из нулей и единиц, в оптимальном процессе $\alpha_k = \frac{T_{k-2}(\gamma)}{T_k(\gamma)}$.

Векторы ошибки в трехчленном процессе удовлетворяют соотношениям

$$Y^{(1)} = BY^{(0)}$$

$$Y^{(k)} = (1 + \alpha_k)BY^{(k-1)} - \alpha_k Y^{(k-2)} \quad (k = 2, 3, \dots),$$

так что

$$Y^{(k)} = P_k(B)Y^{(0)},$$

где $P_k(t)$ последовательность полиномов, построенных по рекуррентным соотношениям

$$P_0(t) = 1, \quad P_1(t) = t,$$

$$P_k(t) = (1 + \alpha_k)tP_{k-1}(t) - \alpha_k P_{k-2}(t).$$

Теорема 91. 1 Полиномы $P_k(t)$ удовлетворяют неравенствам $|P_k(t)| \leq 1$ при $-1 \leq t \leq 1$, $-1 \leq \alpha_k \leq 1$ ($k = 2, 3, \dots$).

Доказательство. Обозначим $P_k(t) = \Phi(t, \alpha_2, \dots, \alpha_k)$. При фиксированном t полином $\Phi(t, \alpha_2, \dots, \alpha_k)$ есть линейная функция от каждого из аргументов $\alpha_2, \dots, \alpha_k$. Поэтому при изменении параметров $\alpha_2, \dots, \alpha_k$ в кубе $-1 \leq \alpha_2 \leq 1, \dots, -1 \leq \alpha_k \leq 1$ функция $\Phi(t, \alpha_2, \dots, \alpha_k)$ принимает экстремальные значения в одной из вершин куба, т. е.

$$|\Phi(t, \alpha_2, \dots, \alpha_k)| \leq |\Phi(t_1, \varepsilon_2, \dots, \varepsilon_k)|,$$

где

$$\varepsilon_2 = \pm 1, \dots, \varepsilon_k = \pm 1.$$

Покажем, что

$$\bar{P}_k(t) = \Phi(t, \varepsilon_2, \dots, \varepsilon_k) = T_{s_k}(t) = \cos s_k \arccos t,$$

причем чебышевский номер s_{k+1} каждого последующего полинома на единицу больше или на единицу меньше номера s_k предшествующего.

Действительно, это верно для $k=0, k=1$, ибо

$$\bar{P}_0(t) = 1 = T_0(t), \quad \bar{P}_1(t) = t = T_1(t).$$

Допустим, что это верно для полиномов $\bar{P}_{k-1}(t)$ и $\bar{P}_k(t)$, т. е. допустим, что $\bar{P}_k(t) = T_{s_k}(t)$ и $\bar{P}_{k-1}(t) = T_{s_{k-1}}(t)$. Тогда

$$\begin{aligned} \bar{P}_{k+1}(t) &= (1 + \varepsilon_{k+1}) t \bar{P}_k(t) - \varepsilon_{k+1} \bar{P}_{k-1} = \\ &= (1 + \varepsilon_{k+1}) t T_{s_k}(t) - \varepsilon_{k+1} T_{s_{k-1}}(t). \end{aligned}$$

При $\varepsilon_{k+1} = -1$ получим

$$\bar{P}_{k+1}(t) = T_{s_{k+1}}(t).$$

При $\varepsilon_{k+1} = 1$ получим

$$\bar{P}_{k+1}(t) = 2tT_{s_k}(t) - T_{s_{k-1}}(t) = T_{s_{k+1}}(t).$$

В обоих случаях полиномы $\bar{P}_{k+1}(t)$ и $\bar{P}_k(t)$ оказались полиномами Чебышева с соседними номерами.

Тем самым

$$|P_k(t)| = |\Phi(t, \alpha_2, \dots, \alpha_k)| \leq |\bar{P}_k(t)| = |T_{s_k}(t)| \leq 1,$$

что и требовалось доказать.

Можно доказать, что если $0 < \alpha_k < \alpha < 1$, то полиномы $P_k(t)$ равномерно стремятся к нулю во всяком промежутке $-\gamma \leq t \leq \gamma < 1$, так что трехчленный итерационный процесс в этих условиях будет сходящимся.

Отметим некоторые интересные частные случаи трехчленных универсальных алгорифмов сверх указанных в начале параграфа.

1. Трехчленный алгорифм с постоянным α . Пусть $\alpha_k = \alpha$, $0 < \alpha < 1$ при всех $k \geq 2$. Тогда, как легко видеть,

$$P_k(t) = \alpha^{\frac{k}{2}} T_k(\gamma t) + \frac{1-\alpha}{2\sqrt{\alpha}} \alpha^{\frac{k}{2}} t U_{k-1}(\gamma t),$$

где

$$\gamma = \frac{1+\alpha}{2\sqrt{\alpha}}, \quad \alpha = (\gamma - \sqrt{\gamma^2 - 1})^2, \quad T_k = \cos k \arccos t,$$

$$U_k(t) = \frac{\sin(k+1) \arccos t}{\sin \arccos t}.$$

Поэтому при $-\frac{1}{\gamma} \leq t \leq \frac{1}{\gamma}$ имеем

$$|P_k(t)| \leq \alpha^{\frac{k}{2}} \left(1 + \frac{1-\alpha}{1+\alpha} k\right) = (\gamma - V\sqrt{\gamma^2 - 1})^k \left(1 + \frac{1-\alpha}{1+\alpha} k\right).$$

Таким образом, если все собственные значения матрицы B заключены в интервале $(-\frac{1}{\gamma}, \frac{1}{\gamma})$, процесс с постоянным $\alpha_k = \alpha = (\gamma - V\sqrt{\gamma^2 - 1})^2$ сходится почти столь же быстро, как оптимальный процесс.

2. Универсальный алгорифм с полиномами Чебышева 2-го рода. Пусть $\alpha_k = \frac{k-1}{k+1}$. В этом случае

$$P_k(t) = \frac{1}{k+1} \frac{\sin(k+1)\arccos t}{\sin \arccos t} = \frac{1}{k+1} U_k(t)$$

есть полином Чебышева второго рода, нормированный условием $P_k(1) = 1$.

Очевидно, что последовательность $P_k(t)$ равномерно стремится к нулю во всяком сегменте, внутреннем для промежутка $(-1, 1)$, хотя сформулированное выше достаточное условие сходимости здесь не выполнено. Процесс будет сходиться довольно медленно, однако его достоинство состоит в том, что даже при небольших k происходит ощутимое погашение компонент в широком интервале.

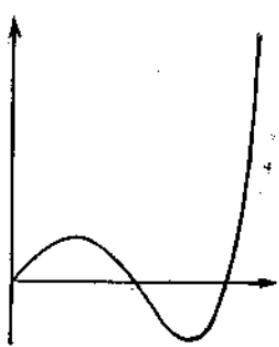


Рис. 23.

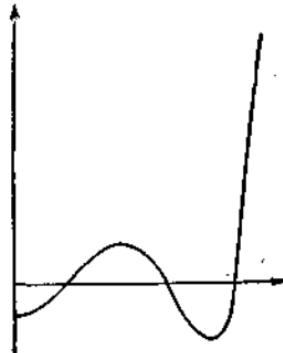


Рис. 24.

Действительно, ближайший к 1 максимум модуля σ_k полинома $P_k(t)$ асимптотически равен $|\cos \rho_1| = 0.217$, где ρ_1 есть наименьший положительный корень трансцендентного уравнения $\operatorname{tg} t = t$. Достигается этот максимум в точке $t_1 = \cos \frac{\rho_1}{k+1} \approx 1 - \frac{\rho_1^2}{2(k+1)^2}$, близкой к единице.

На рис. 23 и 24 даны графики полиномов $P_6(t)$ и $P_6(t)$, в табл. 91.1 приведены значения внутреннего максимума модуля σ_k полинома $P_k(t)$ для $k = 2, 3, \dots, 6$ и границы промежутка $(-\gamma_k, \gamma_k)$, в котором $|P_k(t)| \leq \sigma_k$.

Описанный универсальный алгорифм целесообразно применять циклически при не очень больших k . Применение m циклов подавляет компоненты вектора ошибки из интервала $(-\gamma_k, \gamma_k)$ с интенсивностью σ_k^m .

Приводим результаты решения системы (9) § 23, подготовленной при $h = \frac{2}{2.62}$ ((3) § 31) по трехчленному алгорифму с постоянным α .

Таблица 91.1

Значения σ_k и γ_k

k	σ_k	γ_k
2	0.333	0.707
3	0.272	0.816
4	0.250	0.878
5	0.239	0.913
6	0.233	0.935

$$\alpha = \frac{1}{2}$$

X_0	0.22900763	0.38167939	0.53435115	0.68702290
X_1	-0.4055708	0.0372933	0.3577880	0.5162869
X_2	-0.5442885	0.1987509	0.7684031	1.0678660
X_{18}	-1.2593757	0.0436874	1.0408131	1.4846543
X_{25}	-1.2574775	0.0435276	1.0389918	1.4821385
X_{30}	-1.2578296	0.0434832	1.0391862	1.4824220
X_{40}	-1.2577930	0.0434875	1.0391659	1.4823924
X_{41}	-1.2577939	0.0434874	1.0391664	1.4823932

$$\alpha = \frac{1}{3}$$

X_0	0.22900763	0.38167939	0.53435115	0.68702290
X_1	-0.4055708	0.0372939	0.3577880	0.5162869
X_2	-0.4583667	0.2190763	0.7423973	1.0255501
X_{18}	-1.2577739	0.0434788	1.0391330	1.4823466
X_{25}	-1.2577920	0.0434873	1.0391649	1.4823909
X_{30}	-1.2577939	0.0434873	1.0391662	1.4823929

Приведем еще результаты применения к той же системе алгорифма с полиномами Чебышева 2-го рода.

Циклическое проведение процесса через четыре шага укладывается в общую схему при $\alpha_1 = 0$, $\alpha_2 = \frac{1}{3}$, $\alpha_3 = \frac{1}{2}$, $\alpha_4 = \frac{3}{5}$, $\alpha_5 = 0$, $\alpha_6 = \frac{1}{3}$, $\alpha_7 = \frac{1}{2}$, $\alpha_8 = \frac{3}{5}$, $\alpha_9 = 0$; ... Получим

X_0	0	0	0	0
X_4	-1.2048821	0.0715392	1.0432993	1.4836599
X_8	-1.2594577	0.0409823	1.0371510	1.4800347
X_{12}	-1.2576202	0.0437954	1.0393241	1.4825350
X_{16}	-1.2578018	0.0434441	1.0391507	1.4823855
X_{20}	-1.2577945	0.0434944	1.0391677	1.4823922
X_{24}	-1.2577936	0.0434859	1.0391659	1.4823933
X_{28}	-1.2577938	0.0434876	1.0391663	1.4823927

Применение полиномов более высоких степеней в данном случае дает худший результат. Именно, процесс через шесть шагов при $\alpha_1 = 0$, $\alpha_2 = \frac{1}{3}$, $\alpha_3 = \frac{1}{2}$, $\alpha_4 = \frac{3}{5}$, $\alpha_5 = \frac{2}{3}$, $\alpha_6 = \frac{5}{7}$, $\alpha_7 = \alpha_1$, ... дает

X_0	0	0	0	0
X_6	-1.5437461	0.0461754	1.2796197	1.8191787
X_{12}	-1.1941027	0.0504107	0.9848809	1.4059616
X_{18}	-1.2734516	0.0443237	1.0509812	1.4992437
X_{24}	1.2543485	0.0436402	1.0363966	1.4784862
X_{30}	1.2586028	0.0434955	1.0397846	1.4832740

§ 92. Универсальный алгорифм Ланцоша

Удобный универсальный алгорифм был предложен Ланцошем [6]. Мы изложим его с некоторыми незначительными изменениями, касающимися предварительной подготовки системы. Именно, будем считать, что система подготовлена к виду

$$X = BX + G \quad (1)$$

(вместо $Y = \frac{1}{2}BY + 2c_0$ в обозначениях автора) и все собственные значения матрицы B находятся в промежутке $(-1, 1)$.

В качестве полинома $e_s(t)$ берется

$$e_s(t) = \frac{1 - T_{s+1}(t)}{(s+1)^2(1-t)}, \quad (2)$$

где T_{s+1} полином Чебышева. Очевидно, что $e_s(t)$ действительно есть полином, ибо $T_{s+1}(1) = 1$, и, следовательно, $1 - T_{s+1}(t)$ делится на $1 - t$. Легко проверить, что $e_s(1) = 1$. Действительно, положив $t = \cos \theta$, получим, что

$$e_s(t) = \frac{1 - \cos(s+1)\theta}{(s+1)^2(1 - \cos\theta)} = \frac{\sin^2 \frac{s+1}{2}\theta}{(s+1)^2 \sin^2 \frac{\theta}{2}} \quad (3)$$

и, следовательно,

$$e_s(1) = \lim_{\theta \rightarrow 0} \frac{\sin^2 \frac{s+1}{2}\theta}{(s+1)^2 \sin^2 \frac{\theta}{2}} = 1.$$

Таблица 92.1

Значения σ_s и γ_s

s	σ_s	γ_s
3	0.0741	0.334
4	0.0625	0.541
5	0.0572	0.683
6	0.0543	0.750
7	0.0525	0.806

Из формулы (3) видно, что полином $e_s(t)$ превращается при замене $t = \cos \theta$ в ядро Фейера $K_s(\theta)$, нормированное условием $K_s(0) = 1$. Приведем графики $e_5(t)$ и $e_6(t)$ (рис. 25 и рис. 26).

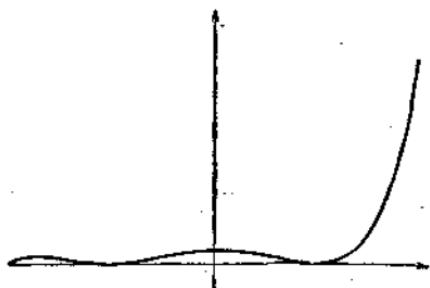


Рис. 25.

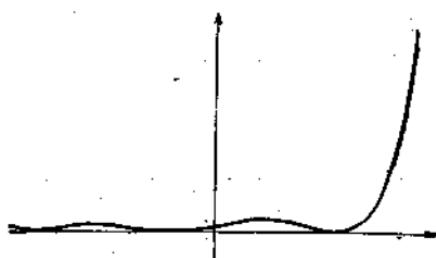


Рис. 26.

Последовательные максимумы полинома $e_s(t)$ убывают при движении по оси абсцисс справа налево. Наибольший внутренний максимум σ_s стремится к пределу 0,0472 при $s \rightarrow \infty$. В табл. 92.1 приведены значения внутреннего максимума σ_s полиномов $e_s(t)$ для $s = 3, \dots, 6$ и граница γ_s промежутка $[-1, \gamma_s]$, в котором $|e_s(t)| \leq \sigma_s$.

Сравнение с табл. 91.1 показывает, что при том же номере s значения максимума модуля в алгоритме Ланцоша значительно меньше, чем в алгоритме с полиномами Чебышева 2-го рода. Числа же γ_s приближаются к 1 в последнем алгоритме более быстро.

Универсальный алгорифм Ланцоша проводится по схеме, использующей рекуррентные соотношения для полинома

$$f_{s-1}(t) = \frac{1 - e_s(t)}{1 - t}.$$

Выведем эти соотношения. Имеем

$$f_{s-1}(t) = \frac{1 - e_s(t)}{1 - t} = \frac{1 - \frac{1 - T_{s+1}(t)}{(s+1)^2(1-t)}}{1 - t} = \frac{(s+1)^2(1-t) - 1 + T_{s+1}(t)}{(s+1)^2(1-t)^2},$$

Поэтому

$$(s+1)^2(1-t)^2 f_{s-1}(t) - (s+1)^2(1-t) + 1 = T_{s+1}(t).$$

На основании соотношения

$$T_{s+2}(t) = 2tT_{s+1}(t) - T_s(t),$$

получим

$$\begin{aligned} (s+2)^2(1-t)^2 f_s(t) - (s+2)^2(1-t) + 1 &= \\ &= 2t(s+1)^2(1-t)^2 f_{s-1}(t) - 2t(s+1)^2(1-t) + \\ &\quad + 2t - s^2(1-t)^2 f_{s-2}(t) + s^2(1-t) - 1. \end{aligned}$$

После приведения подобных членов и сокращения на $(1-t)^2$ мы придем к соотношению

$$(s+2)^2 f_s(t) = 2t(s+1)^2 f_{s-1}(t) - s^2 f_{s-2}(t) + 2(s+1)^2 \quad (4)$$

с начальными полиномами $f_{-1} = 0$, $f_0 = \frac{1}{2}$.

Введем далее в рассмотрение полиномы

$$F_s(t) = \frac{(s+2)^2}{4} f_s(t). \quad (5)$$

Эти полиномы удовлетворяют еще более простому рекуррентному соотношению

$$F_s(t) = 2tF_{s-1}(t) - F_{s-2}(t) + \frac{1}{2}(s+1)^2 \quad (6)$$

при начальных условиях $F_0(t) = \frac{1}{2}$, $F_1(t) = t + 2$. Последние соотношения позволяют провести универсальный алгорифм следующим образом.

Вычислив невязку начального приближения $r = BX + G - X$, образуем последовательность векторов

$$Z_0 = \frac{1}{2}r$$

$$Z_1 = 2BZ_0 + 4Z_0$$

$$Z_i = 2BZ_{i-1} - Z_{i-2} + (1+t)^2 Z_0 \quad (i = 2, 3, \dots, s-1).$$

Тогда

$$X' = X + \frac{4}{(s+1)^2} Z_{s-1}.$$

В обозначениях Ланцоша $(Y = \frac{1}{2} BY + 2c_0)$ формулы приобретают еще более простой вид

$$Z_0 = \frac{1}{2} r, \quad r = \frac{1}{2} BY + 2c_0 - Y$$

$$Z_1 = BZ_0 + 4Z_0$$

$$Z_i = BZ_{i-1} - Z_{i-2} + (i+1)^2 Z_0$$

и

$$Y' = Y + \frac{4}{(s+1)^2} Z_{s-1}.$$

Рассматриваемый универсальный процесс сходится при безграничном увеличении s очень медленно, так как множители затухания $e_{s+1}(\mu_i)$ на каждом фиксированном промежутке, содержащемся внутри $(-1, +1)$, убывают лишь со скоростью $\frac{1}{s^2}$.

Однако при небольших значениях s он дает значительное подавление компонент вектора ошибки на довольно широком интервале $(-1, \gamma_s)$ для собственных значений.

Циклическое повторение процесса обеспечивает хорошую сходимость для систем с P -числом обусловленности, не превосходящим $\frac{1+\gamma_s}{1-\gamma_s}$.

Ланцош рекомендует брать $s=7$. В этом случае (в применении к системе $Y = \frac{1}{2} BY + 2c_0$)

$$Z_6 = \frac{1}{2}(B^6 + 4B^5 + 4B^4 + 4B^3 + 16B + 32)$$

$$Y' = Y + \frac{1}{16} Z_6.$$

Вычисление вектора Z_6 рекомендуется производить не по рекуррентным соотношениям, а по формулам схемы Хорнера (п. 86, п. 4, б).

Ланцош советует употреблять описанный универсальный алгорифм для систем высокого порядка с целью предварительной подготовки начального приближения при применении метода минимальных итераций, так как этот последний при таком выборе начального приближения достаточно быстро становится вырожденным. Приводим в качестве примера результат циклического применения процесса Ланцоша при $s=9$ для системы (9) § 23, подготовленной при $h = 2/2.62$.

Первый цикл дает

$$X = (-1.2574119, \quad 0.0428473, \quad 1.0387672, \quad 1.4829441)',$$

второй цикл:

$$X = (-1.2577870, \quad 0.0434717, \quad 1.0391658, \quad 1.4823988)',$$

третий цикл:

$$X = (-1.2577935, \quad 0.0434868, \quad 1.0391663, \quad 1.4823931)',$$

В результате третьего цикла (27 итераций) мы получаем компоненты решения с точностью до $3 \cdot 10^{-7}$.

В данном примере четыре цикла при $s = 7$ (28 итераций) дают худший результат

$$X = (-1.2577659, \quad 0.0434909, \quad 1.0391514, \quad 1.4823708)',$$

§ 93. Универсальные алгоритмы, наилучшие в среднем

Как мы видели, при выборе полиномов, подавляющих компоненты вектора ошибки решения системы

$$X = BX + G, \quad (1)$$

при условии, что собственные значения матрицы B заключены в промежутке $(-1, 1)$, нужно руководствоваться двумя плохо совместимыми условиями, именно, малостью уклонения полинома от нуля внутри промежутка $(-1, 1)$ и обращением в единицу на его правом конце.

Так как буквально удовлетворить этим двум требованиям невозможно, естественно пытаться удовлетворить первому из них в смысле малости среднего квадратического уклонения.

Именно, будем строить полином $e_s(t)$, минимизирующий

$$\int_{-1}^1 p(t) e_s^2(t) dt$$

в классе полиномов степени s , удовлетворяющих условию $e_s(1) = 1$. Весовая функция $p(t) > 0$ может выбираться различным образом на основе информации о плотности распределения собственных значений данной матрицы B на промежутке $(-1, 1)$ и информации о характере распределения компонент начального вектора ошибок. Если такого рода информация отсутствует, естественно брать $p(t) = 1$ при $-1 \leq t \leq 1$.

Легко установить, что полиномы $e_s(t)$ образуют ортогональную систему по весу $(1-t)\rho(t)$. Действительно, пусть

$$e_s(t) = 1 + a_1(1-t) + a_2(1-t)^2 + \dots + a_s(1-t)^s.$$

$$J = \int_{-1}^1 \rho(t) e_s^2(t) dt. \quad (2)$$

Тогда очевидно, что $\frac{\partial J}{\partial a_i} = 2 \int_{-1}^1 \rho(t) (1-t)^i e_s(t) dt$. Для экстремального полинома все частные производные равны нулю, так что $e_s(t)$ будет ортогонален по весу $\rho(t)$ к полиномам $(1-t)^i$ при $i = 1, \dots, s$, а по весу $(1-t)\rho(t)$ к полиномам $(1-t)^{i-1}$ при $i = 1, \dots, s$ и, следовательно, к любому полиному, степень которого меньше s . В частности,

$$\int_{-1}^1 (1-t)\rho(t) e_s(t) e_i(t) dt = 0 \quad \text{при } i = 0, 1, \dots, s-1. \quad (3)$$

Из теории ортогональных полиномов следует, что при любой весовой функции полиномы $e_i(t)$ связаны трехчленными рекуррентными соотношениями

$$e_i(t) = (\alpha_i t + \beta_i) e_{i-1}(t) - \gamma_i e_{i-2}(t). \quad (4)$$

Это позволяет строить последовательные приближения к решению системы в универсальном алгорифме, наилучшем в среднем при данном выборе весовой функции, по формулам п. 8 § 86, как только вычислены коэффициенты α_i , β_i и γ_i .

Для $\rho(t) = 1$ соотношения (4) имеют вид

$$e_0(t) = 1, \quad e_1(t) = \frac{3}{4}t + \frac{1}{4}$$

$$e_i(t) = \left[\frac{i(2i+1)}{(i+1)^2} t + \frac{i}{(2i-1)(i+1)^2} \right] e_{i-1}(t) - \frac{(2i+1)(i-1)^2}{(2i-1)(i+1)^2} e_{i-2}(t).$$

Поэтому последовательные приближения вычисляются по формулам
 $X_1 = \frac{3}{4}(BX_0 + O) + \frac{1}{4}X_0$

$$X_i = \frac{i(2i+1)}{(i+1)^2}(BX_{i-1} + O) + \frac{i}{(2i-1)(i+1)^2} X_{i-1} - \frac{(2i+1)(i-1)^2}{(2i-1)(i+1)^2} X_{i-2}.$$

На рис. 27 и 28 даны графики

$$e_6(t) = \frac{77}{32}t^6 + \frac{35}{32}t^5 - \frac{35}{32}t^4 - \frac{35}{48}t^3 + \frac{35}{96}t^2 + \frac{5}{96}t + \frac{5}{112}$$

и

$$e_8(t) = \frac{429}{112}t^8 + \frac{99}{56}t^7 - \frac{495}{112}t^6 - \frac{45}{28}t^5 + \frac{135}{112}t^4 + \frac{15}{56}t^3 - \frac{5}{112}.$$

В табл. 93.1 даны значения σ_s и γ_s , где σ_s и γ_s имеют тот же смысл, что и в §§ 91 и 92.

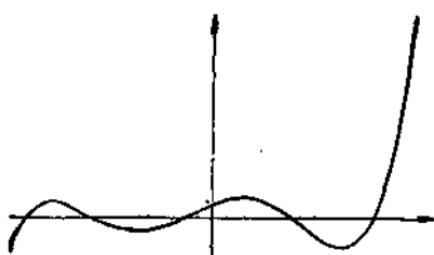


Рис. 27.

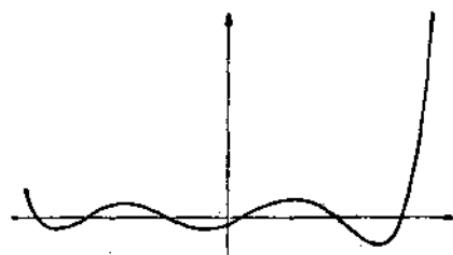


Рис. 28.

Полиномы $e_s(t)$ просто связаны с полиномами Лежандра

Таблица 93.1
Значения σ_s и γ_s

s	σ_s	γ_s
2	$1/3$	0.6
3	$1/4$	0.735
4	$1/6$	0.804
5	$1/8$	0.860
6	$1/7$	0.894

$$L_s(t) = \frac{1}{2^s \cdot s!} \frac{d^s (t^2 - 1)^s}{dt^s}.$$

Именно,

$$e_s(t) = \frac{L_{s+1}(t) - L_s(t)}{(s+1)(t-1)}.$$

При

$$\rho(t) = (1-t)^\alpha (1+t)^\beta (\alpha > -1, \beta > -1)$$

мы придем к полиномам, ортогональным по весу $(1-t)^{\alpha+1}(1+t)^\beta$, т. е. к так называемым полиномам Якоби (гипергеометрическим полиномам). Применение таких полиномов рассматривается в работе Штифеля [5].

Универсальный алгорифм, наилучший в среднем при $\rho(t) = 1$, был применен циклически для нахождения решения системы (9) § 23, подготовленной при $h = 2/2.62$. Было получено

X_0	0	0	0	0
X_6	-1.3813506	0.0338070	1.1209397	1.5985210
X_{12}	-1.2464295	0.0446982	1.0325678	1.4726554
X_{18}	-1.2588065	0.0433614	1.0397375	1.4832186
X_{24}	-1.2577045	0.0434986	1.0391177	1.4823210
X_{30}	-1.2578015	0.0434862	1.0391704	1.4823990
X_{36}	-1.2577939	0.0434873	1.0391663	1.4823930
X_{42}	-1.2577930	0.0434874	1.0391658	1.4823923

Здесь X_{35} оказался ближе к точному решению, чем X_{36} . Это свидетельствует о том, что один из корней полинома $e_5(t)$ оказался близким к тому собственному значению, компонента которого была еще недостаточно подавлена за счет предыдущих циклов.

Вообще, при пользовании специальными полиномами для подавления компонент вектора ошибки целесообразно использовать также „аномальное“ подавление, происходящее от близости одного из корней к собственному значению, подобно тому, как это делается в *ВТ*-процессе „с управлением“.

§ 94. Метод подавления компонент в комплексной области

Все рассмотренные до сих пор универсальные алгоритмы строились для систем с вещественными собственными значениями матрицы коэффициентов.

Идея подавления компонент, в теоретическом плане, легко распространяется и на системы, матрицы которых имеют комплексные собственные значения.

Пусть на плоскости комплексной переменной задано ограниченное замкнутое множество Σ , дополнение к которому есть связная область Δ , содержащая точку $z = 0$.

Рассмотрим класс систем

$$AX = F$$

таких, что все собственные значения матриц A лежат на множестве Σ . Для такого класса систем можно строить (в теоретическом плане) универсальный алгоритм, наилучший в смысле первого критерия.

С этой целью достаточно построить полиномы $g_s(z)$, удовлетворяющие требованию $g_s(0) = 1$ и наименее уклоняющиеся от нуля на множестве Σ , а затем строить приближения к решению системы по формуле

$$X_s = X_0 + h_{s-1}(A)(F - AX_0),$$

$$\text{где } h_{s-1}(t) = \frac{1 - g_s(t)}{t}.$$

Приведем оценку сходимости такого универсального алгоритма при $s \rightarrow \infty$.

Пусть

$$\tau_s(z) = z^s + \dots$$

полином, наименее уклоняющийся от нуля на множестве Σ и такой, что корни полинома $\tau_s(z)$ лежат на множестве Σ . Тогда уклонение от нуля полинома $g_s(z)$ на множестве Σ не более, чем уклонение от нуля полинома $\frac{\tau_s(z)}{\tau_s(0)}$. Обозначим уклонение от нуля полинома $\tau_s(z)$

через τ_s . Тогда, как известно,¹⁾

$$\lim \sqrt[s]{\frac{\tau_s}{|\tau_s(0)|}} = e^{-G(0)},$$

где $G(z)$ функция Грина для области Δ с логарифмической особенностью в точке $z = \infty$. Поэтому, при $z \in \Sigma$

$$|g_s(z)| \leq \frac{\tau_s}{|\tau_s(0)|} \leq C(\varepsilon) e^{-s(G(0)-\varepsilon)},$$

где ε сколь угодно малое положительное число, $C(\varepsilon)$ константа, зависящая от ε .

С другой стороны, пусть $R = \max_{z \in \Sigma} |g_s(z)|$ и Σ^* — совокупность точек, таких, что $|g_s(z)| \leq R$, Σ^* есть ограниченное замкнутое множество, содержащее Σ , и дополнение к нему Δ^* есть связная область, содержащаяся в Δ . Ясно, что $\frac{1}{s} \lg \frac{|g_s(z)|}{R}$ есть функция Грина для области Δ^* . Так как $\Delta^* \subset \Delta$, то будет выполнено неравенство $\frac{1}{s} \lg \frac{|g_s(z)|}{R} \leq G(z)$ для всех $z \in \Delta^*$. Это неравенство будет верно и для всех точек $z \in \Delta$, ибо если $z \in \Delta$ и $z \notin \Delta^*$, то $G(z) \geq 0$, а $|g_s(z)| \leq R$, так что $\frac{1}{s} \lg \frac{|g_s(z)|}{R} \leq 0$. Положив $z = 0$, получим

$$\frac{1}{s} \lg \frac{1}{R} \leq G(0),$$

откуда

$$R = \max_{z \in \Sigma} g_s(z) \geq e^{-sG(0)}.$$

Итак

$$e^{-sG(0)} \leq R \leq C(\varepsilon) e^{-s(G(0)-\varepsilon)}.$$

Таким образом, универсальный алгорифм, наилучший в смысле I-го критерия, сходится с быстротой геометрической прогрессии со знаменателем $e^{-(G(0)-\varepsilon)}$.

Общие способы построения полиномов $g_s(z)$ неизвестны, так что описанный алгорифм может применяться лишь к некоторым частным областям. В § 97 будет установлена возможность построения Σ -универсальных алгорифмов при помощи конформного отображения единичного круга на область Δ (или на ее односвязную накрывающую), которые для односвязных областей Δ обладают почти той же быстротой сходимости.

Технически проще строить универсальный алгорифм, наилучший в среднем. Это сводится к построению полиномов $g_s(t)$, минимизи-

¹⁾ Г. М. Голузин. Геометрическая теория функций комплексного переменного, ГИТТЛ, 1952, гл. VII.

рующих $\int |g_s(z)|^2 d\mu$, где μ некоторая неотрицательная мера, сосредоточенная на множестве Σ .

Нетрудно дать формулы для построения полиномов $g_s(z)$, как только известна система ортогональных полиномов по мере μ . Пусть $\{P_i(z)\}$ — ортогональная система полиномов по мере μ . Будем искать полином $g_s(z)$ в виде

$$g_s(z) = \sum_{i=0}^s c_i P_i(z).$$

Тогда

$$I = \int |g_s(z)|^2 d\mu = |c_0|^2 + |c_1|^2 + \dots + |c_s|^2.$$

Условие $g_s(0) = 1$ примет вид

$$\sum_{i=0}^s c_i P_i(0) = 1.$$

По неравенству Коши — Буняковского

$$\sum_{i=0}^s |c_i|^2 \sum_{i=0}^s |P_i(0)|^2 \geq \left| \sum_{i=0}^s c_i P_i(0) \right|^2 = 1,$$

откуда

$$I \geq \frac{1}{\sum_{i=0}^s |P_i(0)|^2}.$$

Равенство достигается при $c_i = \frac{\overline{P_i(0)}}{\sum_{i=0}^s |P_i(0)|^2}$, где $\overline{P_i(0)}$ число комплексно-сопряженное с $P_i(0)$. Следовательно,

$$g_s(z) = \frac{1}{\sum_{i=0}^s |P_i(0)|^2} \sum_{i=0}^s \overline{P_i(0)} P_i(z).$$

Ортогональные полиномы по данной мере для разных множеств Σ изучались в классических работах Сеге, Бехнера и В. И. Смирнова.

§ 95. Применение конформного отображения к решению линейных систем

До сих пор мы рассматривали универсальные алгоритмы для систем $X = BX + C$ при условии, что все собственные значения матрицы B вещественные и заключены в промежутке $(-1, 1)$.

В. Н. Кублановской¹⁾ разработан метод, использующий аппарат теории функций комплексной переменной, который позволяет строить универсальные алгоритмы для систем $X = BX + G$ при других предположениях о расположении собственных значений матрицы B .

Вместо системы

$$X = BX + G \quad (1)$$

изучается система

$$X = zBX + G, \quad (2)$$

содержащая комплексный параметр z . Решение последней системы есть, очевидно, $X(z) = (E - zB)^{-1}G$. Все компоненты решения являются аналитическими и даже рациональными функциями от z с полюсами только в точках $\frac{1}{\mu_1}, \dots, \frac{1}{\mu_n}$, где μ_1, \dots, μ_n собственные значения матрицы B . Действительно,

$$X(z) = (E - zB)^{-1}G = \frac{1}{|E - zB|}C(z)G, \quad (3)$$

где $C(z)$ союзная с $E - zB$ матрица. Элементы $C(z)$ являются, очевидно, полиномами от z , знаменателя же знаменателя являются числа, обратные к собственным значениям матрицы B . Имеем далее

$$(E - zB)^{-1}G = G + zBG + z^2B^2G + \dots \quad (4)$$

Радиус сходимости этого ряда равен $\frac{1}{\max_i |\mu_i|}$, и если среди собственных значений матрицы B имеются собственные значения, по модулю большие единицы, то интересующая нас точка $z = 1$ окажется за пределами круга сходимости.

Положим теперь

$$z = z(w) = a_1w + a_2w^2 + \dots, \quad (5)$$

где $z(w)$ — функция, мероморфная в единичном круге $|w| < 1$, принимающая значение $z = 1$ в некоторой точке θ этого круга и не принимающая внутри круга значений $\frac{1}{\mu_1}, \dots, \frac{1}{\mu_n}$ (тем самым, если среди чисел μ_i имеется нуль, функция $z(w)$ должна быть регулярной).

Все компоненты вектора $X(z)$ будут регулярными функциями от w в круге $|w| < 1$. Действительно, знаменатель $|E - zB|$ не будет обращаться в нуль при $|w| < 1$. В полюсах же функции $z(w)$, если они есть (что возможно только при $|B| \neq 0$), все компоненты $X(z)$ будут равны нулю, ибо $(E - zB)^{-1}G = \frac{1}{z} \left(\frac{1}{z} E - B \right)^{-1}G \rightarrow 0$ при $z \rightarrow \infty$. Поэтому решение $X(z)$ си-

¹⁾ В. Н. Кублановская [1], [2].

стемы (2) может быть разложено в ряд по степеням w с радиусом сходимости, не меньшим единицы.

Для построения этого ряда разложим функцию $\frac{1}{1-zt}$ по степеням w . Имеем

$$\begin{aligned}\frac{1}{1-zt} &= 1 + tz + t^2 z^2 + \dots = \\ &= 1 + t(a_1 w + a_2 w^2 + \dots) + t^2(a_1 w + a_2 w^2 + \dots)^2 + \dots = \\ &= 1 + a_1 t w + (a_2 t + a_1^2 t^3) w^2 + \dots = \\ &= 1 + b_1(t) w + b_2(t) w^2 + \dots + b_i(t) w^i + \dots\end{aligned}\quad (6)$$

где $b_i(t)$ некоторые полиномы от t степени i . Коэффициенты полиномов $b_i(t)$ могут быть вычислены, как только известно разложение функции $z(w)$.

При $w = \theta$ ряд (6) превращается в разложение

$$\frac{1}{1-t} = 1 + b_1(t) \theta + \dots + b_i(t) \theta^i + \dots \quad (7)$$

Используя теперь равенство (6) для разложения решения $X(z)$ в ряд по степеням w , получим

$$X(z) = (E - zB)^{-1} G = G + wb_1(B) G + w^2 b_2(B) G + \dots$$

Решение исходной системы $X = X(1)$ найдется по формуле

$$X = (E - B)^{-1} G = G + \theta b_1(B) G + \theta^2 b_2(B) G + \dots \quad (8)$$

Этот ряд всегда будет сходящимся, ибо $0 < |\theta| < 1$, и сходимость будет тем быстрее, чем меньше $|\theta|$.

Допустим теперь, что относительно матрицы B известно, что все ее собственные значения принадлежат некоторому ограниченному множеству S , дополнение к которому есть односвязная область, содержащая точку 1. В этих условиях естественно ставить вопрос об S -универсальном алгорифме, т. е. об алгорифме, применимом ко всем системам $X = BX + G$, таким, что собственные значения матрицы B принадлежат S .

Построение S -универсальных алгорифмов может быть осуществлено посредством конформного отображения. Обозначим через D область, которая получается из дополнения к S посредством отображения функцией $\frac{1}{z}$. Так построенная область D будет односвязной, будет содержать точки 0 и 1 и не содержать чисел, обратных собственным значениям всех матриц B рассматриваемого класса. Точка ∞ может как принадлежать, так и не принадлежать области D .

Для применения формулы (8) ко всему классу рассматриваемых систем можно взять в качестве $z(w)$ любую функцию, мероморфную

в единичном круге, все значения которой принадлежат области D и принимающую значение 1 в некоторой точке θ этого круга. Такой функцией является, в частности, функция, осуществляющая конформное отображение единичного круга на область D . Из теории конформного отображения следует, что такой выбор функции $z(w)$ является наивыгоднейшим, так как именно при нем достигается наименьшее возможное значение $|z(w)|$.

Действительно, пусть $\tilde{z}(w) = a_1 w + a_2 w^2 + \dots$ какая-либо допустимая функция, $\tilde{z}(\tilde{\theta}) = 1$. Далее, $z = z(w) = a_1 w + a_2 w^2 + \dots$ функция, осуществляющая конформное отображение единичного круга на область D , $w = F(z)$ — функция, обратная к $z(w)$, $\theta = F(1)$. Покажем, что $|z(w)| \leq |\tilde{z}(w)|$.

С этой целью рассмотрим функцию $W(w) = F(\tilde{z}(w))$. Ясно, что $W(0) = 0$, $W(w)$ регулярна в единичном круге, $|W(w)| < 1$ при $|w| < 1$. Согласно лемме Шварца¹⁾ для всех точек единичного круга будет

$$|W(w)| \leq |w|.$$

Положив $w = \tilde{\theta}$, получим

$$|W(\tilde{\theta})| = |F(1)| = |\theta| \leq |\tilde{\theta}|.$$

Знак равенства может иметь место, только если $\tilde{z}(w) = z(w)$, $|\epsilon| = 1$, т. е. если функция $\tilde{z}(w)$ сама осуществляет конформное отображение единичного круга на область D .

Метод конформного отображения может быть применен и в случае, если дополнение к S , а вместе с ней и D являются многосвязными областями. В этом случае наивыгоднейшей $z(w)$ является функция, осуществляющая конформное отображение единичного круга на универсальную накрывающую область D риманову поверхность.

Перейдем теперь к описанию вычислительных схем метода. В работе [2] В. Н. Кублановской приведены вспомогательные таблицы коэффициентов полиномов

$$L_s(t) = 1 + \theta b_1(t) + \theta^2 b_2(t) + \dots + \theta^s b_s(t) = \\ = l_{s0} + l_{s1}t + \dots + l_{ss}t^s,$$

для ряда отображающих функций $z(w)$ при $s = 5$ и $s = 10$.

Это позволяет вычислять приближения $L_s(B)G$ к решению X как линейную комбинацию последовательных итераций G, BG, B^2G, \dots или посредством применения схемы Хорнера. При этом приходится заранее фиксировать число взятых в формуле (8) членов.

Ограничиваюсь небольшим s трудно рассчитывать на получение достаточной точности. Однако однократное применение приближен-

¹⁾ См. цитированную на стр. 590 книгу Г. М. Голузина, стр. 29

ной формулы $X = L_s(B)G$ можно рассматривать как элементарный шаг итерационного процесса $X^{(k)} = X^{(k-1)} + L_s(B)r_{k-1}$, где $r_{k-1} = BX^{(k-1)} + G - X^{(k-1)}$.

Легко построить вычислительные схемы метода, использующие лишь коэффициенты a_j отображающей функции $z(w)$. Именно, из тождества

$$[1 - t(a_1w + a_2w^2 + \dots)][1 + b_1(t)w + b_2(t)w^2 + \dots] = 1,$$

равносильного формуле (6), получаем рекуррентные соотношения для полиномов $b_i(t)$. Именно,

$$b_1(t) = a_1t$$

$$b_i(t) = a_1tb_{i-1}(t) + a_2tb_{i-2}(t) + \dots + a_{i-1}tb_1(t) + a_it. \quad (9)$$

Обозначив $b_i(B)G = G_i$, $G_0 = G$, получим

$$G_i = B(a_1G_{i-1} + \dots + a_iG_0) \quad (i = 1, 2, \dots)$$

$$X = \sum_{i=0}^{\infty} \theta^i G_i. \quad (10)$$

Компоненты векторов G_i будут или не возрастать, или возрастать очень медленно, ибо радиус сходимости ряда $\sum G_i \theta^i$ равен единице.

Рекуррентные соотношения, аналогичные (9), можно установить и для полиномов $L_i(t)$. Именно, легко проверить, что

$$L_i(t) = 1 + a_1\theta t L_{i-1}(t) + a_2\theta^2 t L_{i-2}(t) + \dots + a_i\theta^i t L_0(t). \quad (11)$$

Это дает возможность строить по рекуррентным формулам сами последовательные приближения $X_i = L_{i-1}(B)G$. Именно,

$$X_1 = G, \quad X_2 = G + a_1\theta BG$$

$$X_i = G + a_1\theta BX_{i-1} + a_2\theta^2 BX_{i-2} + \dots + a_{i-1}\theta^{i-1} BX_1. \quad (12)$$

Очевидно, что формулам (12) можно придать вид

$$X_1 = G, \quad X_2 = (1 - a_1\theta)G + a_1\theta(BX_1 + G)$$

$$X_i = (1 - a_1\theta - a_2\theta^2 - \dots - a_{i-1}\theta^{i-1})G + a_1\theta(BX_{i-1} + G) + \dots + a_{i-1}\theta^{i-1}(BX_1 + G). \quad (13)$$

Вычисления по рекуррентным формулам (10), (12) или (13) требуют несколько большего числа вычислительных операций, чем вычисление при помощи заранее вычисленных коэффициентов l_{ij} , но имеют то преимущество, что нет необходимости заранее фиксировать число членов ряда.

Формулы (13) соответствуют п. 7 § 86. Для получения формул, аналогичных п. 8 § 86, введем в рассмотрение полином

$$L_i(t) = 1 - (1 - t)L_{i-1}(t). \quad (14)$$

Легко проверить, что полиномы $l_i(t)$ связаны рекуррентными соотношениями

$$\begin{aligned} l_i(t) = & a_1 \theta l_{i-1}(t) + \dots + a_{i-1} \theta^{i-1} t l_1(t) + \\ & + (1 - a_1 \theta - \dots - a_{i-1} \theta^{i-1}) t. \end{aligned} \quad (15)$$

Отсюда для i -го приближения $X_i = l_i(B) X_0$ получим

$$\begin{aligned} X_i = & a_1 \theta (B X_{i-1} + G) + \dots + a_{i-1} \theta^{i-1} (B X_1 + G) + \\ & + (1 - a_1 \theta - \dots - a_{i-1} \theta^{i-1}) (B X_0 + G). \end{aligned} \quad (16)$$

Формула (16) превращается в формулу (13) при $X_0 = 0$.

Во многих случаях оказываются более удобными рекуррентные формулы, построенные исходя из коэффициентов d_j разложения в ряд по степеням w функции $\frac{w}{z}$. Эта функция будет регулярной, ибо $z = 0$ только при $w = 0$ (если область D многосвязна, функция $\frac{w}{z}$ будет мероморфной).

Пусть

$$\frac{w}{z} = \frac{w}{a_1 w + a_2 w^2 + \dots} = d_0 + d_1 w + d_2 w^2 + \dots; \quad d_0 \neq 0. \quad (17)$$

Тогда

$$\begin{aligned} \frac{1}{1 - zt} &= \frac{\frac{w}{z}}{\frac{w}{z} - tw} = \frac{d_0 + d_1 w + d_2 w^2 + \dots}{d_0 + d_1 w + d_2 w^2 + \dots - tw} = \\ &= \frac{d_0 + d_1 w + d_2 w^2 + \dots}{d_0 + (d_1 - t) w + d_2 w^2 + \dots} = 1 + b_1(t) w + b_2(t) w^2 + \dots \end{aligned} \quad (18)$$

Отсюда, приравнивая коэффициенты при одинаковых степенях, получим

$$\begin{aligned} d_0 b_1(t) + (d_1 - t) &= d_1 \\ d_0 b_2(t) + (d_1 - t) b_1(t) &= 0 \\ d_0 b_i(t) + (d_1 - t) b_{i-1}(t) + d_2 b_{i-2}(t) + \dots + d_{i-1} b_1(t) &= 0, \end{aligned} \quad (19)$$

так что

$$\begin{aligned} b_1(t) &= \frac{1}{d_0} t; \quad b_2(t) = \frac{t - d_1}{d_0} b_1(t) \\ b_i(t) &= \frac{t - d_1}{d_0} b_{i-1}(t) - \frac{d_2}{d_0} b_{i-2}(t) - \dots - \frac{d_{i-1}}{d_0} b_1(t) \quad (i \geq 3). \end{aligned} \quad (20)$$

Формулы (20) позволяют вычислять векторы G_i в формуле (10) по рекуррентным соотношениям

$$\begin{aligned} G_1 &= \frac{1}{d_0} BG; \quad G_2 = \frac{1}{d_0} BG_1 - \frac{d_1}{d_0} G_1; \\ G_i &= \frac{1}{d_0} BG_{i-1} - \frac{d_1}{d_0} G_{i-1} - \dots - \frac{d_{i-1}}{d_0} G_1 \quad (i \geq 3). \end{aligned} \quad (21)$$

Легко вывести также и рекуррентные формулы для полиномов $L_i(t)$ и $l_i(t)$. Именно

$$\begin{aligned} L_0(t) &= 1; \quad L_1(t) = 1 + \frac{\theta}{d_0} t; \quad L_2(t) = \frac{\theta(t-d_1)}{d_0} L_1(t) + \frac{d_0+d_1\theta}{d_0}; \\ L_i(t) &= \frac{\theta(t-d_1)}{d_0} L_{i-1}(t) - \frac{\theta^2 d_2}{d_0} L_{i-2}(t) - \dots - \frac{\theta^{i-1} d_{i-1}}{d_0} L_1(t) + \\ &\quad + \frac{d_0+d_1\theta+\dots+d_{i-2}\theta^{i-2}-\theta}{d_0}. \end{aligned} \quad (22)$$

Далее,

$$\begin{aligned} l_i(t) &= \frac{\theta(t-d_1)}{d_0} l_{i-1}(t) - \frac{\theta^2 d_2}{d_0} l_{i-2}(t) - \dots - \frac{\theta^{i-2} d_{i-2}}{d_0} l_2(t) + \\ &\quad + \frac{d_0+d_1\theta+\dots+d_{i-2}\theta^{i-2}-\theta}{d_0} l_1(t) \quad (i \geq 3) \end{aligned} \quad (23)$$

при $l_1(t) = t$, $l_2(t) = t - \frac{\theta}{d_0} t + \frac{\theta}{d_0} t^2$.

Поэтому последовательные приближения можно вычислять по формулам

$$\begin{aligned} X_1 &= G; \quad X_2 = \frac{\theta}{d_0} BX_1 + X_1 \\ X_i &= \frac{\theta}{d_0} BX_{i-1} - \frac{\theta d_1}{d_0} X_{i-1} - \frac{\theta^2 d_2}{d_0} X_{i-2} - \dots - \frac{\theta^{i-2} d_{i-2}}{d_0} X_2 + \\ &\quad + \frac{d_0+d_1\theta+\dots+d_{i-2}\theta^{i-2}-\theta}{d_0} X_1 \quad (i \geq 3) \end{aligned} \quad (24)$$

или по формулам

$$\begin{aligned} X_1 &= BX_0 + G; \quad X_2 = \frac{\theta}{d_0} (BX_1 + G) + \frac{d_0-\theta}{d_0} X_1; \\ X_i &= \frac{\theta}{d_0} (BX_{i-1} + G) - \frac{\theta d_1}{d_0} X_{i-1} - \frac{\theta^2 d_2}{d_0} X_{i-2} - \dots - \frac{\theta^{i-2} d_{i-2}}{d_0} X_2 + \\ &\quad + \frac{d_0+d_1\theta+\dots+d_{i-2}\theta^{i-2}-\theta}{d_0} X_1, \quad (i \geq 3). \end{aligned} \quad (25)$$

Рассмотрим теперь метод конформного отображения с точки зрения идеи подавления компонент. Пусть X_0 исходное приближение,

$X' = X_s = X_0 + L_{s-1}(B)r_0$, $r_0 = BX_0 + G - X_0$. Тогда соответствующие векторы ошибок связаны соотношением

$$Y' = Y_0 - L_{s-1}(B)r_0 = l_s(B)Y_0,$$

где $l_s(t) = 1 - (1-t)L_{s-1}(t)$.

Оценим $|l_s(t)|$ для любой точки t , принадлежащей множеству S .

Пусть $0 < \rho_0 < 1$, C_{ρ_0} образ окружности $|w| = \rho_0$ при отображении $z = z(w)$ единичного круга на область D . Имеем

$$|b_j(t)| = \left| \frac{1}{2\pi i} \int_{|w|=\rho_0} \frac{dw}{w^{j+1}(1-tz(w))} \right|.$$

Пока $|w| = \rho_0$, функция $z(w)$ пробегает кривую C_{ρ_0} . Если $t \in S$, то $\frac{1}{t}$ не принадлежат области D . Поэтому $1-tz$ не равно нулю при $t \in S$, $z \in D$ и $|1-tz|$ равномерно ограничен снизу константой d_{ρ_0} , пока $z \in C_{\rho_0}$, $t \in S$. Поэтому имеет место равномерная для $t \in S$ оценка

$$|b_j(t)| \leq \frac{1}{2\pi} \frac{2\pi\rho_0}{\rho_0^{j+1} d_{\rho_0}} = \frac{1}{\rho_0^j d_{\rho_0}}.$$

Таким образом, на основании формулы (7),

$$\left| \frac{1}{1-t} - L_{s-1}(t) \right| \leq \left| \sum_{j=s}^{\infty} b_j(t) \theta^j \right| \leq \sum_{j=s}^{\infty} \left(\frac{|\theta|}{\rho_0} \right)^j \frac{1}{d_{\rho_0}} = \frac{1}{d_{\rho_0}} \frac{\left(\frac{|\theta|}{\rho_0} \right)^s}{1 - \frac{|\theta|}{\rho_0}},$$

откуда

$$|l_s(t)| \leq |1-t| \frac{1}{d_{\rho_0}} \frac{\left(\frac{|\theta|}{\rho_0} \right)^s}{1 - \frac{|\theta|}{\rho_0}} \leq C(\rho_0) [(1+\varepsilon)|\theta|]^s.$$

Здесь $\varepsilon = \frac{1}{\rho_0} - 1$ положительное число, которое можно сделать сколь угодно малым, $C(\rho_0) = \max |(1-t)| \frac{1}{d_{\rho_0} \left(1 - \frac{|\theta|}{\rho_0} \right)}$. Итак, компоненты

вектора ошибки „затухают“ по крайней мере со скоростью $[(1+\varepsilon)|\theta|]^s$.

Сделаем следующее замечание относительно применения метода конформного отображения.

Если в систему $X = BX + G$ ввести параметр z в виде

$$X = zBX + \varphi(z)G,$$

где $\varphi(z)$ регулярная в области D функция, обращающаяся в единицу при $z = 1$, то по методу конформного отображения решение исходной системы получим в виде

$$X = c_0G + \theta c_1(B)G + \theta^2 c_2(B)G + \dots$$

где $c_i(t)$ полиномы, являющиеся коэффициентами в разложении функции $\frac{\sigma(w)}{1-tz(w)} = c_0 + c_1(t)w + c_2(t)w^2 + \dots$. Здесь $\sigma(w) = \rho(z(w))$ функция, регулярная в единичном круге. При любом выборе функции $\sigma(z)$ (или, что то же самое, $\rho(z)$) порядок быстроты сходимости будет одним и тем же.

§ 96. Примеры S-универсальных алгорифмов

Рассмотрим теперь несколько конкретных S-алгорифмов. Как мы видели выше, каждый такой алгорифм определяется ограниченным замкнутым множеством S , содержащим собственные значения матриц B .

1. S — круг радиуса $\frac{1}{\gamma}$, $\gamma > 1$
с центром в начале координат (рис. 29). В этом случае область D есть внутренность круга радиуса γ с центром в начале координат. Отображающая функция есть

$$z(w) = \gamma w, \quad \theta = \frac{1}{\gamma}.$$

Поэтому последовательные приближения вычисляются по формуле

$$X_i = BX_{i-1} + G,$$

т. е. в этом случае мы приходим к классическому методу последовательных приближений.

2. S — отрезок вещественной оси $(-\frac{1}{\gamma}, \frac{1}{\gamma})$, $\gamma > 1$ (рис. 30). В этом случае область D есть плоскость с двумя разрезами вдоль

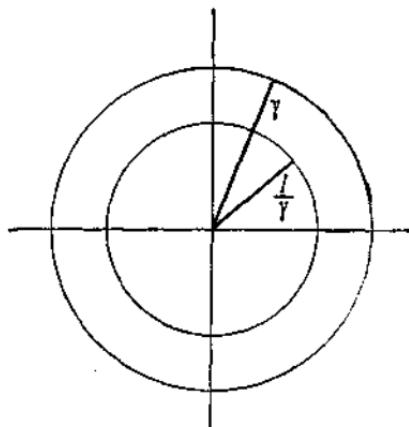


Рис. 29.

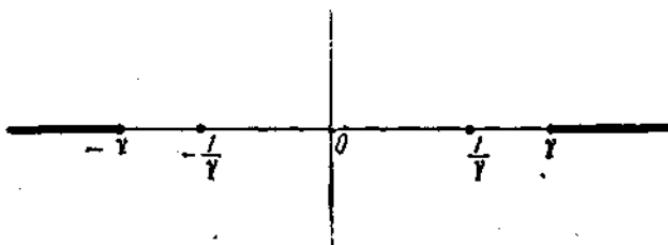


Рис. 30.

вещественной оси, исходящими из точек $-\gamma$ и γ в бесконечность. Отображающая функция есть

$$z(w) = \frac{2\gamma w}{1+w^2}, \quad \theta = \gamma - \sqrt{\gamma^2 - 1}.$$

Однако в этом случае функция $\frac{w}{z(w)}$ будет еще проще. Именно, $\frac{w}{z(w)} = \frac{1}{2\gamma}(1 + w^2)$, так что $d_0 = d_2 = \frac{1}{2\gamma}$, $d_1 = d_3 = d_4 = \dots = 0$. Поэтому последовательные приближения удобно строить по рекуррентным формулам (25), которые обращаются в формулы

$$\begin{aligned} X_i &= 2\gamma\theta(BX_{i-1} + G) - \theta^2 X_{i-2} = \\ &= (1 + \theta^2)(BX_{i-1} + G) - \theta^2 X_{i-2} \quad \text{при } i \geq 3 \\ X_1 &= BX_0 + G, \quad X_2 = 2\gamma\theta(BX_1 + G) + (1 + 2\gamma\theta)X_1 = \\ &= (1 + \theta^2)(BX_1 + G) - \theta^2 X_1. \end{aligned}$$

Процесс почти совпадает с универсальным трехчленным алгорифмом с постоянным множителем $\alpha = \theta^2$ (§ 91), отличаясь от него лишь началом процесса.

3. S — эллипс с фокусами в точках $-\frac{1}{\gamma}$ и $\frac{1}{\gamma}$ и с вершинами в точках $-\frac{1}{\alpha}$, $\frac{1}{\alpha}$, $\gamma > \alpha > 1$ (рис. 31). В этом случае область D

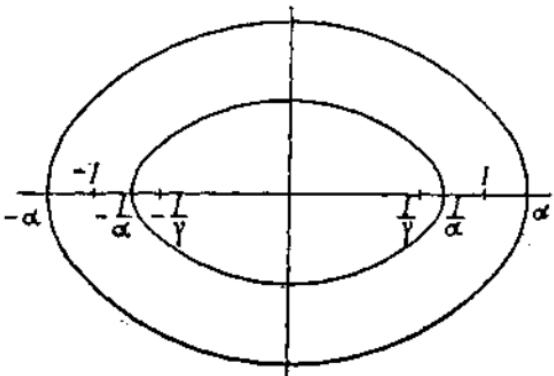


Рис. 31.

будет внутренностью некоторого овала. Отображающая функция есть

$$z(w) = \frac{2\rho\gamma w}{1 + \rho^2 w^2}, \quad \rho^2 = \frac{1}{\alpha} - \sqrt{\frac{\gamma^2}{\alpha^2} - 1}, \quad \theta = \frac{\gamma - \sqrt{\gamma^2 - 1}}{\rho} = \frac{\theta_1}{\rho},$$

где $\theta_1 = \gamma - \sqrt{\gamma^2 - 1}$.

Снова функция $\frac{w}{z(w)}$ оказывается квадратным полиномом

$$\frac{1}{2\rho\gamma} + \frac{\rho}{2\gamma} w^2.$$

Поэтому

$$d_0 = \frac{1}{2\rho\gamma}, \quad d_2 = \frac{\rho}{2\gamma}, \quad d_1 = d_3 = d_4 = \dots = 0.$$

Из формул (25) получим

$$X_1 = BX_0 + G,$$

$$X_2 = 2\gamma\theta_1(BX_1 + G) + (1 - 2\gamma\theta_1)X_1 = (1 + \theta_1^2)(BX_1 + G) - \theta_1^2 X_1,$$

$$X_i = (1 + \theta_1^2)(BX_{i-1} + G) - \theta_1^2 X_{i-2}.$$

Как мы видим, эти формулы совпадают с формулами, полученными выше для отрезка вещественной оси.

4. S — отрезок мнимой оси $(-\frac{1}{\beta}, \frac{1}{\beta})$ (рис. 32). В этом случае область D есть плоскость с двумя разрезами вдоль мнимой оси, исходящими из точек $-\beta i$ и βi в бесконечность. Отображающая функция есть

$$z(w) = \frac{2\beta w}{1-w^2}, \quad \theta = \sqrt{\beta^2 + 1} - \beta.$$



Рис. 32.

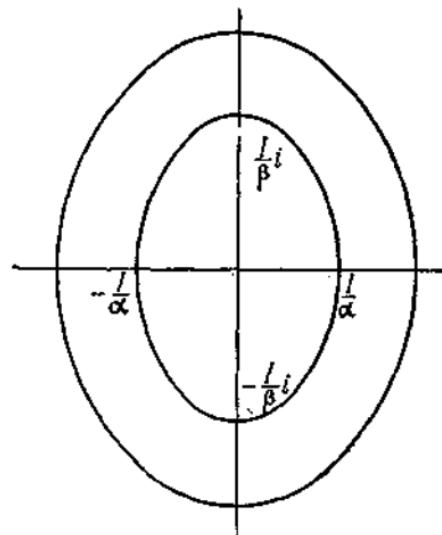


Рис. 33.

Так же как и в предыдущем случае, функция $\frac{w}{z(w)}$ будет квадратным полиномом, именно $\frac{w}{z(w)} = \frac{1}{2\beta}(1-w^2)$, и потому $d_0 = \frac{1}{2\beta}$, $d_2 = -\frac{1}{2\beta}$, $d_1 = d_3 = d_4 = \dots = 0$. Для последовательных приближений из формулы (25) получаются рекуррентные формулы

$$X_i = 2\beta\theta(BX_{i-1} + G) + \theta^2 X_{i-2} = (1 - \theta^2)(BX_{i-1} + G) + \theta^2 X_{i-2} \quad (i \geq 3)$$

$$X_1 = BX_0 + G, \quad X_2 = (1 - \theta^2)(BX_1 + G) + \theta^2 X_1.$$

5. S — эллипс с фокусами в точках $-\frac{1}{\beta}i$ и $\frac{1}{\beta}i$ и мнимой полуосью $\frac{1}{\alpha}$, $\alpha > 1$ (рис. 33).

В этом случае область D будет внутренностью некоторого овала. Отображающая функция есть

$$z(w) = \frac{2\rho\beta w}{1-\rho^2w^2}, \quad \text{где } \rho = \frac{\sqrt{\alpha^2 + \beta^2} - \beta}{\alpha},$$

$$\theta = \frac{\sqrt{\beta^2 + 1} - \beta}{\rho} = \frac{\theta_1}{\rho}, \quad \theta_1 = \sqrt{\beta^2 + 1} - \beta.$$

Так как $\frac{w}{z(w)} = \frac{1}{2\rho\beta} - \frac{\rho}{2\beta}w^2$, получим из формул (25) после простых преобразований

$$X_1 = BX_0 + G, \quad X_2 = (1 - \theta_1^2)(BX_1 + G) + \theta_1^2 X_1$$

$$X_i = (1 - \theta_1^2)(BX_{i-1} + G) + \theta_1^2 X_{i-2}.$$

Вычислительные формулы для приближений совпадают с формулами п. 4 после замены в них θ на θ_1 .

6. S — два касающихся круга, опирающихся на отрезки $(0, \frac{1}{\gamma})$ и $(0, -\frac{1}{\gamma})$ как на диаметры, $\gamma > 1$ (рис. 34). В этом случае

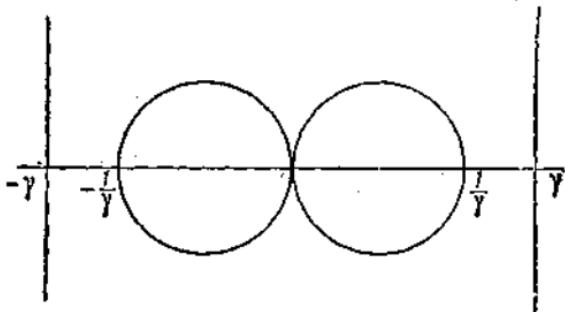


Рис. 34.

область D есть полоса $-\gamma < \operatorname{Re} z < \gamma$, отображающая функция

$$z(w) = \frac{4\gamma}{\pi} \operatorname{arctg} w = \frac{2\gamma}{\pi i} \lg \frac{1+wi}{1-wi} = \frac{4\gamma}{\pi} \left[w - \frac{w^3}{3} + \frac{w^5}{5} - \dots \right], \quad \theta = \operatorname{tg} \frac{\pi}{4\gamma} w.$$

Замечание. При практическом пользовании методом конформного отображения нет необходимости использовать наименьшее известное нам множество S , содержащее собственные значения данной матрицы, в силу имеющейся информации. Множество S следует

подбирать так, чтобы отображающая функция имела наиболее простой вид. Небольшие изменения множества S влечут за собой лишь небольшое увеличение θ .

§ 97. Метод конформного отображения в применении к неподготовленной системе

Метод конформного отображения может быть применен и непосредственно к решению системы

$$AX = F. \quad (1)$$

Пусть Σ — ограниченное замкнутое множество, дополнение к которому Δ (включая бесконечно удаленную точку) есть односвязная область, содержащая точку 0. Мы будем считать, что все собственные значения матрицы A принадлежат множеству Σ . Рассмотрим систему

$$AX - uX = F, \quad (2)$$

зависящую от комплексного параметра u . Исходная система получается из (2) при $u = 0$.

Положим, что функция $u = u(w)$ отображает единичный круг $|w| < 1$ на область Δ так, что $u(0) = \infty$. Обозначим через θ прообраз 0. Решение системы (2) запишется в виде

$$X(u) = (A - uE)^{-1}F. \quad (3)$$

Исследуем функцию $\frac{1}{t - u}$. Эта функция регулярна в области Δ , включая точку $u = \infty$, при любом t , принадлежащем множеству Σ .

Ясно, что функция $u(w)$ имеет следующее разложение в ряд

$$u(w) = -\frac{c_{-1}}{w} - c_0 - c_1 w - c_2 w^2 - \dots \quad (4)$$

причем $c_{-1} \neq 0$. Следовательно,

$$\begin{aligned} \frac{1}{t - u(w)} &= \frac{w}{c_{-1} + (t + c_0)w + c_1 w^2 + \dots} = \\ &= \frac{w}{c_{-1}} \left(1 + \frac{c_0 + t}{c_{-1}} w + \frac{c_1}{c_{-1}} w^2 + \dots \right)^{-1} = \\ &= w [d_0(t) + d_1(t)w + d_2(t)w^2 + \dots], \end{aligned} \quad (5)$$

где $d_i(t)$ полиномы степени i . Радиус сходимости этого ряда будет не меньше единицы.

В соответствии с разложением (5) решение системы (2) будет

$$X(u) = w [d_0(A)F + wd_1(A)F + w^2d_2(A)F + \dots]. \quad (6)$$

Решение же исходной системы представляется в виде сходящегося ряда

$$X = \theta [d_0(A)F + \theta d_1(A)F + \theta^2 d_2(A)F + \dots], \quad (7)$$

отрезки которого дают приближенные решения системы.

Метод допускает следующее видоизменение. Вместо функции $\frac{1}{t-u(w)}$ можно ввести в рассмотрение функцию $\frac{\rho(w)}{w[t-u(w)]}$, где $\rho(w)$ любая регулярная в единичном круге функция, удовлетворяющая условию $\rho(0) = \theta$. Тогда

$$\frac{\cdot \rho(w)}{w[t-u(w)]} = \alpha_0(t) + \alpha_1(t)w + \alpha_2(t)w^2 + \dots, \quad (8)$$

где $\alpha_i(t)$ — некоторые полиномы от t . Радиус сходимости последнего ряда по-прежнему равен единице. Соответственно решение системы (1) представится в виде

$$X = \alpha_0(A)F + \theta\alpha_1(A)F + \theta^2\alpha_2(A)F + \dots. \quad (9)$$

Сравним теперь решение системы $AX = F$, найденное по ряду (7), с решением той же системы, найденным по ряду (8) § 95 после предварительной подготовки системы (1) к виду $X = BX + G$ при $B = E - hA$, $G = Fh$.

Собственные значения матрицы B связаны с собственными значениями матрицы A соотношением $\mu_i = 1 - h\lambda_i$. Поэтому за множество S для матрицы B можно принять множество точек $\frac{1}{1-ht}$ при $t \in \Sigma$. Тогда область D будет дополнением к множеству $\frac{1}{1-ht}$, $t \in \Sigma$, так что (в силу однолистности функции $\frac{1}{1-\xi h}$ на всей плоскости комплексного переменного, включая бесконечно далекую точку) D получается из Δ отображением посредством функции $\frac{1}{1-\xi h}$.

Функция $u(w)$ отображает единичный круг на область Δ , следовательно $z(w) = \frac{1}{1-hu(w)}$ отображает единичный круг на D , причем $z(0) = 0$, $z(\theta) = 1$. Таким образом, функция $z(w)$ удовлетворяет требованиям § 95.

Положив $t_1 = 1 - ht$ (здесь t является „представителем“ матрицы A , t_1 „представителем“ матрицы B), имеем

$$X = b_0(B)G + \theta b_1(B)G + \theta^2 b_2(B)G + \dots,$$

где $b_i(t)$ коэффициенты в разложении функции $\frac{1}{1-t_1 z(w)}$ по степеням w . Подставив $B = E - hA$ и $G = hF$, получим

$$X = h\beta_0(A)F + \theta h\beta_1(A)F + \theta^2 h\beta_2(A)F + \dots,$$

где $\beta_i(t) = b_i(1 - ht)$.

Ясно, что полиномы $\beta_i(t)$ будут коэффициентами в разложении функции

$$\frac{1}{1 - (1 - ht)z(w)} = \frac{1}{1 - \frac{1 - ht}{1 - hu(w)}} = \frac{1 - hu(w)}{h(t - u(w))}.$$

а полиномы $h\beta_i(t)$ будут коэффициентами в разложении функции

$$\frac{1 - hu(w)}{t - u(w)}.$$

Таким образом, сопоставляя с формулой (8), получим, что

$$h\beta_i(t) = \alpha_i(t)$$

при

$$\rho(w) = w(1 - hu(w)).$$

Отсюда мы делаем заключение, что порядок быстроты сходимости рядов, дающих решение системы $AX = F$ при различных ее подготовках, не зависит от числа h , определяющего данную подготовку, так как этот порядок определяется лишь числом θ . Интересно отметить, что решение, даваемое рядом (7), получается как предельный случай решения подготовленной системы при $h \rightarrow 0$.

Функция Грина $G(u)$ для области Δ и функция $u(w)$ связаны, как известно, следующим образом:

$$G(u) = -\lg |w(u)|,$$

где $w(u)$ функция, обратная $u(w)$.

Поэтому

$$|\theta| = |w(0)| = e^{-G(0)}.$$

Таким образом, метод конформного отображения имеет такой же порядок быстроты сходимости, какой имеет метод подавления компонент полиномами, наименее уклоняющимися от нуля (см. § 94).

Мы не будем останавливаться на разборе конкретных Σ -алгорифмов, которые могут быть построены так же, как это делалось при построении конкретных S -алгорифмов.

Остановимся теперь на случае, когда область Δ многосвязна.

Метод конформного отображения распространяется на этот случай почти без изменений. Берется мероморфная в единичном круге функция $z(w)$, имеющая простой полюс в точке $w = 0$, принимающая значение 0 в некоторой точке θ и не принимающая значений из множества Σ . Тогда

$$\frac{1}{t - z(w)} = d_0 + d_1(t)w + d_2(t)w^2 + \dots$$

есть сходящийся ряд в круге $|w| < 1$ при любом $t \in \Sigma$. Решение системы $AX = F$ представляется в виде

$$X = d_0F + \theta d_1(A)F + \theta^2 d_2(A)F + \dots \quad (10)$$

Ряд (10) будет сходиться тем быстрее, чем меньше $|\theta|$, так что в качестве θ следует брать наименьший по модулю прообраз точки $z = 0$, если их много.

Наилучшей функцией $z(w)$ является функция $z = \varphi(w)$, отображающая единичный круг на односвязную накрывающую область Δ Риманову поверхность.

Действительно, пусть $z(w)$ какая-либо функция рассматриваемого класса. Каждая ветвь функции $\varphi^{-1}(z(w))$ будет регулярна внутри единичного круга, так что многозначная функция $\varphi^{-1}(z(w))$ в действительности распадается на однозначные регулярные ветви. Выберем из них ту $\Phi(w)$, для которой $\Phi(0) = 0$. Ясно, что $|\Phi(w)| < 1$ при $|w| < 1$. По известной лемме Шварца

$$|\Phi(w)| \leq |w|. \quad (11)$$

Пусть θ наименьший по модулю прообраз точки $z = 0$ при отображении $z = z(w)$ и $\theta_0 = \Phi(\theta)$. Ясно, что $\varphi(\theta_0) = 0$, так что θ_0 является одним из прообразов точки $z = 0$ при отображении $z = \varphi(w)$. Положив $w = \theta$ в неравенстве (11) получим:

$$|\theta_0| \leq |\theta|.$$

Если обозначить через θ^* наименьший по модулю прообраз точки $z = 0$ при отображении $z = \varphi(w)$, то подавно

$$|\theta^*| \leq |\theta|.$$

Из той же леммы Шварца следует, что знак равенства возможен, только если $\Phi(w) = \nu w$ при $|\nu| = 1$, т. е. если $z(w) = \varphi(\nu w)$. Тем самым доказано, что наименьшее по модулю число θ доставляет функция $z = \varphi(w)$.

Покажем теперь, что в случае многосвязной области метод конформного отображения дает худший результат, чем метод подавления компонент при помощи полиномов, наименее уклоняющихся от нуля.

Пусть $G(z)$ функция Грина для области Δ с логарифмической особенностью в точке $z = \infty$. Обозначим через $\psi(z)$ наименьшее по модулю значение функции $\varphi^{-1}(z)$ (если их несколько, выбираем какое-нибудь одно) и рассмотрим функцию $H(z) = -\lg |\psi(z)|$. Ясно, что $H(z)$ есть гармоническая функция в окрестности точки $z = \infty$, исключая эту точку, в которой $H(z)$ имеет логарифмическую особенность, ибо $\psi(z)$ в окрестности бесконечно далекой точки совпадает с той ветвью функции $\varphi^{-1}(z)$, для которой $\varphi^{-1}(\infty) = 0$.

Пусть z_0 произвольная точка области Δ , δ круг с центром в точке z_0 , содержащийся со своей границей γ внутри области Δ . Возьмем ту ветвь функции $\varphi^{-1}(z)$, для которой $\varphi^{-1}(z_0) = \psi(z_0)$. Тогда $|\psi(z)| \leq |\varphi^{-1}(z)|$ во всех точках круга δ , включая его границу γ . Ясно, что

$$H(z_0) = -\lg |\varphi^{-1}(z_0)| = \frac{1}{2\pi r} \int_{\gamma} -\lg |\varphi^{-1}(z)| d\sigma,$$

ибо $-\lg |\varphi^{-1}(z)|$ есть гармоническая функция в круге δ (через r обозначен радиус этого круга). Поэтому

$$H(z_0) \leq \frac{1}{2\pi r} \int_{\Gamma} -\lg |\psi(z)| d\sigma = \frac{1}{2\pi r} \int_{\Gamma} H(z) d\sigma.$$

Таким образом, $H(z)$ есть субгармоническая функция в области Δ . Далее, $H(z) - G(z)$ есть субгармоническая функция, ограниченная в окрестности точки $z = \infty$, и ее наибольшее значение не может достигаться внутри области Δ . Допустим, что Δ ограничена конечным числом аналитических кривых. Тогда на границе Δ как $G(z)$, так и $H(z)$, определены и равны нулю. Следовательно, $H(z) - G(z) < 0$ во всех точках области Δ или $H(z) = G(z)$ в Δ . Вторая возможность отпадает, ибо для многосвязной области $H(z)$ не является гармонической функцией в точках, для которых имеется более одного прообраза с наименьшим модулем. Итак, $H(z) - G(z) < 0$ в Δ . В частности, $H(0) = -\lg |\theta| < G(0)$, откуда

$$|\theta| > e^{-G(0)}. \quad (12)$$

Предположение о границе области Δ , при котором выведено неравенство (12), снимается посредством предельного перехода от областей Δ_ϵ , ограниченных ϵ -линиями уровня для функции Грина исходной области Δ .

Неравенство (12) означает, что быстрота сходимости метода конформного отображения уступает быстроте сходимости метода подавления компонент при помощи полиномов, наименее уклоняющихся от нуля.

Приведем один пример, иллюстрирующий это обстоятельство.

Пусть Σ есть совокупность двух отрезков $(-b, -a)$ и (a, b) вещественной оси (рис. 35). В этом случае область Δ двусвязна.

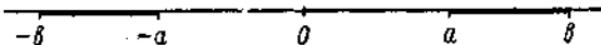


Рис. 35.

Функция, отображающая единичный круг на универсальную накрывающую, задается уравнением

$$\frac{4k_1}{\pi} \operatorname{arctg} w = \int_{-\infty}^{\infty} \frac{dz}{V(z^2 - a^2)(z^2 - b^2)}.$$

Здесь

$$k_1 = \int_b^{\infty} \frac{dz}{V(z^2 - a^2)(z^2 - b^2)} = \int_0^{\frac{\pi}{2}} \frac{d\varphi}{Vb^2 - a^2 \sin^2 \varphi}.$$

При этом

$$|\theta| = \frac{1 - e^{-\frac{\pi}{2} \frac{k_2}{k_1}}}{1 + e^{-\frac{\pi}{2} \frac{k_2}{k_1}}},$$

где

$$k_2 = \int_a^b \frac{dz}{V(z^2 - a^2)(b^2 - z^2)} = \int_0^{\frac{\pi}{2}} \frac{d\varphi}{V b^2 - (b^2 - a^2) \sin^2 \varphi}.$$

Если взять $a = 1$, $b = \sqrt{2}$, то $k_1 = k_2$ и потому

$$|\theta| = \frac{1 - e^{-\frac{\pi}{2}}}{1 + e^{-\frac{\pi}{2}}} \approx 0.6558.$$

Покажем, что метод подавления компонент приведет к процессу с более быстрой сходимостью.

Рассмотрим полином

$$g_{2s}(t) = \frac{T_s(3 - 2t^2)}{T_s(3)},$$

где $T_s(t) = \cos s \arccos t$. При $t \in \Sigma$

$$|g_{2s}(t)| \leq \frac{1}{T_s(3)} \approx \frac{2}{(3 + \sqrt{8})^s} = \frac{2}{(\sqrt{2} + 1)^{2s}} = 2(\sqrt{2} - 1)^{2s}.$$

Далее, $g_{2s}(0) = 1$. Положим

$$h_{2s-1}(t) = \frac{1 - g_{2s}(t)}{t}$$

и найдем приближение

$$X_{2s} = X_0 + h_{2s-1}(A)(F - AX_0).$$

Тогда

$$Y_{2s} = X^* - X_{2s} = Y_0 - h_{2s-1}(A)AY_0 = g_{2s}(A)Y_0.$$

Следовательно, компоненты вектора Y_{2s} в разложении по собственным векторам умножаются на множители $g_{2s}(\lambda_i)$. Но

$$|g_{2s}(\lambda_i)| \leq \frac{1}{T_s(3)} \approx 2(\sqrt{2} - 1)^{2s}.$$

Итак, используя $2s$ итераций начального вектора, мы получили для погрешности оценку порядка $(\sqrt{2} - 1)^{2s} \approx (0.4142)^{2s}$, в то время как метод конформного отображения с использованием такого же числа итераций дает быстроту сходимости порядка $\theta^{2s} \approx (0.6558)^{2s}$.

§ 98. Применение идеи подавления компонент к решению частичной проблемы собственных значений

Степенной метод для определения собственного вектора, принадлежащего наибольшему по модулю собственному значению, основан на том, что последовательность векторов $A^k Y_0$, при произвольном начальном векторе Y_0 , сходится по направлению к указанному собственному вектору.

Действительно, компоненты собственных векторов U_1, \dots, U_n в разложении начального вектора

$$Y_0 = c_1 U_1 + \dots + c_n U_n$$

в результате k -кратного применения итерации матрицей A получают множители $\lambda_1^k, \dots, \lambda_n^k$, среди которых λ_i^k преобладает над остальными. Если нормировать процесс так, чтобы коэффициент при U_1 оставлять равным единице, то остальные компоненты „подавляются“ стремящимися к нулю множителями $\left(\frac{\lambda_2}{\lambda_1}\right)^k, \dots, \left(\frac{\lambda_n}{\lambda_1}\right)^k$. Эта идея „подавления“ компонент может быть обобщена следующим образом.

Пусть известно, что все собственные значения матрицы A , кроме одного λ_i , подлежащего определению, лежат на некотором ограниченном множестве Σ , дополнение к которому Δ есть связная область плоскости комплексной переменной z . Пусть $\tau_k(t) = t^k + \dots$ полином k -й степени, наименее уклоняющийся от нуля на множестве Σ такой, что его корни, в свою очередь, лежат на множестве Σ .

Построим вектор $\frac{1}{\tau_k(\lambda_i)} \tau_k(A) Y_0$. Будем предполагать, что Y_0 имеет ненулевую i -ю компоненту в разложении по собственным векторам. Ясно, что i -я компонента построенного вектора будет прежней, а остальные компоненты приобретут множителей $\frac{\tau_k(\lambda_j)}{\tau_k(\lambda_i)}$ ($j \neq i$). Модули этих „множителей подавления“ удовлетворяют неравенству

$$\left| \frac{\tau_k(\lambda_j)}{\tau_k(\lambda_i)} \right| \leq \frac{\tau_k}{|\tau_k(\lambda_i)|}.$$

Здесь $\tau_k = \max_{t \in S} |\tau_k(t)|$. Известно¹⁾, что

$$\lim \sqrt[k]{\frac{\tau_k}{|\tau_k(\lambda_i)|}} = e^{-G(\lambda_i)} < 1,$$

где $G(t)$ есть функция Грина для области Δ . Поэтому последовательность векторов $\tau_0(A) Y_0, \tau_1(A) Y_0, \dots$ сходится по направлению к собственному вектору, принадлежащему собственному значению λ_i .

¹⁾ См. цитированную на стр. 590 книгу Г. М. Голузина, гл. VII.

с быстрой $e^{-k[G(\lambda_i) - \epsilon]}$, где ϵ сколь угодно малое число. После определения собственного вектора собственное значение λ_i определяется без труда.

§ 99. Применение конформного отображения к решению частичной проблемы собственных значений

Степенной метод может быть интерпретирован также следующим образом. Рассмотрим вектор $Y(w) = (E - wA)^{-1} Y_0$. Его компоненты являются аналитическими функциями от комплексной переменной w , имеющими полюса в точках $\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n}$, где $|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|$ — собственные значения матрицы A . Наименьшим по модулю полюсом будет число, обратное к наибольшему по модулю собственному значению. Разложим вектор $Y(w)$ в ряд по степеням w

$$Y(w) = (E - wA)^{-1} Y_0 = Y_0 + wAY_0 + w^2 A^2 Y_0 + \dots \quad (1)$$

Радиус сходимости этого ряда будет, очевидно, равен $\frac{1}{|\lambda_1|}$ и на его круге сходимости будет существовать единственный простой полюс $\frac{1}{\lambda_1}$. Компоненты вектора $Y(w)$ будут аналитическими функциями, имеющими, вообще говоря, простой полюс $\frac{1}{\lambda_1}$ на границе сходимости ряда (1). Для любой такой компоненты $y(w)$ имеем

$$y(w) = y_0 + y_1 w + y_2 w^2 + \dots \quad (2)$$

Здесь y_k есть выбранная компонента вектора $A^k Y_0$.

Предельная формула

$$\lambda_1 = \lim_{k \rightarrow \infty} \frac{y_{k+1}}{y_k} \quad (3)$$

может быть истолкована как результат применения теоремы Кёнига о коэффициентах разложения функции, имеющей единственный простой полюс на границе круга сходимости.

Оценка быстроты сходимости процесса (порядка $\left|\frac{\lambda_2}{\lambda_1}\right|^k$, где λ_2 следующее по модулю собственное значение матрицы A) также следует из теоремы Кёнига.

Такое рассмотрение степенного метода позволяет обобщить его в следующем направлении. Пусть

$$u(w) = -\frac{c_{-1}}{w} - c_0 - c_1 w - c_2 w^2 - \dots$$

есть мероморфная в единичном круге функция, $u(0) = \infty$, $u'(0) = \lambda_1$, $0 < |\theta| < 1$, причем при $|w| \leq |\theta|$, $w \neq 0$, $u(w)$ не принимает значений, равных собственным значениям матрицы A .

Рассмотрим вектор $(A - u(w)E)^{-1} Y_0$. Согласно формуле (6) § 97 $(A - u(w)E)^{-1} Y_0 = w [d_0(A) Y_0 + d_1(A) Y_0 w + d_2(A) Y_0 w^2 + \dots]$,

где $d_i(t)$ некоторые полиномы степени i . Любая компонента вектора $(A - u(w)E)^{-1}$ будет мероморфна в единичном круге и регулярна в круге $|w| \leq |\theta|$, кроме точки $w = \theta$, в которой она, вообще говоря, будет иметь простой полюс.

Следовательно, по теореме Кёнига отношения выбранных компонент векторов $d_{k+1}(A)Y_0$ и $d_k(A)Y_0$ будут стремиться к θ , а сами векторы будут сходиться по направлению к собственному вектору, принадлежащему собственному значению λ_i . Собственное значение найдется как $u(\theta)$. Быстрота сходимости процесса будет иметь порядок $\left|\frac{\theta}{\theta^*}\right|^k$, где $|\theta^*| \geq 1$, если внутри единичного круга нет прообразов собственных значений, кроме θ . Если же в единичном круге имеются также прообразы собственных значений матрицы A , то θ^* есть наименьший из них по модулю.

Допустим теперь, что относительно собственных значений матрицы A известно, что все они, кроме одного, подлежащего определению, лежат на ограниченном замкнутом множестве Σ , дополнение к которому Δ есть односвязная область. В этом случае в качестве функции $u(w)$ следует взять функцию, осуществляющую конформное отображение единичного круга на область Δ . По соображениям, изложенным в § 95, такой выбор функции будет наивыгоднейшим в смысле быстроты сходимости полученного процесса.

При проведении вычислений можно пользоваться вспомогательными таблицами для коэффициентов последовательных полиномов $d_k(t)$. Векторы $d_k(A)Y_0$ можно также строить по рекуррентным соотношениям.

Отметим также, что так как векторы $d_k(A)Y_0$ сходятся по направлению к собственному вектору матрицы A , принадлежащему определяемому собственному значению λ_i , то последнее можно находить как отношение компонент векторов $Ad_k(A)Y_0$ и $d_k(A)Y_0$ (а не как образ θ).

Впервые использование конформного отображения для нахождения собственных значений было предложено В. Н. Кублановской [1], [2], которая вместо функции $u(w)$ рассматривала функцию $z(w) = \frac{1}{u(w)}$, отображающую единичный круг на область $\tilde{\Delta}$, полученную из области Δ отображением посредством $z = \frac{1}{u}$. В этих работах исследуется функция

$$(E - zA)^{-1} = -u(A - uE)^{-1},$$

полюса которой совпадают с полюсами рассмотренной выше функции $A - uE$. В работе [2] приведены таблицы коэффициентов соответствующих полиномов для ряда областей $\tilde{\Delta}$.

ЗАКЛЮЧЕНИЕ

Значительное число методов, рассмотренных нами в настоящей книге, далеко не исчерпывает всего многообразия приемов, предложенных для численного решения основных задач линейной алгебры. Так, нами совершенно опущены методы „Монте-Карло“, обоснование которых имеет более теоретико-вероятностный, чем алгебраический характер. Из многочисленных схем исключения рассмотрены лишь немногие наиболее употребительные, описаны далеко не все итерационные процессы. Почти не рассматривались схемы, приспособленные для решения задач частного вида, имеющих узкую область применения. Наконец, в книге не отражены приемы, опубликованные в самое последнее время.

И хотя уже рассмотренный материал дает основание поставить вопрос о том, какие из описанных методов должны быть рекомендованы для практических расчетов предпочтительнее перед другими, на этот вопрос трудно и даже невозможно дать определенный ответ, так как в различных конкретных условиях к методам должны предъявляться разные требования.

Важнейшим из критериев оценки качества численного метода является его надежность, т. е. способность перенести в решение задачи почти всю информацию, содержащуюся в ее условиях. Однако кроме критерия надежности имеются и другие, достаточно существенные.

Это — простота вычислительной схемы. Далее, минимальность числа вычислительных операций; минимальная загрузка памяти для машин с программным управлением или компактность записи при пользовании настольными машинами. Часто важной оказывается возможность использования индивидуальных особенностей задачи, облегчающих ее решение (преобладание диагональных элементов матрицы, наличие большего числа нулевых элементов и т. д.). Наконец, иногда важна приспособленность метода к серийному решению однотипных задач.

Каждый заслуживающий внимания численный метод должен в той или иной мере удовлетворять всем указанным требованиям. Но в различных конкретных условиях удельный вес каждого требования может быть различным. Даже такое, казалось бы, необходимое требование, как надежность, может отступать на задний план, например, при решении большой серии однотипных задач, не требующих значительной точности результата. Часто из двух методов, в одном из которых проще вычислительная схема, но требуется большее число

вычислительных операций, чем в другом, следует предпочесть первый. Однако в задачах, связанных с матрицами высоких порядков, критерий минимальности числа вычислительных операций может оказаться решающим.

Сравнение методов по критериям простоты вычислительной схемы, загрузке памяти, приспособленности к индивидуальным особенностям задачи и серийности не представляет труда. Учет количества вычислительных операций тоже не сложен как в точных методах, так и в итерационных, если заранее известна быстрота сходимости.

Значительно сложнее оценивать методы по критерию надежности. По сути дела, при приближенной постановке вычислительных задач в их решении всегда имеется неопределенность в исходных данных. Эта неопределенность может быть значительной, например, при решении плохо обусловленной системы. Численный метод должен считаться надежным, если при его применении решение получается с погрешностью, не превосходящей существенно указанной выше неизбежной погрешности, обусловленной неопределенностью исходных данных.

Основными факторами, снижающими надежность метода, являются ошибки, происходящие от округления в промежуточных вычислениях. Таким образом, строгий учет надежности должен основываться на оценке влияния ошибок округления. Практически интересными являются вероятностные оценки, так как оценки „на максимум“ почти всегда практически завышены, в силу малой вероятности их реализации.

В литературе имеется довольно много работ, посвященных исследованию влияния ошибок округления в основных арифметических операциях и в некоторых численных методах линейной алгебры. Сюда относятся работы: Нейман и Голдстайн [1], Тюринг [1], Голдстайн и Нейман [1], Дуайр и Уо [1], Абрамов [2], [3], Хаусхольдер [3], [11], Голдстайн, Меррей и Нейман [1], Карр [1] и др. В частности, детальному анализу подвергнута Нейманом и Голдстайном схема главных элементов Гаусса. Однако в настоящее время далеко не все основные методы подвергнуты такого рода исследованию.

Более того, иногда такое исследование оказывается принципиально невозможным, в силу зависимости влияния ошибок округления от факторов, которые заранее нельзя учесть.

Так, при проведении схемы единственного деления при решении системы линейных уравнений с заранее фиксированным порядком исключения неизвестных, надежность результата определяется в значительной степени тем, будет ли происходить уничтожение значащих цифр по ходу процесса или нет, что, в свою очередь, зависит от существования или non-existence малых главных миноров у матрицы коэффициентов. Для одной и той же системы может оказаться, что при различном выборе порядка исключения неизвестных различной будет и надежность результата. Наиболее надежным будет выбор порядка исключения, совпадающий с порядком в схеме главных

элементов. Таким образом, схема единственного деления с фиксированным порядком исключения неизвестных оказывается не безусловно надежным методом и во всяком случае менее надежным, чем схема главных элементов. Однако отсюда еще не следует, что схема главных элементов всегда предпочтительнее, так как ее реализация значительно сложнее, чем реализация схемы с фиксированным порядком исключения. Многие другие численные методы линейной алгебры оказываются не безусловно надежными. Такими оказываются биортогональный алгорифм и метод минимальных итераций. Их надежность зависит во многом от удачного выбора начального вектора. Среди методов определения собственных значений наиболее надежными являются те, в которых собственные значения определяются минуя вычисление коэффициентов характеристического полинома, так как незначительные ошибки в определении этих коэффициентов могут повлечь значительные ошибки в вычислении корней. Но вместе с тем итерационные методы для решения полной проблемы собственных значений оказываются значительно более трудоемкими, чем точные, связанные с вычислением коэффициентов характеристического полинома.

Нам представляется, что в оценке применимости не безусловно надежных методов основное значение имеет опыт их использования и лишь по мере обобщения этого опыта будет возможно высказать более определенные суждения по этому вопросу.

Скажем еще о некоторых приемах, повышающих надежность методов. Теоретически говоря, каждый „точный“ метод линейной алгебры имеет в себе неограниченный запас надежности, который может быть реализован за счет точности проведения промежуточных вычислений. Значительное повышение точности на всем протяжении вычислительного процесса часто оказывается невозможным, и повышение точности следует применять в наиболее „уязвимых“ ситуациях. Источником значительных ошибок от округления является

вычисление сумм произведений $\sum_{k=1}^n a_k b_k$ с округлением до данного количества цифр в каждом слагаемом. Эту операцию часто целесообразно проводить с двойной точностью, вычисляя каждое слагаемое без округления и отбрасывая лишние знаки после выполнения сложения. Конечно, применение двойной точности иногда оказывается излишним, например, при проведении самоисправляющегося итерационного процесса.

При пользовании не безусловно надежными методами часто может быть целесообразно проводить алгорифм в двух неэквивалентных вариантах; например, проводить схему единственного деления при двух выборах порядка исключения неизвестных; применять биортогональный алгорифм исходя из двух начальных векторов и т. д.

Неплохой косвенной проверкой надежности вычислений может служить применение контролей по ходу процесса.

ДОПОЛНЕНИЕ

Изложим еще один итерационный процесс для решения полной проблемы собственных значений, сообщенный авторам В. Н. Кублановской. Процесс заключается в следующем.

Пусть матрица A симметрична и ее квадрат не имеет кратных собственных значений. Строим последовательность матриц

$$\begin{aligned} AP_1 &= \Lambda_1; \quad P'_1 \Lambda_1 = P_1^{-1} AP_1 = A_1; \\ A_1 P_2 &= \Lambda_2; \quad P'_2 \Lambda_2 = P_2^{-1} A_1 P_2 = A_2; \\ &\dots \dots \dots \dots \\ A_{k-1} P_k &= \Lambda_k; \quad P'_k \Lambda_k = P_k^{-1} A_{k-1} P_k = A_k; \\ &\dots \dots \dots \dots \end{aligned}$$

Здесь P_i ортогональные матрицы, Λ_i левые треугольные. Матрицы P_i могут быть вычислены, например, как произведения матриц вращения или матриц отражения, подобно тому, как это делалось в § 16 при решении линейных систем.

Покажем, что последовательность матриц A_k сходится к диагональной матрице, составленной из собственных значений матрицы A , расположенных в порядке убывания модулей, а матрица $Q_k = P_1 \dots P_k$ при достаточно большем k имеет столбцы сколь угодно близкие к нормированным собственным векторам матрицы A .

Для доказательства установим связь процесса с LR -алгорифмом, примененным к матрице A^2 .

Имеем

$$\begin{aligned} \Lambda_1 \Lambda'_1 &= AP_1 P'_1 A = A^2 \\ \Lambda'_1 \Lambda_1 &= P'_1 A^2 P_1 = A_1^2 = \Lambda_2 \Lambda'_2 \\ &\dots \dots \dots \dots \\ \Lambda'_k \Lambda_k &= P'_k A_{k-1}^2 P_k = A_k^2 = \Lambda_{k+1} \Lambda'_{k+1} \\ &\dots \dots \dots \dots \end{aligned}$$

Пусть Δ_i диагональная матрица, составленная из диагональных элементов Λ_i . Положим

$$\begin{aligned} L_k &= \Delta_{k-1} \dots \Delta_1 \Delta_k \Delta_1^{-1} \dots \Delta_{k-1}^{-1} \Delta_k^{-1} \\ R_k &= \Delta_k \Delta_{k-1} \dots \Delta_1 \Delta_k' \Delta_1^{-1} \dots \Delta_{k-1}^{-1}. \end{aligned}$$

Ясно, что L_k есть левая треугольная матрица с единичной диагональю, а R_k есть правая треугольная матрица с диагональю Δ_k^2 .

Легко проверяется, что $L_1 R_1 = A^2$ и $R_{k-1} L_{k-1} = L_k R_k$, т. е. матрицы L_i и R_i совпадают с одноименными матрицами LR -алгорифма, примененного к матрице A^2 .

Ввиду того, что матрица A^2 положительно определена и ее собственные значения попарно различны, LR -алгорифм для нее сходится, в частности, диагональные элементы матриц R_k сходятся к квадратам собственных значений матрицы A . Следовательно,

$$\operatorname{Sp} R_k = \operatorname{Sp} \Delta_k^2 \rightarrow \operatorname{Sp} A^2.$$

Далее, положив $\Lambda_k = (l_{ij, k})$, имеем

$$\begin{aligned}\operatorname{Sp} A^2 &= \operatorname{Sp} A_k^2 = \operatorname{Sp} \Lambda'_k \Lambda_k = \sum_{i,j} l_{ij, k}^2 = \\ &= \sum_i l_{ii, k}^2 + \sum_{i>j} l_{ij, k}^2 = \operatorname{Sp} \Delta_k^2 + \sum_{i>j} l_{ij, k}^2.\end{aligned}$$

Поэтому, при $k \rightarrow \infty$, $\sum_{i>j} l_{ij, k}^2 \rightarrow 0$, так что все недиагональные элементы матрицы Λ_k стремятся к нулю. Следовательно,

$$A_k^2 = \Lambda'_k \Lambda_k \rightarrow [\lambda_1^2, \dots, \lambda_n^2] \quad \text{при } k \rightarrow \infty.$$

Матрицы A_k при достаточно больших k становятся сколь угодно близкими к диагональным матрицам $[\pm \lambda_1, \dots, \pm \lambda_n]$. Но так как матрицы A_k при всех k подобны матрице A , то

$$A_k \rightarrow [\lambda_1, \dots, \lambda_n].$$

Далее, $A_k = Q'_k A Q_k$ и, следовательно, при достаточно большом k столбцы матрицы Q_k сколь угодно близки к собственным векторам матрицы A , нормированным, в силу ортогональности Q_k . В силу неоднозначности выбора матриц P_k , последовательность матриц Q_k может не быть сходящейся. Она будет сходящейся лишь с точностью до знаков столбцов. Сходимость будет иметь место, если, начиная с некоторого места, брать на каждом шагу матрицу P_k возможно более близкой к единичной.

Процесс позволяет использовать для ускорения сходимости как сдвиги (подобно LR -алгорифму), так и уточняющие формулы метода Якоби.

ЛИТЕРАТУРА

- Абрамов А. А. [1] Ускорение сходимости в итеративных процессах. *Докл. АН СССР*, 1950, 74, 1051—1052; М. Р., 12, 861.
- [2] О влиянии ошибок округления при решении уравнения Лапласа. *Вычисл. матем. и вычисл. техника*, 1953, 1, № 1, 37—40; М. Р., 16, 1156.
- [3] Об ошибке округлений при решении систем линейных уравнений. *Докл. АН СССР*, 1954, 97, № 2, 189—191; Р. Ж. М., 1955, 6107.
- [4] Об ошибке округлений при решении систем линейных уравнений. *Ber. Internat. Math.-Kolloq.*, 1955 (1957), Nov., 151—153; Р. Ж. М., 1958, 6232.
- Азбелев Н. и Виноград Р. [1] Процесс последовательных приближений для отыскания собственных чисел и собственных векторов. *Докл. АН СССР*, 1952, 83, № 2, 173—174; М. Р., 14, 126.
- Алберт (Albert A. A.). [1] A rule for computing the inverse of a matrix. *Amer. Math. Monthly*, 1941, 48, 198—199; М. Р., 2, 100.
- Алекскеров С. С. [1] К вопросу решения системы линейных численных уравнений. *Тр. Азербайдж. индустр. ин-та*, 1957, 16, 5—10; Р. Ж. М., 1958, 4254.
- Аллен (Allen D. N. de G.). [1] *Relaxation methods*. Mc-Graw-Hill Book Company, Inc., New York — Toronto — London, 1954, 257 pp; М. Р., 15, 831.
- Аллен (Allen D. W.). [1] Numerical solution of „n“ linear equations in „n“ unknowns, and the evaluation of „n“ th order determinant (complex coefficients). *J. Roy. Aeronaut. Soc.*, 1956, 60, 350—353; М. Р., 17, 1137.
- Альтман (Altman M.). [1] On the solution of linear algebraic equations. *Bull. Acad. Polon. Sci. cl. 3*, 1957, 5, № 2, 93—97, IX; Р. Ж. М., 1959, 7478.
- [2] On the approximate solution of linear algebraic equations. *Bull. Acad. Polon. Sci. cl. 3*, 1957, 5, № 4, 365—370, XXIX; Р. Ж. М., 1959, 8560.
- Ангелич (Angelitch T. P.). [1] Résolutions des systèmes d'équations linéaires algébriques par la méthode de Banachiewicz. *Srpska Akad. Nauka. Zbornik Radova*, 1952, 18, *Mat. Inst.* 2, 71—92; М. Р., 14, 501.
- Андерсен (Andersen E.). [1] Solution of great systems of normal equations together with an investigation of Andrae's dotfigure. An arithmetic-technical investigation. *Geodætisk Inst. Skriften (Mém. Inst. Géod. Danemark)*, 1947, 3, 11, 65 pp; М. Р., 9, 622.
- [2] Solution of great systems of normal equations. *Bull. Géod.*, 1950, 15, 19—29; М. Р., 11, 693.
- Андре (Andree R. V.). [1] Computation of the inverse of a matrix. *Amer. Math. Monthly*, 1951, 58, 87—92; М. Р., 12, 639.
- Апаро (Aparo Enzo). [1] Sulle equazioni algebriche matriciali. *Atti Accad. naz. Lincei. Rend. Cl. Sci. fis., mat. e natur. Ser. 8*, 1957, 22, 20—23; М. Р., 19, 685.
- Аржаных И. С. [1] Распространение метода А. Н. Крылова на полиномиальные матрицы. *Докл. АН СССР*, 1951, 81, № 5, 749—752; М. Р., 14, 92.
- Армс, Гейтс и Зондек (Arms R. J., Gates L. D. and Zondek B.). [1] A method of block iteration. *J. Soc. Industr. and Appl. Math.*, 1956, 4, № 4, 220—229; Р. Ж. М., 1958, 2434.

Арнольди (Arnoldi W. E.). [1] The principle of minimized iteration in the solution of the matrix eigenvalue problem. *Quart. Appl. Math.*, 1951, 9, 17—29; M. R., 13, 163.

Аспельтия (Azpeitia A. G.). [1] Un método para el cálculo de la matriz inversa. *Rev. Real Acad. Cienc. Exactas, fis. y natur.*, Madrid, 1956, 50, № 4, 463—470; P. Ж. М., 1957, 8962.

Атта (Atta Susie A.). [1] Effect of propagated error on inverse of Hilbert matrix. *J. Assoc. Comput. Machinery*, 1957, 4, № 1, 36—40; P. Ж. М., 1959, 3244.

Африат (Afriat S. N.). [1] An iterative process for the numerical determination of characteristic values of certain matrices. *Quart. J. Math.*, Oxford, Ser. (2), 1951, 2, 121—122; M. R., 12, 861—862.

Бабушка (Babuška Ivo). [1] O jednom numerickém řešení úplne regulárních systémů lineárních rovnic a o jeho aplikaci na statické řešení patrových rámů. *Časop. pěstov. mat.*, 1955, 80, № 1, 60—88; P. Ж. М., 1956, 4107.

Базиль (Basile R.). [1] Résolution de systèmes d'équations linéaires algébriques et inversions de matrices au moyen des machines de mécanographie comptable. *Complément pratique par R. Janin. Office National d'Études et de Recherches Aéronautiques, Paris*, 1949, 28; M. R., 11, 692.

Базиль и Жанен (Basile R. et Janin R.). [1] Résolution de systèmes d'équations linéaires algébriques et inversions de matrices au moyen des machines de mécanographie comptable. *Office National d'Études et de Recherches Aéronautiques, Paris*, 1949, 28; M. R., 12, 208.

Баллантайн (Ballantine J. P.). [1] Numerical solutions of linear equations by vectors. *Amer. Math. Monthly*, 1931, 38, 275—277.

Бакман (Backman G.). [1] Rekursionsformeln zur Lösung der Normalgleichungen auf Grund der Krakowianenmethodik. *Ark. Mat., Astr. Fys.*, 1946, 33 A, № 1, 1—14; M. R., 8, 287.

Банахевич (Banachiewicz T.). [1] Zur Berechnung der Determinanten, wie auch der Inversen, und zur darauf basierten Auflösung der Systeme linearer Gleichungen. *Acta Astron.*, Ser. C, 1937, 3, 42—67.

[2] Calcul des déterminants par la méthode des cracoviens. *Bull. intern. Acad. Polon. Sci. A.*, 1937, 109—120.

[3] Sur la résolution numérique d'un système d'équations linéaires. *Bull. Intern. Acad. Polon. Sci. A.*, 1937, 350—354.

[4] Principes d'une nouvelle technique de la méthode des moindres carrés. *Bull. Intern. Acad. Polon. Sci. A.*, 1938, 134—135.

[5] Méthode de résolution numérique des équations linéaires, du calcul des déterminants et des inverses et de réduction des formes quadratiques. *Bull. Intern. Acad. Polon. Sci. A.*, 1938, 393—401.

[6] La règle de Chio, cracoviens et matrices. *Bull. intern. Acad. Polon. Sci. A.*, 1939, 405—412.

[7] An outline of the Cracovian algorithm of the method of least squares. *Astr. J.*, 1942, 50, 38—41; M. R., 4, 90—91.

[8] Fragmentos de novo algoríthmo de metodo de mínimo quadratos. *Rocznik Astr. Obserw. Krakow. Suppl. Internat.*, 1949, 20, 87—98; M. R., 11, 403.

[9] Sur la résolution des équations normales de la méthode des moindres carrés. *Soc. Sci. Lett. Varsovie. C. R. Cl. III, Sci. Math., Phys.*, 1948, 41, 63—68 (1950); M. R., 13, 285.

[10] Résolution d'un système d'équations linéaires algébriques par division. *Ens Igneum Math.*, 1951, 39, (1942—1950), 34—45; M. R., 12, 861.

Бандемер (Bandemer Hans). [1] Berechnung der reellen Eigenwerte einer reellen Matrix mit dem Verfahren von Rutishauser. *Wiss. Z. Martin-Luther Univ. Halle-Wittenberg. Math.-naturwiss. Reihe*, 1957, 6, № 5, 807—814; P. Ж. М., 1959, 2015.

Бандъопадхъай и Нарасимхан (Bandyopadhyay G. and Narasimhan R. K.). [1] Special types of group relaxation for simultaneous linear equations. *Quart. J. Mech. and Appl. Math.*, 1956, 9, № 1, 122—128; Р. Ж. М., 1957, 2683.

Баранкин (Barankin Edward W.). [1] Bounds for the characteristic roots of a matrix. *Bull. Amer. Math. Soc.*, 1945, 51, 767—770; M. R., 7, 107.

[2] Bounds on characteristic values. *Bull. Amer. Math. Soc.*, 1948, 54, 728—735.

Бартлетт (Bartlett M. S.). [1] An inverse matrix adjustment arising in discriminant analysis. *Ann. Math. Statistics*, 1951, 22, 107—111; M. R., 12, 639.

Барч (Bartsch Helmut). [1] Ein Einschließungssatz für die charakteristischen Zahlen allgemeiner Matrizen-Eigenwertaufgaben. *Arch. Math.*, 1953, 4, № 2, 133—136; Р. Ж. М., 1954, 3235.

[2] Abschätzungen für die kleinste charakteristische Zahl einer positiv-definiten hermiteschen Matrix. *Z. angew. Math. und Mech.*, 1954, 34, № 1—2, 72—74; Р. Ж. М., 1955, 4710.

Бауи (Bawie O. L.). [1] Practical solution of simultaneous linear equations. *Quart. Appl. Math.*, 1951, 8, 369—373; M. R., 12, 538.

Баукер (Bowker A. H.). [1] On the norm of a matrix. *Ann. Math. Statistics*, 1947, 18, 285—288; M. R., 9, 75.

Бауэр (Bauer Friedrich L.). [1] Der Newton-Prozeß als quadratisch konvergente Abkürzung des allgemeinen linearen stationären Iterationsverfahrens. I. Ordnung (Wittmeyer-Prozeß). *Z. Angew. Math. und Phys.*, 1955, 35, № 12, 469—470; Р. Ж. М., 1957, 912.

[2] Das Verfahren der abgekürzten Iteration für algebraische Eigenwertprobleme, insbesondere zur Nullstellenbestimmung eines Polynoms. *Z. angew. Math. und Phys.*, 1956, 7, № 1, 17—32; Р. Ж. М., 1958, 752.

[3] Zur numerischen Behandlung von algebraischen Eigenwertproblemen höherer Ordnung. *Z. angew. Math. und Mech.*, 1956, 36; M. R., 18, 766.

[4] Iterationsverfahren der linearen Algebra vom bernoullischen Konvergenztyp. *Nachrichtentechn. Fachber.*, 1956, 4, 171—175, 221. Diskuss; Р. Ж. М., 1957, 8266.

[5] Zusammenhänge zwischen einigen Iterationsverfahren der linearen Algebra. *Ber. Internat. Math. Kolloq.*, 1955 (1957), Nov., 99—111; Р. Ж. М., 1958, 10300.

[6] Beiträge zum Danilewski-Verfahren. *Ber. Internat. Math. Kolloq.*, 1955 (1957), Nov., 133—139; Р. Ж. М., 1959, 5229.

[7] Das Verfahren der Treppeniteration und verwandte Verfahren zur Lösung algebraischer Eigenwertprobleme. *Z. angew. Math. und Phys.*, 1957, 8, № 3, 214—235; Р. Ж. М., 1958, 10299.

[8] On modern matrix iteration process of Bernoulli and Greffe type. *J. Assoc. Comput. Machinery*, 1958, 5, № 3, 246—258; Р. Ж. М., 1959, 8573.

Беджарано и Розенблатт (Bejaramo Gabriel G. and Rosenblatt Bruce R.). [1] A solution of simultaneous linear equations and matrix inversion with high speed computing devices. *Math. Tables and other Aids Comput.*, 1958, 7, № 42, 77—81; Р. Ж. М., 1954, 3837.

Белл (Bell W. D.). [1] Punched card techniques for the solution of simultaneous equations and other matrix operations. *Proc. Scient. Comput. Forum*, 1948. N. Y., I. B. M. Corp., 1950, 28—31; M. R., 13, 887.

Бендиксон (Bendixson I.). [1] Sur les racines d'une équation fondamentale. *Acta Math.*, 1902, 25, 359—365.

Берардино (Di Berardino V.). [1] Risoluzione dei sistemi di equazioni algebriche lineari per incrementi successivi delle incognite. Nuovo metodo di calcolo. *Riv. catastro e serv. tecn. eraralti*, 1956, 11, № 5, 6, 334—338; Р. Ж. М., 1958, 7200.

[2] Il metodo di Hardy Cross e la sua giustificazione mediante un nuovo procedimento di risoluzione dei sistemi di equazioni algebriche lineari. *Ingegneria ferroviaria*, 1957, 12, № 10, 821—831; Р. Ж. М., 1959, 876.

Берардино и Джирарделли (Di Berardino V. e Girardelli L.). [1] Nuovo metodo di risoluzione di particolari sistemi di equazioni algebriche lineari. Sistemi a catena. *Riv. catasti e serv. tech. erariali*, 1957, 12, № 1, 46—50; Р. Ж. М., 1958, 7201.

Берардино и Франди (Di Berardino V. e Frandi P.). [1] Formule ricorrenti per la risoluzione graduale dei sistemi di equazioni algebriche lineari. *Archimede*, 1950, 2, 108—113; M. R., 13, 586.

[2] Formule ricorrenti per la risoluzione graduale dei sistemi di equazioni algebriche lineari. *Ricerca Sci.*, 1950, 20, 662—666; M. R., 13, 587.

Бергер (Berger A. P.). [1] Inversion of matrices with the aid of punched card machines. *Statistica*, Rijswijk, 1952, 6, 121—133; M. R., 14, 1128.

Берджер и Сейбл (Berger E. J. and Saibel Edward). [1] On the inversion of continuant matrices. *J. Franklin Inst.*, 1953, 256, № 3, 249—253; Р. Ж. М., 1954, 5067.

Бёрджесс (Burgess H. T.). [1] On the matrix equation $BX = C$. *Amer. Math. Monthly*, 1916, 23, 152—155.

Берри (Berry C. E.). [1] A criterion of convergence for the classical iterative method of solving linear simultaneous equations. *Ann. Math. Statistics*, 1945, 16, 398—400; M. R., 7, 338.

Берш-Зупан и Боттенбрюх (Borsch-Supan W. und Bottendorf H.). [1] Eine Methode zur Eingrenzung sämtlicher Eigenwerte einer hermiteschen Matrix mit überwiegender Hauptdiagonale. *Z. angew. Math. und Mech.*, 1958, 38, № 5—6, 169—171; Р. Ж. М., 1958, 10301.

Биденхарн и Блэйт (Biedenharn L. C. and Blatt J. M.). [1] A variation principle for eigenfunctions. *Physical Rev.*, 1954, (2) 93, 230—232; M. R., 15, 745.

Бильт (Bil'y J.). [1] Solution of a system of linear equations with large coefficients in the diagonal. *Aktuarske Vedy*, 1949, 5, № 3, 114—127; M. R., 11, 403.

Бингхэм (Bingham M. D.). [1] A new method for obtaining the inverse matrix. *J. Amer. Statist. Assoc.*, 1941, 36, 530—534; M. R., 3, 154.

Бирман М. Ш. [1] Некоторые оценки для метода наискорейшего спуска. *Успехи матем. наук*, 1950, 5, № 3, 152—155.

[2] Об одном варианте метода последовательных приближений. *Вестн. Ленингр. ун-та*, серия матем., физ. и хим., 1952, 9, 69—76.

[3] О вычислении собственных чисел методом наискорейшего спуска. *Л. Записки Горн. ин-та*, 1952, 27, № 1, 209—216.

Бицено и Боттема (Biezeno C. B. and Bottema O.). [1] The convergence of a specialized iterative process in use in structural analysis. *Proc. Koninkl. nederl. akad. wetensch.*, 1946, 49, 489—499 (Also *Indagationes math.*, 8); M. R., 9, 104.

Бишоп (Bishopp K. E.). [1] The inverse of a stiffness matrix. *Quart. Appl. Math.*, 1945, 3, 82—84; M. R., 6, 218.

Блан и Линигер (Blanc Ch. et Liniger W.). [1] Erreurs de chute dans la résolution de systèmes algébriques linéaires. *Comment. math. helv.*, 1956, 30, № 4, 257—264; Р. Ж. М., 1957, 1839.

Блэк (Black A. N.). [1] Further notes on the solutions of algebraic linear simultaneous equations. *Quart. J. Mech. and Appl. Math.*, 1949, 2, 321—324; M. R., 11, 743.

Блюменталь (Blumenthal O.). [1] Über die Genauigkeit der Wurzeln linearer Gleichungen. *Z. Math. und Phys.*, 1914, 62, 359—362.

Бодеиг (Bodewig E.). [1] Comparison of some direct methods for computing determinants and inverse matrices. *Proc. Koninkl. nederl. akad. wetensch.*, 1947, A, 50, 49—57 (Also *Indagationes math.*, 9); M. R., 8, 407.

[2] Bericht über die verschiedenen Methoden zur Lösung eines Systems linearer Gleichungen mit reellen Koeffizienten I, II, III, IV, V. *Proc. Koninkl. nederl. akad. wetensch.*, 1947, 50, 930—941; 1104—1116; 1285—1295; 1948, 51, 53—64; 211—219 (Also *Indagationes math.*, 9; 10); M. R., 9, 251, 382, 621.

[3] Bericht über die Methoden zur numerischen Lösung von algebraischen Eigenwertproblemen. I, II. *Atti. Sem. Mat. Fis. Univ. Modena*, 1951, № 3, 3—39; № 4, 133—193; M. R., 13, 991.

[4] A practical refutation of the iteration method for the algebraic eigenproblem. *Math. Tables and Other Aids Comput.*, 1954, 8, № 48, 237—240; P. Ж. М., 1956, 762.

[5] Zum Matrizenkalkül. I—V. *Proc. Koninkl. nederl. akad. wetensch.*, 1955, A 58, № 1, 95—106; 1956, A 59, № 3, 301—304; 1956, A 59, № 3, 305—312; 1957, A 60, № 1, 82—87; 1957, A 60, № 3, 242—247. (Also *Indagationes math.*, 17; 18; 19); P. Ж. М., 1955, 4876; 1957, 4614; 1958, 3298, 3299; 1959, 881.

[6] *Matrix calculus*. Amsterdam, North-Holl. Publ. Co., 1956, xli, 334; P. Ж. М., 1957, 4623.

[7] Zu Stiefels Berechnung der Eigenwerte aus den Schwarzschen Konstanten. *Z. angew. Math. und Mech.*, 1958, 38, № 1—2, 72—73; P. Ж. М., 1959, 880.

Боде виг и Цюрмюль (Bodewig E. und Zurmühl R.). [1] Zu R. Zurmühl. Zur numerischen Auflösung linearer Gleichungssysteme nach dem Matrixverfahren von Banachiewicz. *Z. angew. Math. und Mech.*, 1950, 30, 130—132; M. R., 11, 743.

Боли (Bolie Victor W.). [1] Minimum-storage matrix inversion. *Z. angew. Math. und Mech.*, 1958, 38, № 9—10, 369—372; P. Ж. М., 1959, 5223.

Больц (Boltz H.). [1] Entwicklungsverfahren zur Ausgleichung Geodätischer Netze nach der Methode der kleinsten Quadrate. *Veröff. Preussischen Geod. Inst.*, 1928, 90.

Бондарь Н. Г. [1] О точности некоторых приближенных методов вычисления собственных чисел квадратных матриц. *Tr. Днепропетровского ин-та инж. ж.-д. транспорта*, 1953, 23, 61—69; P. Ж. М., 1954, 3468.

Бородянский М. Я. [1] Приведение некоторого типа матриц к диагональному виду. *Tr. Киевск. Технол. ин-та пищ. пром-ти*, 1953, 13, 195—196.

Боттема (Bottema O.). [1] A geometrical interpretation of the relaxation method. *Quart. Appl. Math.*, 1950, 7, 422—423; M. R., 11, 403.

Бошан (Boschan Paul). [1] The consolidated Doolittle technique. *Ann. Math. Statistics*, 1946, 17, 503.

Бранстеттер (Branstetter R. D.). [1] A round-off theory for scalar products. *Iowa State Coll. J. Sci.*, 1954, 28, № 3, 283—284; P. Ж. М., 1955, 4715.

Браун (Brown E. T.). [1] The characteristic equation of a matrix. *Bull. Amer. Math. Soc.*, 1928, 34, 363—368.

[2] The characteristic roots of a matrix. *Bull. Amer. Math. Soc.*, 1930, 36, 705—710.

[3] Limits to the characteristic roots of a matrix. *Amer. Math. Monthly*, 1939, 46, 252—265.

Браун и Бассетт (Brown R. D. and Bassett J. M.). [1] A method for calculating the first order perturbation of an eigenvector of a finite matrix, with applications to molecular-orbital theory. *Proc. Phys. Soc.*, 1958, 71, № 5, 724—732; P. Ж. М., 1959, 4249.

Браузэр (Brauer Alfred). [1] Limits for the characteristic roots of a matrix. I—VI (VI совм. с Л. Борд (La Borde H. T.)). *Duke Math. J.*, 1946, 13, 387—395; 1947, 14, 21—26; 1948, 15, 871—877; 1952, 19, 75—91; 1952, 19, 553—562; 1955, 22, 253—261; M. R., 8, 192; 8, 559; 10, 231; 13, 813; 14, 836; 17, 1044.

- [2] On the characteristic equations of certain matrices. *Bull. Amer. Math. Soc.*, 1947, **53**, 605—607; M. R., 8, 559.
- [3] Matrices with all their characteristic roots in the interior of the unit circle. *J. Elisha Mitchell Soc. Sci.*, 1952, **68**, 188—183; M. R., 14, 836.
- [4] Über die Lage der charakteristischen Wurzeln einer Matrix. *J. reine und angew. Math.*, 1953, **192**, № 2, 113—116; P. Ж. М., 1954, 5451.
- [5] Bounds for characteristic roots of matrices. *Nat. Bur. Standards Appl. Math. Ser.*, 1953, **29**, 101—106.
- [6] Bounds for the ratios of the coordinates of the characteristic vectors of a matrix. *Proc. Nat. Acad. U. S. A.*, 1955, **41**, № 3, 162—164; P. Ж. М., 1956, 2780.
- [7] The theorem of Ledermann and Ostrowski on positive matrices. *Duke Math. J.*, 1957, **24**, № 2, 265—274; P. Ж. М., 1958, 2733.
- [8] A new-proof of theorems of Perron and Frobenius on nonnegative matrices. I. Positive matrices. *Duke Math. J.*, 1957, **24**, № 3, 367—378.
- [9] A method for the computation of the greatest root of a positive matrix. *J. Soc. Indust. Appl. Math.*, 1957, **5**, 250—253; M. R., 19, 1797.
- Бреннер и Рейтвайснер (Brenner J. L. and Reitwiesner G. W.). [1] Remark on determination of characteristic roots by iteration. *Math. Tables and Other Aids Comput.*, 1955, **9**, № 51, 117—118; P. Ж. М., 1957, 903.
- Бродский М. Л. [1] Вероятностные оценки погрешностей при определении собственных значений и собственных векторов варьирующейся матрицы. *Успехи матем. наук*, 1952, **7**, № 5, 205—214; M. R., 14, 692.
- Брок (Brock John E.). [1] Variation of coefficients of simultaneous linear equations. *Quart. Appl. Math.*, 1953, **11**, № 2, 234—240; P. Ж. М., 1954, 2351.
- Брукер и Сумнер (Brooker R. A. and Sumner F. H.). [1] The method of Lanczos for calculating the characteristic roots and vectors of a real symmetric matrix. *Proc. Inst. Electr. Engrs.*, 1956, В 103 Suppl. № 1, 114—119. Discuss., 120—122; P. Ж. М., 1958, 2439.
- Брунер (Bruner N.). [1] Note on the Doolittle solution. *Econometrica*, 1947, **5**, 43—44; M. R., 8, 407.
- Бурдина В. И. [1] К одному методу решения систем линейных алгебраических уравнений. *Докл. АН СССР*, 1958, **120**, № 2, 235—238; P. Ж. М., 1959, 5218.
- Буркхард (Burkhardt Felix). [1] Über spezielle lineare Gleichungssysteme mit der Eigenschaft $\lim_{n \rightarrow \infty} (\mathfrak{F} - \mathfrak{A})^n = 0$. *Wiss. Z. Univ. Leipzig*, 1952/53, № 5, 187—192; P. Ж. М., 1956, 4106.
- Бърхамар (Вјерхамар А.). [1] Rectangular reciprocal matrices with special reference to geodetic calculations. *Bull. géod.*, 1951, 188—220; M. R., 13, 312.
- [2] Triangular matrices for adjustment of triangular networks. *Kungl. Tekn. Högsk. Handl.*, Stockholm, 1956, **105**, 82.
- Бэтсле (Baetslé P. L.). [1] Sur les méthodes itératives de calcul numérique des vecteurs propres d'une matrice. *III-e Congrès National des Sciences*, Bruxelles, 1950, **2**, 104—106; M. R., 17, 666.
- [2] Systématisation des calculs numériques de matrices. *Bull. géod.*, 1951, **22**—**41**; M. R., 12, 861.
- Бюкнер (Bückner H.). [1] Über ein unbeschränkt anwendbares Iterationsverfahren für Systeme linearer Gleichungen. *Arch. Math.*, 1950, **2**, 172—177; M. R., 11, 743.
- Важевский (Ważewski Tadeusz). [1] Sur l'algorithmitation des méthodes d'éliminations successives. *Ann. Soc. polon.*, 1953, **24**, № 2, 157—164; P. Ж. М., 1956, 4108.

- Вазов (Wasow W. R.). [1] A note on the inversion of matrices by random values. *Math. Tables and Other Aids Comput.*, 1952, 6, 78—81.
- Варга (Varga Richard). [1] Eigenvalues of circulant matrices. *Pacif. J. Math.*, 1954, 4, № 1, 151—160; Р. Ж. М., 1956, 4918.
- [2] A comparison of the successive overrelaxation method and semi-iterative methods using Chebyshev polynomials. *J. Soc. Industr. and Appl. Math.*, 1957, 5, 39—46; Р. Ж. М., 1958, 9266.
- Васидзу (Washizu K.). [1] On the bounds of eigenvalues. *Quart. J. Mech. and Appl. Math.*, 1955, 8, № 3, 311—325; Р. Ж. М., 1956, 6864.
- Васильевский С. [1] Схема для решения нормальных уравнений на вычислительных машинах. *Тр. Латвийского ун-та*, 1940, 3, № 11, 3—12; M. R., 3, 154.
- Вебер (Weber R.). [1] Sur les méthodes de calcul employées pour la recherche des valeurs et vecteurs propres d'une matrice. *Rech. aéronaut.*, 1949, № 10, 57—60; M. R., 11, 266.
- Вегнер (Wegner Udo). [1] Bemerkungen zur Matrizentheorie. *Z. angew. Math. und Mech.*, 1953, 33, 262—264; M. R., 15, 388.
- [2] Contributi alla teoria dei procedimenti iterativi per la risoluzione numerica dei sistemi di equazioni lineari algebriche. *Atti Accad. naz. Lincei Mem., cl. sci. fis., mat. e natur.*, 1953, 4, № 1, 1—48; Р. Ж. М., 1954, 5781.
- Вейленд (Wayland H.). [1] Expansion of determinantal equations into polynomial form. *Quart. Appl. Math.*, 1945, 2, 277—306; M. R., 6, 218. (Есть перевод. Успехи матем. наук, 1947, 2, № 4, 128—158.)
- Вейнер (Weiner B. L.). [1] Variations of coefficients of simultaneous linear equations. (With Discussion). *Trans. Amer. Soc. Civil Engrs*, 1948, 113, 1349—1390.
- Вайссингер (Weissinger Johannes). [1] Über das Iterationsverfahren. *Z. angew. Math. und Mech.*, 1951, 31, 245—246.
- [2]. Zur Theorie und Anwendung des Iterationsverfahrens. *Math. Nachr.*, 1952, 8, 193—212.
- [3] Verallgemeinerung des Seidelschen Iterationsverfahrens. *Z. angew. Math. und Mech.*, 1953, 33, 155—163.
- Венке (Wenke Klaus). [1] Erfahrungen und Probleme bei der Lochkartenmässigen Berechnung von Kehrmatrizen. *Nachrichtentechn. Fachber.*, 1956, 4, 198—201, 227; Р. Ж. М., 1958, 1596.
- Верзух (Verzuh F. M.). [1] The solution of simultaneous linear equations with the aid of the 60-r calculating punch. *Math. Tables and Other Aids Comput.*, 1949, 3, 453—462; M. R., 11, 57.
- Взорова А. И. [1] О решении системы линейных алгебраических уравнений способом Ю. А. Шрейдера. *Вычисл. матем. и вычисл. техника*, 1953, № 1, 90—94; Р. Ж. М., 1954, 5266.
- Виландт (Wielandt Helmut). [1] Ein Einschliessungssatz für charakteristische Wurzeln normaler Matrizen. *Arch. Math.*, 1949, 1, 348—352; M. R., 11, 4.
- [2] Die Einschliessung von Eigenwerten normaler Matrizen. *Math. Ann.*, 1949, 121, 234—241; M. R., 11, 307.
- [3] Inclusion theorems for eigenvalues. *Nat. Bur. Standards Appl. Math. Ser.*, 1953, 29, 75—78; Р. Ж. М., 1956, 1975.
- [4] Einschliessung von Eigenwerten hermitescher Matrizen nach dem Abschnittsverfahren. *Arch. Math.*, 1954, 5, № 1—3, 108—114; Р. Ж. М., 1956, 4103.
- [5] An extremum property of sums of eigenvalues. *Proc. Amer. Math. Soc.*, 1955, 6, № 1, 106—110; Р. Ж. М., 1956, 6394.
- Виноград (Vinograd B.). [1] Note on the escalator method. *Proc. Amer. Math. Soc.*, 1950, 1, 162—164; M. R., 11, 618.
- Витмейер (Wittmeyer Helmut). [1] Einfluss der Änderung einer Matrix auf die Lösung des zugehörigen Gleichungssystems, sowie auf die charakteristischen Zahlen und die Eigenvektoren. Dissertation, Darmstadt, 1934.

[2] Einfluss der Änderung einer Matrix auf die Lösung des Zugehörigen Gleichungssystems, sowie auf die charakteristischen Zahlen und die Eigenvektoren. *Z. angew. Math. und Mech.*, 1936, 16, № 5, 287—300.

[3] Über die Lösung von linearen Gleichungssystemen durch Iteration. *Z. angew. Math. und Mech.*, 1936, 16, № 5, 301—310.

[4] Berechnung einzelner Eigenwerte eines algebraischen linearen Eigenwertproblems durch „Störiteration“. *Z. angew. Math. und Mech.*, 1955, 35, № 12, 441—452; Р. Ж. М., 1957, 2686.

Вольта (Volta E.). [1] Un nuovo metodo per la risoluzione rapida di sistemi di equazioni lineari. *Atti Accad. naz. Lincei. Rend. Cl. sci. fis., mat. e natur.* ser. 8, 1949, 7, № 50, 203—207; M. R., 11, 743.

Ворх (Worch G.). [1] Über die zweckmässigste Art, lineare Gleichungen durch Elimination aufzulösen. *Z. angew. Math. und Mech.*, 1932, 12, 175—181.

Вриес (Vries D. de). [1] Eigenwaarden van matrices. *Tijdschr. Kadaster en Landmeetkunde*, 1953, 69, № 5, 316—322; Р. Ж. М., 1956, 4104.

Вучкович (Вучковић Милорад). [1] Систем линеарних једначина и његова примена у решавању статички неодређених конструкција. *Изградна*, 1956, 10, № 7, 8, 3—13; Р. Ж. М., 1958, 2437.

Вујаклија (Vuјаклија Г.). [1] Sur le calcul des déterminants. *Godišnjak Techn. Fak. Univ. Beograd*, 1946—47, № 1—4, 1949; M. R., 11, 154.

Гаврилов Ю. М. [1] Про збіжність простих ітерацій та критерії знаковизначеності квадратичних форм. *Доповіді АН УРСР*, 1953, № 6, 389—393; Р. Ж. М., 1955, 436.

[2] О сходимости итерационных процессов и критериях знакопределённости квадратичных форм. *Изв. АН СССР*, сер. мат., 1954, 18, № 1, 87—94; Р. Ж. М., 1955, 437.

[3] Про збіжність простих, а також групових ітерацій при розв'язуванні систем нормальних рівнянь. *Наук. зап. Львовск. політехн. ин-та*, 1955, 29, 114—120; Р. Ж. М., 1956, 8348.

Гавурин М. К. [1] Применение полиномов наилучшего приближения для улучшения сходимости итеративных процессов. *Успехи матем. наук*, 1950, 5, № 3, 156—160; M. R., 12, 209.

Гантмахер Ф. Р. [1] К алгебраическому анализу метода ак. А. Н. Крылова преобразования векового уравнения. *Пр. 2-го Всесоюзного матем. съезда*, 1937, 45—48.

[2] Теория матриц. Гостехиздат, 1953.

Гарса (Garza A. de la). [1] Error bounds on approximate solutions to systems of linear algebraic equations. *Math. Tables and Other Aids Comput.*, 1953, 7, № 42, 81—84; Р. Ж. М., 1954, 3838.

[2] Error bounds for a numerical solution of a recurring linear system. *Quart. Appl. Math.*, 1956, 13, № 4, 453—456, Р. Ж. М., 1958, 4239.

Гастинель (Gastinel Noël). [1] Procédé itératif pour la résolution numérique d'un système d'équations linéaires. *C. r. Acad. sci.*, 1958, 246, № 18, 2571—2574; Р. Ж. М., 1959, 4244.

Гаттман (Guttmann L.). [1] Enlargement methods for computing the inverse matrix. *Ann. Math. Statistics*, 1946, 17, 336—343; M. R., 8, 171.

Гатто (Gatto F.). [1] Sulla risoluzione numerica dei sistemi di equazioni lineari. *Ricerca Sci.*, 1949, 19, 1385—1388; M. R., 11, 743.

Гатшолл (Gutshall W. D.). [1] Practical inversion of matrices of high order. *Proc. Comput. Sem. Dec.* 1949, N. Y., IBM Corp., 1951, 171—173; M. R., 13, 387.

Гаусс (Gauss C. F.). [1] Supplementum theoriae combinationis observationum erroribus minimis obnoxiae. *Werke*, Göttingen, 1873, 4, 55—93.

Гауччи (Gautschi Werner). [1] The asymptotic behaviour of powers of matrices. *Duke Math. J.*, 1953, 20, 127—140; M. R., 15, 94.

[2] The asymptotic behaviour of powers of matrices. II. *Duke Math. J.*, 1953, 20, № 3, 375—379; Р. Ж. М., 1954, 3620.

- [3] Bounds of matrices with regard to an Hermitian metric. *Compositio Math.*, 1954, 12, 1–16.
- [4] On norms of matrices and some relations between norms and eigenvalues. *Philosoph.-Naturwiss. Fak. der Universität Basel*, Basel, Buchdruckerei Birkhäuser AG, 1954, 39 S.; P. Ж. М., 1959, 5559.
- Гейрингер (Geiringer H.). [1] Zur Praxis der Lösung linearer Gleichungen in der Statik. *Z. angew. Math. und Mech.*, 1928, 8, 446–447.
- [2] On the numerical solution of linear problems by group iteration. *Bull. Amer. Math. Soc.*, 1942, 48, p. 370.
- [3] On the solution of systems of linear equations by certain iteration methods. *Reissner Anniversary Volume, Contrib. Appl. Mech.*, 1948, 365–393; Ann. Arbor, Mich.; M. R., 10, 574.
- Гельфанд И. М. [1] Лекции по линейной алгебре. Изд. 2-е, М. Л. 1951.
- Герман (Hermann A.). [1] Bestimmung der höheren Eigenwerte einer Matrix durch Iteration. *Ann. Univ. Saravensis*, 1952, 1, 220–223; M. R., 14, 1129.
- Германский (Germansky Boris). [1] Zur angenäherten Auflösung linearer Gleichungssysteme mittels Iteration. *Z. angew. Math. und Mech.*, 1936, 16, 57–58.
- Гершгорин С. А. [1] Ueber die Abgrenzung der Eigenwerte einer Matrix. *Изв. АН СССР, сер. матем.*, 1931, 7, 749–754.
- Гест (Guest I.). [1] The solution of linear simultaneous equations by matrix iteration. *Austral. J. Phys.*, 1955, 8, № 4, 425–439; P. Ж. М., 1957, 6710.
- Гивенс (Givens Wallace). [1] A method of computing eigenvalues and eigenvectors suggested by classical results on symmetric matrices. *Nat. Bur. Standards. Appl. Math. Ser.*, 1953, 29, 117–122; P. Ж. М., 1956, 6861.
- [2] Numerical computation of the characteristic values of a real symmetric matrix. *U. S. Atomic Energy Comm. Repts.*, 1954, ORNL—1574, 107 pp.; P. Ж. М., 1956, 8347.
- [3] The characteristic value-vector problem. *J. Assoc. Comput. Machinery*, 1957, 4, № 3, 298–307; P. Ж. М., 1959, 2014.
- [4] Computation of plane rotations transforming a general matrix to triangular form. *J. Soc. Industr. and Appl. Math.*, 1958, 6, № 1, 26–50; M. R., 19, 1081.
- Глодан (Gloden A.). [1] La méthode du Luxembourgeois B. I. Clasen pour la résolution d'un système d'équations linéaires (méthode des coefficients égaux). *Rev. histtoire sci.*, 1953, 6, № 2, 168–170; P. Ж. М., 1954, 2353.
- Голдстайн и Нейман (Goldstine H. and Neumann J.). [1] Numerical inverting of matrices of high order, II. *Proc. Amer. Math. Soc.*, 1951, 2, 188–202; M. R., 12, 861.
- Голдстайн, Меррей и Нейман (Goldstine H. H., Murray F. J. and J. von Neumann). [1] The Jacobi method for real symmetric matrices. *J. Assoc. Comput. Machinery*, 1959, 6, № 1, 59–97.
- Голдстайн и Хорвиц (Goldstine H. H. and Hotz L. P.). [1]. A procedure for the diagonalization of normal matrices. *J. Assoc. Comput. Machinery*, 1959, 6, № 2, 176–195.
- Гопштейн Н. М. [1] О решении однородных линейных уравнений методом итерации. *Докл. АН СССР*, 1944, 43, 332–395; M. R., 6, 218.
- Готхардт (Gotthardt E.). [1] Boltzsches Entwicklungsverfahren und Gaußscher Algorithmus. *Z. Vermessungswesen*, 1953, 78, № 4, 97–104.
- Гофман (Hoffman A. J.). [1] Lower bounds for the rank and location of the eigenvalues of a matrix. *Nat. Bur. Standards. Appl. Math. Ser.*, 1954, 39, 117–130.
- Гофман и Вилендт (Hoffman A. J. and Wielandt H. W.). [1] The variation of the spectrum of a normal matrix. *Duke Math. J.*, 1953, 20, № 1, 37–39; P. Ж. М., 1953, 587.
- Гофман Ш. М. [1] К вопросу о решении системы линейных алгебраических уравнений. *Докл. АН Уз. ССР*, 1949, № 2, 7–10.

[2] Об оценке одного определителя и о некоторых неравенствах, связанных с итерационными процессами Зейделя. Тр. Ташкентск. ин-та ж.-д. трансп., 1956, вып. 5, 178—186; Р. Ж. М., 1957, 902.

[3] Об одном варианте решения системы трёхчленных матричных уравнений. Тр. Ташкентск. ин-та ж.-д. трансп., 1956, вып. 5, 204—206; Р. Ж. М., 1957, 2684.

Грандалл (Grandall S. H.). [1] On a relaxation method for eigenvalue problems. *J. Math. and Phys.*, 1951, 30, 140—145; M. R., 13, 496.

[2] Iterative procedures related to relaxation methods for eigenvalue problems. *Proc. Roy. Soc., Ser. A*, 1951, 207, 416—423; M. R., 13, 163.

Грегори (Gregory Robert T.). [1] Computing eigenvalues and eigenvectors of a symmetric matrix on the ILLIAC. *Math. Tables and Other Aids Comput.*, 1953, 7, 44, 215—221.

[2] On the convergence rate of an iterative process. *Math. Mag.*, 1955, 29, № 2, 63—68; Р. Ж. М., 1955, 7646.

[3] Results using Lanczos method for finding eigenvalues of arbitrary matrices. *J. Soc. Industr. and Appl. Math.*, 1958, 6, № 2, 182—188.

Грей (Gray H. J. Jr.). [1] Numerical methods in digital real-time simulation. *Quart. Appl. Math.*, 1954, 12, 133—140; M. R., 15, 991.

Гринспан (Greenspan Donald). [1] Methods of matrix inversion. *Amer. Math. Monthly*, 1955, 62, № 5, 303—318; Р. Ж. М., 1956, 6870.

Гринштадт (Greenstadt J.). [1] A method for finding roots of arbitrary matrices. *Math. Tables and Other Aids Comput.*, 1955, 9, № 50, 47—52.

Грохне (Grohne D.). [1] Rechenverfahren zur Auflösung von Gleichungssystemen. *Veröffentlichungen Math. Inst. Tech. Hochschule Braunschweig*, 1946, № 4, i + 28p.

Гроссман Д. П. [1] К проблеме численного решения системы однородных линейных алгебраических уравнений. *Успехи матем. наук*, 1950, 5, № 3, 87—103; M. R., 13, 586.

Грушка (Hruska V.). [1] Lösung von Gleichungssystemen durch das Iterationsverfahren. *Acad. Tchéque. Sci. Bull. Cl. Sci. Math. Nat.*, 1943, 44, 239—304, 399—422.

Гуарне (Gouarné René). [1] Méthodes algébriques de la physique et de la chimie. *Résolution rapide des systèmes d'équations linéaires*. Thèse Doct. Sci. math. Ann. Univ. Paris, 1956, 26, № 4, 588—591; Р. Ж. М., 1958, 1599.

[2] Calcul automatique des déterminants. *C. r. Acad. sci.*, 1957, 245, № 8, 824—826; Р. Ж. М., 1959, 3246.

[3] Remarques sur le calcul automatique des déterminants et polynomes caractéristiques par la méthode des cycles. *C. r. Acad. sci.*, 1957, 245, № 23, 1998—2000; Р. Ж. М., 1959, 3247.

[4] Calcul automatique des polynomes caractéristiques. *C. r. Acad. sci.*, 1957, 245, № 14, 1114—1117.

Губерман (Губерман И. О.). [1] Спрощена схема розв'язання систем лінійних алгебраїчних рівнянь. *Пракл. механіка*, 1957, 3, № 1, 108—112; Р. Ж. М., 1957, 8268.

Гуди (Goodey W. J.). [1] Note on the improvement of approximate latent roots and modal columns of a symmetrical matrix. *Quart. J. Mech. and Appl. Math.*, 1955, 8, № 4, 452—453; Р. Ж. М., 1956, 6865.

Давыдов В. В. [1] Решение трехчленных уравнений, встречающихся в строительной механике корабля. Тр. Горьковск. ин-та инж. водн. трансп., 1957, 14, 10—24; Р. Ж. М., 1958, 3297.

Данилевский А. М. [1] О численном решении векторного уравнения. *Матем. сб.*, 1937, 2, [44], 169—171.

Данциг и Орчард-Хейс (Dantzig George B. and Orchard-Hays Wm.). [1] The product form for the inverse in the simplex method. *Math. Tables and Other Aids Comput.*, 1954, 8, № 46, 64—67; Р. Ж. М., 1956, 6149.

Даунинг и Хаусхолдер (Downing A. C. Jr. and Hausholder A. S.). [1] Some inverse characteristic value problems. *J. Assoc. Comput. Machinery*, 1956, 3, № 3, 203—207; Р. Ж. М., 1957, 8963.

Деминг (Deming H. O.). [1] A systematic method for the solution of simultaneous linear equations. *Amer. Math. Monthly*, 1928, 35, 360—363.

Дени-Папен и Кауфман (Denis-Papin M. et Kaufmann A.). [1] *Cours de calcul matriciel appliqu *. Electro-techn., hydraul., radio, 1958, 237, Suppl., 150—157; Р. Ж. М., 1959, 5555.

Дервидюэ (Derwidue L.). [1] La m thode de L. Couffignal pour la r solution num rique des syst mes d'equations lin aires. *Mathesis*, 1954, 63, № 1—2, 9—12; Р. Ж. М., 1955, 4712.

[2] Une m thode m canique de calcul des vecteurs propres d'une matrice quelconque. *Bull. Soc. roy. sci., Li ge*, 1955, 24, № 5, 150—171; Р. Ж. М., 1956, 9115.

Джери (Geri Gero). [1] Il simbolismo delle matrici nella soluzione del sistema normale gaussiano. *Riv. catastro e serv. tecn. erariali*, 1954, 9, № 2, 117—120; Р. Ж. М., 1955, 6111.

Диаш Агууду (Dias Agudo Rold o Fernando). [1] On the characteristic equation of a matrix. *Univ. Lisboa Rev. Fac. Cl. A. Cl. Mat.*, 1954, (2) 3, 87—136; Р. Ж. М., 1957, 6854.

Димсдэль (Dimsdale B.). [1] The non-convergence of a characteristic root method. *J. Soc. Industr. and Appl. Math.*, 1958, 6, № 1, 23—25; Р. Ж. М., 1959, 5227.

Дмитриев Н. и Диинкин Е. [1] Характеристические корни стохастических матриц. *Изв. АН СССР*, сер. матем., 1946, 10, № 2, 167—184; М. Р., 8, 129.

Дойл (Doyle Thomas C.). [1] Inversion of symmetric coefficient matrix of positive-definite quadratic form. *Math. Tables and Other Aids Comput.*, 1957, 11, № 58, 55—58; Р. Ж. М., 1958, 6214.

Дуайр (Dwyer P. S.). [1] The solution of simultaneous equations. *Psychometrika*, 1941, 6, 101—129; М. Р., 2, 367.

[2] The evaluation of the determinants. *Psychometrika*, 1941, 6, 191—204; М. Р., 2, 367.

[3] The evaluation of linear forms. *Psychometrika*, 1941, 6, 355—365; М. Р., 3, 154.

[4] The Doolittle technique. *Ann. Math. Statistics*, 1941, 12, 449—458; М. Р., 3, 276.

[5] Recent developments in correlation technique. *J. Amer. Statist. Assoc.*, 1942, 37, 441—460; М. Р., 4, 164.

[6] A matrix presentation of least squares and correlation theory with matrix justification of improved methods of solution. *Ann. Math. Statistics*, 1944, 15, 82—89; М. Р., 5, 245.

[7] The square root method and its use in correlation and regression. *J. Amer. Statist. Assoc.*, 1945, 40, 493—503; М. Р., 7, 338.

[8] *Linear Computations*, 1951, pp. 344, New York, Wiley; Р. Ж. М., 1954, 1829.

[9] Errors of matrix computations. *Nat. Bur. Standards. Appl. Math. Ser.*, 1953, 29, 49—58; Р. Ж. М., 1956, 6868.

Дуайр и Уо (Dwyer Paul S. and Waugh Frederick V.). [1] On errors in matrix inversion. *J. Amer. Statist. Assoc.*, 1953, 48, № 262, 289—319; Р. Ж. М., 1954, 5785.

Дункан (Duncan W. J.). [1] Some devices for the solution of large sets of simultaneous linear equations. *Philos. Mag.*, 1944, (7) 35, 660—670; М. Р., 7, 84.

Дюранд (Durand David). [1] A note on matrix inversion by the square root method. *J. Amer. Statist. Assoc.*, 1956, 51, № 274, 288—292; Р. Ж. М., 1958, 6215.

Дюло (Duleau Jacques). [1] Résolution numérique de certains systèmes d'équations linéaires vectorielles. *C. r. Acad. sci.*, 1956, 242, № 7, 870—873; Р. Ж. М., 1956, 9130.

Енне (Jenne W.). [1] Zur Auflösung linearer Gleichungssysteme. *Astr. Nachr.*, 1949, 278, 79—95.

Енсен (Jensen H.). [1] An attempt at a systematic classification of some methods for the solution of normal equations. *Geod. Inst. Medd.*, 1944, № 18, 45 pp; M. R., 7, 488.

Ершов А. П. [1] Об одном методе обращения матриц. *Докл. АН СССР*, 1955, 100, № 2, 209—211; Р. Ж. М., 1955, 4242.

Жанен (Janin R.). [1] Résolution de systèmes d'équations algébriques linéaires d'ordre élevé, à l'aide des méthodes mécanographiques (Emploi du calculateur électronique). *Rech. aéronaut.*, 1955, № 44, 47—50; Р. Ж. М., 1956, 4914.

Живоглядов В. Г. [1] О некоторых численных методах решения системы линейных алгебраических уравнений, их приспособлении для других вычислений и об ошибках округления в этих процессах. Автографат дисс. канд. физ. матем. н. Казанский уч-т, Казань, 1955; Р. Ж. М., 1956, 1689.

Задунайский (Zadunaisky Pedro E.). [1] Un metodo de iteracion para la resolucion de sistemas de ecuaciones lineales algebraicas. *Rev. Union mat. argent. y Asoc. fis. argent.*, 1955, 17, 335—343; Р. Ж. М., 1957, 8265.

Зассенфельд (Sassenfeld H.). [1] Ein hinreichendes Konvergenzkriterium und eine Fehlerabschätzung für die Iteration in Einzelschritten bei linearen Gleichungen. *Z. angew. Math. und Mech.*, 1951, 31, 92—94; M. R., 14, 692.

Зейдель (Seidel L.). [1] Über ein Verfahren, die Gleichungen, auf welche die Methode der kleinsten Quadrate führt, sowie lineare Gleichungen überhaupt, durch successive Annäherung aufzulösen. *Abh. math.-phys. Kl. Bayrische Akad. Wiss., München*, 1874, 11, № 3, 81—108.

Зинден (Sinden Frank W.). [1] An oscillation theorem for algebraic eigenvalue problems and its applications. *Mitt. Inst. angew. Math.*, Zürich, 1954, 4, 57 pp; M. R., 16, 666.

Зыльев В. П. [1] Признаки сходимости и оценки погрешностей решений системы линейных алгебраических уравнений способом итераций в матричном изложении. М.—Л., Инжен. строит. ин-т им. Куйбышева. *Сб. трудов*, 1939, 2, 232—245.

Иванов В. К. [1] О сходимости процессов итерации при решении систем линейных алгебраических уравнений. *Изв. АН СССР*, сер. матем., 1939, № 4, 477—483; M. R., 2, 118.

Идельсон Н. И. [1] Вычисление весов неизвестных в методе наименьших квадратов. *Астр. журнал*, 1943, 20, 11—13; M. R., 6, 51.

Йосса (Jossa F.). [1] Risoluzione progressiva di un sistema di equazioni lineari. Analogia con un problema meccanico. *Rend. Accad. Sci. fis. e mat.*, Napoli, Ser. 4, 1940, 10, 346—352; M. R., 8, 535.

Исхак (Ishaq M.). [1] Sur les spectres des matrices. *Sémin. P. Dubreil et Ch. Pisot. Fac. sci. Paris*, 1955—1956, 9, № 14, 1—14; Р. Ж. М., 1958, 2730.

Ито (Ito M.). [1] A geometrical study of the characteristic equation. *Tôhoku Math. J.*, 1932, 35, 294—303.

Кайрони (Caironi Mario). [1] Osservazioni sui procedimenti di approssimazioni successive nel metodo delle forze. *Ingegnerie*, 1957, 31, № 5, 410—420; Р. Ж. М., 1958, 2436.

Калиновская (Каліновська С. С.). [1] Оцінка швидкості збіжності деяких ітераційних процесів. *Наук. зап. Луцьк. держ. пед. ін-ту*, 1955, 3, № 2, 11—17; Р. Ж. М., 1956, 7649.

Камела (Kamela C.). [1] Die Lösung der Normalgleichungen nach der Methode von Prof. Dr. T. Banachiewicz (sogenannte „Krakowianenmethode“). *Schweiz. Z. Vermessungswesen Kulturtech.*, 1943, 41, 225—232, 265—275; M. R., 7, 488.

Кан (Khan N. A.). [1] A theorem on the characteristic roots of matrices. *J. Univ. Bombay, sect. A.*, 1955, 38, 13—18.

Каторович Л. В. [1] Об одном эффективном методе решения экстремальных задач для квадратичных функционалов. *Докл. АН СССР*, 1945, 48, № 7, 455—460; M. R., 8, 30.

[2] О методе наискорейшего спуска. *Докл. АН СССР*, 1947, 56, № 3, 233—236; M. R., 1, 308.

[3] Функциональный анализ и прикладная математика. *Успехи матем. наук*, 1948, 3, № 6, 89—185; M. R., 10, 380.

[4] Метод Ньютона. *Тр. Матем. ин-та АН СССР*, 1949, 28, 104—144; M. R., 12, 419.

Карпюс (Karpus R.). [1] L'algorithme de Gauss modernisé et son application à des systèmes d'équations linéaires dégénérés ou mal ordonnés. *Note techn. O. N. E. R. A.*, 1953, № 11, 133 p. ill.; Р. Ж. М., 1959, 5215.

Капролли (Caprioli Luigi). [1] Sulla risoluzione dei sistemi di equazioni lineari con il metodo di Cimmino. *Boll. Unione mat.*, 1953, 8, № 3, 260—265; Р. Ж. М., 1955, 6110.

Карпелевич Ф. И. [1] Характеристические корни матриц с неотрицательными коэффициентами. *Успехи матем. наук*, 1949, 4, № 5(33), 177—178; M. R., 11, 154.

[2] О характеристических корнях матриц с неотрицательными элементами. *Изв. АН СССР, сер. матем.*, 1951, 15, 361—383; M. R., 13, 201.

Карр (Carr John W.). [1] Error analysis in floating point arithmetic. *Comm. Assoc. Comput. Machinery*, 1959, 2, № 5, 10—15.

Карри (Currie J. C.). [1] Cassini ovals associated with a second order matrix. *Amer. Math. Monthly*, 1948, 55, 487—489; M. R., 10, 177.

Карри (Curry H. B.). [1] The method of steepest descent for non linear minimization problems. *Quart. Appl. Math.*, 1944, 2, 258—261; M. R., 6, 52.

Каруш (Karush W.). [1] An iterative method for finding characteristic vectors of a symmetric matrix. *Pacif. J. Math.*, 1951, 1, № 2, 233—247; M. R., 13, 388.

[2] Convergence of a method of solving linear problems. *Proc. Amer. Math. Soc.*, 1952, 3, 839—851; M. R., 14, 1127.

Касадзима (Kasajima Tomomi). [1] A note on a theorem of Rutishauser. *Comment. math. Univ. St. Pauli*, 1957, 6, № 1, 89—91; Р. Ж. М., 1959, 7488.

Кассина (Cassina U.). [1] Sul numero delle operazioni elementari necessarie per risoluzione dei sistemi di equazioni lineari. *Boll. Unione mat. Ital.*, ser. III, 1948, 3, 142—147; M. R., 10, 405.

Кассинис (Cassinis Gino). [1] I metodi di H. Boltz per risoluzione dei sistemi di equazioni lineari e il loro impiego nelle compensazioni della triangolazione. *Riv. catastro e ser. tecn. eraralti*, 1944, 1.

[2] Risoluzione dei sistemi di equazioni algebriche lineari. *Rend. Sem. mat. e fis.*, *Milano*, 1946, 17, 62—78; M. R., 9, 622.

Като (Kato T.). [1] On the upper and lower bounds of eigenvalues. *J. Phys. Soc. Japan*, 1949, 4, 334—339; M. R., 12, 447.

Кауден (Cowden D. J.). [1] Correlation concepts and the Doolittle method. *J. Amer. Statist. Assoc.*, 1943, 38, 327—334; M. R., 5, 42.

Качмаж (Kaczmarz S.). [1] Angenäherte Auflösung von Systemen linearer Gleichungen. *Bull. intern. Acad. Polon. Sci. A*, 1937, 355—357.

Кашанин (Kašanin R.). [1] Interprétation géométrique du schéma de Banachiewicz. *Srpska Akad. Nauka. Zbornik Radowa*, 1952, 18, Mat. Inst., 2, 93—96; M. R., 14, 501.

Кваде (Quade W.). [1] Auflösung linearer Gleichungen durch Matrizeniteration. *Ber. Math.-Tagung Tübingen*, 1946, 1947, 123—124; M. R., 9, 104.

Келли (Kelly J.). [1] Matrix multiplication on the IBM Card—Programmed Electronic Calculator. *Proc. Comput. Sem. Dec. 1949, N. Y. IBM Corp.*, 1951, 47—48; M. R., 13, 387.

Кенуи (Quenouille M. H.). [1] A further note on discriminatory analysis. *Ann. Eugenics*, 1949, 15, 11—14; M. R., 11, 743.

Керкхофс (Kerkhofs W.). [1] Résolution de systèmes d'équations simultanées à un grand nombre d'inconnues. *Ossature Métallique*, 1947, 12, 187—195; M. R., 10, 70.

Кертисс (Curtiss J. H.). [1] A generalization of the method of conjugate gradients for solving systems of linear algebraic equations. *Math. Tables and Other Aids Comput.*, 1954, 8, № 48, 189—193; P. Ж. М., 1956, 4105.

[2] A theoretical comparison of the efficiencies of two classical methods and a Monte Carlo method for computing one component of the solution of a set of linear algebraic equations. *Symp. Monte Carlo Methods*, New York, John Wiley and Sons Inc., 1956, 191—233; P. Ж. М., 1958, 8272.

Керубино (Cherubino Salvatore). [1] *Calcolo delle matrici*. Montr. mat. Consiglio naz. ricerche, 4, Roma. Ed. Cremonese, 1957, VII, 322 p.; P. Ж. М., 1958, 9554.

Кикукава (Kikukawa M.). [1] A numerical method for multiplication, reciprocation and division of matrices and for the solution of simultaneous linear algebraic equations. *Tech. Rep. Osaka Univ.*, 1952, 2, 11—30; M. R., 14, 501.

Кимбэлл (Kimbball Bradford F.). [1] Note on computation of orthogonal predictors. *Ann. Math. Statistics*, 1953, 24, 299—303; M. R., 14, 1019.

Кинкейд (Kincaid W. M.). [1] Numerical methods for finding characteristic roots and vectors of matrices. *Quart. Appl. Math.*, 1947, 5, 320—345; M. R., 9, 210.

Кио (Chio F.). [1] *Mémoire sur les Fonctions Connues sous le Nom de Résultants ou de Déterminants*, Turin, 1853.

Клазен (Clasen B. I.). [1] Sur une nouvelle méthode de résolution des équations linéaires et sur l'application de cette méthode au calcul des déterminants. *Ann. Soc. scient., Bruxelles*, 1888, 12, A 50—59, B 251—281.

Клер (Clerc D.). [1] Sur le calcul par itération des modes propres d'ordre supérieur. I. part. *Rech. aéronaut.*, 1956, № 54, 39—48; P. Ж. М., 1957, 7400.

Кнёдель (Knödel W.). [1] Lineare Gleichungen. Beispiele aus den Anwendungen, Lösungsmethoden, Lochkartenverfahren. *MTW-Mitt.*, 1956, 138—140; P. Ж. М., 1957, 1838.

Когбетлянц (Cogbetliantz Ervand George). [1] Diagonalization of general complex matrices as a new method for solution of linear equations. *Proc. Internat. Congr. Math.*, 1954, 2, Amsterdam, 1954, 356—357; P. Ж. М., 1955, 6109.

[2] Solution of linear equations by diagonalization of coefficients matrix. *Quart. Appl. Math.*, 1955, 13, № 1, 123—132; P. Ж. М., 1958, 1595.

Кози (Causey Robert L.). [1] Computing eigenvalues of non-hermitian matrices by methods of Jacobi type. *J. Soc. Industr. and Appl. Math.*, 1958, 6, № 2, 172—181.

[2] On some error bounds of Givens. *J. Assoc. Comput. Machinery*, 1958, 5, № 2, 127—131; P. Ж. М., 1959, 5228.

Коллар (Collar A. R.). [1] On the reciprocation of certain matrices. *Proc. Roy-Soc. Edinburgh*, 1939, 59, 95—206.

[2] Some notes on Jahn's method for the improvement of approximate latent roots and vectors of a square matrix. *Quart. J. Mech. and Appl. Math.*, 1948, 1, 145—148; M. R., 10, 152.

Коллац (Collatz L.). [1] Fehlerabschätzung für das Iterationsverfahren zur Auflösung linearer Gleichungssysteme. *Z. angew. Math. und Mech.*, 1942, 22, 357—361; M. R., 5, 50.

[2] Einschließungssatz für die charakteristischen Zahlen von Matrizen. *Math. Z.*, 1942, 48, 221—226; M. R., 5, 30.

[3] Eigenwertaufgaben mit technischen Anwendungen. Mathematik und ihre Anwendungen in Physik und Technik, Reihe A, 1949, 19, Akademische Verlagsgesellschaft, Leipzig, XVII + 466 pp; M. R., 11, 137.

[4] Über die Konvergenzkriterien bei Iterationsverfahren für lineare Gleichungssysteme. *Math. Z.*, 1950, 53, 149—161; M. R., 12, 361.

[5] Zur Herleitung von Konvergenzkriterien für Iterationsverfahren bei linearen Gleichungssystemen. *Z. angew. Math. und Mech.*, 1950, 30, 278—280.

[6] Zur Fehlerabschätzung bei linearen Gleichungssystemen. *Z. angew. Math. und Mech.*, 1954, 34, № 1—2, 71—72; P. Ж. М., 1955, 439.

Кон (Kohn W.). [1] A variational iteration method for solving secular equations. *Chem. Phys.*, 1949, 17, 670; M. R., 11, 136.

Конференция. Conference on matrix computations (Wayne State Univ. sept. 3rd—6th, 1957). *J. Assoc. Comput. Machinery*, 1957, 4, № 4, 520—523.

Коско (Kosko E.). [1] Reciprocal of triply partitioned matrices. *J. Roy. Aeronaut. Soc.*, 1956, 60, 490—491; M. R., 18, 418.

[2] Matrix inversion by partitioning. *Aeronaut. Quart.*, 1957, 8, № 2, 157—184; P. Ж. М., 1958, 751.

Костарчук В. Н. [1] Об одном методе решения систем линейных уравнений и отыскания собственных векторов матрицы. *Докл. АН СССР*, 1954, 98, № 4, 531—534; P. Ж. М., 1955, 4709.

[2] Застосування методу мінімальних нев'язок до знаходження власних чисел матриці. *Наук. зап. Житомирськ. держ. пед. ін-ту, сер. фіз.-мат.*, 1957, 3, 63—76; P. Ж. М., 1958, 1598.

Костарчук В. Н. и Пугачев Б. П. [1] Точная оценка уменьшения погрешности на одном шаге метода наискорейшего спуска. *Пр. семин. по функциональному анализу. Воронежск. ун-т*, 1956, 2, 25—30.

Кострикин Борр (Kostrikin Borro). [1] Resolução de sistemas de equações lineares pelo método de Cross. *Técnica*, 1953, 28, № 230, 365—370; P. Ж. М., 1953, 1407.

Котелянский Д. М. [1] О расположении точек матричного спектра. *Украин. матем. журн.*, 1955, 7, № 2, 131—133; M. R., 17, 228.

[2] О некоторых достаточных признаках вещественности и простоты матричного спектра. *Матем. сб.*, 1955, 36 (78), 163—168; M. R., 16, 894.

[3] О влиянии преобразования Гаусса на спектры матриц. *Успехи матем. наук*, 1955, 10, № 1, 117—121; P. Ж. М., 1956, 1055.

Коши (Cauchy A. L.). [1] Méthode générale pour la résolution des systèmes d'équations simultanées. *C. r. Acad. sci.*, 1847, 25, 536—538.

Коюар (Coüard A.). [1] Résolution d'un système d'équations linéaires par approximations successives. *Génie civil*, 1953, 130, № 6, 114; P. Ж. М., 1953, 454.

Крандалл (Crandall S. H.). [1] Engineering analysis; A survey of numerical procedures. McGraw-Hill Book Co., Inc. New York — Toronto — London, 1956; M. R., 18, 674—675.

Крапури Свейгард (Krapur T. and Svejgaard B.). [1] A method for matrix multiplication, matrix inversion, and problems of adjustment by punched card equipment. *Geodet. Inst. Kobenhavn. Medd.*, 1956, 31, 31; M. R., 18, 337.

Красносельский М. А. [1] О некоторых приемах приближенного вычисления собственных значений и собственных векторов положительно-определенной матрицы. *Успехи матем. наук*, 1956, 11, № 3, 151—158; P. Ж. М., 1957, 1841.

Красносельский М. А. и Крейн С. Г. [1] Итеративный процесс с минимальными невязками. *Матем. сб.*, 1952, 31(73), 315—334; M. R., 14, 692.

[2] Замечание о распределении ошибок при решении системы линейных уравнений при помощи итерационного процесса. *Успехи матем. наук*, 1952, 7, № 4, 157—161; M. R., 14, 501.

Краут (Crout P. D.). [1] A short method for evaluating determinants and solving systems of linear equations with real or complex coefficients. *Trans. Amer. Inst. Elec. Engng*, 1941, 60, 1235—1240.

Крейг (Craig Edward J.). [1] The N-step iteration procedures. *J. Math. and Phys.*, 1955, 34, № 1, 64—73; Р. Ж. М., 1956, 6866.

Крейзиг (Kreyszig E.). [1] Die Einschliessung von Eigenwerten hermitescher Matrizen beim Iterationsverfahren. *Z. angew. Math. and Mech.*, 1954, 34, № 12, 459—469; Р. Ж. М., 1956, 4919.

[2] Die Ausnutzung zusätzlicher Vorkenntnisse für die Einschliessung von Eigenwerten beim Iterationsverfahren. *Z. angew. Math. and Mech.*, 1955, 35, № 3, 89—95; Р. Ж. М., 1956, 4920.

[3] Einschliessung von Eigenwerten und Mohrsches Spannungs diagramm. *Z. angew. Math. und Phys.*, 1958, 9a, № 2, 202—206; Р. Ж. М., 1959, 7483.

Крокетт и Чернов (Crockett Jean Bronfenbrenner and Chernoff Herman). [1] Gradient methods of maximization. *Pacif. J. Math.*, 1955, 5, № 1, 33—50; Р. Ж. М., 1956, 4925.

Крон (Kron Gabriel). [1] Detailed example of interconnecting piece-wise solutions. *J. Franklin Inst.*, 1955, 259, № 4, 307—333; Р. Ж. М., 1956, 6872.

[2] Inverting a 256×256 matrix. *Engineering*, 1955, 179, № 4 650, 309—312; Р. Ж. М., 1956, 4915.

[3] Improved procedure for interconnecting piece-wise solutions. *J. Franklin Inst.*, 1956, 262, 385—392; M. R., 19, 64.

[4] Factorized inverse of partitioned matrices. *Matrix Tensor Quart.*, 1957, 8, 39—41; M. R., 19, 1198.

Крылов А. Н. [1] О численном решении уравнения, которым в технических вопросах определяются частоты малых колебаний материальных систем. *ИАН ОМЕН*, 1931, 4, 491—539.

Крылов Н. М. и Богоявленов Н. Н. [1] Новые методы для решения некоторых математических проблем, встречающихся в технике. *Укр. научн. ин-т сооружений*, 1933, 78, 78—95.

Кублановская В. Н. [1] Применение аналитического продолжения в численных методах анализа. Автореферат дисс. канд. физ. матем. н., ЛГУ, Л., 1955; Р. Ж. М., 1956, 1688.

[2] Применение аналитического продолжения методом замены переменных в численном анализе. *Tr. Матем. ин-та АН СССР*, 1959, 53, 145—185.

Кунц (Kunk K. S.). [1] Matrix methods. *Proc. Comput. Sem. Dec.*, 1949, № 9, IBM Corp., 1951, 37—42; M. R., 13, 496.

Купер (Cooper J. L. B.). [1] The solution of natural frequency equations by relaxation methods. *Quart. Appl. Math.*, 1948, 6, 179—183; M. R., 10, 70.

Куффиньяль (Couffignal Louis). [1] Recherches de mathématiques utilisables. La résolution numérique des systèmes d'équations linéaires. I. L'opération fondamentale de réduction d'un tableau. *Revue sci.*, 1944, 82, 67—78; M. R., 8, 128.

[2] Sur la précision des solutions approchées d'un système d'équations linéaires. *C. r. Acad. sci.*, 1948, 227, 30—32; M. R., 10, 212.

[3] Sur la résolution numérique des systèmes d'équations linéaires. II. *Rev. sci.*, 1951, 89, 3—10; M. R., 13, 284.

[4] Méthodes pratiques de réalisation des calculs matriciels. *Rend. mat. e appl.*, 1954, 14, № 1—2, 85—97; Р. Ж. М., 1956, 6867.

Лаврентьев М. М. [1] К вопросу об улучшении точности решения системы линейных уравнений. *Докл. АН СССР*, 1953, № 5, 885—886; Р. Ж. М., 1954, 3839.

[2] О точности решения линейных уравнений. *Матем. сб.*, 1954, 34 (76) № 2, 259—268; Р. Ж. М., 1954, 5784.

[3] Об оценке точности решения систем линейных уравнений. *Докл. АН СССР*, 1954, 95, № 3, 447—448; Р. Ж. М., 1955, 4714.

Лавут А. В. [1] Расположение собственных чисел преобразований Зейделя для систем нормальных уравнений. *Успехи матем. наук*, 1952, 7, № 6, 197—202; М. Р., 14, 1128.

Ладерман (Laderman J.). [1] The square root method for solving simultaneous linear equations. *Math. Tables and Other Aids Comput.*, 1948, 3, 13—16; М. Р., 9, 622.

Ламин (Lamine T.). [1] Diagonalisation des matrices. *Bull. Assoc. Ingrs issus École appliq., artill et génie*, 1958, 36, № 1, 49—74; Р. Ж. М., 1959, 1252.

Лангефорд (Langefors B.). [1] Approximate solution of simultaneous equations by means of transformation of variables. Applications to aeronautical problems. *SAAB TN*, 1953, № 7, 26 pp; Р. Ж. М., 1959, 3241.

[2] Ill-conditioned matrices. *SAAB TN*, 1953, № 22; Р. Ж. М., 1956, 8349.

[3] On the practical solution of linear equations. *SAAB TN*, 1955, № 35, 24 pp; Р. Ж. М., 1959, 9505.

Ланцош (Lanczos Cornelius). [1] A simple recursion method for solving a set of linear equations. *Bull. Amer. Math. Soc.*, 1936, 42, 325; М. Р., 1, 97.

[2] An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Standards*, 1950, 45, № 4, 255—288; М. Р., 13, 163.

[3] An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Proc. Second. Symp. Large-Scale Digital Calcul. Mach.*, (1949), 1951, 164—206; М. Р., 13, 589.

[4] An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Proc. Symp. Spectral Theory and Differential Problems. Oklahoma Agricultural Mech. College*, 1951, 301—316; М. Р., 13, 497.

[5] Solution of systems of linear equations by minimized iterations. *J. Res. Nat. Bur. Standards*, 1952, 49, № 1, 33—53; М. Р., 14, 501.

[6] Chebyshev polynomials in the solution of large-scale linear systems. *Proc. Assoc. Comput. Mach., Meeting at Toronto Ont.*, 1952, Sept. 1953, 124—133; Р. Ж. М., 1954, 5262.

[7] Spectroscopic eigenvalue analysis. *J. Wash. Acad. Sci.*, 1955, 45, № 10, 315—323; Р. Ж. М., 1956, 9114.

[8] Applied analysis. Prentice Hall, Inc., Englewood Cliffs, N. Y., 1956, XX + 539 pp.; Р. Ж. М., 1958, 4265.

[9] Iterative solution of large-scale linear systems. *J. Soc. Indust. and Appl. Math.*, 1958, 6, 1, 91—109; М. Р., 20, 70.

Леверье (Leverrier U. J. J.). [1] Sur les variations séculaires des éléments des orbites. *J. Math.*, 1840.

[2] Recherches Astronomiques. *Ann. l'Obser., Paris*, 1856, 11, 128.

Легра (Legras Jean). [1] La méthode de relaxation. *Age nucléaire*, 1957, № 6, 65—66; Р. Ж. М., 1958, 6208.

Ледерман (Ledermann Walter). [1] Bounds for the greatest latent-roots of a positive matrix. *J. London Math. Soc.*, 1950, 25, 265—268; М. Р., 12, 312.

Лейдерман Ю. Р. [1] Об одном способе решения системы линейных алгебраических уравнений, когда обычные методы последовательного приближения неприменимы. *Докл. АН Узб. ССР*, 1953, № 1, 8—11; Р. Ж. М., 1953, 902.

[2] К вопросу о решении системы линейных алгебраических уравнений методом последовательных приближений. *Докл. АН Узб. ССР*, 1953, № 10, 6—9; Р. Ж. М., 1954, 4919.

[3] Об одном методе решения совместных линейных алгебраических уравнений. *Тр. Ин-та матем. и механ. АН Узб. ССР*, 1954, в. 13, 153—158; Р. Ж. М., 1957, 6711.

Ленти [Lenti Raum o]. [1] Eine Methode von sukzessiven Projektionen zur Lösung der linearen algebraischen Vektorgleichung und ihre Anwendungen für Inversion von Matrizen. *Soc. Sci. Fennica. Jr. Phys.-Math.*, 1958, XXI s.

Лепперт (Leppert E. L.). [1] A fraction series solution for characteristic values useful in some problems of airplane dynamics. *J. Aeronaut. Sci.*, 1955, 22, № 5, 326—328; Р. Ж. М., 1956, 5510.

Ливенс (Leavens D. H.). [1] Accuracy in the Doolittle solution. *Econometrica*, 1947, 15, 45—50; М. Р., 8, 407.

Лидский В. Б. [1] О собственных значениях суммы и произведения симметричных матриц. *Докл. АН СССР*, 1950, 75, 769—772.

Линецкий В. Д. [1] Решение системы трехчленных уравнений с помощью фокусных отношений. *Научн. тр. Ленингр. инж.-строит. ин-та*, 1954, 17, 185—190; Р. Ж. М., 1956, 765.

Литвинов (Літвінов М. В.). [1] Обчислення коефіцієнтів впливу для складених сіткових областей за допомогою матричних перетворень. *Доповіді АН УРСР*, 1955, № 3, 222—226; Р. Ж. М., 1956, 6871.

Ловарш-Надь (Lowass-Nagy Viktor). [1] *Mátrixszámítás* Müszaki matematikai gyakorlatok. С. IV. Budapest, Tankönyvkiadó, 1956; Р. Ж. М., 1958, 6487.

Ломан (Lohman J. B.). [1] An iterative method for finding the smallest eigenvalue of a matrix. *Quart. Appl. Math.*, 1949, 7, 234; М. Р., 10, 743.

Лонсетт (Lonseth A. T.). [1] Systems of linear equations with coefficients subject to error. *Ann. Math. Statistics*, 1942, № 3, 332—337; М. Р., 4, 90.

[2] On relative errors in systems of linear equations. *Ann. Math. Statistics*, 1944, 15, 323—325; М. Р., 6, 51.

[3] An extension of an algorithm of Hotelling. *Berkeley Symp. Math. Stat. Prob.*, 1945, 1946, 353—357; М. Р., 10, 627.

[4] The propagation of error in linear problems. *Trans. Amer. Math. Soc.*, 1947, 62, 193—212; М. Р., 9, 192.

Лопшиц А. М. [1] Численный метод нахождения собственных значений и собственных плоскостей линейного оператора. *Тр. сем. по векторн. и тензорн. анализу*, 1949, 7, 233—259; М. Р., 13, 991.

[2] Экстремальная теорема для гиперэллипсоида и ее применение к решению системы линейных алгебраических уравнений. *Тр. сем. по векторн. и тензорн. анализу*, 1952, 9, 183—197; М. Р., 14, 1127.

Лоткин (Lotkin Mark). [1] A set of test matrices. *Math. Tables and Other Aids Comput.*, 1955, 9, № 52, 153—161; Р. Ж. М., 1957, 2685.

[2] Characteristic values of arbitrary matrices. *Quart. Appl. Math.*, 1956, 14, № 3, 267—275; Р. Ж. М., 1957, 8964.

[3] The diagonalization of skew-Hermitian matrices. *Duke Math. J.*, 1957, 24, 9—14; М. Р., 19, 685.

Лоткин и Ремидж (Lotkin Mark and Remage Russell). [1] Scaling and error analysis for matrix inversion by partitioning. *Ann. Math. Statistics*, 1953, 24, 428—439; Р. Ж. М., 1954, 5265.

[2] Matrix inversion by partitioning. *Proc. Assoc. Comput. Mach., Meeting at Toronto, Ont.*, 1952, Sept. 1953, 36—41; Р. Ж. М., 1954, 5264.

Лоу (Lowe J.). [1] Solution of simultaneous linear algebraic equations using the IBM Type 604 Electronic Calculating Punch. *Proc. Comput. Sem., Dec. 1949, N. Y. IBM Corp.*, 1951, 54—56; М. Р., 13, 388.

Люстерник Л. А. [1] Замечания к численному решению краевых задач уравнения Лапласа и вычислению собственных значений методом сеток. *Тр. Матем. ин-та АН СССР*, 1947, 20, 49–64; М. Р., 10, 71.

[2] О сходимости при случайных начальных данных и накоплении ошибок итерационного процесса решения системы алгебраических уравнений. *Вычисл. матем. и вычисл. техника*, 1953, № 1, 41–45; Р. Ж. М., 1955, 435.

[3] Решение задач линейной алгебры методом непрерывных дробей. *Тр. семин. по функцион. анализу. Воронежск. ун-т*, 1956, 2, 85–90; Р. Ж. М., 1957, 5971.

Лузин Н. Н. [1] О методе ак. А. Н. Крылова составления векового уравнения. *ИАН ОМЕН*, 1931, 903–958.

[2] О некоторых свойствах перемещающегося множителя в методе акад. А. Н. Крылова. I, II, III. *ИАН ОМЕН*, 1932, 596–638, 735–762, 1065–1102.

Лукашевич и Вармус (Lukaszewicz Józef i Warmus Mieczysław). [1] *Metody numeryczne i graficzne*. I. Państwowe Wydawn. Nauk., Warszawa, 1956, 429 pp.; М. Р., 18, 235.

Лэн Сень-мин (Leng Sen-ming). [1] The characteristic roots of a matrix. *Duke Math. J.*, 1952, 19, № 1, 139–154; М. Р., 14, 7.

Лю и Кван (Loo W. et Kwan Chao-Chih). [1] La méthode de col dans le problème de relaxation. *Acta Math. sinica*, 1955, 5, 497–504; М. Р., 17, 791.

Мадзарелла (Mazzarella F.). [1] Semplificazione alla soluzione di un sistema di equazioni lineari. *Rend. Accad. sci. fisi. e mat., Napoli, Ser. 4*, 1945, 13, 197–201; М. Р., 8, 287.

Мадич (Мадил Петаг). [1] L'étude de solubilité des systèmes des équations algébriques linéaires. *Bull. Inst. Nuclear Sci., Boris Kidritch*, 1952, 18, 13–15; М. Р., 17, 1008.

[2] Домен грешке у решењима система линеарних алгебарских једначина. *Весн. Друшт. матем. и физ. Н. Р. Србије*, 1956, 8, № 3–4, 191–194; Р. Ж. М., 1958, 1594.

[3] Sur une méthode de résolution des systèmes d'équations algébriques linéaires. *C. r. Acad. sci.*, 1956, 242, № 4, 439–441; Р. Ж. М., 1956, 9129.

Майер и Шмидтマイер (Mayer Daniel a Schmidt Mayer Josef). [1] Výjádření inversní matic konvergentní geometrickou řadou. *Aplikace mat.*, 1957, 2, № 1, 24–37; Р. Ж. М., 1958, 2438.

Мак-Миллан (Mac Millan R. H.). [1] A new method for the numerical evaluation of determinants. *J. Roy. Aeronaut. Soc.*, 1955, 59, № 539, 772–773; Р. Ж. М., 1956, 6151.

Маккензи (Mackenzie J. K.). [1] A least squares solution of linear equations with coefficients subject to a special type of error. *Austral. J. Phys.*, 1957, 10, 1, 103–109; Р. Ж. М., 1958, 7204.

Мальцев А. И. [1] Основы линейной алгебры. Изд. 2-е, Гостехиздат, 1956.

Манье (Magnier A.). [1] Sur le calcul numérique des matrices. *C. r. Acad. sci.*, 1948, 226, 464–465; М. Р., 9, 471.

Марков (Марков О. О.). [1] До питання про метод релаксації. *Наук. зап. Херсонськ. держ. пед. ін-т*, 1956, 7, 45–48; Р. Ж. М., 1958, 9265.

Маркус (Marcus M.). [1] An eigenvalue inequality for the product of normal matrices. *Amer. Math. Monthly*, 1956, 63, 3, 173–174; Р. Ж. М., 1957, 2063.

[2] On the optimum gradient method for systems of linear equations. *Proc. Amer. Math. Soc.*, 1956, 7, № 1, 77–81; Р. Ж. М., 1957, 2681.

Маслов П. Г. [1] К определению обратных матриц потенциальной энергии многоатомных молекул (Метод определителей). *Докл. АН СССР*, 1949, 67, 819–822.

[2]. К определению обратных матриц потенциальной энергии многоатомных молекул. *Л. Зап. Горн. ин-та*, 1949, 24, 185.

[3] К определению обратных матриц потенциальной энергии многоатомных молекул (Метод комбинированного наискорейшего спуска). *Докл. АН СССР*, 1950, 71, 867—870; *M. R.*, 12, 152.

[4] К методике определения колебаний многоатомных молекул. Метод комбинированного наискорейшего спуска. *ЖЭТФ АН СССР*, 1950, 20, 609—618; *M. R.*, 12, 640.

[5] К определению обратных матриц потенциальной энергии многоатомных молекул (Приближенные методы нахождения коэффициентов влияния). *Докл. АН СССР*, 1950, 70, 985.

[6] Метод решения системы однородных линейных уравнений при расчете колебаний полиатомических молекул. *ЖЭТФ АН СССР*, 1952, 22, 276—283; *M. R.*, 14, 322.

Масуяма (Masuyama M.). [1] On a numerical method of solution of the equation $|K - M| = 0$. *Rep. Statist. Appl. Res. Union Jap. Sci. Eng.*, 1951, 1, 26—28; *M. R.*, 13, 283.

Мачинский (Matschinski M.). [1] Über eine form der Lösung linearer Gleichungssysteme. *Portugaliae math.*, 1955, 14, № 3—4, 133—139; *P. Ж. М.*, 1958, 5187.

Маяниц Л. С. [1] Метод возмущений с применением двойной итерации. *Докл. АН СССР*, 1945, 48, № 5, 334—337; *M. R.*, 8, 54.

[2] Метод уточнения корней вековых уравнений высоких степеней и численного анализа их зависимости от параметров соответствующих матриц. *Докл. АН СССР*, 1945, 50, 121—124; *M. R.*, 14, 1129.

Медлин (Medlin Gene W.). [1] Bounds for the characteristic roots of matrices with real elements. *Duke Math. J.*, 1952, 19, 563—565; *M. R.*, 14, 836.

[2] On bounds for the greatest characteristic root of a matrix with positive elements. *Proc. Amer. Math. Soc.*, 1953, 4, 769—771; *P. Ж. М.*, 1955, 85.

[3] On limits of the real characteristic roots of matrices with real elements. *Proc. Amer. Math. Soc.*, 1956, 7, № 5, 912—917; *P. Ж. М.*, 1957, 6154.

Мейер и Холлингсуэрт (Meyer H. L. and Hollingsworth B. J.). [1] A method of inverting large matrices of special form. *Math. Tables and Other Aids Comput.*, 1957, 11, № 58, 94—97; *P. Ж. М.*, 1958, 6213.

Мейзлик (Mejzlík Ladislav). [1] Riešenie systémov lineárnych rovnic priamymi metódami. *Inžen. stavby*, 1954, 2, № 3, 110—116; *P. Ж. М.*, 1955, 438.

Мемке (Memke R.). [1] Praktische Lösung der Grundaufgaben über Determinanten, Matrizen und lineare Transformationen. *Math. Ann.*, 1892, 103, 303—318.

[2] К способу Зейделя, служащему для решения системы линейных уравнений с весьма большим числом неизвестных посредством последовательных приближений. *Матем. сб.*, 1892, 16, № 2, 342—345.

Мемке Р. и Некрасов П. А. [1] Решение линейной системы уравнений посредством последовательных приближений. *Матем. сб.*, 1892, 16, 437—459.

Мендельсон (Mendelsohn N. S.). [1] Some elementary properties of ill-conditioned matrices and linear equations. *Amer. Math. Monthly*, 1956, 63, № 5, 285—295; *P. Ж. М.*, 1957, 6850.

[2] Some properties of approximate inverses of matrices. *Trans. Roy. Soc. Canada*, sec. 3, 1956, 50, 53—59; *P. Ж. М.*, 1959, 5221.

[3] An iterative method for the solution of linear equations based on the power method for proper vectors. *Math. Tables and Other Aids Comput.*, 1957, 11, № 58, 88—91; *P. Ж. М.*, 1958, 2435.

[4] The computation of complex proper values and vectors of a real matrix with application to polynomials. *Math. Tables and Other Aids Comput.*, 1957, 11, № 58, 91—94; *P. Ж. М.*, 1958, 2440.

- Менли (Manley R. G.). [1] Roots of frequency equations. Nature and distribution of roots. *Aircraft Engng.*, 1944, 16, 203; M. R., 6, 74.
- Меррей (Murray F. J.). [1] Simultaneous linear equations. *Proc. Scient. Comput. Forum.*, 1948, IBM Corp., 1950, 105—106; M. R., 13, 496.
- Мизес и Гейрингер (Mises R. and Oeringer H.). [1] Praktische Verfahren der Gleichungsauflösung. *Z. angew. und Mech.*, 1929, 9, 58—77, 152—164.
- Микеладзе Ш. Е. [1] О разложении определителя, элементами которого служат полиномы. *Прикл. матем. и механика*, 1948, 18, 219—222; M. R., 9, 622.
- Милн (Milne W. E.). [1] *Numerical calculus*. Princeton 1949, London 1950; M. R., 10, 483.
- Миро (Miroux Jean). [1] Une nouvelle machine analogique à itération matricielle donnant les racines des équations algébriques. L'itération des matrices ayant des valeurs propres de modules voisins. *Ann. télécommun.*, 1956, II, 226—232; M. R., 19, 769.
- Мирский (Mirsky L.). [1] The norms of adjugate and inverse matrices. *Arch. Math.*, 1956, 7, № 4, 276—277; P. Ж. М., 1958, 5498.
- [2] Inequalities for normal and Hermitian matrices. *Duke Math. J.*, 1957, 24, № 4, 591—599; P. Ж. М., 1958, 8578.
- Митани. [1] Численное решение системы линейных алгебраических уравнений на вычислительной машине. *Bull. Electrotechn. Lab.*, 1955, 19, № 8, 576—581; P. Ж. М., 1957, 5959.
- Митра (Mitra S. K.). [1] On an orthogonalisation method of evaluating the reciprocal and the determinant of a matrix and its Gaussian transform. *Proc. Second Congr. Theor. and Appl. Mech., New Delhi*, Octob., 1956, 261—268; M. R., 19, 1080.
- Митчелл (Mitchell H. F. Sr.). [1] Inversion of a matrix of order 38. *Math. Tables and Other Aids Comput.*, 1948, 3, 161—166; M. R., 10, 152.
- Митчелл и Рутерфорд (Mitchell A. R. and Rutherford D. E.). [1] On the theory of relaxation. *Proc. Glasgow Math. Assoc.*, 1953, 1, 101—110; M. R., 15, 353.
- Михкович (Michkowitch V. V.). [1] Résolution des systèmes d'équations linéaires algébriques à l'aide des cracoviens. *Srpska Akad. Nauka. Zbornik Radova*, 1952, 18, *Mat. Inst.*, 2, 53—70; M. R., 14, 501.
- Моррис (Morris J.). [1] On a simple method for solving simultaneous linear equations by means of successive approximations. *J. Roy. Aeronaut. Soc.*, 1935, 39, 349.
- [2] A successive approximation process for solving simultaneous linear equations. *Aeronaut. Res. Comm.*, 1936, Rep. 1711, 1—12.
- [3] Frequency equations. *Aircraft Engng.*, 1942, 14, 108—110; M. R., 4, 90.
- [4] An escalator process for the solution of linear simultaneous equations. *Philos. Mag.*, 1946, (7) 37, 106—120; M. R., 8, 287.
- [5] The escalator process for the solution of damped Lagrangian frequency equations. *Philos. Mag.*, 1947, (7) 38, 275—287; M. R., 9, 210.
- [6] An application of the escalator process. Solution thereby of quasi-Hermitian frequency equations encountered in specific practical problems. *Aircraft Engng.*, 1951, 23, 136—137; M. R., 12, 862.
- Моррис и Хед (Morris J. and Head J. W.). [1] Lagrangian frequency equations. An "escalator" method for numerical solution. *Aircraft Engng.*, 1942, 14, 312—314, 316; M. R., 4, 148.
- [2] The "escalator" process for the solution of Lagrangian frequency equations. *Philos. Mag.*, 1944, (7) 35, 735—759; M. R., 7, 84.
- Моррисон (Morrison J. F.). [1] The solution of three-term simultaneous linear equation by the use of submatrices. *Engineering J.*, 1946, 29, 80—83; M. R., 8, 128.

Моултон (Moulton F. R.). [1] On the solutions of linear equations having small determinants. *Amer. Math. Monthly*, 1913, 20, 242—249.

Мункельт (Munkelt Karl). [1] Formeln zur maschinellen Berechnung von Kehrmatrizen. *Dtsch. hydrogr. Z.*, 1956, 9, № 3, 143—146; Р. Ж. М., 1959, 4248.

Мюлиг и Копперман (Mühlig F. and Koppermann E.). [1] Die Rechenvorschrift des „modernisierten“ Gaußschen Algorithmus in ihrer einfachsten Form. *Z. Vermessungswesen*, 1953, 78, № 12, 389—393; Р. Ж. М., 1955, 4713.

Мюллер (Müller E.). [1] Genauigkeit der bei Reduktion von Fehlergleichungen eliminierten Unbekannten. *Z. Vermessungswesen*, 1942, 71, 186—190; M. R., 5, 161.

Наглер (Nagler H.). [1] On the simultaneous numerical inversion of a matrix and all its leading submatrices. *Math. Tables and Other Aids Comput.*, 1956, 10, № 56; 225—226; Р. Ж. М., 1957, 8271.

Натансон И. П. [1] К теории приближенного решения уравнений. Уч. зап. Ленингр. гос. пед. ин-та им. А. И. Герцена, 1948, 64, 3—8.

Невилл (Neville E. H.). [1] Ill-conditioned sets of linear equations. *Philos. Mag.*, 1948, (7) 39, 35—48; M. R., 9, 382.

Нейман (von Neumann J.). [1] Some matrix inequalities and metrization of matrix-space. Изв. ин-та матем. и механики. Томский ун-т, 1937, 1, № 3, 286—299.

Нейман и Голдстайн (von Neumann J. and Goldstine H. H.). [1] Numerical inverting of matrices of high order. *Bull. Amer. Math. Soc.*, 1947, 53, 1021—1099; M. R., 9, 471.

Некрасов П. А. [1] Определение неизвестных по способу наименьших квадратов при весьма большом числе неизвестных. *Матем. сб.*, 1885, 12, 189—204.

[2] К вопросу о решении линейной системы уравнений с большим числом неизвестных посредством последовательных приближений. *Приложение к т. LXIX. Записки Ак. Наук*, 1892, № 5, 1—18.

Николаева М. В. [1] О методе релаксации. *Тр. Матем. ин-та АН СССР*, 1949, 28, 160—182; M. R., 12, 539.

Ньюмарк (Newmark N. M.). [1] Bounds and convergence of relaxation and iteration procedures. *Proc. First U. S. Nat. Congr. Appl. Mech. Chicago, 1951, N. Y. Amer. Soc. Mech. Eng.*, 1952; M. R., 15, 353.

Олт (Alt F. L.). [1] Machine methods for finding characteristic roots of a matrix. *Proc. Comput. Sem. Dec.*, 1949, N. Y. IBM. Corp., 1951, 49—53; M. R., 13, 496.

Ольденбургер (Oldenburger Rufus). [1] Infinite powers of matrices and characteristic roots. *Duke Math. J.*, 1940, 6, 357—361; M. R., 1, 324.

Орлов (Orloff Constantin). [1] Méthode spectrale pratique d'évaluation numérique des déterminants et de résolution du système d'équations linéaires. *Вестн. Друштва матем. и физ. Н. Р. Србије*, 1953, 5, № 1—2, 17—30; Р. Ж. М., 1956, 7651.

Осборн (Osborne Elmer E.). [1] On acceleration and matrix deflation processes used with the power method. *J. Soc. Indust. and Appl. Math.*, 1958, 6, № 3, 279—287.

Островский (Ostrowski A. M.). [1] Sur la variation de la matrice inverse d'une matrice donnée. *C. r. Acad. sci.*, 1950, 231, 1019—1021; M. R., 12, 396.

[2] Un nouveau théorème d'existence pour les systèmes d'équations. *C. r. Acad. sci.*, 1951, 232, 786—788; M. R., 12, 596.

[3] Sur les matrices peu différentes d'une matrice triangulaire. *C. r. Acad. sci.*, 1951, 233, 1558—1560; M. R., 13, 900.

[4] Sur les conditions générales pour la régularité des matrices. *Univ. Roma Ist. Naz. Alta. Mat. Rend. Mat. e Appl.*, 1951, (5) 10, 156—168; M. R., 14, 125.

[5] Ueber das Nichtverschwinden einer Klasse von Determinanten und die

Lokalisierung der charakteristischen Wurzeln von Matrizen. *Compositio Math.*, 1951, 9, 209—226; M. R., 13, 524.

[6] Bounds for the greatest latent root of a positive matrix. *J. London Math. Soc.*, 1952, 27, № 106, 253—256; M. R., 14, 126.

[7] On over and under relaxation in the theory of the cyclic single step iteration. *Math. Tables and Other Aids Comput.*, 1953, 7, 152—159; P. Ж. М., 1954, 5783.

[8] On the linear iteration procedures for symmetric matrices. *Rend. mat. e applic.*, 1954, 14, № 1—2, 140—163; P. Ж. М., 1956, 8346.

[9] On nearly triangular matrices. *J. Res. Nat. Bur. Standards*, 1954, 52, № 6, 319—345; P. Ж. М., 1955, 4232.

[10] Über Normen von Matrizen. *Math. Z.*, 1955, 63, № 1, 2—18; P. Ж. М., 1956, 6395.

[11] Über Verfahren von Steffensen und Householder zur Konvergenzverbesserung von Iterationen. *Z. angew. Math. und Phys.*, 1956, 7, № 3, 218—229; P. Ж. М., 1957, 911.

[12] Determinanten mit überwiegender Hauptdiagonale und die absolute Konvergenz von linearen Iterationsprozessen. *Comment. math. helv.*, 1956, 30, № 3, 175—210; P. Ж. М., 1957, 5960.

[13] On Gauss's speeding up device in the theory of single step iteration. *Math. Tables and Other Aids Comput.*, 1958, 18, № 62, 116—132.

Панов Д. Ю. [1] Решение систем линейных уравнений. Добавление к книге Д. Скарборо. «Численные методы математического анализа», М.—Л., 1934.

Панц (Рапс Vladimír). [1] Upravená relaxační metoda. *Aplikace mat.*, 1957, 2, № 3, 184—201; P. Ж. М., 1958, 7206.

Паркер (Parker W. V.). [1] The characteristic roots of a matrix. *Duke Math. J.*, 1937, 3, 484—487.

[2] Limits to the characteristic roots of a matrix. *Duke Math. J.*, 1943, 10, 479—482; M. R., 5, 30.

[3] The characteristic roots of matrices. *Duke Math. J.*, 1945, 12, 519—526; M. R., 7, 107.

[4] Characteristic roots and the field of values of a matrix. *Duke Math. J.*, 1948, 15, 439—442; M. R., 10, 4.

[5] Sets of complex numbers associated with a matrix. *Duke Math. J.*, 1948, 15, 711—715; M. R., 10, 230.

[6] Characteristic roots and field of values of a matrix. *Bull. Amer. Math. Soc.*, 1951, 57, 103—108; M. R., 12, 581.

Паркес (Parkes E. W.). [1] Linear simultaneous equations. Some practical aspects of their solution in respect to the time involved with a series and the relative accuracy of the results. *Aircraft Engrg.*, 1950, 22, 48, 56; M. R., 11, 618.

Пароди (Parodi Maurice). [1] Remarque sur la stabilité. *C. r. Acad. sci.*, 1949, 228, 51—52; M. R., 10, 501.

[2] Complément à un travail sur la stabilité. *C. r. Acad. sci.*, 1949, 228, 1198—1200; M. R., 10, 501.

[3] Sur les limites des modules des racines des équations algébriques. *Bull. Sci. Math.*, (2), 1949, 73, 135—144; M. R., 11, 307.

[4] Quelques propriétés des matrices H. *Ann. Soc. sci. Bruxelles.*, ser. I, 1950, 64, 22—25; M. R., 12, 234.

[5] Sur une limite supérieure du rapport des valeurs caractéristiques de deux matrices symétriques définies positives, à éléments réels, dont les éléments correspondants diffèrent peu. *C. r. Acad. sci.*, 1950, 230, 705—707; M. R., 11, 413.

[6] Sur des familles de matrices auxquelles est applicable une méthode d'itération. *C. r. Acad. sci.*, 1951, 232, 1053—1054; M. R., 12, 639.

[7] Sur un théorème de M. Ostrowski. *C. r. Acad. sci.*, 1952, 234, 282—284; M. R., 14, 126.

- [8] Sur quelques propriétés des valeurs caractéristiques des matrices carrées. *Mem. sci. Math.*, 1952, 118, 64 pp.; M. R., 14, 236.
- [9] Sur la localisation des valeurs caractéristiques des matrices dans le plan complexe. *C. r. Acad. sci.*, 1956, 242, 2617—2618; M. R., 18, 4.
- [10] Sur une méthode de localisation des valeurs caractéristiques de certaines matrices. *C. r. Acad. sci.*, 1957, 244, 1597—1598; M. R., 19, 379.
- Пельтье (Peltier Jean). [1] Détermination de vecteurs propres de certaines matrices à déterminant faible. *C. r. Acad. sci.*, 1955, 240, № 23, 2201—2203; Р. Ж. М., 1957, 901.
- [2] Mécanisation des problèmes linéaires sur machines électroniques. *C. r. Acad. sci.*, 1957, 244, № 8, 1003—1005; Р. Ж. М., 1959, 3243.
- Пенроуз (Penrose R.). [1] On best approximation solutions of linear matrix equations. *Proc. Cambridge Philos. Soc.*, 1956, 52, 17—19; M. R., 17, 536.
- Перес (Peres M.). [1] On solution of systems of simultaneous linear equations. *Las Ciencias*. Madrid, 1952, 17, 443—449; M. R., 17, 1137.
- Персэлл (Purcell Everett W.). [1] The vector method of solving simultaneous linear equations. *J. Math. and Phys.*, 1953, 32, 180—183; Р. Ж. М., 1954, 5787.
- Петри (Petrie George W.). [1] Matrix inversion and solution of simultaneous linear algebraic equations with the IBM 604 electronic calculating punch. *Nat. Bur. Standards. Appl. Math. Ser.*, 1953, 29, 107—112; Р. Ж. М., 1956, 3332.
- Планкетт (Plankett R.). [1] On the convergence of matrix iteration processes. *Quart. Appl. Math.*, 1950, 7, 419—421; M. R., 11, 464.
- Покорна (Pokorná O.). [1] Řešení soustav lineárních algebraických rovnic minimisací součtu čtverců residui. *Sbor. Českosl. akad. věd. Lab. mat. strojů*, 1954, № 2, 111—116; Р. Ж. М., 1956, 3329.
- [2] Řešení soustav lineárních algebraických rovnic průhled a srovnání metod. *Stroje zpracov. inform.*, 1955, 3, 139—196; Р. Ж. М., 1957, 6151.
- [3] Schema pro řešení soustav lineárních algebraických rovnic eliminaci. *Aplikace mat.*, 1957, 2, № 3, 235—241; Р. Ж. М., 1958, 6209.
- Поллачек-Гейрингер (Pollaczek-Geiringer H.). [1] Zur Praxis der Lösung linearer Gleichungen in der Statistik. *Z. angew. Math. und Mech.*, 1928, 8, 446—447.
- Поп и Томпкинс (Pop David A. and Tompkins C.). [1] Maximizing functions of rotations-experiments concerning speed of diagonalization of symmetric matrices using Jacobi's method. *J. Assoc. Comput. Machinery*, 1957, 4, № 4, 459—466; Р. Ж. М., 1959, 2013.
- Попович (Popovici Constatin C.). [1] Asupra metodei iterativiei, aplicată la un sistem de ecuații liniare. *Studii și cercetări mat.*, 1953, 4, № 1—2, 233—247; Р. Ж. М., 1955, 3987.
- Портер (Porter R. E.). [1] Single order reduction of a complex matrix. *Proc. Comput. Sem. Dec.*, 1949, N. Y., IBM Corp., 1951, 138—140; M. R., 13, 387.
- Потрон (Potron l'Abbé). [1] Sur les matrices non négatives, et les solutions positives de certains systèmes linéaires. *Bull. Soc. math. France*, 1939, 67, 56—61; M. R., 1, 97.
- Пугачев Б. П. [1] О двух приемах приближенного вычисления собственных значений и собственных векторов. *Докл. АН СССР*, 1956, 110, № 3, 334—337; Р. Ж. М., 1957, 5181.
- [2] Об одном методе приближенного отыскания собственных значений. *Тр. 3-го Всесоюзного матем. съезда*, 1956, 2, 153—154; Р. Ж. М., 1956, 9113.
- Райли (Riley James D.). [1] Solving systems of linear equations with a positive definite, symmetric, but possibly ill-conditioned matrix. *Math. Tables and Other Aids Comput.*, 1955, 9, № 51, 96—101; Р. Ж. М., 1957, 5958.

Раймонди (Raymondi C.). [1] Contributo allo studio dei sistemi elasticci staticamente indeterminanti. *Atti. Ist. Costruzioni Univ. Pisa*, 1949, 13, 1—15; M. R., 13, 587.

Рейс (Rice Leptine Hall). [1] The rank of a matrix, the value of a determinant, and the solution of a system of linear equations. *J. Math. and Phys.*, 1932, 11, 146—149.

Райт (Wright L. T., Jr.). [1] The solution of simultaneous linear equations by an approximation method. *Cornell Univ., Engrg. Exper. Station Bull.*, 1943, 31, 6 pp; M. R., 5, 110.

Растон (Rushton S.). [1] On least squares fitting by orthonormal polynomials using the Choleski method. *J. Roy. Statist. Soc., Ser. B*, 1951, 13, 92—99; M. R., 13, 990.

Ратерфорд (Rutherford D. E.). [1] On the rational solution of the matrix equation $sx = xt$. *Proc. Koninkl. nederl. akad. wetensch.*, 1933, 36, 432—442.

Рахман (Rahman A.). [1] Numerical evaluation of determinants. *Bull. cl. sci. Acad. roy. Belgique*, 1954, 40, № 8, 798—801; P. Ж. М., 1956, 764.

Реджини (Reggini Horacio C.). [1] Resolucion de sistemas de ecuaciones lineales. *Cienc. y tecn.*, 1952, 125, № 628, 158—167; P. Ж. М., 1959, 4245.

Редхеффер (Redheffer R.). [1] Errors in simultaneous linear equations. *Quart. Appl. Math.*, 1948, 6, 342—343; M. R., 10, 152.

Рейерсоль (Reiersøl O.). [1] A method for recurrent computation of all the principal minors of a determinant, and its application in confluence analysis. *Ann. Math. Statistics*, 1940, 11, 193—198; M. R., 2, 61.

Рейх (Reich E.). [1] On the convergence of the classical iterative method of solving linear simultaneous equations. *Ann. Math. Statistics*, 1949, 20, 448—451; M. R., 11, 136.

Рихтер (Richter Hans). [1] Bemerkung zur Norm der Inversen einer Matrix. *Arch. Math.*, 1954, 5, № 4—6, 447—448; P. Ж. М., 1956, 1054.

Ричардсон (Richardson L. E.). [1] A purification method for computing the latent columns of numerical matrices and some integrals of differential equations. *Phil. Trans. Roy. Soc. London, Ser. A*, 1950, 242, 439—491; M. R., 12, 133.

Риччи (Ricci L.). [1] Confronto fra i metodi di Banachiewicz, Roma e Volta per la risoluzione dei sistemi di equazioni algebriche lineari. *Atti. Accad. naz. Lincei. Rend. Cl. sci. fis., mat. e natur. ser. 8*, 1949, 7, 72—76; M. R., 11, 743.

Рой (Roy S. W.). [1] A useful theorem in matrix theory. *Proc. Amer. Math. Soc.*, 1954, 5, 635—638; M. R., 16, 4.

Ролл (Roll L. B.). [1] Error bounds for iterative solutions of Fredholm integral equations. *Pacif. J. Math.*, 1955, 5, Suppl. № 2, 977—986; P. Ж. М., 1957, 5178.

Рома (Roma M. S.). [1] Il metodo dell'ortogonalizzazione per la risoluzione numerica dei sistemi di equazioni lineari algebriche. *Ricerca Sci.*, 1946, 16, 309—312; M. R., 8, 171.

[2] Il metodo dell'ortogonalizzazione per la risoluzione numerica dei sistemi di equazioni algebriche. *Pubblicazioni Ist. Appl. Calcolo*, 1947, 189, 12 pp; M. R., 10, 574.

[3] Sulla risoluzione numerica dei sistemi di equazioni algebriche lineari col metodo della ortogonalizzazione. *Pubblicazioni Ist. Appl. Calcolo*, 1950, 283; M. R., 13, 691.

Россер (Rosser J. Barkley). [1] A general iteration scheme for solving simultaneous equations. *Bull. Amer. Math. Soc.*, 1950, 56, 176—177.

[2] A method of computing exact inverses of matrices with integer coefficients. *J. Res. Nat. Bur. Standards*, 1952, 49, 349—353; M. R., 14, 1128.

- [3] Rapidly converging iterative methods for solving linear equations. *Nat. Bur. Standards Appl. Math. Ser.*, 1953, 29, 59–64; M. R., 15, 651.
- Россер, Хестингс, Каруш и Ланцос (Rosser J. B., Hestenes M. R., Karush W. and Lanczos C.). [1] Separation of close eigenvalues of a real symmetric matrix. *J. Res. Nat. Bur. Standards*, 1951, 47, № 4, 291–297; M. R., 14, 92.
- Рот и Скотт (Roth J. P. and Scott D. S.). [1] A vector method for solving linear equations and inverting matrices. *J. Math. and Phys.*, 1956, 35, № 3, 312–317; Р. Ж. М., 1958, 749.
- Рубинштейн и Ратледж (Rubinstein H. and Rutledge Y. D.). [1] High order matrix computations on the Univac. *Proc. Assoc. Comput. Mach.*, 1952, 181–186; M. R., 14, 1019.
- Руджьерио (Ruggiero R. J.). [1] Investigation of three methods for solving the flutter equations and their relative merits. *J. Aeronaut. Sci.*, 1946, 13, 3–22; M. R., 7, 338.
- Ружичка (Růžička Miroslav). [1] Zkrácení iteračního způsobu výpočtu systému lineárních rovnic. *Inžen. stavby*, 1957, 5, № 9, 490–492; Р. Ж. М., 1958, 8288.
- Рутисхаузер (Rutishauser Heinz). [1] Beiträge zur Kenntnis des Biorthogonalisierungs-Algorithmus von Lanczos. *Z. angew. Math. und Phys.*, 1953, 4, № 1, 35–56; Р. Ж. М., 1953, 453.
- [2] Der Quotienten-Differenzen-Algorithmus. *Z. angew. Math. und Phys.*, 1954, 5, № 3, 233–251; Р. Ж. М., 1955, 5316.
- [3] Anwendungen des Quotienten-Differenzen-Algorithmus. *Z. angew. Math. und Phys.*, 1954, 5, № 6, 496–503; Р. Ж. М., 1956, 4926.
- [4] Bestimmung der Eigenwerte und Eigenvektoren einer Matrix mit Hilfe des Quotienten-Differenzen-Algorithmus. *Z. angew. Math. und Phys.*, 1955, 6, № 5, 387–401; Р. Ж. М., 1956, 6862.
- [5] Une méthode pour la détermination des valeurs propres d'une matrice. *C. r. Acad. sci.*, 1955, 240, № 1, 34–36; Р. Ж. М., 1956, 4916.
- [6] Der Quotienten-Differenzen-Algorithmus. *Mitt. Inst. angew. Math. Eidgenoss. techn. Hochschule Zürich*, 1957, № 7, 745; Р. Ж. М., 1958, 760.
- [7] Solution of eigenvalue problems with the LR-transformation. *Nat. Bur. Standards Appl. Math. Ser.*, 1958, 49, 47–81; M. R., 19, 770.
- [8] Zur Bestimmung der Eigenwerte schiefssymmetrischer Matrizen. *Z. Angew. Math. und Phys.*, 1958, 9b, № 5–6, 586–590; Р. Ж. М., 1959, 7482.
- Рутисхаузер и Баэр (Rutishauser Heinz et Bauer Friedrich L.). [1] Détermination des vecteurs propres d'une matrice par une méthode itérative avec convergence quadratique. *C. r. Acad. sci.*, 1955, 240, № 17, 1680–1681; Р. Ж. М., 1956, 4917.
- Сабров и Хиггинс (Sabroff R. R. and Higgins T. J.). [1] A critical study of Kron's method of "tearing". *Matrix Tensor Quart.*, 1957, 7, 107–113; M. R., 19, 64.
- Самсонов К. В. [1] Прибор для решения системы линейных алгебраических уравнений методом итерации. *Прикл. матем. и механика*, 1935, 2, 309–313.
- Самокиш Б. А. [1] Исследование быстроты сходимости метода наискорейшего спуска. *Успехи матем. наук*, 1957, 12, № 1, 238–240.
- Самуэлсон (Samuelson P. A.). [1] A method of determining explicitly the coefficients of the characteristic equation. *Ann. Math. Statistics*, 1942, 13, 424–429; M. R., 4, 148.
- [2] Efficient computation of the latent vectors of a matrix. *Proc. Nat. Acad. Sci. U. S. A.*, 1943, 29, 393–397; M. R., 5, 161.
- [3] A convergent iterative process. *J. Math. and Phys.*, 1945, 24, 131–134.
- [4] Solving linear equations by continuous substitution. *Bull. Amer. Math. Soc.*, 1950, 56, 159.

- Сэттерсвайт (Satterthwaite F. E.). [1] Error control in matrix calculation. *Ann. Math. Statistics*, 1944, 15, 373—389; M. R., 6, 218.
- Сауселл (Southwell R. V.). [1] *Relaxation Methods in engineering Science, a Treatise on Approximate Computation*. Oxford Univ. Press, 1940.
- [2] *Relaxation Methods in Theoretical Physics*. Oxford Univ. Press, 1946.
- Сейбл (Saibel Edward). [1] A modified treatment of the iterative method. *J. Franklin Inst.*, 1943, 235, 163—166; M. R., 4, 148.
- [2] A rapid method of inversion of certain types of matrices. *J. Franklin Inst.*, 1944, 237, 197—201; M. R., 5, 245.
- Сейбл и Берджер (Saibel Edward and Berger W. J.). [1] On finding the characteristic equation of a square matrix. *Math. Tables and Other Aids Comput.*, 1953, 7, № 49, 228—236; Р. Ж. М., 1954, 5786.
- Семеняев К. А. [1] О нахождении собственных значений и инвариантных многообразий матриц посредством итераций. *Прикл. матем. и механика*, 1943, 7, 193—222; M. R., 6, 51.
- Серебряников С. В. [1] О решении численных уравнений методом итераций. *Новочеркасск. Тр. инж. мелиорат. ин-та*, 1939, 3, 168—172.
- Синг (Syng J. L.). [1] A geometrical interpretation of the relaxation method. *Quart. Appl. Math.*, 1944, 2, 87—89; M. R., 6, 50.
- Синго (Shingo Takaichi). [1] An exact method of solving the linear simultaneous equations with the principal diagonal coefficients and those adjacent to them only. *Trans. Japan soc. Civill. Engrs*, 1954, № 19, 1—7; Р. Ж. М., 1956, 1681.
- Синомия (Shinomiya Tetsuo). [1] Обращение матриц методом последовательных приближений. *Res. Repts Fac. Engng Gifu Prefect. Univ.*, 1955, № 5, 5—10; Р. Ж. М., 1958, 4252.
- [2] Практический метод обращения матриц при помощи вычисления определителей. *Res. Repts Fac. Engng Gifu Prefect. Univ.*, 1956, 6, 1—6; Р. Ж. М., 1958, 4253.
- [3] Some notes on the process of reversing a matrix. *Proc. 5th Japan Nat. Congr. Appl. Mech.*, 1955, Tokyo, 1956, 481—484; Р. Ж. М., 1958, 10302.
- Синомия и Оти. [1] Несколько формул для обратных матриц. *Trans. Japan. Soc. Civil. Engrs*, 1955, 24, 78—82; Р. Ж. М., 1956, 7109.
- Скотто (Scotto L. G.). [1] Sul calcolo ed affinamento delle caratteristiche delle vibrazioni dei sistemi elastici ad n gradi di libertà. *Rend. Ist. Lombardo sci. e lettere Cl. sci. mat. e natur. Ser. 3*, 1956, 21 (90), 89—106; M. R., 18, 676.
- Снейдер (Snyder James N.). [1] On the improvement of the solutions to a set of simultaneous linear equations using the ILLIAC. *Math. Tables and Other Aids Comput.*, 1955, 9, № 52, 177—184; Р. Ж. М., 1957, 2682.
- Соколов (Sokoloff N. P.). [1] Sur l'application des déterminants supérieurs à la résolution de certains systèmes d'équations linéaires. *Ann. Soc. Scient. Bruxelles. Sér. I*, 1937, 57, 60—66.
- Станкевич (Stankiewicz L.). [1] Sur les opérations arithmétiques dans le calcul des inverses d'après la méthode de M. T. Banachiewicz. *Bull. intern. Acad. Polon. Sci. A.*, 1937, 363—376.
- [2] Sur les méthodes de Cesari et Kaczmarz relatives à la résolution de systèmes d'équations linéaires à l'aide d'approximations successives. *Bull. intern. Acad. Polon. Sci. A*, 1937, 521—529.
- Стейн (Stein M.). [1] Gradient methods in the solution of systems of linear equations. *J. Res. Nat. Bur. Standards*, 1952, 48, № 6, 407—413; M. R., 14, 322.
- [2] Determining the mode shapes and frequencies of low aspect ratio multispar wings. *J. Aeronaut. Sci.*, 1955, 22, № 2, 137—138; Р. Ж. М., 1958, 8287.

Стейн (Stein P.). [1] The convergence of Seidel iterants of nearly symmetric matrices. *Math. Tables and Other Aids Comput.*, 1951, 5, 237—239.

[2] A note on inequalities for the norm of a matrix. *Amer. Math. Monthly*, 1951, 58, 558—559; M. R., 13, 717.

[3] A note on bounds of multiple characteristic roots of a matrix. *J. Res. Nat. Bur. Standards*, 1952, 48, 59—60; M. R., 13, 813.

[4] A note on the bounds of the real parts of the characteristic roots of a matrix. *J. Res. Nat. Bur. Standards*, 1952, 48, 106—108; M. R., 14, 8.

Стейн и Розенберг (Stein P. and Rosenberg R. L.). [1] On the solution of linear simultaneous equations by iteration. *J. London Math. Soc.*, 1948, 23, 111—118; M. R., 10, 485.

Стерн (Stearn J. L.). [1] Iterative solutions of normal equations. *Bull. Géod.*, 1951, 331—339; M. R., 13, 990.

Стесин И. М. [1] Вычисление собственных значений при помощи непрерывных дробей. *Успехи матем. наук*, 1954, 9, 191—198; Р. Ж. М., 1957, 4384.

Стоякович (Стојаковић Мирко). [1] О неким поступцима за решавање система линеарних алгебарских једначина. Зб. *Маш. фак. Ун-т Београду*, 1954—1955 (1956), 19—27; Р. Ж. М., 1957, 8957.

Свинглхерст (Swindlchurst Beverly). [1] On the solution of simultaneous linear equations. *Proc. Montana Acad. Sci.*, 1956, 16, 59—60; Р. Ж. М., 1959, 5220.

Сурьо (Souriau J. M.). [1] Une méthode pour la décomposition spectrale et l'inversion des matrices. *C. r. Acad. sci.*, 1948, 227, 1010—1011; M. R., 10, 348.

Сурьо и Боннар (Souriau J. M. et Bonnard R.). [1] Théorie des erreurs en calcul matriciel. *Rech. aéronaut.*, 1951, № 19, 41—48; M. R., 12, 638.

Тага (Taga Y.). [1] О линейных вычислениях на автоматической релейной вычислительной машине Института математической статистики. *Proc. Inst. Statist. Math.*, 1957, 5, № 1, 32—48; Р. Ж. М., 1958, 4251.

Таккер (Tucker L. R.). [1] The determination of successive principal components without computation of tables of residual correlation coefficients. *Psychometrika*, 1944, 9, 149—153; M. R., 6, 51.

Таккерман (Tuckerman L. B.). [1] On the mathematically significant figures in the solution of simultaneous linear equations. *Ann. Math. Statistics*, 1941, 12, 307—316; M. R., 3, 154.

Тауски (Taussky Olga). [1] Bounds for characteristic roots of matrices. *Duke Math. J.*, 1948, 15, 1043—1044; M. R., 10, 501.

[2] A recurring theorem on determinants. *Amer. Math. Monthly*, 1949, 56, 672—676; M. R., 11, 307.

[3] Notes on numerical analysis. II. Notes on the condition of matrices. *Math. Tables and Other Aids Comput.*, 1950, 4, 111—112; M. R., 12, 361.

[4] Bounds for characteristic roots of matrices. II. *J. Res. Nat. Bur. Standards*, 1951, 46, 124—125; M. R., 13, 311.

Тауски и Тодд (Taussky Olga and Todd John). [1] Systems of equations, matrices and determinants. *Math. Mag.*, 1952, 26, 9—20, 71—78; M. R., 14, 715.

Темпл (Temple G.). [1] The general theory of relaxation methods applied to linear systems. *Proc. Roy. Soc. Ser. A.*, 1939, 169, 476—500.

[2] The accuracy of Rayleigh's method of calculating the natural frequencies of vibrating systems. *Proc. Roy. Soc. Ser. A.*, 1952, 211, 204—224; M. R., 13, 691.

Тернбулл Эйткен (Turnbull H. W. and A. C. Aitken). [1] *An Introduction to the theory of canonical matrices*. London, Blackie & Son, Ltd, 1932.

Тернер (Terner L. R.). [1] Improvement in the convergence of methods of successive approximation. *Proc. Comput. Sem., Dec. 1949, N. Y. IBM Corp.*, 1951, 135—137; M. R., 13, 586.

Терракини (Terracini Alessandro). [1] Un procedimento per la risoluzione numerica dei sistemi di equazioni lineari. *Ric. Ingegn.*, 1935, 3, 40—48.

Тодд (Todd J.). [1] The condition of certain matrices. I. *Quart. J. Mech. and Appl. Math.*, 1949, 2, 469—472; M. R., 11, 619.

[2] The condition of a certain matrix. *Proc. Cambridge Philos. Soc.*, 1950, 46, 116—118; M. R., 11, 403.

[3] Experiments on the inversion of a 16×16 matrix. *Nat. Bur. Standards Appl. Math. Ser.*, 1953, 29, 113—115; P. Ж. М., 1956, 3333.

[4] The condition of certain matrices. II. *Arch. Math.*, 1954, 5, № 4—6, 249—257; P. Ж. М., 1956, 763.

[5] The condition of matrices. *Proc. Internat. Congr. Math.*, 1954, 2, Amsterdam, 1954, 385—386; P. Ж. М., 1955, 4244.

[6] The condition of the finite segments of the Hilbert matrix. *Nat. Bur. Standards Appl. Math. Ser.*, 1954, 39, 109—116; M. R., 16, 861.

[7] The condition of certain matrices. III. *J. Res. Nat. Bur. Standards*, 1958, 60, № 1, 1—7; P. Ж. М., 1958, 6217.

Тодоров (Todorow Marko). [1] Über die iterative Behandlung linearer Gleichungssysteme. *Bautechnik*, 1958, 35, № 4, 136—138; P. Ж. М., 1959, 877.

Турецкий (Turetsky R.). [1] The least square solution for a set of complex linear equations. *Quart. Appl. Math.*, 1951, 9, 108—110; M. R., 12, 641.

Тьюринг (Turing A. M.). [1] Rounding-off errors in matrix processes. *Quart. J. Mech. and Appl. Math.*, 1948, 1, 287—308; M. R., 10, 405. (Есть перевод. Успехи матем. наук, 1951, 6, № 1, 138—162.)

Тартон (Turton F. J.). [1] On the solution of the numerical simultaneous equations arising in the analysis of redundant structures. *J. Roy. Aeronaut. Soc.*, 1945, 49, 104—111; M. R., 6, 218.

Уагнер (Wagner Harvey M.). [1] A partitioning method of inverting symmetric definite matrices on a card-programmed calculator. *Math. Tables and Other Aids Comput.*, 1954, 8, № 47, 132—139; P. Ж. М., 1956, 3334.

Уиллер и Наш (Wheeler D. J. and Nash J. P.). [1] Digital computer methods for solving linear algebraic equations and finding eigenvalues and eigenvectors. *Digital and Analog Computers and Computing Methods. Symposium at the 18 th Appl. Mech. Div. Conf. of the Asme held at the Univ. of Minnesota, June 18—20, 1953*, N. Y. 1953, 21—35; P. Ж. М., 1955, 1519.

Уилкинсон (Wilkinson J. N.). [1] The calculation of the latent roots and vectors of matrices on the pilot model of the A. C. E. *Proc. Cambridge Philos. Soc.*, 1954, 50, № 4, 536—566; P. Ж. М., 1955, 6108.

[2] The use of iterative methods for finding the latent roots and vectors of matrices. *Math. Tables and Other Aids Comput.*, 1955, 9, № 52, 184—191; P. Ж. М., 1957, 8965.

Уилкс (Wilkes M. V.). [1] Solution of linear algebraic and differential equations by the long-division algorithm. *Proc. Cambridge Philos. Soc.*, 1956, 52, № 4, 758—763; P. Ж. М., 1957, 3524.

Улиг (Ullig J.). [1] Untersuchung der Genauigkeit einer Reihe von Verfahren zur Bestimmung von charakteristischen Zahlen und der Eigenvektoren einer Matrix. *Z. angew. Math. und Mech.*, 1957, 37, 7—8, 265; P. Ж. М., 1959, 7486.

[2]. Über ein inverses Eigenwertproblem. *Z. angew. Math. und Mech.*, 1958, 38, 7—8, 284; P. Ж. М., 1959, 5226.

Ульман (Ullman J.). [1] The probability of convergence of an iterative process of inverting a matrix. *Ann. Math. Statistics*, 1944, 15, 205—213; M. R., 6, 51.

Уманский А. А. [1] Курс строительной механики. 1935 г.

Унгер (Unger H.). [1] Orthogonalisierung von Matrizen nach E. Schmidt und ihre praktische Durchführung. *Z. angew. Math. und Mech.*, 1951, 31, 53—54; M. R., 14, 692.

[2] Zur Aufführung umfangreicher linearer Gleichungssysteme. *Z. angew. Math. und Mech.*, 1952, 32, 1—9; M. R., 14, 92.

[3] Zur Praxis der Biorthonormierung von Eigen-und-Hauptvektoren. *Z. angew. Math. und Mech.*, 1953, 33, 319—331; M. R., 15, 560.

[4] Über direkte Verfahren bei Matrizeneigenwertproblemen. *Wiss. Z. Techn. Hochschule Dresden*, 1952, 1953, 2, № 3, 449—456; Р. Ж. М., 1958, 10298.

Уо (Waugh F. V.). [1] A note concerning Hotelling's method of inverting a partitioned matrix. *Ann. Math. Statistics*, 1945, 16, 216—217; M. R., 7, 84.

[2] Inversion of the Leontief matrix by power series. *Econometrica*, 1950, 18, 142—154; M. R., 12, 133.

Уо и Дуайр (Waugh F. V. and Dwyer P. S.). [1] Compact computation of the inverse of a matrix. *Ann. Math. Statistics*, 1945, 16, 259—271; M. R., 7, 218.

Уокер и Уэстон (Walker A. G. and Weston J. D.). [1] Inclusion theorems for the eigenvalues of a normal matrix. *J. London Math. Soc.*, 1949, 24, 28—31; M. R., 10, 501.

Уэбб (Webb John). [1] Matrices. I. *Electr. Rev.*, 1956, 159, № 6, 237—240.

[2] Matrices. II. Solution of simultaneous equations. *Electr. Rev.*, 1956, 159, № 9, 397—400; Р. Ж. М., 1958, 750.

Фабиан (Fabian Václav). [1] Zufälliges Abrunden und die Konvergenz des linearen (seidelschen) Iterationsverfahrens. *Math. Nachr.*, 1957, 16, № 5—6, 265—270; Р. Ж. М., 1958, 7203.

Фаддеев Д. К. [1] О преобразовании характеристического уравнения матрицы. *Л. Тр. ин-та инж. пром. строит.*, 1937, № 4, 78—96.

[2] О некоторых последовательностях полиномов, полезных для построения итерационных методов решения систем линейных алгебраических уравнений. *Вестн. Ленингр. ун-та*, 1958, № 7, 155—159; Р. Ж. М., 1959, 875.

[3] К вопросу о верхней релаксации при решении систем линейных уравнений. *Изв. высших учебн. заведений. Математика*, 1958, № 5, 122—125; Р. Ж. М., 1959, 5219.

[4] Об обусловленности матриц. *Тр. Матем. ин-та АН СССР*, 1959, 53, 387—391.

Фаддеев Д. К. и Фаддеева В. Н. [1] Вычислительные методы линейной алгебры. *Тр. 3-го Всесоюзного матем. съезда*, 1958, 3, 434—445.

Фаддеева В. Н. [1] Вычислительные методы линейной алгебры. Гос. техиздат, 1950; M. R., 13, 872.

Фальк (Falk Sigurd). [1] Ein übersichtliches Schema für die Matrizenmultiplikation. *Z. angew. Math. und Mech.*, 1951, 31, 152—153; M. R., 12, 751.

[2] Neue Verfahren zur direkten Lösung des allgemeinen Matrizeneigenwertproblems. *Z. angew. Math. und Mech.*, 1954, 34, № 8/9, 289—291; Р. Ж. М., 1956, 761.

[3] Das Ersatzwertverfahren als Hilfsmittel bei der iterativen Bestimmung von Matrizen-Eigenwerten. *Abhandl. Braunschweig. wiss. Ges.*, 1956, 8, 99—110; Р. Ж. М., 1958, 6219.

Фань Цуй (Fan Ky). [1] Note on circular disks containing the eigenvalues of a matrix. *Duke Math. J.*, 1958, 25, 3, 441—445; Р. Ж. М., 1959, 7485.

Фань Цуй и Гофман (Fan Ky and Hoffman A. J.). [1] Lower bounds for the rank and location of the eigenvalues of a matrix. *Nat. Bur. Standards Appl. Math. Ser.*, 1954, 39, 117—130; Р. Ж. М., 1958, 5190.

- Фарнелл (Farnell A. B.). [1] Limits for the characteristic roots of a matrix. *Bull. Amer. Math. Soc.*, 1944, 50, 789—794; M. R., 6, 113.
 [2] Limits for the field of values of a matrix. *Amer. Math. Monthly*, 1945, 52, 488—493.
- Феллер и Форсайт (Feller W. and Forsythe G. E.). [1] New matrix transformation for obtaining characteristic vectors. *Quart. Appl. Math.*, 1951, 8, 325—331; M. R., 12, 538.
- Фельберг (Fehlberg E.). [1] Bemerkungen zur Konvergenz des Iterationsverfahrens bei linearen Gleichungssystemen. *Z. angew. Math. und Mech.*, 1951, 31, 387—389; M. R., 13, 990.
- Фёттер (Voetter H.). [1] Über die numerische Berechnung der Eigenwerte von Säkulargleichungen. *Z. angew. Math. und Phys.*, 1952, 3, 314—316; M. R., 14, 501.
- Феттис (Fettis H.). [1] A method for obtaining the characteristic equation of a matrix and computing the associated modal columns. *Quart. Appl. Math.*, 1950, 8, 206—212; M. R., 12, 209.
- Фидлер и Птак (Fiedler M. und Ptak V.). [1] Über die Konvergenz des verallgemeinerten Seidelschen Verfahrens zur Lösung von Systemen linearer Gleichungen. *Math. Nachr.*, 1956, 15, 31—38; P. Ж. М., 1959, 104, 84.
- Филипowskiй (Filipowsky R.). [1] Numerical calculations in electrical engineering and electronics. I. Calculation of determinants of higher order and the solution of simultaneous algebraic equations. *J. Madras Inst. Tech.*, 1952, 1, 64—88; M. R., 14, 692.
- Фишбах (Fischbach Joseph W.). [1] Some applications of gradient methods. *Proc. Sympos. Appl. Math.*, 6, New York — Toronto — London, 1956, 52—72; P. Ж. М., 1957, 5962.
- Фишер и Фуллер (Fisher Michael E. and Fuller A. T.). [1] On the stabilization of matrices and the convergence of linear iterative processes. *Proc. Cambridge Philos. Soc.*, 1958, 54, 417—425; P. Ж. М., 1959, 8561.
- Фландерс и Шортли (Flanders D. and Shortley G.). [1] Numerical determination of fundamental modes. *J. Appl. Phys.*, 1950, 21, 1326—1332; M. R., 12, 640.
- Фломенхофт (Flomenhoft H. J.). [1] A method for determining mode shapes and frequencies above the fundamental by matrix iteration. *J. Appl. Mech.*, 1950, 17, 249—256; M. R., 12, 287.
- Фокс (Fox L.). [1] Some improvements in the use of relaxation methods for the solution of ordinary and partial differential equations. *Proc. Roy. Soc. Ser. A*, 1947, 190, 31—59.
 [2] A short account of relaxation methods. *Quart. J. Mech. and Appl. Math.*, 1948, 1, 253—280; M. R., 10, 574.
 [3] Practical methods for the solution of linear equations and the inversion of matrices. *J. Roy. Statist. Soc. Ser. B*, 1950, 12, 120—136; M. R., 12, 538.
 [4] Escalator methods for latent roots. *Quart. J. Mech. and Appl. Math.*, 1952, 5, 178—190; M. R., 14, 92.
 [5] Practical solution of linear equations and inversion of matrices. *Nat. Bur. Standards Appl. Math. Ser.*, 1954, 39, 1—54; P. Ж. М., 1957, 8956.
- Фокс, Хаски и Уилкинсон (Fox L., Huskey H. D. and Wilkinson F. N.). [1] Notes on the solution of algebraic linear simultaneous equations. *Quart. J. Mech. and Appl. Math.*, 1948, 7, 149—173; M. R., 10, 152. (Есть перевод. *Успехи матем. наук*, 1950, 5, № 3, 60—86.)
- Фокс и Хейс (Fox L. and Hayes J. G.). [1] More practical methods for the inversion of matrices. *J. Roy. Statistics Soc. Ser. B*, 1951, 13, 83—91; M. R., 13, 990.
- *Фор, Симон-Сuisse и Рона (Faure G., Simon-Suisse J. et Rona Th.). [1] Deux circuits analogiques pour l'inversion des matrices symétriques et la recherche de la vitesse critique de flutter. *Proc. Seventh Internat. Congress Appl. Mech.*, 1948, 4, 81—95; M. R., 11, 403.

- Форсайт (Forsythe G. E.). [1] Gauss to Gerling on relaxation. *Math. Tables and Other Aids Comput.*, 1951, 5, 255—258.
- [2] Alternative derivations of Fox's escalator formulae for latent roots. *Quart. J. Mech. and Appl. Math.*, 1952, 5, 191—195; M. R., 14, 92.
- [3] Tentative classification of methods and bibliography on solving systems of linear equations. *Nat. Bur. Standards. Appl. Math. Ser.*, 1953, 29, 1—28; M. R., 15, 164.
- [4] Solving linear algebraic equations can be interesting. *Bull. Amer. Math. Soc.*, 1953, 59, № 4, 299—329; P. Ж. М., 1954, 3840.
- Форсайт и Лейблер (Forsythe G. E. and Leibler R. A.). [1] Matrix inversion by a Monte Carlo method. *Math. Tables and Other Aids Comput.*, 1950, 4, 127—129; M. R., 12, 361.
- Форсайт и Моткин (Forsythe G. E. and Motzkin T. S.).
- [1] Asymptotic properties of the optimum gradient method. *Bul. Amer. Math. Soc.*, 1951, 57, 183.
- [2] Acceleration of the optimum gradient method. *Bul. Amer. Math. Soc.*, 1951, 57, № 4, 304.
- [3] An extension of Gauss' transformation for improving the condition of systems of linear equations. *Math. Tables and Other Aids Comput.*, 1952, 6, 9—17; M. R., 13, 991.
- Форсайт и Страус (Forsythe G. E. and Straus Ernst O.).
- [1] On best conditioned matrices. *Proc. Internat. Congr. Math.*, 1954, 2, Amsterdam, 1954, 102—103; P. Ж. М., 1956, 2791.
- [2] On best conditioned matrices. *Proc. Amer. Math. Soc.*, 1955, 6, № 3, 340—345; P. Ж. М., 1956, 7882.
- Форсайт и Страус (Forsythe G. E. and Straus Louse W.).
- [1] The Souriau-Frame characteristic equation algorithm on a digital computer. *J. Math. and Phys.*, 1955, 34, № 3, 152—156; P. Ж. М., 1957, 900.
- Форсайт и Форсайт (Forsythe A. I. and Forsythe G. E.).
- [1] Punched-card experiments with accelerated gradient methods for linear equations. *Nat. Bur. Standards. Appl. Math. Ser.*, 1954, 39, 55—69; P. Ж. М., 1958, 10296.
- Форсайт, Хестинс и Рассер (Forsythe G. E., Hestenes M. R. and Rosser J. B.). [1] Iterative methods for solving linear equations. *Bul. Amer. Math. Soc.*, 1951, 57, 480.
- Франк (Frank W. L.). [1] Computing eigenvalues of complex matrices by determinant evaluation and by methods of Danilewski and Wielandt. *J. Soc. Indust. and Appl. Math.*, 1958, 6, № 4, 378—392.
- Франкс (Franckx E d.). [1] Résolution pratique des systèmes linéaires par la méthode des matrices de relaxation. *Bull. Soc. roy. sci. Liège*, 1957, 26, № 7—12, 390—395; P. Ж. М., 1959, 3242.
- Фрёберг (Fröberg Carl-Eric). [1] Solutions of linear systems of equations on a relay machine. *Nat. Bur. Standards. Appl. Math. Ser.*, 1953, 29, 39—42; P. Ж. М., 1956, 3331.
- Фрезер (Frazer R. A.). [1] Note on the Morris escalator process for the solution of linear simultaneous equations. *Philos. Mag.*, 1947, (7) 38, 287—289; M. R., 9, 250.
- [2] Some problems in aerodynamics and structural engineering related to eigenvalues. *Nat. Bur. Standards. Appl. Math. Ser.*, 1953, 29, 65—74; M. R., 15, 164.
- Фрезер, Дункан и Коллар (Frazer R. A., Duncan W. J. and Collar A. R.) [1]. *Elementary matrices and some application to dynamics and differential equations*. 1946; M. R., 8, 365. (Имеется перевод. *Теория матриц и ее приложения к дифференциальным уравнениям и динамике*. Изд. иностран. литер., Москва, 1950.)
- [2]. *Elementary Matrices*. Oxford Univ. Press, 1951.

- Фрейм (Frame J. S.). [1] A simple recursion formula for inverting a matrix. *Bull. Amer. Math. Soc.*, 1949, 55, 1045.
- Фрейре (Freire Rémy). [1] A matricial method for the solution of certain systems of linear equations. *Soc. Paraná Mat. Anuario*, 1956, 3, 54–59.
- Фридрих и Енне (Friedrich K. und Jenne W.). [1] Geometrisch-analytische Auflösung linearer mit Nullkoeffizienten ausgestatteter Gleichungssysteme. *Deutsche Akad. Wiss. Berlin. Veröff. Geodät. Inst. Potsdam*, 1951, 5, 68; M. R., 13, 387.
- Фриман (Freiman G. F.). [1] On the iterative solution of linear simultaneous equations. *Philos. Mag.*, 1943, (7), 34, 409–416; M. R., 5, 50.
- Фрухтер (Fruchter B.). [1] Note on the computation of the inverse of a triangular matrix. *Psychometrika*, 1949, 14, 89–93; M. R., 11, 403.
- Фуруя (Furuya Shigeru). [1] Methods of numerical calculation for simultaneous linear equations and inverse matrices. *Sugaku*, 1957/1958, 9, 240–249; M. R., 20, 1406.
- Халлерт (Hallert B.). [1] Über einige Verfahren zur Lösung von Normalgleichungen. *Z. Vermessungswesen*, 1943, 72, 238–244; M. R., 8, 171.
- Хаммерсли (Hammersley J. M.). [1] The numerical reduction of nonsingular matrix pencils. *Philos. Mag.*, 1949, (7) 40, 783–807; M. R., 11, 464.
- Харман (Harman Harry H.). [1] The square root method and multiple group methods of factor analysis. *Psychometrika*, 1954, 19, 39–55; M. R., 16, 177.
- Хаусхольдер (Householder Alston S.). [1] Some numerical methods for solving systems of linear equations. *Amer. Math. Monthly*, 1950, 57, 453–459; M. R., 12, 538.
- [2] Errors in iterative solution of linear systems. *Proc. Assoc. Comput. Mach. Meeting at Toronto, Ont.*, 1952, Sept. 1953, 30–33; P. Ж. М., 1954, 5263.
- [3] *Principles of numerical analysis*. McGraw-Hill Book Co., 1953; M. R., 15, 470. (Имеется перевод. *Основы численного анализа*. Изд. иностранной литер., Москва, 1956.)
- [4] The geometry of some iterative methods of solving linear systems. *Nat. Bur. Standards Appl. Math. Ser.*, 1953, 29, 35–37; P. Ж. М., 1956, 3157.
- [5] On norms of vectors and matrices. *Oak Ridge Nat. Lab. Oak Ridge Tenn. Rep. ORNL* 1756, 1954, 18 pp.; P. Ж. М., 1957, 6855.
- [6] Terminating and nonterminating iterations for solving linear systems. *J. Soc. Industr. and Appl. Math.*, 1955, 3, № 2, 67–72; P. Ж. М., 1958, 1593.
- [7] On the convergence of matrix iterations. *J. Assoc. Comput. Machinery*, 1956, 3, № 4, 314–324; P. Ж. М., 1958, 6211.
- [8] A survey of some closed methods for inverting matrices. *J. Soc. Industr. and Appl. Math.*, 1957, 5, 155–169; M. R., 19, 982.
- [9] A class of methods for inverting matrices. *J. Soc. Indust. and Appl. Math.*, 1958, 6, № 2, 189–195.
- [10] The approximate solution of matrix problems. *J. Assoc. Comput. Machinery*, 1958, 5, № 3, 205–243.
- [11] Generated error in rotational tridiagonalization. *J. Assoc. Comput. Machinery*, 1958, 5, № 4, 335–338.
- [12] Unitary triangularization of a nonsymmetric matrix. *J. Assoc. Comput. Machinery*, 1958, 5, № 4, 339–342.
- Хед и Оултон (Head Y. W. and Oulton G. M.). [1] The solution of ill-conditioned linear simultaneous equations. *Aircraft Engng*, 1958, 30, 356, 309–312; M. R., 20, 343.
- Хейнрих (Heinrich Helmut). [1] Bemerkungen zu den Verfahren von Hessenberg und Voetter. *Z. angew. Math. und Mech.*, 1956, 36, 250–252; P. Ж. М., 1959, 7481.

- [2] Zur Eingrenzung der charakteristischen Zahlen einer beliebigen Matrix. *Technik*, 1958, 13, № 2, 82—86; P. Ж. М., 1958, 9264.
- Хейс и Виккерс (Hayes J. G. and Vickers T.). [1] The fitting of polynomials to unequally-spaced data. *Philos. Mag.*, 1951, (7) 42, 1387—1400; M. R., 13, 990.
- Хеллер (Heller J.). [1] Ordering properties of linear successive iteration schemes applied to multi-diagonal type linear systems, *J. Soc. Industr. and Appl. Math.*, 1957, 5, 238—243; M. R., 19, 1080.
- Хенричи (Henrici P.). [1] On the speed of convergence of cyclic and quasicyclic Jacobi methods for computing eigenvalues of Hermitian matrices. *J. Soc. Industr. and Appl. Math.*, 1958, 6, № 2, 144—162; M. R., 20, 343.
- [2] The quotient-difference algorithm. *Nat. Bur. Standards. Appl. Math. Ser.*, 1958, 49, 23—46; M. R., 20, 233.
- Херцбергер (Herzberger M.). [1] The normal equations of the method of least squares and their solution. *Quart. Appl. Math.*, 1949, 7, 217—223; M. R., 11, 57.
- Херцбергер и Моррис (Herzberger M. and Morris R. H.). [1] A contribution to the method of least squares. *Quart. Appl. Math.*, 1947, 5, 354—357; M. R., 9, 210.
- Хестинс (Hestenes Magnus R.). [1] Determination of eigenvalues and eigenvectors of matrices. *Nat. Bur. Standards. Appl. Math. Ser.*, 1953, 29, 89—94; P. Ж. М., 1956, 6863.
- [2] Iterative computational methods. *Communic. Pure and Appl. Math.*, 1955, 8, 8595; M. R., 16, 863.
- [3] The conjugate-gradient method for solving linear systems. *Proc. Sympos. Appl. Math.*, 6, New York — Toronto — London, 1956, 83—102; P. Ж. М., 1959, 878.
- [4] Inversion of matrices by biorthogonalization and related results. *J. Soc. Industr. and Appl. Math.*, 1958, 6, 51—90.
- Хестинс и Кауш (Hestenes Magnus R. and Karush W.). [1] A method of gradients for the calculation of the characteristic roots and vectors of a real symmetric matrix. *J. Res. Nat. Bur. Standards*, 1951, 47, 45—61; M. R., 13, 283.
- Хестинс и Штифель (Hestenes Magnus R. and Stiefel Eduard). [1] Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Standards*, 1952 (1953), 49, 409—436; P. Ж. М., 1956, 7645.
- Хечт (Hecht Josef). [1] Poznámka k řešení soustav algebraických lineárních rovnic. *Aplikace mat.*, 1958, 3, 233—237; P. Ж. М., 1959, 5217.
- Хлодовский И. Н. [1] К теории общего случая преобразования векторного уравнения методом акад. А. Н. Крылова. *ИАН ОМЕН*, 1933, 8, 1077—1102.
- Хол (Hoel P. G.). [1] The errors involved in evaluating correlation determinants. *Ann. Math. Statistics*, 1940, 11, 58—65.
- [2] On methods of solving normal equations. *Ann. Math. Statistics*, 1941, 12, 354—359; M. R., 3, 154.
- Хольдт (Holdt Richard Elton). [1] An iterative procedure for the calculation of the eigenvalues and eigenvectors of a real symmetric matrix. *J. Assoc. Comput. Machinery*, 1956, 3, № 3, 223—238; P. Ж. М., 1957, 8966.
- Хольцер и Мелан (Holzer L. und Melan E.). [1] Ein Beitrag zur Auflösung linearer Gleichungssysteme mit positiv definiter Matrix mittels Iteration. *Akad. Wiss. Wien, S.-B. IIA*, 1942, 151, 249—254; M. R., 8, 407.
- Хорви (Horvay G.). [1] Solution of large equation systems and eigenvalue problems by Lanczos' matrix iteration method. *Gen. Electr. Co., Knolls Atomic Power Lab. Schenectady, N. Y., Rept.*, 1953, № KAPL-1004, 113 pp; P. Ж. М., 1957, 8958.
- Хорст (Horst Paul). [1] A method for determining the coefficients of a characteristic equation. *Ann. Math. Statistics*, 1935, 6, 83—84.

Хотеллинг (Hotelling H.). [1] Analysis of a complex of statistical variables into principal components. *J. Educ. Phys.*, 1933, 24, 417—441, 498—520.

[2] Simplified calculation of principal components. *Psychometrika*, 1936, 1, 27—35.

[3] Some new methods in matrix calculation. *Ann. Math. Statistics*, 1943, 14, 1—34; M. R., 4, 202.

[4] Further points on matrix calculation and simultaneous equations. *Ann. Math. Statistics*, 1943, 14, 440—441; M. R., 5, 245.

[5] Practical problems of matrix calculation. *Proc. Berkeley Symp. Math. Stat. Prob.*, 1945, 1946, 275—293; M. R., 10, 574.

Хубларова С. Л. [1] К вопросу о механизации решения больших систем нормальных уравнений. *Сб. реф. Центр. н.-и. ин-та геод. аэро-съёмки и картогр.*, 1954, № 1, 15—16; Р. Ж. М., 1958, 8286.

Цурмюль (Zurmühl R.). [1] Das Eliminationsverfahren von Gauss zur Auflösung linearer Gleichungssysteme. *Ber. Inst. Prakt. Math.*, T. H. Darmstadt, Prof. Dr. A. Walther, Z. W. B. Unters. u. Mitt., 1944, 774, 11—14.

[2] Zur numerischen Auflösung linearer Gleichungssysteme nach dem Matrizenverfahren von Banachiewicz. *Z. angew. Math. und Mech.*, 1949, 29, 76—84; M. R., 10, 743.

[3] *Matrizen*. Berlin—Göttingen—Heidelberg, 1950; M. R., 12, 73.

[4] Zur Iteration einzelner Eigenwerte von Matrizen. *Z. angew. Math. und Mech.*, 1957, 37, № 5—6, 228; Р. Ж. М., 1958, 1597.

Чанселор, Шелдон и Татум (Chancellor Justus, Sheldon J. W. and Tatum G. L.). [1] The solution of simultaneous linear equations using the IBM Card-Programmed Electronic Calculator. *Proc. Indus. Comput. Sem.*, N. Y., IBM Corp., 1951, 57—61; M. R., 13, 587.

Чезари (Cesari Lamberto). [1] Sulla risoluzione dei sistemi di equazioni lineari, per approssimazioni successive. *Atti Acad. naz. Lincei, Rend. Ser.* 6, 1937, 25, 422—428.

[2] Sulla risoluzione dei sistemi di equazioni lineari per approssimazioni successive. *Estratta della Rass. Poste, Telegr. e Telef.*, 1937, 4, 37.

Черенков Ф. С. [1] О решении систем линейных уравнений методом итерации. *Матем. сб.*, 1936, I (43), № 6, 953—958.

Чжао-Фан-Сюн (Chao F. H.). [1] A gradient method for solving simultaneous equations. *Acta math. sinica*, 1953, 3, 328—342; M. R. 17, 194.

[2] Конечно-разностный метод для решения систем линейных алгебраических уравнений. *Acta math. sinica*, 1955, 5, № 2, 149—159; Р. Ж. М., 1957, 8264.

[3] Метод табулирования для решения системы линейных алгебраических уравнений. *Chinese J. Civil Engng*, 1956, 3, № 4, 463—474; Р. Ж. М., 1957, 8267.

[4] Сравнение градиентных методов. *Acta math. sinica*, 1957, 7, № 1, 63—78; Р. Ж. М., 1958, 10295.

Чжень и Уиллоуби (Chen T. C. and Willoughby R. A.). [1] A note on the computation of eigenvalues and vectors of Hermitean matrices. *IBM J. Res. and Developm.*, 1958, 2, № 2, 169—170; Р. Ж. М., 1959, 5230.

Чжоу (Chow C.). [1] Gradual developing method. *Bull. Géod.*, 1951, 221—229; M. R., 13, 496.

Чикала (Cicala P.). [1] Determination of modes and frequencies above the fundamental by matrix iteration. *J. Aeronaut. Sci.*, 1952, 19, 719—720; M. R., 14, 587.

Чиммино (Cimmino Gianfranco). [1] Calcolo approssimato per risoluzioni dei sistemi di equazioni lineari. *Ricerca Scien. Roma*, (2), 1938, 9, 326—333.

Шанкс (Shanks Daniel). [1] On analogous theorems of Fredholm and Frame and on the inverse of a matrix. *Quart. Appl. Math.*, 1955, 13, 95—98; Р. Ж. М., 1957, 8270.

Шварц (Schwarz H. R.). [1] Ein Verfahren zur Stabilitätsfrage bei Matrizen-Eigenwertproblemen. *Z. angew. Math. und Phys.*, 1956, 7, 473—500; M. R., 18, 676.

Шерман (Sherman Jack). [1] Computations relating to inverse matrices. *Nat. Bur. Standards Appl. Math. Ser.*, 1953, 29, 123—124; Р. Ж. М., 1956, 6150.

Шерман и Моррисон (Sherman J. and Morrison W. S.). [1] Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Ann. Math. Statistics*, 1950, 21, 124—127; M. R., 11, 693.

Шмейдлер (Schmeidler W.). [1] *Vorträge über Determinanten und Matrizen mit Anwendungen in Physik und Technik*. Berlin, 1949.

Шмидт (Schmidt R. J.). [1] On the numerical solution of linear simultaneous equations by an iterative method. *Philos. Mag.*, 1941, (7) 32, 369—383; M. R., 3, 276.

Шмиттмайер (Schmidtmaier Josef). [1] Über die Auflösung des Systems linearer algebraischer Gleichungen mit komplexen Koeffizienten. *Z. angew. Math. und Mech.*, 1958, 38, 74—77; M. R., 19, 1080.

Шмидтмайер и Майер (Schmidtmaier Josef a Mayer Daniel). [1] Výhodné řešení lineárních problémů v oboru komplexních čísel. *Saboproudý obzor*, 1958, 19, № 7, 472—477; Р. Ж. М., 1959, 4246.

Шмульян Ю. Л. [1] Замечание по поводу статьи Ю. М. Гаврилова «О сходимости итерационных процессов». *Изв. АН СССР, сер. матем.*, 1955, 19, № 2, 191; Р. Ж. М., 1956, 5507.

Шнейдер (Schneider Hans). [1] Regions of exclusion for the latent roots of a matrix. *Proc. Amer. Math. Soc.*, 1954, 5, № 2, 320—322; Р. Ж. М., 1956, 760.

Шоу (Shaw T. S.). [1] *An introduction to relaxation methods*. N. Y. Dover, 1953, 396 pp.; M. R. 15, 353.

Шперль (Spoerl Ch. A.). [1] A fundamental proposition in the solution of simultaneous linear equations. *Trans. Actuar. Soc. Amer.*, 1943, 44, 276—288; M. R., 5, 161.

[2] On solving simultaneous linear equations. *Trans. Actuar. Soc. Amer.*, 1944, 45, 18—32, 67—69; M. R., 6, 50.

Шрёдер (Schröder Johann). [1] Eine Bemerkung zur Konvergenz der Iterationsverfahren für lineare Gleichungssysteme. *Arch. Math.*, 1953, 4, № 4, 322—326; Р. Ж. М., 1954, 4191.

[2] Neue Fehlerabschätzungen für verschiedene Iterationsverfahren. *Z. angew. Math. und Mech.*, 1956, 36, № 5—6, 168—181; Р. Ж. М., 1957, 1827.

Шрейдер Ю. А. [1] Решение системы линейных совместных алгебраических уравнений. *Докл. АН СССР*, 1951, 76, 651—654; M. R., 12, 639.

Штифель (Stiefel E.). [1] Über einige Methoden der Relaxationsrechnung. *Z. angew. Math. und Phys.*, 1952, 3, 1—33; M. R., 13, 874.

[2] Zur Interpolation von tabellierten Funktionen durch Exponentialsummen und zur Berechnung von Eigenwerten aus den schwarzschén Konstanten. *Z. angew. Math. und Mech.*, 1953, 33, № 8—9, 260—262; Р. Ж. М., 1954, 4195.

[3] Some special methods of relaxation technique. *Nat. Bur. Standards Appl. Math. Ser.*, 1953, 29, 43—48; Р. Ж. М., 1956, 7643.

[4] Ausgleichung ohne Aufstellung der Gaußschen Normalgleichungen. *Wiss. Z. Techn. Hochschule Dresden*, 1953, 2, 441—442; M. R., 16, 1155.

[5] Relaxationsmethoden bester Strategie zur Lösung linearer Gleichungssysteme. *Comment. math. helv.*, 1955, 29, 157—179; Р. Ж. М., 1956, 7644.

[6] Kernel polynomials in linear algebra and their numerical applications. Four lectures on solving linear equations and determining eigenvalues. *Nat. Bur. Standards*, Washington, 1955, 52 pp.; M. R., 17, 790.

[7] Kernel polynomials in linear algebra and their numerical applications. *Nat. Bur. Standards Appl. Math. Ser.*, 1958, 49, 1—22; M. R., 19, 1080.

Шурп (Schur J.). [1] Über Potenzreihen, die im Inneren des Einheitskreises beschränkt sind. *J. reine und angew. Math.*, 1917, 147, 205—232.

Шура-Бура М. Р. [1] Оценка погрешностей при вычислении обратной матрицы для матрицы высокого порядка. *Успехи матем. наук*, 1951, 6, № 4, 121—150; M. R., 13, 284.

Эгервари (Egerváry Jenő). [1] Über die Faktorisation von Matrizen und ihre Anwendung auf die Lösung von linearen Gleichungssystemen. *Z. angew. Math. und Mech.*, 1955, 35, № 3, 111—118; P. Ж. М., 1956, 6389.

[2] Az inverz matrix általánosítása. *Magyar tud. akad. Mat. kutató int. közl.*, 1956, 1, № 3, 315—324; P. Ж. М., 1959, 3513.

[3] Régi és új módszerek lineáris egyenletrendszer megoldására. *Magyar tud. akad. Mat. kutató int. közl.*, 1956, 1, 1—2, 109—123; P. Ж. М., 1959, 7479.

[4] Über eine Verallgemeinerung der Purcellschen Methode zur Auflösung linearer Gleichungssysteme. *Osterr. Ingr.-Arch.*, 1957, 11, 4, 249—251; P. Ж. М., 1959, 5216.

Эйземан (Eisemann Kurt). [1] Removal of ill-conditioning for matrices. *Quart. Appl. Math.*, 1957, 15, № 3, 225—230; P. Ж. М., 1958, 6216.

Эйзен (Eisen Axel). [1] Beitrag zur Lösung linearer Gleichungen. *Internat. Vereinig. Brücken. und Hochbau*, 1935, 3, 56—66.

Эйткен (Aitken A.). [1] On Bernoulli's numerical solution of algebraic equations. *Proc. Roy. Soc. Edinburgh*, 1926, 46, 289.

[2] Further numerical studies in algebraic equations and matrices. *Proc. Roy. Soc. Edinburgh*, 1931, 51, 80.

[3] On the evaluation of determinants, the formation of their adjugates, and the practical solution of simultaneous linear equations. *Proc. Edinburgh Math. Soc.*, II, 1933, 3, 207—219.

[4] Studies in practical mathematics. I. The evaluation with applications of a certain triple product matrix. *Proc. Roy. Soc. Edinburgh*, 1937, 57, 172—181.

[5] Studies in Practical Mathematics. II. The evaluation of the latent roots and latent vectors of a matrix. *Proc. Roy. Soc. Edinburgh*, Ser. A., 1936, 1937, 57, 269—304.

[6] Studies in practical mathematics. V. On the iterative solution of a system of linear equations. *Proc. Roy. Soc. Edinburgh*, Ser. A, 1950, 63, 52—60; M. R., 12, 56.

[7] *Determinants and matrices*. 9 th ed. Edinburgh—London, Oliver and Boyd; New York, Interscience, 1956, vii, 144 pp.; P. Ж. М., 1957, 6159.

Юргенс (Jürgens E.). [1] Zur Auflösung linearer Gleichungssysteme und numerischen Berechnung von Determinanten. Festgabe. Aachen Palm., 1886.

Якоби (Jacobi C. G. J.). [1] Ueber eine neue Auflösungsart der bei der Methode der kleinsten Quadrate vorkommenden lineären Gleichungen. *Astr. Nachr.*, 1845, 22, № 523, 297—306; Jacobis Werké 3, 467.

Ян (Jahn H. A.). [1] Improvement of an approximate set of latent roots and modal columns of a matrix by methods akin to those of classical perturbation theory. *Quart. J. Mech. and Appl. Math.*, 1948, 1, 131—144; M. R., 10, 152.

Янг (Young David). [1] On Richardson's method for solving linear systems with positive definite matrices. *J. Math. and Phys.*, 1954, 32, 243—255; P. Ж. М., 1954, 5782.

[2] Iterative methods for solving partial difference equations of elliptic type. *Trans. Amer. Math. Soc.*, 1954, 76, № 1, 92—111; P. Ж. М., 1955, 4, 1953.

[3] On the solution of linear systems by iteration. *Proc. Sympos. Appl. Math.*, 6, New York — Toronto — London, 1956, 283—298; P. Ж. М., 1957, 5961.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

- Анзорге (Ansorge R.). [1] Über ein Iterationsverfahren von G. Schulz zur Ermittlung der Reziproken einer Matrix. *Z. angew. Math. und Mech.*, 1959, 39, № 3/4, 164—165.
- [2] Bemerkungen zu einem Iterationsverfahren von Bodewig zur Auflösung linearer Gleichungssysteme. *Z. angew. Math. und Mech.*, 1959, 39, № 3/4, 165.
- [3] Das Hertwigsche Iterationsverfahren zur Auflösung linearer Gleichungssysteme als Gesamt- und Einzelschrittverfahren. *Z. angew. Math. und Mech.*, 1959, 39, № 5/6, 248—249.
- Бауэр (Bauer Friedrich L.). [9] Sequential reduction to tridiagonal form. *J. Soc. Industr. and Appl. Math.*, 1959, 7, № 1, 107—113.
- Бейкер (Baker George A.). [1] A new derivation of Newton's identities and their application to the calculation of the eigenvalues of a matrix. *J. Soc. Industr. and Appl. Math.*, 1959, 7, № 2, 143—148.
- Бессмертных Г. А. [1] Об одновременном отыскании двух собственных чисел самосопряженного оператора. *Докл. АН СССР*, 1959, 128, № 6, 1106—1109.
- Браун (Brown J.). [1] Propagation in coupled transmission line systems. *Quart. J. Mech. and Appl. Math.*, 1958, 11, 236—243.
- Брёдер и Смит (Broeder George O. and Smith Harry J.). [1] A property of semi-definite Hermitian matrices. *J. Assoc. Comput. Machinery*, 1958, 5, № 3, 244—245.
- Вильсон (Wilson L. B.). [1] Solution of certain large sets of equations on Pegasus using matrix methods. *Comput. J.*, 1959, 2, № 3, 130—133.
- Вилф (Wilf Herbert S.). [1] Matrix inversion by the annihilation of rank. *J. Soc. Industr. and Appl. Math.*, 1959, 7, № 2, 149—151.
- Винн (Wynn P.). [1] A sufficient condition for the instability of the q - d -algorithm. *Numerische Math.*, 1959, 1, № 4, 203—207.
- [2] On the propagation of error in certain non-linear algorithms. *Numerische Math.*, 1959, 1, № 3, 142—149.
- Ган (Gan G.). [1] Limits for the characteristic roots of a matrix. I. *Advancement in Math.*, 1958, 4, 450—456; M. R., 20, 6, 3893.
- Гольдбаум Я. С. [1] К преобразованию векторного уравнения. *Прикл. матем. и механика*, 1958, 22, № 4, 539—541.
- Гринспан (Greenspan Donald). [2] On popular methods and extent problems in the solution of polynomial equations. *Math. Mag.*, 1958, 31, № 5, 239—253; Р. Ж. М., 1959, 8572.
- Дюк (Dück W.). [1] Eine Fehlerabschätzung zum Einzelschrittverfahren bei linearen Gleichungssystemen. *Numerische Math.*, 1959, 1, № 1, 73—77.
- Катхил и Варга (Cuthill Elizabeth H. and Varga Richard S.). [1] A method of normalized block iteration. *J. Assoc. Comput. Machinery*, 1959, 6, № 2, 236—244.
- Ланцос (Lanczos C.). [10] Linear systems in self-adjoint form. *Amer. Math. Monthly*, 1958, 65, № 9, 665—679.
- Лоткин (Lotkin Mark). [4] Note on the method of contractants. *Amer. Math. Monthly*, 1959, 66, № 6, 476—479.
- [5] Determination of characteristic values. *Quart. Appl. Math.*, 1959, 17, № 3.
- Мак-Гинн (McGinn Laurence C.). [1] The matrix math compiler. *J. Franklin Inst.*, 1957, 264, № 5, 415—416; Р. Ж. М., 1959, 11619.
- Манайра (Manaira Mario). [1] L'inversione delle matrici con L'UNIVAC. *Idee e sist.*, 1958, № 24—25, 9—11; Р. Ж. М., 1959, 8562.
- Маратхе (C. R. Marathe). [1] Note on some semimoduli of a rectangular matrix. *Amer. Math. Monthly*, 1958, 65, № 4, 259—263.
- Мирский (Mirskey L.). [3] On the minimization of matrix norms. *Amer. Math. Monthly*, 1958, 65, № 2, 106—107; Р. Ж. М., 1959, 8823.

- [4] Diagonal elements of orthogonal matrices. *Amer. Math. Monthly*, 1959, 66, № 1, 19–22.
- Монжаллон (Monjalon Albert). [1] *Initiation au calcul matriciel. Matrices. Déterminants. Applications à l'algèbre et à la géométrie analytique*. Paris, Librairie Vuibert, 1955, 131 pp; Р. Ж. М., 1959, 8834.
- Мостовский и Штарк (Mostowski Andrzej a Stark Marcel). [1] *Algebra liniowa* (Bibliot. mat., 19), Warszawa, PWN, 1958, 188 s.; Р. Ж. М., 1959, 9789.
- Нитше (Nitsche J.). [1] Einfache Fehlerschranken beim Eigenwertproblem symmetrischer Matrizen. *Z. angew. Math. und Mech.*, 1959, 39, № 7/8, 322–325.
- Нобл (Noble B.). [1] The numerical solution of an infinite set of linear simultaneous equations. *Quart. Appl. Math.*, 1959, 17, № 1, 98–102.
- Ньюмен и Тодд (Newman Morris and Todd John). [1] The evaluation of matrix inversion programs. *J. Soc. Industr. and Appl. Math.*, 1959, 6, № 4, 466–476.
- Пароди (Parodi Maurice). [1] Sur une méthode de localisation des valeurs caractéristiques de certaines matrices. *C. r. Acad. sci.*, 1958, 247, № 5, 571–573; Р. Ж. М., 1959, 10834.
- Пугачев Б. П. [3] Об одном способе одновременного вычисления двух границ спектра. *Тр. семин. по функцион. анализу*. Воронежск. ун-т, 1957, 5, 52–70.
- [4] К вопросу о быстроте сходимости метода нормальных хорд. *Тр. семин. по функцион. анализу*. Ростовск. н/Д и Воронежск. гос. ун-ты, 1960, 3–4, 77–80.
- [5] Исследование одного метода приближенного вычисления собственных чисел и векторов. *Тр. семин. по функцион. анализу*. Ростовск. н/Д и Воронежск. гос. ун-ты, 1960, 3–4, 81–97.
- Рутисхаузер (Rutishauser Heinz). [9] Zur Matrizeninversion nach Gauss-Jordan. *Z. angew. Math. und Phys.*, 1959, 10, № 3, 281–291.
- [10] Deflation bei Bandmatrizen. *Z. angew. Math. und Phys.*, 1959, 10, № 3, 314–319.
- Райхль (Raichl Jiří). [1] The economical coding of high-order matrices for automatic computers. *Stroje na zpracování informací*, 1956, 4, 257–271; M. R., 20, 6, 4345.
- Розенблум (Rosenblum Marvin). [1] On the Hilbert matrix I. *Proc. Amer. Math. Soc.*, 1958, 9, № 1, 137–140.
- [2] On the Hilbert Matrix II. *Proc. Amer. Math. Soc.*, 1958, 9, № 4, 581–585.
- Сархан и Гринберг (Sarhan A. E. and Greenberg B. G.). [1] Inverting patterned matrices. *Abstr. Short communis Internat. Congress Math. in Edinburgh*, Edinburgh, Univ. Edinburgh, 1958, 128; Р. Ж. М., 1959, 8563.
- Таппер (Tupper S. J.). [1] Ill-conditioned linear equations. *Math. Gaz.*, 1958, 42, № 342, 299–300; Р. Ж. М., 1959, 9507.
- Уайт (White Paul A.). [1] The computation of eigenvalues and eigenvectors of a matrix. *J. Soc. Industr. and Appl. Math.*, 1958, 6, № 4, 393–437.
- Уилкинсон (Wilkinson J. H.). [3] Linear algebra on the Pilot A.C.E. Automatic Digital Comput. at N. P. L., 1955, 129–137.
- [4] The calculation of the eigenvectors of codiagonal matrices. *Comput. J.*, 1958, 1, 90–96.
- [5] The calculation of eigenvectors by the method of Lanczos. *Comput. J.*, 1958, 1, № 3, 148–152.
- [6] The evaluation of the zeros of ill-conditioned polynomials. Part I. *Numerische Math.*, 1959, 1, № 3, 150–166.
- [7] The evaluation of the zeros of ill-conditioned polynomials. Part II. *Numerische Math.*, 1959, 1, № 3, 167–180.

[8] Stability of the reduction of a matrix to almost triangular and triangular forms by elementary similarity transformations. *J. Assoc. Comput. Machinery*, 1959, 5, № 3, 336—359.

Уиндли (Windley P. F.). [1] Transposing matrices in a digital computer. *Comput. J.*, 1959, 2, № 1, с. 47—48.

Форсайт (Forsythe G. E.). [5] Singularity and near singularity in numerical analysis. *Amer. Math. Monthly*, 1959, 65, № 4, 229—240.

Франк (Frank W. L.). [2] Finding zeros of arbitrary functions. *J. Assoc. Comput. Machinery*, 1958, 5, № 2, 154—160.

Фрёберг (Fröberg C. E.). [2] Diagonalization of Hermitian matrices. *Math. Tables and Other Aids Comput.*, 1958, 12, 219—220.

Хаусхольдер и Бауэр (Householder Alston S. and Bauer Friedrich L.). [1] On certain methods for expanding the characteristic polynomial. *Numerische Math.*, 1959, 1, № 1, 29—37.

Хенричи (Henrici P.). [3] On the speed of convergence of cyclic and quasicyclic Jacobi methods for computing eigenvalues of Hermitian matrices. *Abstr. Short communs Internat. Congress Math. in Edinburgh*. Edinburgh, Univ. Edinburgh, 1958, 160; Р. Ж. М., 1959, 8565.

Хорник (Hornick S. D.). [1] IBM 709 Tape Matrix Compiler. *Comm. Assoc. Comput. Machinery*, 1959, 2, № 9, 31—32.

Шехтер (Schechter S.). [1] On the inversion of certain matrices. *Math. Tables and Other Aids Comput.*, 1959, 13, 73—77.

Шелдон (Sheldon J. W.). [1] On the spectral norms of several iterative processes. *J. Assoc. Comput. Machinery*, 1959, 6, № 4, 494—505.

Шмидтмайер (Schmidtmaier Josef). [2] Linear computations over a complex field. *J. Roy. Aeronaut. Soc.*, 1958, 62, № 570, 451—455; Р. Ж. М., 1959, 10483.

Шрейдер Ю. А. [2] Решение систем линейных алгебраических уравнений по методу Монте-Карло. *Вопр. теории матем. машин.* 1. М., Физматгиз, 1958, 167—171; Р. Ж. М., 1959, 8559.