Р. Тьюарсон

РАЗРЕЖЕННЫЕ МАТРИЦЫ

MATHEMATICS IN SCIENCE AND ENGINEERING, V. 99
EDITED BY RICHARD BELLMAN

Department of Applied Mathematics and Statistics
State University of New York
Stony Brook, New York

SPARSE MATRICES

Reginald P. Tewarson

ACADEMIC PRESS
New York and London 1973

Р. Тьюарсон

РАЗРЕЖЕННЫЕ МАТРИЦЫ

Перевод с английского
Э. М. ПЕЙСАХОВИЧА
Под редакцией
Х. Д. ИКРАМОВА

ИЗДАТЕЛЬСТВО „МИР“ МОСКВА 1977

УДК 512.83

Первая в мировой литературе книга, специально посвященная разреженным матрицам, — матрицам с большим числом нулевых элементов. В ней в доступной форме излагается техника применения разреженных матриц в широких классах задач, использующих вычислительные методы линейной алгебры и математического программирования. Учет разреженности матриц позволяет экономить время решения на электронных вычислительных машинах, увеличить размерность задач.

Книга будет полезна математикам-вычислителям, специалистам по прикладной математике и исследованию операций, а также инженерам различных специальностей.

Редакция литературы по математическим наукам

Т $\frac{20203\text{-}028}{041(01)\text{-}77}$ 28-77

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА

Разреженной называют матрицу, имеющую малый процент ненулевых элементов. При этом относительно местоположения нулей никаких предположений не выдвигается: они могут быть расположены совершенно случайным образом.

Журнальная литература последних пятнадцати лет, относящаяся к таким матрицам, насчитывает сотни названий, по этой проблематике состоялось по крайней мере два международных симпозиума. И вместе с тем до последнего времени не было ни одной книги, специально посвященной разреженным матрицам. Понятно поэтому, что выход в свет книги Р. Тьюарсона, в которой основное внимание уделяется вопросам реализации прямых методов для решения систем линейных уравнений очень высокого порядка с разреженными матрицами, — событие примечательное.

Представляя эту книгу советскому читателю, хотелось бы вкратце обрисовать круг вопросов, в ней затрагиваемых, и ее место в нынешней литературе по вычислительной алгебре.

Хотя в численной алгебре за два минувших десятилетия был достигнут значительный прогресс, вопросы, связанные с разреженными матрицами, ускользнули от внимания большинства вычислителей-профессионалов. Наиболее интенсивные разработки велись в это время в других направлениях. С одной стороны, это были задачи для заполненных матриц средних размеров, которые можно целиком разместить в быстродействующей памяти вычислительной машины. С другой стороны, рассматривались и матрицы

высокого порядка с большим количеством нулей. Однако для них выдвигалось условие (выполняющееся во многих приложениях) о закономерном характере расположения ненулевых элементов, которое может быть учтено программой метода заранее. Таковы ленточные матрицы, матрицы Хессенберга, матрицы со свойством A и т. д.

Вследствие сказанного, вопросами, связанными с разреженными матрицами произвольной структуры, занялись главным образом математики прикладных дисциплин, а также специалисты в некоторых других областях численного анализа, например, в области математического программирования, линейного и нелинейного. Так как многие задачи для разреженных матриц естественным образом формулируются на языке теории графов, то к исследованиям примкнули и специалисты в этой области. Книга Р. Тьюарсона фиксирует достигнутый уровень разработки проблемы.

Можно отметить следующие особенности полученных до сих пор результатов. Не делалось серьезных попыток создать принципиально новые алгебраические алгоритмы, рассчитанные именно на разреженные матрицы. Считалось, что вполне достаточно найти разумную модификацию существующих методов. Вопросам устойчивости не отводится столь первостепенная роль, как в численной алгебре «средних размеров». Основное внимание уделяется тому, чтобы при данных возможностях вычислительной машины решить задачу максимального порядка, быть может, за счет некоторой потери точности результатов. Поэтому главными объектами исследования были наиболее целесообразное хранение информации, заключенной в разреженной матрице, и поддержание наибольшей степени ее разреженности на всех этапах вычислительного процесса.

Как известно, большинство прямых методов решения линейных систем основано на приведении матрицы системы к одной из более простых форм — диагональной, треугольной и т. д. Каждая из этих форм характеризуется наличием большого количества нулей, расположенных в заранее определенных

позициях. Прямой метод представляет собой последовательность шагов, на каждом из которых получают нули в нужных позициях очередного обрабатываемого столбца матрицы. При этом сохраняются нули, полученные ранее в предыдущих столбцах. Однако операции по получению нулей, вообще говоря, приводят к появлению новых ненулевых элементов в еще не приведенной части матрицы, ее *заполнению*. Минимизировать это заполнение — вот основная задача, рассматриваемая в книге.

Возможность такой минимизации заложена в самой структуре прямых методов. Каждый из них допускает выбор для проведения очередного шага любого из еще не обработанных столбцов и (или) любой из оставшихся строк. Каждому такому выбору соответствует свое значение последующего заполнения. Как оказывается, можно заранее промоделировать весь процесс решения системы в целочисленной или булевой арифметике и выбрать оптимальный в том или ином смысле порядок строк и столбцов. Эта, вообще говоря довольно значительная, предварительная работа вполне оправданна, если решается целый класс задач с одинаковым расположением ненулевых элементов.

В книге имеется также глава, посвященная вычислению собственных значений и собственных векторов разреженных матриц. Читатель заметит, что эта глава, опирающаяся только на работы самого автора, носит менее систематический характер, чем предыдущее изложение, и в основном демонстрирует отдельные приемы, которые можно использовать для минимизации локального заполнения при приведении матрицы к форме Хессенберга. В действительности Дафф и Рейд [1]) в своей недавней работе на основании численных экспериментов делают вывод, что при преобразовании к хессенберговой форме разреженную матрицу целесообразно рассматривать как за-

[1]) Duff I. S., Reid J. K. On the reduction of sparse matrices to condensed form by similarity transformation. «J. Inst. Math. and Appl.», 1975, 15, No 2, 217—224.

полненную, если исключить случай сверхразреженности.

Дзвая оценку книге в целом, нужно сказать, что она предоставляет читателю краткое и вместе с тем очень ясное изложение указанных выше вопросов. Можно надеяться на то, что она окажется полезной широкому кругу математиков-вычислителей, прикладников и инженеров и вызовет еще больший интерес к этой важной области.

Х. Икрамов

ПРЕДИСЛОВИЕ

Эта книга написана с целью представить в унифицированной и доступной форме обширный материал по исследованиям в области вычислений, связанных с разреженными матрицами, который разбросан по специальным журналам. До сих пор не было книги, в которой излагались бы результаты в этом направлении, в частности по прямым методам обращения больших разреженных матриц и вычислению их собственных значений и собственных векторов.

Разреженные матрицы встречаются при решении многих важных практических задач: структурного анализа, теории электрических сетей и энергосистем распределения энергии, численного решения дифференциальных уравнений, теории графов, а также генетики, социологии и поведенческих наук, программирования для ЭВМ. В связи с развитием современной техники можно ожидать, что и дальше большие разреженные матрицы будут встречаться во многих прикладных задачах, включающих большие системы: например, при планировании работы городской пожарной службы и службы скорой помощи, при моделировании системы сигнализации, управляющей движением транспорта, в распознавании образов и при планировании городов.

Возникновение моего интереса к разреженным матрицам относится к 1962—1964 гг., когда я участвовал в разработке системы команд вычислительной машины для решения задач линейного программирования для крупного изготовителя электронных

вычислительных машин. Матрицы, встречающиеся в задачах линейного программирования, обычно имеют большие размеры и разрежены (они содержат небольшое число ненулевых элементов). Поэтому для повышения эффективности системы команд предусматриваются хранение в памяти и обработка только ненулевых элементов таких матриц.

Тогда и обнаружилось, что имеется очень мало работ, посвященных разложениям обратных матриц на разреженные множители, необходимых для алгоритмов линейного программирования. Это привело к ряду публикаций в этой области.

Весной 1968 г. я был приглашен в качестве лектора на Симпозиум по разреженным матрицам и их приложениям, организованный IBM в Йорк таун Хейтс, Нью-Йорк, в сентябре того же года. Летом 1969 г. последовало другое предложение — написать статью для Конференции по большим разреженным системам линейных уравнений в Оксфордском университете (апрель 1970 г.). Летом 1969 г. по просьбе SIAM я написал обзорную статью о вычислениях с разреженными матрицами; она была опубликована в сентябрьском выпуске 1970 г. SIAM Review. В том же самом году проф. Р. Беллман предложил мне написать книгу на эту тему. По счастливому совпадению тогда же проф. Л. Фокс пригласил меня вести семинар по разреженным матрицам в Оксфордском университете. Из этих лекций и возникла эта книга.

Книга предназначена для специалистов по численному анализу, математическому обеспечению, исследованию операций, в общем для всех, кому приходится иметь дело с большими разреженными матрицами. Она доступна для студентов старших курсов; предполагается, что читатели знакомы с линейной алгеброй. Я старался избегать сжатого математического стиля изложения, иногда допуская некоторое многословие, а также стремился соблюдать нужный баланс между требованиями строгости изложения и потребностями приложений. Я полагаю, что именно прикладные задачи должны приводить к обобщениям и абстракциям. Насколько это возможно,

я придерживался алгоритмического и конструктивного подхода к рассматриваемым проблемам.

В книге описаны основные результаты и последние достижения в области прямых методов обращения больших разреженных матриц, а также вычисления их собственных значений и собственных векторов. Я старался не включать в нее сведения, которые легко найти в хорошо известных книгах по численному анализу, за исключением тех, которые необходимы в качестве основы излагаемого в книге материала.

Книга построена следующим образом.

В гл. 1 описываются некоторые общеупотребительные схемы хранения больших разреженных матриц и приводится метод масштабирования матриц, при котором ошибки округления при вычислениях остаются малыми.

В гл. 2 обсуждается хорошо известный метод исключения Гаусса. Показывается, каким образом гауссово исключение может быть использовано для получения обратной для данной разреженной матрицы в факторизованной форме, которая называется элиминативной формой обратной матрицы (EFI)[1]), излагаются способы, позволяющие для заданной матрицы получать форму EFI настолько разреженной, насколько это возможно.

Описываются некоторые методы минимизации общего числа арифметических операций при вычислении EFI. Рассматриваются также вопросы хранения и использования EFI при практических расчетах.

В гл. 3 даются некоторые методы получения приемлемой разреженности EFI. Эти методы не требуют таких затрат труда, как методы, изложенные в гл. 2. Рассматривается также преобразование заданной матрицы к одной из нескольких форм (например, к ленточной форме), подходящих для получения разреженной EFI.

---

[1]) EFI являются начальными буквами английского названия элиминативной формы обратной матрицы — Elimination Form of Inverse. — *Прим. перев.*

Методы Краута, Дулитла и Холецкого, тесно связанные с методом исключения Гаусса, излагаются в гл. 4.

Для этих методов даются способы минимизации числа ненулевых элементов, создающихся на каждом шаге вычислений. Естественно, эти способы аналогичны тем, которые приводятся в гл. 2 и 3 для гауссова исключения.

В гл. 5 исследуется хорошо известный метод исключения Гаусса — Жордана и показывается, каким образом можно получить другую форму разложения на множители обратной матрицы, называющуюся мультипликативной формой обратной матрицы (PFI)[1]), рассматривается связь между формами EFI и PFI и даются способы нахождения разреженной формы PFI.

Ортонормализация заданного множества разреженных векторов с помощью методов Грама — Шмидта, Хаусхолдера или Гивенса излагается в гл. 6. Последние два метода используются также в гл. 7 для вычисления собственных значений и собственных векторов разреженных матриц. В другом методе гл. 7 для преобразования заданной матрицы используется способ, подобный гауссову исключению. В обеих главах описываются методы, способствующие сведению к минимуму общего числа новых ненулевых элементов (создаваемых в процессе вычислений).

Наконец, в гл. 8 рассматриваются изменения в формах EFI и PFI, вызванные изменением одного или более столбцов в заданной матрице. Это имеет место во многих приложениях, например в линейном программировании. Приводится также еще одна форма разложения на множители обратной матрицы, подобная EFI.

После гл. 8 приводится обширная библиография по разреженным матрицам.

---

[1]) PFI — начальные буквы английского названия мультипликативной формы обратной матрицы — Product Form of Inverse. — *Прим. перев.*

БЛАГОДАРНОСТЬ

Я хотел бы поблагодарить следующих лиц: проф. Л. Фокса за то, что он пригласил меня на год в Оксфордский университет, где я и написал бóльшую часть этой книги; проф. Р. Беллмана, побудившего меня написать ее и поддержавшего меня своими советами; проф. Р. Джозефа и моих аспирантов — гг. Даффа, Ченя и Чена за чтение рукописи и предложения об улучшении отдельных мест.

Р. Тьюарсон

Глава 1

ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

1.1. Введение

В этой вводной главе мы вначале перечислим некоторые области приложения, использующие разреженные матрицы, затем опишем ряд широко распространенных схем хранения больших разреженных матриц в памяти электронной вычислительной машины — внутренней и внешней. Далее излагается также простой метод масштабирования матриц для обеспечения малости ошибок округления. Глава заканчивается библиографией и соответствующими комментариями.

1.2. Разреженные матрицы

Матрица, имеющая небольшой процент ненулевых элементов, называется *разреженной*. Практически матрицу размеров $n \times n$ можно считать разреженной, если количество ее ненулевых элементов имеет порядок n; скажем, от двух до десяти ненулевых элементов в каждой строке при больших n. Разреженными являются матрицы широкого класса задач, относящихся к совокупности людей, связанных совместной работой. Например, матрица, отражающая связи между служащими крупного учреждения, будет разреженной, если предполагать, что элемент i-й строки и j-го столбца матрицы отличен от нуля тогда и только тогда, когда i-й и j-й служащие взаимодействуют друг с другом. Разреженные матрицы встречаются в задачах линейного программирования, структурного анализа, теории цепей и систем распределения энергии, при численном решении дифференциальных уравнений, в задачах теории графов,

генетики, социологии и поведенческих наук[1]), при системном программировании.

В настоящее время проявляется интерес к задачам социологии, бихевиоральных наук и экологии (в частности, к таким задачам, которые возникают для больших городских массивов) (см., например, Роджерс (1971)). Во многих случаях попытки формулировки и решения таких задач приводят к системе уравнений, матрица коэффициентов которой является разреженной и имеет большие размеры. Если уравнения оказываются нелинейными, то в результате их линеаризации (которая часто является первым шагом к решению) получаются разреженные матрицы еще больших размеров.

Интересные и важные задачи часто не могут быть решены потому, что их решение связано с обращением матриц больших размеров, которое либо неосуществимо при имеющемся объеме памяти ЭВМ, либо требует больших затрат. Так как такие матрицы, как правило, разрежены, то полезно знать существующие методы обработки разреженных матриц, что позволяет выбрать лучший метод для каждой разреженной матрицы. Затрата времени и усилий на создание различных методов, пригодных для разреженных матриц, особенно оправданны в тех случаях, когда рассматриваются несколько матриц с одинаковой структурой распределения нулей и с различными численными значениями ненулевых элементов. Это имеет место во многих уже упомянутых областях приложения.

1.3. Упакованная форма хранения

Большие разреженные матрицы обычно хранятся в ЭВМ в упакованном виде. Другими словами, хранятся только ненулевые элементы таких матриц вме-

[1]) В США под поведенческими (бихевиоральными) науками (behavioral sciences) понимают науки, изучающие поведение (behavior) — реакции и действия человека и животных, выражающие их взаимоотношения с внешней средой. — *Прим. перев.*

сте с необходимой информацией об их положении в матрице. Можно указать четыре причины использования упакованной формы хранения. Во-первых, такая форма позволяет хранить и обрабатывать в оперативной памяти ЭВМ матрицы бóльших размерностей, чем при обычном хранении. Во-вторых, могут встретиться случаи, когда даже в упакованном виде матрица не размещается в оперативной памяти (например, при работе ЭВМ в режиме с разделением времени) и требуется использовать внешнюю память (например, магнитные ленты или диски). Ввод данных из внешней памяти ЭВМ в оперативную обычно происходит значительно медленнее обработки этих данных в оперативной памяти. Поэтому упакованная форма хранения предпочтительнее также и при использовании внешней памяти. В-третьих, существенно экономится время благодаря тому, что программой предусматривается исключение тривиальных операций, т. е. вычисления с нулевыми элементами матрицы опускаются. Это часто является единственной возможностью обработки больших матриц. В-четвертых, можно добиться экономии в памяти при хранении обратных матриц, если их представлять в виде произведения элементарных матриц и в упакованной форме хранить только нетривиальные элементы таких матриц. Такие мультипликативные формы обратной матрицы особенно предпочтительны в тех случаях, когда они многократно используются для умножения на ряд вектор-строк и столбцов, как, например, в линейном программировании.

Существует много различных схем упаковки. Некоторые из них описаны ниже — они были признаны эффективными и внедрены в программы для ЭВМ.

Пусть квадратная матрица A порядка n содержит τ ненулевых элементов, причем $\tau \ll n^2$. Ясно, что матрица A является разреженной. Обозначим элемент i-й строки и j-го столбца матрицы через a_{ij}. Для того чтобы хранить в памяти только ненулевые элементы $a_{ij} \neq 0$, необходимо запомнить i, j и a_{ij}. Если используется одна ячейка памяти для каждой из этих величин, то для хранения всех ненулевых

элементов матрицы A требуется 3τ ячеек. Очевидно, 3τ должно быть существенно меньше n^2, чтобы имело смысл тратить на введение упаковки дополнительные усилия и машинное время.

Многие алгоритмы, преобразующие матрицу A в какую-либо другую желаемую форму, порождают на различных этапах вычислений дополнительные ненулевые элементы. Поэтому при хранении в упакованной форме должна быть каким-то образом предусмотрена возможность добавления новых ненулевых элементов в различные столбцы (или строки) матрицы A, если в процессе вычислений элементы матрицы изменяются. Идеальным хранением будет такое, при котором минимизируются одновременно и общий объем используемой памяти, и общее затраченное машинное время. Вообще говоря, требования минимума памяти и минимума времени являются несовместными и необходим компромисс.

Использование связных списков при упаковке

Одним из способов хранения ненулевых элементов данной разреженной матрицы A является использование *связных списков*

Каждому ненулевому элементу a_{ij} соответствует в памяти запись. Записи хранятся по столбцам — элементу a_{ij} будет соответствовать некоторая запись j-го столбца матрицы (рис. 1.3.1) Каждая *запись* представляет собой упорядоченную тройку значений (i, a, p), где i — номер строки, a — значение элемента a_{ij} и p — адрес следующего ненулевого элемента j-го столбца. Значение p равно нулю, если запись соответствует последнему ненулевому элементу столбца. Память для хранения всей матрицы состоит из двух частей: BC-памяти для начальных адресов столбцов и SI-памяти для записей. Первая часть (BC) является массивом из n последовательно расположенных ячеек, содержащих адреса записей первых ненулевых элементов соответствующих столбцов. Например, j-я ячейка BC содержит адрес SI(α) записи, соответствующей первому ненулевому элементу j-го столбца (рис. 1.3.2). Вторая часть (SI) состоит из

записей, связанных с ненулевыми элементами матрицы A. Так как матрица A содержит τ ненулевых элементов и каждому из них соответствует запись, включающая три параметра, то SI потребует для

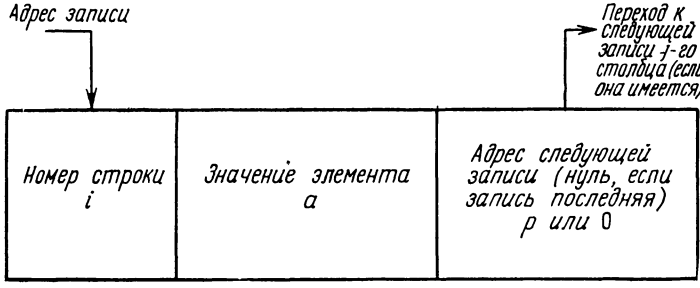

Рис. 1.3.1. Запись, соответствующая a_{ij}.

своего хранения 3τ ячеек, не обязательно непрерывно следующих друг за другом. Таким образом, если мы применяем связные списки, то для хранения матрицы A в упакованной форме потребуется объем памяти в $n + 3\tau$ ячеек.

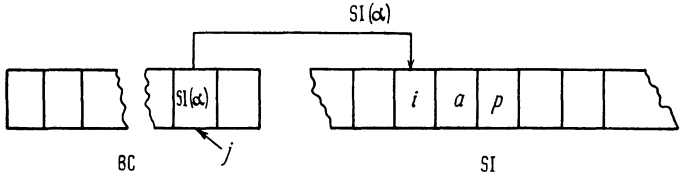

Рис. 1.3.2. Связный список в упакованной форме хранения.

Главное преимущество такой схемы хранения заключается в том, что новые ненулевые элементы, образующиеся в столбцах в процессе вычислений, могут быть легко размещены в SI. Для этого нет необходимости смещать все последующие элементы, как это имеет место в обычных схемах хранения, когда производится вставка нового элемента. Более того, все записи в SI не обязательно должны быть сосредоточены

в одной области памяти, а могут быть разбросаны по всей доступной оперативной памяти ЭВМ (группами ячеек, число которых кратно 3).

Приведем простой пример, показывающий, каким образом появление нового ненулевого элемента влияет на BC и SI. Предположим, что $a_{13} = a_{33} = 0$, $a_{23} = 0{,}5$ и $a_{43} = 1{,}5$. Пусть BC размещается в памяти ЭВМ начиная со 101-й ячейки, а записи для a_{23} и a_{43} начинаются с 200-й и 203-й ячеек соответственно. Если в дальнейшем a_{33} вместо нуля принимает ненулевое значение (скажем, 2,5) и запись для a_{33} должна быть размещена начиная с 300-й ячейки, то необходимые изменения в памяти могут быть представлены следующим образом:

Адреса	103	200	201	202	203	204	300	301	302
Текущее содержимое ячеек	200	2	0,5	203	4	1,5	—	—	—
Новое содержимое ячеек	200	2	0,5	300	4	1,5	3	2,5	203

Таким образом, включение нового ненулевого элемента потребовало только изменения содержимого 202-й ячейки в списке, отражающем текущее состояние матрицы.

Если вместо элемента a_{33} ненулевым стал бы элемент a_{13} (пусть, например, 3,5) и соответствующая ему запись хранилась (как и в предыдущем случае) начиная с ячейки 300, то мы имели бы:

Адрес	103	200	201	202	203	204	300	301	302
Текущее содержимое ячеек	200	2	0,5	203	4	1,5	—	—	—
Новое содержимое ячеек	300	2	0,5	203	4	1,5	1	3,5	200

Как видно, и в этом случае для того, чтобы включить новый ненулевой элемент, в исходном связном списке необходимо изменить только содержимое одной ячейки.

Если в процессе вычислений некоторые ненулевые элементы становятся равными нулю, то ячейки памяти, занятые записями, соответствующими этим элементам, освобождаются и могут быть использованы для хранения записей новых ненулевых элементов. Начальные адреса таких освободившихся записей можно хранить в памяти в виде связного списка свободных записей, для чего использовать третьи ячейки каждой записи. Только начальный адрес первой свободной записи должен где-нибудь запоминаться отдельно. Третья ячейка каждой свободной записи должна содержать адрес следующей свободной записи. Если данная свободная запись является последней в списке свободных записей, то третья ячейка должна содержать нуль. Когда освобождается новая запись, она присоединяется к началу списка. Аналогично для включения в список записей новых ненулевых элементов используются свободные записи, расположенные в начале списка свободных записей.

Рассмотрим два простых примера, иллюстрирующих изложенные выше методы. Предположим, что свободными являются две записи с начальными адресами 101 и 201 и мы хотим добавить к списку еще одну свободную запись с начальным адресом 301. Если ячейка 50 содержит адрес первой свободной записи, то требуемые изменения в содержимом ячеек памяти могут быть представлены следующей таблицей:

Адрес	50	101	102	103	201	202	203	301	302	303
Текущее содержимое ячеек	101	—	—	201	—	—	0	—	—	—
Новое содержимое ячеек	301	—	—	201	—	—	0	—	—	101

С другой стороны, для хранения новой записи используется первая свободная запись в списке, а именно ячейки 301, 302, 303. Затем изменяется содержимое ячеек таким образом, чтобы оно соответствовало строке таблицы, помеченной «Текущее содержимое ячеек».

Иногда полезны упаковки, не использующие связных списков. Для них требуется меньшая память, но добавочный ненулевой элемент может вводиться только путем сдвига всех следующих за ним элементов на одну запись. Эти схемы пригодны в тех случаях, когда только небольшая часть матрицы в процессе вычислений может храниться в оперативной памяти ЭВМ, и поэтому потребовалось бы значительное время для ввода и вывода данных при обращении к внешней памяти. Ниже описываются четыре такие схемы и для первых трех показывается, каким образом хранится матрица A_5 с ненулевыми элементами a_{21}, a_{41}, a_{52}, a_{13}, a_{33}, a_{24} и a_{45}. Последняя схема предназначена для симметричных матриц, и поэтому она отличается от предыдущих. В первых трех схемах матрицы хранятся по столбцам, а в последней — по строкам.

Схема I

Каждому ненулевому элементу матрицы соответствует запись, занимающая две ячейки памяти. Первая ячейка содержит номер строки, вторая — значение элемента. Нуль в первой ячейке означает конец данного столбца. Вторая ячейка в этом случае содержит номер следующего столбца. Нули в обеих ячейках указывают на конец массива, хранящего матрицу. Таким образом, общее число записей будет равно $n + \tau + 1$, из них n — для столбцов, τ — для ненулевых элементов матрицы A и одна запись — для указания конца матрицы. Так как каждая запись использует две ячейки памяти, то для хранения матрицы A потребуется $2(n + \tau + 1)$ ячеек.

Матрица A_5, для которой $\tau = 7$ и $n = 5$, будет храниться в виде массива

(0, 1; 2, a_{21}; 4, a_{41}; 0, 2; 5, a_{52}; 0, 3; 1, a_{13}; 3, a_{33}; 0, 4; 2, a_{24}; 0, 5; 4, a_{45}; 0, 0).

Схема II

Информация о данной матрице хранится в трех массивах: VE — значений ненулевых элементов, RI — индексов строк и CIP — указателей индексов столбцов.

Элемент RI(α) — α-й элемент массива RI — содержит индекс строки α-го элемента VE — VE(α). Если первый ненулевой элемент β-го столбца данной матрицы размещается в VE(t_β), то t_β хранится в β-м элементе CIP, т. е. CIP(β) $= t_\beta$. Очевидно, VE и RI состоят из τ элементов, а CIP из n элементов. Следовательно, эта схема требует общее число в $2\tau + n$ ячеек.

Матрица A_5 будет храниться следующим образом:

$$\begin{aligned} \mathrm{VE} &= (a_{21},\ a_{41},\ a_{52},\ a_{13},\ a_{33},\ a_{24},\ a_{45}),\\ \mathrm{RI} &= (2,\ 4,\ 5,\ 1,\ 3,\ 2,\ 4),\\ \mathrm{CIP} &= (1,\ 3,\ 4,\ 6,\ 7). \end{aligned}$$

Вышеизложенной схемой легко пользоваться. Например, a_{33} может быть найдено следующим образом. Так как CIP(3) $= 4$, то RI(4) даст индекс строки первого ненулевого элемента третьего столбца. Если $a_{33} \neq 0$, то RI(4) или один из следующих за ним элементов RI, предшествующих первому ненулевому элементу четвертого столбца, должен быть равен 3.

В нашем случае RI(5) $= 3$, так как VE(5) содержит a_{33}.

Схема III

Каждому ненулевому элементу данной матрицы однозначно ставится в соответствие целое число $\lambda(i, j)$ вида

$$\lambda(i, j) = i + (j - 1)n, \quad a_{ij} \neq 0.$$

Хранение ненулевых элементов обеспечивается двумя массивами: VE — значений ненулевых элементов и LD, в каждом из которых содержится τ элементов. В LD(α) находится $\lambda(i, j)$, соответствующее a_{ij} из VE(α), где $\alpha = 1, 2, \ldots, \tau$.

Матрица A_5 хранится в виде

$$\begin{aligned} \mathrm{VE} &= (a_{21},\ a_{41},\ a_{52},\ a_{13},\ a_{33},\ a_{24},\ a_{45}),\\ \mathrm{LD} &= (2,\ 4,\ 10,\ 11,\ 13,\ 17,\ 24). \end{aligned}$$

Исходная матрица может быть восстановлена по этой схеме хранения следующим образом. В соответ-

ствии с данным выше определением $\lambda(i, j)$ является очевидным, что j есть наименьшее целое число, большее или равное $\frac{\lambda(i, j)}{n}$, и

$$i = \lambda(i, j) - (j - 1)n.$$

Например, если $\lambda(i, j) = \mathrm{LD}(5) = 13$, тогда $\frac{\lambda(i, j)}{n} = \frac{13}{5}$ и наименьшее целое число, большее или равное $\frac{\lambda(i, j)}{n}$, будет 3. Следовательно, $j = 3$ и $i = \lambda(i, j) - (j - 1)n = 13 - 10 = 3$.

Схема IV

Если A — симметричная матрица, в которой для всех $i \geqslant j$

$$a_{ij} = 0, \quad i - j > \theta_i,$$

где θ_i намного меньше n и в общем случае может иметь различные значения для каждого i, то такая матрица называется *симметричной ленточной матрицей с локально изменяющейся шириной ленты.* Подробное описание ленточных матриц дано в разд. 3.3. Поскольку матрица A симметрична, должна запоминаться только ее нижняя треугольная часть, содержащая элементы, которые лежат на главной диагонали или ниже. Хранение производится с помощью двух массивов: VE — значений ненулевых элементов и PD — положений диагональных элементов в массиве VE. Для каждой строки в VE хранятся крайний левый ненулевой элемент и все следующие элементы, расположенные справа от него вплоть до диагонального включительно. Поэтому i-я строка матрицы A требует $\theta_i + 1$ ячеек для хранения и VE будет состоять из $\sum_{i=1}^{n}(\theta_i + 1)$ элементов. Добавляя n элементов, необходимых для PD, получим общее требуемое количество в $\sum_{i=1}^{n}\theta_i + 2n$ ячеек для хранения A. Если лента является *полной*, т. е. $a_{ij} \neq 0$ для всех $i - j \leqslant \theta_i$ при

$i > j$, тогда $\sum_{i=1}^{n} \theta_i = \frac{\tau - n}{2}$ и требуемый объем памяти будет составлять $\frac{\tau + 3n}{2}$ ячеек.

Проиллюстрируем эту схему хранения на примере матрицы с ненулевыми элементами a_{11}, a_{21}, a_{22}, a_{32}, a_{33}, a_{42}, a_{44}, a_{52}, a_{53}, a_{55}, лежащими на диагонали и под ней. Заметим вначале, что крайние левые ненулевые элементы в третьей, четвертой и пятой строках расположены во втором столбце. Поэтому нулевые элементы a_{43} и a_{54} должны также храниться в VE. Матрица будет запоминаться в виде

$$VE = (a_{11}, a_{21}, a_{22}, a_{32}, a_{33}, a_{42}, 0, a_{44}, a_{52}, a_{53}, 0, a_{55}),$$
$$PD = (1, 3, 5, 8, 12).$$

Элемент a_{ij} исходной матрицы может быть восстановлен по этой схеме следующим образом. Положение a_{ij} в VE определяется значением $PD(i) - (i - j)$, если только $PD(i) - (i - j) > PD(i - 1)$. Последнее условие означает, что a_{ij} не будет лежать влево от первого ненулевого элемента i-й строки, иначе $a_{ij} = 0$ и не хранится в VE. Например, чтобы найти элемент a_{53} в массиве VE, вычисляем

$$PD(5) - (5 - 3) = 12 - 2 = 10 > 8 = PD(4)$$

и, следовательно, a_{53} хранится в VE(10).

Главное преимущество этой схемы упаковки состоит в следующем. Если в процессе вычислений (например, при исключениях по методу Гаусса, гл. 2) создаются дополнительные ненулевые элементы только вправо от крайнего левого элемента каждой строки, то они могут запоминаться в VE без перемещений всех следующих за ними элементов.

Некоторые замечания о схемах упаковки

Теперь укажем вкратце возможные пути дополнительной экономии памяти для приведенных выше схем упаковки.

Если матрица A симметрична, то хранятся, как в схеме IV, только ненулевые элементы нижней

(или верхней) треугольной ее части, включая диагональ.

В связном списке можно объединять две или более записи в блоки последовательно расположенных ячеек. В этом случае каждая запись блока, кроме последней, может состоять из двух, а не из трех ячеек. Само собой разумеется, что включение или исключение записи становится при такой форме хранения довольно сложным.

Если длина (число двоичных разрядов) ячейки памяти ЭВМ достаточно велика, то для экономии памяти во всех описанных выше схемах можно хранить в ячейке по два и более индекса строк (или столбцов). Это требует знакомства с программированием на уровне языка машин и, вообще говоря, не является приемлемым при использовании алгоритмических языков высокого уровня, подобных ФОРТРАНу или АЛГОЛу.

Во всех схемах упаковки, за исключением схемы IV, описывалось хранение матрицы по столбцам. Во многих приложениях предпочтительнее хранение по строкам. Нет необходимости его описывать, так как оно вполне аналогично хранению по столбцам (это то же самое, что и хранение транспонированной к A матрицы по столбцам).

1.4. Масштабирование

Матрица A часто связана с системой линейных уравнений $Ax = b$, где x и b являются вектор-столбцами n-го порядка. Часто элементы x_j и b_i векторов x и b соответственно измеряются в единицах, которые значительно отличаются друг от друга порядком величины. Например, b_1 измеряется в сантиметрах, а b_2 — в километрах, так что в результате первая строка матрицы A и b_1 значительно больше (в какой-либо норме), чем вторая строка и b_2. Положение может быть исправлено, если обе строки сделать в какой-либо норме равными. Для этого обычно рекомендуется строки и столбцы заданной матрицы преобразовать так, чтобы они были величинами одного порядка. Такое преобразование называется *масштабированием*.

Простой способ масштабирования матрицы A состоит в делении каждой строки на наибольший по абсолютному значению элемент этой строки. Масштабированию строк может предшествовать масштабирование столбцов, а именно деление каждого столбца матрицы A на наибольший по абсолютному значению элемент этого столбца. Многие программы задач линейного программирования предусматривают масштабирование, так как полученная в ЭВМ обратная матрица содержит меньшие ошибки округления, если исходная матрица до обращения масштабируется.

Мы можем следующим образом описать влияние масштабирования на решение уравнения $Ax = b$. Пусть e_j обозначает j-й столбец единичной матрицы I_n n-го порядка. Тогда решение x для $Ax = b$ является тем же, что и для

$$D_2AD_1D_1^{-1}x = D_2b, \tag{1.4.1}$$

где D_1 и D_2 — диагональные матрицы, такие, что

$$e_j'D_1e_j = [\max_i |a_{ij}|]^{-1};$$

$$e_i'D_2e_i = [\max_j |e_i'AD_1e_j|]^{-1}. \tag{1.4.2}$$

Решением (1.4.1) является

$$x = D_1(D_2AD_1)^{-1}D_2b. \tag{1.4.3}$$

Как видно, вместо вычисления A^{-1} производится вычисление обращенной масштабированной матрицы D_2AD_1. Если $D_1 = I_n$, то имеет место только *масштабирование строк* и (1.4.3) будет иметь вид

$$x = (D_2A)^{-1}D_2b. \tag{1.4.4}$$

1.5. Библиография и комментарии

Ниже приводятся некоторые ссылки на литературу по разреженным матрицам в различных приложениях.

а) Линейное программирование: Маркович (1957), Ларсон (1962), Вольф и Катлер (1963), Данциг (1963 а, б), Смит и Орчард-Хейс (1963), Диксон

(1965), Тьюарсон (1966), (1967а), Орчард-Хейс (1968), Данциг и др. (1969), Орчард-Хейс (1969), Смит (1969), Вулф (1969), Томлин (1970), Бил (1971), Де Бюше (1971), Карре (1971), Форрест и Томлин (1972).

б) Структурный анализ: Ливсли (1960—1961), Стьюард (1962), Олвей и Мартин (1965), Дженнингс (1968), Розен (1968), Катхил и Мак-Ки (1969), Мак-Кормик (1969), Пэлекол (1969), Олвуд (1971), Катхил (1971), Джордж (1972).

в) Теория цепей и систем распределения энергии: Брейнин (1959), Рот (1959), Крон (1963), Сато и Тинни (1963), Тинни и Уокер (1967), Эдельман (1968), Чен (1969), Тинни (1969), Бети и Стьюарт (1971), Бауман (1971), Черчилл (1971), Огбуобири (1971).

г) Численное решение дифференциальных уравнений: Варга (1962), Карре (1966), Лайниджер и Уиллегби (1969), Гир (1971), Ивенс (1972), Гимон и Кинг (1972).

д) Теория графов: Басейкер и Саати (1965), Далмейдж и Мендельсон (1967), Харари (1967, 1971 а, б), Беллман и др. (1970).

е) Теория генетики: Фалкерсон и Гросс (1965).

ж) Поведенческие науки: Харари (1960), Меримонт (1959), Росс и Харари (1959).

3) Программирование для ЭВМ: Меримонт (1960).

и) Другие области: Ашкенази (1971), Глейзер (1972), Гимон и Кинг (1972).

Связные списки описали Маурзер (1968), Кеттлер и Вейл (1969), Огбуобири (1970), Черчилл (1971), Густавсон (1972). Другие схемы хранения дали Дженнингс (1966), Хименес (1969), Иенсон и Паркс (1970), Надинг и Калерт-Уормболд (1970), Берри (1971), Де Бюше (1971), Густавсон (1972), Дженнингс и Тафф (1971). Общее введение к методам разреженности дано Тинни и Огбуобири (1970).

Масштабирование строк (даже если ему предшествует масштабирование столбцов) не является решением проблемы масштабирования в целом. Для дальнейшего знакомства с вопросами масштабирования, которое иногда называют также *уравновешиванием*,

читатель может обратиться к Уилкинсону (1965), Форсайту и Молеру (1967) и Уэстлейку (1968). Более совершенные методы масштабирования рассмотрели Фалкерсон и Вулф (1962), Бауер (1963), Ван-дер-Слюис (1969) и Кертис и Рейд (1971 б).

В этой книге мы будем заниматься главным образом прямыми методами обращения разреженных матриц. Непрямые (или итеративные) методы для разреженных матриц не будут описаны, так как они хорошо известны (Варга (1962)). В общих руководствах по численному анализу и в периодической литературе обычно рассматриваются прямые методы обращения для небольших полных матриц. Однако во многих прикладных задачах (например, линейного программирования, структурного анализа, электрических цепей и систем генерирования и распределения энергии) получили широкое развитие и внедрены в программы для ЭВМ прямые методы обращения больших разреженных матриц: Ливсли (1960—1961), Смит и Орчард-Хейс (1963), Диксон (1965), Дженнингс (1968), Данциг и др. (1969), Ли (1969), Густавсон и др. (1970), Берри (1971), Де Бюше (1971), Кантин (1971), Форрест и Томлин (1972).

Глава 2
МЕТОД ИСКЛЮЧЕНИЯ ГАУССА

2.1. Введение

В этой главе описывается метод исключения Гаусса для решения систем линейных уравнений. Мы покажем, каким образом матрицы, связанные с различными стадиями процесса исключения, могут быть использованы для представления в факторизованной форме матрицы, обратной для матрицы коэффициентов линейных уравнений. Доказываются некоторые теоремы, с помощью которых определяются разреженные множители разложения обращенных разреженных матриц.

2.2. Основной метод

Наиболее известным методом решения системы уравнений

$$Ax = b \qquad (2.2.1)$$

(где, как и в гл. 1, x и b — n-мерные векторы-столбцы, а A — неособенная матрица n-го порядка) является метод исключения Гаусса (Уилкинсон (1965)). Он состоит из двух частей: *прямого исключения*, в котором с помощью ряда элементарных преобразований (операций над строками) матрица A приводится к верхней треугольной матрице U с единичной диагональю, и так называемой *обратной подстановки*, которая приводит к обращению U.

Прямое гауссово исключение состоит из n шагов. Пусть $A^{(k)}$ обозначает матрицу в начале k-го шага, причем $A^{(1)} \equiv A$ и $A^{(n+1)} \equiv U$. Пусть $a_{ij}^{(k)}$ является элементом i-й строки и j-го столбца ((i, j)-м элементом) матрицы $A^{(k)}$. Другими словами, пусть $a_{ij}^{(k)} = e_i' A^{(k)} e_j$, где e_i является i-м столбцом единичной мат-

рицы I_n n-го порядка, как уже указывалось в разд. 1.3. Для первых $k-1$ столбцов матрица $A^{(k)}$ уже имеет форму верхней треугольной матрицы (рис. 2.2.1). На

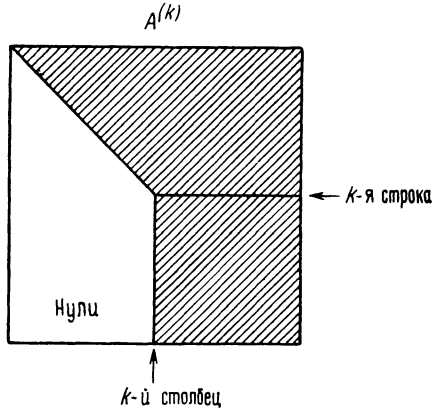

Рис. 2.2.1. Матрица в начале k-го шага.

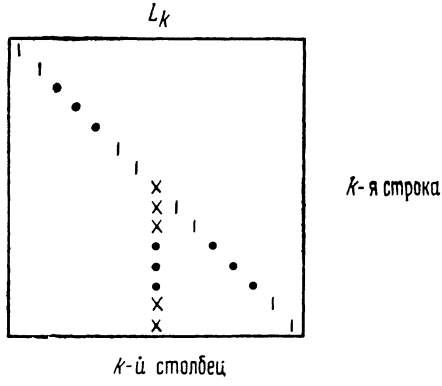

Рис. 2.2.2. Элементарная матрица на k-м шаге.

k-м шаге k-я строка матрицы $A^{(k)}$ делится на ее (k, k)-й элемент, умножается на различные коэффициенты и вычитается из всех следующих за ней строк так, чтобы все ненулевые элементы k-го столбца, лежащие ниже k-й строки, становились нулями.

Получающаяся в результате матрица обозначается через $A^{(k+1)}$. В матричных обозначениях этот процесс может быть по-другому сформулирован следующим образом.

Прямое гауссово исключение состоит в вычислении

$$A^{(k+1)} = L_k A^{(k)}, \quad k = 1, 2, \ldots, n, \tag{2.2.2}$$

где элементарная нижняя треугольная матрица L_k (рис. 2.2.2) задается в виде

$$L_k = I_n + (\eta^{(k)} - e_k) e_k', \tag{2.2.3}$$

с элементами вектора-столбца $\eta^{(k)}$, определяемыми следующим образом:

$$\eta_i^{(k)} = 0, \quad i < k,$$

$$\eta_k^{(k)} = \frac{1}{a_{kk}^{(k)}}, \quad \eta_i^{(k)} = -\frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}, \quad i > k. \tag{2.2.4}$$

Таким образом, диагональные элементы матрицы L_k равны единицам во всех столбцах, за исключением k-го, в котором элементы, лежащие на диагонали и под ней, равны $\eta_i^{(k)}$ ($i = k, k+1, \ldots, n$). Все остальные элементы матрицы L_k равны нулю.

Теперь из (2.2.2) получаем

$$A^{(n+1)} = L_n \ldots L_2 L_1 A^{(1)}. \tag{2.2.5}$$

Если положить

$$L = L_n \ldots L_2 L_1 \tag{2.2.6}$$

и учесть, что $A^{(1)} \equiv A$ и $A^{(n+1)} \equiv U$, то вместо формул (2.2.5) и (2.2.6) получим

$$U = LA. \tag{2.2.7}$$

Таким образом, прямое гауссово исключение состоит в нахождении нижней треугольной матрицы L (произведение нижних треугольных матриц дает также нижнюю треугольную матрицу), которая преобразует матрицу A в верхнюю треугольную матрицу U. Имея в виду уравнение (2.2.7) и то, что операторы L_k применяются к обеим частям уравнения (2.2.1), приходим

в результате гауссова исключения к уравнению

$$Ux = Lb. \tag{2.2.8}$$

Обратная подстановка метода Гаусса состоит в решении уравнения (2.2.8), которое производится следующим образом. Пусть x_i означает i-й элемент вектора x. Тогда последний элемент x_n равен последнему элементу вектора-столбца Lb, так как в послед-

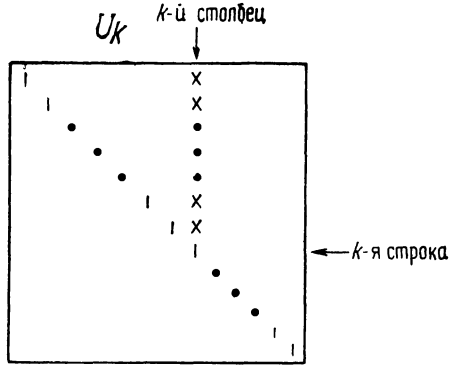

Рис. 2.2.3. Элементарная матрица на k-шаге обратной подстановки.

ней строке матрицы U равны нулю все элементы, кроме последнего, равного единице. Это значение x_n подставляется в предыдущее уравнение, что позволяет легко вычислить x_{n-1}. Подстановка x_n и x_{n-1} в $(n-2)$-ю строку уравнения (2.2.8) дает x_{n-2} и так далее.

Для того чтобы описанную выше обратную подстановку выразить в матричных обозначениях, отметим прежде всего, что $a_{ij}^{(i+1)}$ является (i, j)-м элементом матрицы U. Это следует из того факта, что i-я строка матрицы A в формуле (2.2.2) видоизменяется только до $k = i$ и затем остается неизменной. Другими словами, i-е строки матриц $A^{(i+1)}$ и U совпадают.

Обратная подстановка метода исключения Гаусса может быть теперь определена следующим образом:

$$U_2 \ldots U_{n-1}U_nU = I_n, \tag{2.2.9}$$

где

$$U_k = I_n + \xi^{(k)} e_k', \quad k = n,\ n-1 \ldots 2, \quad (2.2.10)$$

и элементы вектора-столбца $\xi^{(k)}$ задаются в виде

$$\xi_i^{(k)} = -a_{ik}^{(i+1)}, \quad i < k \quad \text{и} \quad \xi_i^{(k)} = 0, \quad i \geqslant k \quad (2.2.11)$$

(рис. 2.2.3). Таким образом, все диагональные элементы матрицы U_k равны единице, а в k-м столбце наддиагональные элементы принимают значения $\xi_i^{(k)} (i=1, 2, \ldots, k-1)$. Все остальные элементы U_k являются нулями.

Теперь можно описать общий результат применения процессов прямого исключения и обратной подстановки к матрице A. Из уравнения (2.2.9) получим

$$U^{-1} = U_2 \ldots U_{n-1} U_n, \quad (2.2.12)$$

и из (2.2.8), (2.2.6) и (2.2.12) следует

$$x = U^{-1} L b = U_2 \ldots U_{n-1} U_n L_n \ldots L_2 L_1 b. \quad (2.2.13)$$

2.3. Выбор главного элемента и ошибки округления

Элемент $a_{kk}^{(k)}$ в формулах (2.2.4) называется *главным элементом* k-го шага исключения. Так как в памяти ЭВМ числа хранятся в ячейках конечной длины, то при вычислениях, вообще говоря, вносятся *ошибки округления*. Для того чтобы минимизировать влияние ошибок округления при гауссовом исключении, Уилкинсон (1965) предлагает для полных (неразреженных) матриц изложенные ниже способы. Его рекомендации основаны на том обстоятельстве, что эти способы позволяют получить границы для ошибок, и, кроме того, было показано, что процесс вычислений устойчив.

Первый способ, называемый *частичным упорядочением* (частичным выбором главного элемента), заключается в следующем. На k-м шаге выбирают наибольший по абсолютному значению из элементов k-го столбца матрицы $A^{(k)}$, расположенных в k-й строке или

ниже ее, а именно

$$|a_{sk}^{(k)}| = \max_i |a_{ik}^{(k)}|, \quad k < i \leqslant n, \tag{2.3.1}$$

и переставляют s-ю и k-ю строки матрицы $A^{(k)}$ перед выполнением вычислений k-го шага по формуле (2.2.2). Конечно, все такие перестановки строк должны запоминаться для дальнейшего использования.

Второй способ минимизации влияния ошибок округления, называемый *полным упорядочением* (полным выбором главного элемента), может быть описан следующим образом. На k-м шаге (при каждом k) выбирают наибольший по абсолютному значению из элементов, расположенных в последних $n-k+1$ строках и столбцах матрицы $A^{(k)}$, а именно

$$|a_{s\sim}^{(k)}| = \max_{i,\,j} |a_{ij}^{(k)}|, \quad k \leqslant i, j \leqslant n, \tag{2.3.2}$$

и переставляют s-ю и k-ю строки и t-й и k-й столбцы матрицы $A^{(k)}$ перед выполнением вычислений k-го шага по формуле (2.2.2). Эти перестановки строк и столбцов запоминаются для дальнейшего использования при нахождении решения.

Во многих практических приложениях, в которых встречаются большие разреженные матрицы, вместо частичного или полного упорядочения обычно достаточно убедиться в том, что все главные элементы больше некоторого числа ε, называемого *допустимым значением главного элемента* (Тьюарсон (1969а)). Это в особенности справедливо для линейного программирования (Класен (1966)), в котором матрицы, как правило, больших размеров и разрежены. При значениях главных элементов больше допустимого ε устраняется возможность выбора главного элемента с таким малым значением, которое привело бы к затруднениям в вычислениях из-за ошибок округления. На практике было найдено, что для большинства задач, включающих большие разреженные матрицы, значение $\varepsilon = 10^{-3}$ является удовлетворительным, если при вычислениях в памяти хранятся 9 или 10 десятичных разрядов ненулевых элементов. Конечно, для полных (неразреженных) матриц необходимо произ-

водить частичное или полное упорядочение. Если некоторые элементы становятся в процессе вычислений очень малыми (менее так называемого *критического значения*), то рекомендуется приравнивать их нулю (Класен (1966)). Вулф (1965) предлагал принимать критическое значение равным 10^{-7}.

2.4. Элиминативная форма обратной матрицы

Разложение матрицы A^{-1} на множители может быть получено из разд. 2.2 следующим образом. Обычное решение уравнения (2.2.1) для произвольного b дается в виде $x = A^{-1}b$. Поэтому, сравнивая это с решением, которое дается формулой (2.2.13), мы заключаем, что

$$A^{-1} = U_2 \ldots U_{n-1}U_nL_n \ldots L_2L_1. \qquad (2.4.1)$$

Такое представление матрицы A^{-1} называется *элиминативной формой обратной матрицы* и обозначается через EFI. Таким образом, матрица A^{-1} выражается в виде произведения n нижних и $n-1$ верхних треугольных матриц.

Одним из главных преимуществ формы EFI является легкость, с которой она может быть слева или справа умножена на данный вектор, как это показано ниже.

Пусть ρ_i обозначает i-й элемент вектора-столбца ρ. Тогда из формул (2.2.10)и (2.2.11) следует, что (см. также рис. 2.2.3)

$$\begin{aligned} e_i'(U_k\rho) &= \rho_i + \xi_i^{(k)}\rho_k, & i < k, \\ e_i'(U_k\rho) &= \rho_i, & i \geqslant k. \end{aligned} \qquad (2.4.2)$$

Таким образом, умножение вектора-столбца слева на матрицу U_k эквивалентно добавлению к данному вектору-столбцу его k-го элемента, умноженного на вектор-столбец $\xi^{(k)}$. Имеем также

$$\begin{aligned} (\rho'U_k)\,e_i &= \rho_i, & i \neq k, \\ (\rho'U_k)\,e_k &= \rho'\xi^{(k)} + \rho_k, \end{aligned} \qquad (2.4.3)$$

и, следовательно, умножение вектора-строки ρ' справа на матрицу U_k оставляет все элементы ρ' без изменений, за исключением k-го элемента, к которому добавляется скалярное произведение векторов ρ' и $\xi^{(k)}$. Из формул (2.2.3) и (2.2.4) имеем (см. также рис. 2.2.2)

$$\begin{aligned} e_i'(L_k\rho) &= \rho_i, & i < k, \\ e_k'(L_k\rho) &= \eta_k^{(k)}\rho_k, & \\ e_i'(L_k\rho) &= \eta_i^{(k)}\rho_k + \rho_i, & i > k. \end{aligned} \tag{2.4.4}$$

Таким образом, произведение $L_k\rho$ может быть получено, исходя из вектора ρ, посредством замены его k-го элемента нулем и добавления к нему вектора-столбца $\rho_k\eta^{(k)}$. Также имеем

$$\begin{aligned} (\rho'L_k)\,e_i &= \rho_i', \quad i \neq k, \\ (\rho'L_k)\,e_k &= \rho'\eta^{(k)}. \end{aligned} \tag{2.4.5}$$

Следовательно, умножение вектора-строки ρ' справа на матрицу L_k оставляет все его элементы без изменения, за исключением k-го элемента, который заменяется скалярным произведением векторов ρ' и $\eta^{(k)}$. Вычисление произведения $A^{-1}\pi$, где π есть вектор-столбец, а матрица A^{-1} задана в элиминативной форме EFI выражением (2.4.1), может быть организовано таким образом, чтобы на первых n шагах применялись формулы (2.4.4), а на последних $n-1$ шагах применялись формулы (2.4.2). Вектор-столбец, образованный на очередном шаге, используется на следующем шаге таким образом:

$$A^{-1}\pi = U_2(\ldots(U_n(L_n\ldots(L_2(L_1\pi))\ldots).$$

Подобным же образом может быть вычислено произведение $\pi'A^{-1}$ с помощью формул (2.4.3) и (2.4.5). Вектор-строка, образованная на очередном шаге, используется на следующем шаге таким образом:

$$\pi'A^{-1} = (\ldots((\pi'U_2)U_3)\ldots U_n)L_n)\ldots)L_1.$$

Теперь, в силу соотношений (2.2.3), (2.2.4), (2.2.10), (2.2.11) и (2.4.1), очевидно, что для вычисления EFI

требуются только ненулевые элементы векторов-столбцов $\eta^{(k)}$ и $\xi^{(k)}$. Поэтому должны храниться только эти элементы (вместе с соответствующей информацией об их местонахождении в памяти). Для большой разреженной матрицы особенно требуется экономия в объеме памяти, необходимой для их хранения, поэтому важно определить элиминативную форму обратной матрицы EFI так, чтобы этот объем был минимальным. Другими словами, необходимо минимизировать количество ненулевых элементов всех векторов-столбцов $\eta^{(k)}$ и $\xi^{(k)}$. Покажем теперь, каким образом можно это осуществить.

2.5. Минимизация общего числа ненулевых элементов в EFI

Из разд. 2.2 мы вспоминаем, что на каждом шаге гауссова исключения умноженная на различные коэффициенты строка вычитается из ряда других строк матрицы. Это обычно приводит к образованию новых ненулевых элементов вместо нулевых. Например, если к началу k-го шага ненулевыми являются элементы $a_{kk}^{(k)}$, $a_{kj}^{(k)}$ и $a_{ik}^{(k)}$, а $a_{ij}^{(k)}=0$, где $i, j>k$, то из формул (2.2.2), (2.2.3) и (2.2.4) следует, что в конце k-го шага

$$a_{ij}^{(k+1)}=-\frac{a_{ik}^{(k)}a_{kj}^{(k)}}{a_{kk}^{(k)}}. \qquad (2.5.1)$$

Ясно, что элемент $a_{ij}^{(k+1)}$ ненулевой. Таким образом, при упомянутых выше условиях нуль в (i, j)-й позиции матрицы $A^{(k)}$ превратился в ненулевой элемент матрицы $A^{(k+1)}$. Общее число всех таких элементов, которые были нулями в матрице $A^{(k)}$ и приняли ненулевые значения в матрице $A^{(k+1)}$, называется *локальным заполнением*.

Если вместо элемента $a_{kk}^{(k)}$ в качестве главного на k-м шаге мы выбираем другой ненулевой элемент $a_{st}^{(k)}$, $s\geqslant k$, $t\geqslant k$, необходимо переставить k-ю и s-ю строки так же, как k-й и t-й столбцы в матрице $A^{(k)}$, прежде чем вычислять $A^{(k+1)}$ по формуле (2.2.2). Тогда

вместо формулы (2.2.2) имеем

$$A^{(k+1)} = L_k \hat{A}^{(k)}, \quad k = 1, 2, \ldots, n, \tag{2.5.2}$$

где

$$\hat{A}^{(k)} = P_k A^{(k)} Q_k, \tag{2.5.3}$$

а матрица P_k и матрица Q_k получены из единичной матрицы I_n путем перестановки соответственно ее k-й и s-й строк и k-го и t-го столбцов. Все матрицы L_k также выражаются формулой (2.2.3), но элементы векторов-столбцов $\eta^{(k)}$ имеют теперь вид

$$\eta_i^{(k)} = 0, \quad i < k,$$
$$\eta_k^{(k)} = \frac{1}{\hat{a}_{kk}^{(k)}}, \quad \eta_i^{(k)} = -\frac{\hat{a}_{ik}^{(k)}}{\hat{a}_{kk}^{(k)}}, \quad i > k, \tag{2.5.4}$$

где $\hat{a}_{ij}^{(k)}$ — (i, j)-й элемент матрицы $\hat{A}^{(k)}$.

Принимая во внимание последний абзац разд. 2.3, вспомним, что $|a_{st}^{(k)}| > \varepsilon$, где ε — допустимое значение главного элемента. Следовательно, из всех наличных кандидатов на роль главного элемента, а именно для всех элементов, для которых $|a_{ij}^{(k)}| > \varepsilon$ при $i \geqslant k$, $j \geqslant k$, необходимо выбрать такой, который обеспечил бы наименьшее локальное заполнение. Это может быть сделано следующим образом.

Определение. Обозначим через B_k матрицу, полученную из последних $n - k + 1$ строк и столбцов матрицы $A^{(k)}$ путем замены ненулевых элементов единицами.

Следующая теорема (Тьюарсон (1967 б)) может быть использована для определения главного элемента, обеспечивающего наименьшее заполнение.

Теорема 2.5.5. *Если $a_{i+k-1,\, j+k-1}^{(k)}$ выбран главным элементом k-го шага прямого гауссова исключения, то локальное заполнение дается (i, j)-м элементом матрицы G_k, равной*

$$G_k = B_k \bar{B}_k' B_k, \tag{2.5.6}$$

где $\overline{B}'_k$ — транспонированная к матрице, полученной из матрицы B путем замены всех ее нулевых элементов единицами, а всех единиц нулями.

Доказательство. Если в матрице $A^{(k)}$ $(p+k-1, q+k-1)$-й элемент равен нулю, но и $(i+k-1, q+k-1)$-й элемент и $(p+k-1, j+k-1)$-й элемент ненулевые, то из соотношений (2.5.2), (2.2.3), (2.5.3) и (2.5.4) следует, что $(p+k-1, q+k-1)$-й элемент в $A^{(k+1)}$ будет ненулевым. Это равносильно утверждению, что если

$$b_{pq}^{(k)}=0 \quad \text{и} \quad b_{iq}^{(k)}=b_{pj}^{(k)}=1,$$

где $b_{pq}^{(k)}$ есть (p, q)-й элемент матрицы B_k, то создается новый ненулевой элемент. Если через $g_{ij}^{(k)}$ обозначить общее число таких вновь созданных ненулевых элементов (локальное заполнение) на k-м шаге гауссова исключения, то

$$g_{ij}^{(k)}=\sum_p \sum_q b_{iq}^{(k)}\left(1-b_{pq}^{(k)}\right) b_{pj}^{(k)}, \quad p \neq i, \quad q \neq j.$$

Или

$$g_{ij}^{(k)}=\sum_p \sum_q e_i' B_k e_q \left(1-e_p' B_k e_q\right) e_p' B_k e_j, \qquad (2.5.7)$$

где ограничения $p \neq i$, $q \neq j$ опущены, так как для $p=i$ или $b_{iq}^{(k)}=0$, или $1-b_{pq}^{(k)}=0$ и для $q=j$ или $1-b_{pq}^{(k)}=0$, или $b_{pj}^{(k)}=0$. Теперь, если M является матрицей $(n-k+1)$-го порядка со всеми элементами, равными единице, то

$$1-e_p' B_k e_q = e_p' M e_q - e_p' B_k e_q = e_p'\left(M-B_k\right) e_q = \\ = e_p' \overline{B}_k e_q = e_q' \overline{B}_k' e_p, \qquad (2.5.8)$$

где последний знак равенства следует из того, что величина $e_p' \overline{B}_k e_q$ является скалярной. Из выражений (2.5.7) и (2.5.8) имеем

$$g_{ij}^{k}=\sum_p \sum_q e_i' B_k e_q e_q' \overline{B}_k' e_p e_p' B_k e_j = \\ = e_i' B_k \sum_q e_q e_q' \overline{B}_k' \sum_p e_p e_p' B_k e_j = e_i' B_k \overline{B}_k' B_k e_j,$$

так как $\sum_q e_q e_q' = \sum_p e_p e_p' = I_{n-k+1}$. Этим завершается доказательство теоремы.

Теперь мы, наконец, в состоянии выбрать главный элемент, который обеспечит наименьшее локальное заполнение. Это может быть осуществлено на основании приводимого ниже следствия, доказательство которого мы опускаем, так как оно непосредственно вытекает из теоремы 2.5.5.

Следствие 2.5.9. *Локальное заполнение будет минимальным, если на k-м шаге гауссова исключения в качестве главного выбрать элемент* $a_{st}^{(k)}$, *где* $s = \alpha + k - 1$, $t = \beta + k - 1$ *и* α, β *даны формулой*

$$g_{\alpha\beta}^{(k)} = \min_{i,j} e_i' G_k e_j \quad \text{для всех} \quad |a_{i+k-1,\, j+k-1}^{(k)}| > \varepsilon \tag{2.5.10}$$

(ε — *некоторое подходящим образом выбранное допустимое значение главного элемента*).

Принимая во внимание формулу (2.2.11), получим все ненулевые элементы $\xi_i^{(k)}$ путем изменения знаков у ненулевых наддиагональных элементов матрицы U. Поэтому из формул (2.2.9) и (2.2.10) следует, что обратная подстановка в методе Гаусса не порождает новых ненулевых элементов. Таким образом, новые ненулевые элементы создаются только при прямом исключении. В конце k-го шага последние $n-k$ строк и столбцов матрицы $A^{(k+1)}$ содержат, вообще говоря, некоторое число ненулевых элементов там, где в матрице $A^{(k)}$ соответствующие элементы были нулями. Так как такие строки и столбцы используются на последующих шагах для вычисления векторов $\eta^{(k)}$ и $\xi^{(k)}$, то минимизация локального заполнения будет минимизировать и число ненулевых элементов векторов $\eta^{(k)}$ и $\xi^{(k)}$ при условии, что достижение локальных минимумов приводит к глобальному минимуму. Это условие может и выполняться для некоторых матриц, но не является обязательным для произвольных разреженных матриц. Во всяком случае, минимизация локального роста таких ненулевых элементов с помощью следствия 2.5.9 все же приводит

к существенному уменьшению числа ненулевых элементов во всех векторах $\xi^{(k)}$ и $\eta^{(k)}$.

Выбор главных элементов $a_{st}^{(k)}$ для обеспечения минимума локального заполнения не влечет за собой особых сложностей. Необходимые изменения в формулах могут быть описаны следующим образом. Из формул (2.5.2) и (2.5.3) имеем

$$A^{(n+1)} = L_n P_n \dots L_2 P_2 L_1 P_1 A Q_1 Q_2 \dots Q_n, \quad (2.5.11)$$

и если положить

$$\hat{L} = L_n P_n \dots L_2 P_2 L_1 P_1,$$

$$Q_1 Q_2 \dots Q_n = Q \quad \text{и} \quad A^{(n+1)} = \hat{U}, \quad (2.5.12)$$

получим

$$A^{-1} = Q\hat{U}^{-1}\hat{L}. \quad (2.5.13)$$

Матрицы перестановок Q и P_k в совокупности требуют для своего хранения объем памяти порядка n ячеек (а не n^2), так как в каждом случае необходимо запомнить только позиции нетривиальных элементов.

Более простой, хотя и менее точный метод нахождения главного элемента, при котором локальное заполнение было бы малым, основан на следующей теореме (Маркович (1957)).

Теорема 2.5.14. *Если на k-м шаге гауссова исключения в качестве главного выбран элемент $a_{i+k-1,\,j+k-1}^{(k)}$, то максимальное возможное заполнение (не обязательно совпадающее с действительным заполнением) дается (i, j)-м элементом матрицы G_k, причем*

$$\hat{G}_k = (B_k - I_{n-k+1})\, M\, (B_k - I_{n-k+1}), \quad (2.5.15)$$

где M — матрица, все элементы которой единицы.

Доказательство. Если обозначить через $\hat{g}_{ij}^{(k)}$ максимальное возможное заполнение на k-м шаге, то так же, как и при доказательстве теоремы 2.5.5, имеем

$$\hat{g}_{ij}^{(k)} = \sum_p \sum_q b_{iq}^{(k)} b_{pj}^{(k)}, \quad p \neq i \quad \text{и} \quad q \neq j.$$

Или

$$\hat{g}_{ij}^{(k)} = \Big(\sum_q b_{iq}^{(k)} - b_{ij}^{(k)}\Big)\Big(\sum_p b_{pj}^{(k)} - b_{ij}^{(k)}\Big) =$$
$$= \Big(\sum_q b_{iq}^{(k)} - 1\Big)\Big(\sum_p b_{pj}^{(k)} - 1\Big), \quad \text{так как} \quad b_{ij}^{(k)} = 1.$$

Отсюда

$$\hat{g}_{ij}^{(k)} = \Big(e_i' B_k \sum_q e_q - 1\Big)\Big(\sum_p e_p' B_k e_j - 1\Big) =$$
$$= (e_i' B_k V_k - e_i' V_k)(V_k' B_k e_j - V_k' e_j), \qquad (2.5.16)$$

где V_k — $(n-k+1)$-мерный вектор-столбец с единичными элементами.

Таким образом,

$$\hat{g}_{ij}^{(k)} = e_i'(B_k - I_{n-k+1}) V_k V_k' (B_k - I_{n-k+1}) e_j =$$
$$= e_i'(B_k - I_{n-k+1}) M (B_k - I_{n-k+1}) e_j,$$

так как $V_k V_k' = M$. Этим завершается доказательство теоремы.

Для того чтобы воспользоваться приведенной выше теоремой, будем выбирать на k-м шаге главный элемент по следующей формуле (вместо формулы 2.5.10):

$$\hat{g}_{\alpha\beta}^{(k)} = \min_{i,\,j} \hat{g}_{ij}^{(k)} \quad \text{для всех} \quad |a_{i+k-1,\,j+k-1}^{(k)}| > \varepsilon, \qquad (2.5.17)$$

где, как и прежде, $\alpha + k - 1 = s$ и $\beta + k - 1 = t$. Заметим, что выбор главного элемента $a_{st}^{(k)}$ в соответствии с формулой (2.5.17) не обязательно приводит к наименьшему локальному заполнению.

Интересно применение изложенного выше метода выбора главного элемента в случае, когда $B_k = B_k'$ и только диагональные элементы выбираются в качестве главных.

Из формулы (2.5.16) в этом случае следует, что

$$\hat{g}_{ii}^{(k)} = (e_i' B_k V_k - 1)(V_k' B_k e_i - 1) =$$
$$= (e_i' B_k V_k - 1)^2, \quad \text{так как} \quad B_k' = B_k.$$

Поэтому

$$\hat{g}_{\alpha\alpha}^{(k)} = \min_i (e_i' B_k V_k - 1)^2.$$

Но индекс α будет тем же для $\min_i (e_i' B_k V_k - 1)$, что и для $\min_i e_i' B_k V_k$, так как $e_i' B_k V_k \geqslant 1$. Другими словами, для выбора главного среди диагональных элементов применяется формула

$$\min_i e_i' B_k V_k = e_\alpha' B_k V_k \qquad (2.5.18)$$

при

$$|a_{\alpha+k-1,\,\alpha+k-1}^{(k)}| > \varepsilon.$$

Заметим, что $e_i' B_k V_k$ есть общее число ненулевых элементов $(i+k-1)$-й строки матрицы $A^{(k)}$. Таким образом, на каждом шаге выбирается строка (и соответствующий столбец), которая содержит наименьшее число ненулевых элементов. Это относительно просто сделать и поэтому рекомендуется во многих практических приложениях (Тинни и Уокер (1967), Спиллерс и Хикерсон (1968), Черчилл (1971)). Одно из главных оснований для выбора главного элемента только среди диагональных элементов заключается в следующем. Если матрица A симметричная, то очень часто хранится в памяти только ее верхняя треугольная часть вместе с главной диагональю, и выбор главного среди диагональных элементов при прямом гауссовом исключении не нарушает симметрии. Кроме того, все векторы $\eta^{(k)}$ могут быть легко получены из верхней треугольной матрицы в конце прямого исключения. Формулируем эти положения в виде теоремы.

Теорема 2.5.19. *Если матрица A — симметричная и только диагональные элементы выбираются в качестве главных, то*

а) *матрица, состоящая из последних $n-k$ строк и столбцов матрицы $A^{(k+1)}$ в формуле* (2.5.2) *при $k = 1, 2, \ldots, n-1$, тоже симметричная,*

б) *элементы* $\eta_i^{(k)}$ *при* $i > k$ *в уравнении* (2.5.4) *определяются равенством*

$$\eta_i^{(k)} = -a_{ki}^{(k+1)} = \xi_k^{(i)}, \quad i > k. \tag{2.5.20}$$

Доказательство. Если мы сможем показать, что $a_{ij}^{(k+1)} = a_{ji}^{(k+1)}$ для $i, j > k$ всякий раз, когда $a_{ij}^{(k)} = a_{ji}^{(k)}$ для $i, j \geqslant k$, то посредством индукции по k и из равенства $a_{ji}^{(1)} = a_{ij}^{(1)}$ очевидным образом следует часть (а) теоремы. Так как только диагональные элементы выбираются в качестве главных, то в равенстве 2.5.3 очевидным образом следует $Q_k = P'_k$ и $\hat{a}_{ij}^{(k)} = \hat{a}_{ji}^{(k)}$ для $i, j \geqslant k$. Теперь из формул (2.5.2), (2.2.3) и (2.5.4) при $i, j > k$ имеем

$$a_{ij}^{(k+1)} = \hat{a}_{ij}^{(k)} - \frac{\hat{a}_{ik}^{(k)}\hat{a}_{kj}^{(k)}}{\hat{a}_{kk}^{(k)}},$$

$$a_{ji}^{(k+1)} = \hat{a}_{ji}^{(k)} - \frac{\hat{a}_{jk}^{(k)}\hat{a}_{ki}^{(k)}}{\hat{a}_{kk}^{(k)}}.$$

Отсюда следует

$$a_{ij}^{(k+1)} = a_{ji}^{(k+1)}, \quad i, j > k,$$

так как

$$\hat{a}_{ij}^{(k)} = \hat{a}_{ji}^{(k)}, \quad i, j \geqslant k.$$

Этим завершается доказательство части (а) теоремы.

Теперь из формул (2.5.2), (2.2.3) и (2.5.4) для $i > k$ следует

$$a_{ki}^{(k+1)} = \frac{\hat{a}_{ki}^{(k)}}{\hat{a}_{kk}^{(k)}} \quad \text{и} \quad \eta_i^{(k)} = -\frac{\hat{a}_{ik}^{(k)}}{\hat{a}_{kk}^{(k)}},$$

и, имея в виду формулу (2.2.11) и условие $\hat{a}_{ik}^{(k)} = \hat{a}_{ki}^{(k)}$, получим равенство (2.5.20), что завершает доказательство теоремы.

Закончим этот раздел несколькими замечаниями.

Если в формуле (2.5.10) минимум достигается более чем для одной пары значений (i, j), то следует

выбрать пару, для которой $\hat{g}_{ij}^{(k)}$, вычисленное по формуле (2.5.16), имеет наибольшее значение. Из рассмотрения на следующем шаге, таким образом, будет исключено максимальное число ненулевых элементов.

Вместо (2.5.15) иногда используют матрицу

$$\tilde{G}_k = B_k M B_k \tag{2.5.21}$$

для выбора главного элемента на k-м шаге гауссова исключения. Покажем сейчас, что если на k-м шаге в качестве главного выбран элемент $a_{i+k-1,\,j+k-1}^{(k)}$, то $e_i' \tilde{G}_k e_j$ есть общее число умножений и делений. Из доказательства теоремы 2.5.14 и соотношений (2.5.2), (2.5.3) и (2.5.4) очевидно, что одно деление требуется для вычисления $1/a_{i+k-1,\,j+k-1}^{(k)}$ и $V_k' B_k e_j - 1$ и $e_i' B_k V_k - 1$ умножений нужно для вычисления соответственно $\eta^{(k)}$ и $e_i' A^{(k+1)}$. Кроме того, для исключения $a_{p+k-1,\,j+k-1}^{(k)} \neq 0$, $p \neq i$, требуется общее число в $(e_i' B_k V_k - 1)(V_k' B_k e_j - 1)$ умножений. Таким образом, на k-м шаге гауссова исключения общее число делений и умножений равно

$$1 + (V_k' B_k e_j - 1) + (e_i' B_k V_k - 1) + \\ + (e_i' B_k V_k - 1)(V_k' B_k e_j - 1) = e_i' B_k V_k V_k' B_k e_j = \\ = e_i' B_k M B_k e_j = e_i' \tilde{G}_k e_j.$$

В свете изложенного, если для выбора главного элемента на k-м шаге вместо формулы (2.5.17) мы пользуемся формулой

$$\tilde{g}_{\alpha\beta}^{(k)} = \min_{i,\,j} e_i' \tilde{G}_k e_j \quad \text{для всех} \quad |a_{i+k-1,\,j+k-1}^{(k)}| > \varepsilon, \tag{2.5.22}$$

то минимизируется общее число умножений и делений.

Пусть в качестве меры вычислительных затрат для каждого шага k взято общее число делений и умножений на этом шаге. Тогда для минимизации как заполнения, так и вычислительных затрат следует

пользоваться взвешенным средним значением G_k и $\tilde{G}_k$ при выборе главного элемента. Весовые коэффициенты определят относительную важность обоих критериев. Из формул (2.5.21) и (2.5.6) видно, что взвешенное среднее G_k и $\tilde{G}_k$ есть матрица $B_k\left(M-\delta B_k'\right)B_k$, где $0 \leqslant \delta \leqslant 1$, а δ и $1-\delta$ соответственно весовые коэффициенты. Значение δ зависит от характеристик вычислительной машины и также от ее математического обеспечения. Следует заметить, что для больших разреженных матриц минимизация локального заполнения более важна, чем минимизация локальных вычислительных затрат, потому что первая, сохраняя разреженность матрицы B_k, приводит к минимизации вычислений на следующих шагах.

Если систему уравнений (2.2.1) требуется решить лишь для небольшого числа правых частей, то матрицы L_k, определенные формулами (2.2.3) и (2.5.4), не запоминаются (правые части преобразуются на каждом шаге). Тогда, имея в виду тот факт, что на k-м шаге последние $n-k$ элементов k-го столбца матрицы $\hat{A}^{(k)}$ обращаются в нуль, увеличение числа ненулевых элементов в матрице $A^{(k+1)}$ по сравнению с матрицей $\hat{A}^{(k)}$ можно выразить разностью

$$g_{ij}^{(k)}-\left(V_k'B_k\hat{e}_j-1\right)=$$

$$=e_i'B_k\bar{B}_k'B_ke_j-e_i'MB_ke_j+1=e_i'\left(B_k\bar{B}_k'-M\right)B_ke_j+1.$$

Минимальное значение этого выражения может быть использовано для выбора главного элемента.

2.6. Хранение и использование элиминативной формы обратной матрицы

Все $\eta^{(k)}$, необходимые для элиминативной формы обратной матрицы EFI, хранятся следующим образом. На k-м шаге прямого гауссова исключения все $\hat{a}_{ik}^{(k)} \neq 0$, $i>k$, преобразуются в нуль, а $\hat{a}_k^{(k)}$ преобразуется в единицу. Это значит, что $a_{ik}^{(k+1)}=0$, $i>k$, и $a_{kk}^{(k+1)}=1$. Поэтому, как это ясно из формул

(2.5.4), каждый элемент $\eta_i^{(k)} \neq 0$, $i > k$, может храниться на месте соответствующего элемента $\hat{a}_{ik}^{(k)} \neq 0$, $i > k$. Так же и элемент $\eta_k^{(k)}$ может храниться вместо элемента $a_{kk}^{(k+1)}$, так как нет необходимости хранить значение $a_{kk}^{(k+1)} = 1$ (это относится ко всем k; другими словами, диагональные элементы матрицы U все равны единице).

Матрицы перестановок P_k и Q_k в выражении (2.5.3) могут быть легко построены, если s и t известны. Поэтому для каждого k требуется всего две ячейки для хранения соответствующих матриц P_k и Q_k, что составляет общее число в $2n$ ячеек для всех P_k и Q_k, которые требуются в выражениях (2.5.12).

Из формул (2.2.12), (2.2.10) и (2.2.11) видно, что для вычисления всех U_k требуются только ненулевые элементы всех $\xi^{(k)}$ и что, кроме того, эти элементы могут быть получены путем изменения знаков у тех ненулевых элементов матрицы U, которые лежат над диагональю. Поэтому ненулевые элементы всех $\xi^{(k)}$ могут храниться в области памяти, занятой ненулевыми элементами матрицы U.

Ненулевые элементы каждой матрицы $A^{(k)}$ и соответствующих векторов $\eta^{(k)}$ и $\xi^{(k)}$ хранятся, конечно, в одной из упакованных форм, описанных в разд. 1.3. На каждом шаге в матрице $A^{(k)}$ создаются новые ненулевые элементы, и связные списки особенно пригодны для хранения таких элементов (Огбуобири (1970)).

Заключим этот раздел несколькими примерами, в которых дополнительная работа, связанная с получением разреженной элиминативной формы обратной матрицы, является оправданной.

Во многих практических приложениях система уравнений (2.2.1) должна решаться многократно при различных значениях правых частей и (или) коэффициентов уравнений, но при сохранении структуры разреженности матрицы коэффициентов, т. е. при одном и том же расположении нулевых и ненулевых элементов в ней. Например, стандартный метод Ньютона для решения нелинейных уравнений приводит к

системе (2.2.1), где матрица A имеет фиксированную структуру разреженности, а b изменяется от случая к случаю (Лайниджер и Уиллогби (1969), Черчилл (1971)). В структурном анализе решение системы (2.2.1) требуется для многих правых частей (Олвуд (1971)). В упомянутых выше случаях наиболее пригодна форма EF1. Поэтому стоимость рассмотренного в разд. 2.5 исследования для минимизации числа ненулевых элементов EFI должна быть разложена на все повторные решения системы (2.2.1). Применение EFI для решения задач энергетических систем приводило на практике к выигрышу в скорости, памяти и точности, который приблизительно пропорционален степени разреженности (Тинни (1969)).

2.7. Библиография и комментарии

Основы метода исключения Гаусса и ошибок округления рассмотрены в трудах Фокса (1965), Уилкинсона (1965) и Форсайта и Молера (1967).

EFI и способы сохранения ее разреженности путем минимизации локального заполнения были предметом, которому уделяли большое внимание; см., например, Маркович (1957), Данциг (1963 б), Карпентьер (1963), Эдельман (1963), Сато и Тинни (1963), Тьюарсон (1967 б), Спиллерс и Хикерсон (1968), Брейтон и др. (1969), Огбуобири (1970), Томлин (1970), Берри (1971), Бертеле и Бриоши (1971), Форрест и Томлин (1972) и несколько статей в трудах конференций под редакцией Уиллогби (1969) и Рейда (1971).

Впервые EFI была предложена Марковичем (1957) и позже — Данцигом (1963 б). Во многих практических приложениях применение методов минимизации заполнения не создает трудностей, связанных с ошибками округления. Например, Черчилл (1971) отмечает, что, когда применялись методы минимизации заполнения при решении сложных задач по расходу энергии, не приходилось сталкиваться с трудностями из-за ошибок округления.

Карре (1971) показал, что можно пользоваться способами минимизации заполнения и при решении задач сетевых потоков минимальной стоимости.

В гл. 3 обсуждаются дополнительные методы создания разреженных форм EFI. Некоторые из них требуют предварительной перестановки строк и столбцов матрицы A для приведения ее к форме, при которой заполнение ограничено только определенными областями этой матрицы.

Глава 3

ДОПОЛНИТЕЛЬНЫЕ МЕТОДЫ МИНИМИЗАЦИИ ПАМЯТИ ДЛЯ ХРАНЕНИЯ EFI

3.1. Введение

В этой главе нас будут интересовать главным образом методы, обеспечивающие небольшое заполнение при прямом гауссовом исключении и не требующие в то же время больших затрат. Эти методы обычно именуются «априорными» методами, так как информация для выбора главного элемента на каждом шаге преимущественно получается из исходной матрицы, а не из преобразованных форм на каждом шаге. Они обычно приводят к значительной экономии времени и усилий. Однако такие априорные методы, вообще говоря, менее эффективны для минимизации локального заполнения, чем методы, приведенные в гл. 2. Методы этой главы преследуют цель получения разреженной EFI.

Они могут быть разбиты на две категории: первая включает методы, в которых предполагается главным образом априорное упорядочение столбцов, вторая содержит методы, состоящие из таких априорных перестановок строк и столбцов, которые преобразуют матрицу A к различным формам, удобным для гауссова исключения. При прямом гауссовом исключении такие формы или обеспечивают отсутствие заполнения или ограничивают заполнение некоторыми известными областями матрицы.

3.2. Методы, основанные на априорных перестановках столбцов

Наша цель состоит в том, чтобы определить одну матрицу перестановок Q прежде, чем приступить к гауссовому исключению, а другую матрицу перестановок P в процессе исключения; причем они должны

удовлетворять условию

$$PAQ = \hat{A}, \tag{3.2.1}$$

где матрица $\hat{A}$ имеет диагональные элементы, которые обеспечат наименьшее значение общего заполнения, если их, начиная с верхнего левого угла, последовательно брать в качестве главных элементов. Желательно, конечно, минимизировать затраты усилий на определение матриц P и Q. Опишем некоторые априорные методы нахождения аппроксимации для Q, затем изложим метод аппроксимации для P, основанный частично на информации, полученной при каждом шаге прямого гауссова исключения.

Из формулы (3.2.1) ясно, что матрица Q определяет порядок, в котором столбцы матрицы A следуют друг за другом при выборе главного элемента. Таким образом, определение матрицы Q эквивалентно предварительному упорядочению столбцов матрицы A. Опишем три метода упорядочения столбцов матрицы A, при котором заполнение будет приемлемо малым.

Определение. Будем обозначать через r_i^k и c_j^k соответственно общее число ненулевых элементов в i-й строке и j-м столбце матрицы B_k, определенной в разд. 2.5.

Из этого определения следует, что

$$r_i^{(k)} = e_i' B_k V_k \quad \text{и} \quad c_j^{(k)} = V_k' B_k e_j, \tag{3.2.2}$$

где V_k — $(n-k+1)$-мерный вектор-столбец, все элементы которого единицы, и e_i — i-й столбец единичной матрицы $(n-k+1)$-го порядка. Теперь из формул (2.5.16) и (3.2.2) имеем

$$\hat{g}_{ij}^{(k)} = \left(r_i^{(k)} - 1\right)\left(c_j^{(k)} - 1\right). \tag{3.2.3}$$

Поэтому для данного j-го столбца минимальное значение $\hat{g}_{ij}^{(k)}$ будет

$$\gamma_j^{(k)} = \min_i \hat{g}_{ij}^{(k)} = \left(c_j^{(k)} - 1\right) \min_i \left(r_i^{(k)} - 1\right)$$

для всех i, для которых $b_{ij}^{(k)} = 1$, или

$$\gamma_j^{(k)} = \left(c_i^{(k)} - 1\right)\left(r_\alpha^{(k)} - 1\right), \quad \text{причем} \quad b_{\alpha j}^{(k)} = 1. \tag{3.2.4}$$

В силу теоремы 2.5.14 очевидно, что из всех элементов $(j+k-1)$-го столбца матрицы $A^{(k)}$, которые могли бы играть роль главных, тот, что находится в $(\alpha+k-1)$-й строке, обеспечивает наименьшее значение максимума локального заполнения. Следовательно, чтобы заполнение было малым, необходимо априорно упорядочить столбцы матрицы A по возрастающим значениям всех $\gamma_j^{(1)}$. Такое упорядочение столбцов матрицы ведет к последовательному ухудшению оценок максимумов локальных заполнений при изменении k от единицы до n. Это следует из того, что для данного столбца может оказаться невозможным выбрать главный элемент в той строке, которая требуется из условия наименьшего значения максимума локального заполнения, если ранее эта строка была использована при выборе главного элемента для других столбцов.

Для получения одного из разложений на множители обратной матрицы A^{-1}, которое рассматривается в гл. 5, Орчард-Хейс (1968) рекомендует пользоваться набором $\gamma_j^{(1)}$ для выбора подмножества столбцов матрицы A и выполнить для этих столбцов прямое гауссово исключение. После того как исключение произведено для текущего подмножества столбцов, следующее подмножество выбирается исходя из набора $\gamma_j^{(p)}$, где p — общее число столбцов матрицы A, для которого осуществлено прямое гауссово исключение.

Второй метод упорядочения столбцов, оказавшийся полезным на практике, заключается в следующем (Тьюарсон (1967 б)). В j-м столбце имеется $c_j^{(k)}$ ненулевых элементов, и каждый из них является потенциальным главным элементом. Поэтому, учитывая равенство (3.2.3), среднее значение максимума локального заполнения (если каждый ненулевой элемент столбца имеет ту же вероятность стать главным, что и другие) имеет вид

$$\lambda_j^{(k)} = \frac{\sum_i (r_i^{(k)} - 1)(c_j^{(k)} - 1)}{c_j^{(k)}}$$

(для всех i, при которых $b_{ij}^{(k)} = 1$), или

$$\lambda_j^{(k)} = \left(d_j^{(k)} - c_j^{(k)}\right)\left(1 - \frac{1}{c_j^{(k)}}\right), \tag{3.2.5}$$

где

$$d_j^{(k)} = \sum_i r_i^{(k)} \tag{3.2.6}$$

для всех значений i, при которых $b_{ij}^{(k)} = 1$. Поэтому если столбцы матрицы A упорядочить по возрастающим значениям всех $\lambda_j^{(k)}$, то, очевидно, заполнение будет сохраняться малым.

Третий метод, очень простой, но практически намного менее точный, состоит в упорядочении столбцов матрицы A по возрастающим значениям $c_j^{(1)}$. Практика показала, однако, что использование $\lambda_j^{(k)}$ или $\gamma_j^{(k)}$ приводит к значительно лучшим результатам лишь при небольшом увеличении затрат труда на начальном этапе вычислений (Тьюарсон, 1967 б; Орчард-Хейс, 1968).

Можно также выразить $\gamma_j^{(k)}$ в формуле (3.2.4) и $\lambda_j^{(k)}$ в формуле (3.2.5) следующим образом. Из теоремы 2.5.14 имеем

$$\gamma_j^{(k)} = \min_i e_i' \hat{G}_k e_j \tag{3.2.7}$$

для всех значений i, для которых $e_i' B_k e_j = 1$. Так же, учитывая формулу (3.2.2), имеем

$$\lambda_j^{(k)} = \frac{\sum_i e_i' \hat{G}_k e_j}{V_k' B_k e_j} \quad \text{для всех} \quad e_i' B_k e_j = 1,$$

или

$$\lambda_j^{(k)} = \frac{e_j' B_\kappa' \hat{G}_k e_j}{V_k' B_\kappa e_j}, \tag{3.2.8}$$

так как $\sum_i e_i'$ для всех значений i, для которых $e_i' B_k e_j = 1$, идентично выражению

$$\sum_i e_j' B_k' e_i e_i' = e_j' B_k' \sum_i e_i e_i' = e_j' B_k' I_{n-k+1} = e_j' B_k'.$$

Перегруппировка столбцов матрицы A с использованием $c_j^{(1)}$, $\gamma_j^{(1)}$ или $\lambda_j^{(1)}$ (для всех j) приводит к матрице $\tilde{A}$, такой, что

$$\tilde{A} = AQ. \tag{3.2.9}$$

Если p-й столбец матрицы A становится k-м столбцом матрицы $\tilde{A}$, тогда, имея в виду формулы (3.2.9), получим

$$Ae_p = \tilde{A}e_k = AQe_k,$$

и, как следствие, $e_p = Qe_k$, т. е. p-й столбец единичной матрицы I_n является k-м столбцом матрицы Q. Таким образом, перестановка столбцов, которая преобразует матрицу A в матрицу $\tilde{A}$, примененная к единичной матрице I_n, даст матрицу Q.

Получив матрицу $\tilde{A}$ с помощью всех $c_j^{(1)}$, $\gamma_j^{(1)}$ или $\lambda_j^{(1)}$, мы затем производим прямое гауссово исключение для последовательно взятых столбцов $\tilde{A}$ и в каждом столбце ищем главный элемент, который лежит в строке с наименьшим числом ненулевых элементов. Очевидно, такой выбор главного элемента должен обеспечить небольшое заполнение на каждом шаге гауссова исключения. Все сказанное можно математически описать таким образом. Пусть $A^{(1)} = \tilde{A}$ и вместо формулы (2.5.3) матрицу $\hat{A}^{(k)}$ определяет выражение

$$\hat{A}^{(k)} = P_k A^{(k)}, \tag{3.2.10}$$

где матрица P_k получена из единичной матрицы I_n путем перестановки ее $(\alpha + k - 1)$-й и k-й строк, причем значение α определяется из условия

$$\tilde{r}_\alpha^{(k)} = \min_i \tilde{r}_i^{(k)} \tag{3.2.11}$$

для всех значений i, для которых $|\hat{a}_{i+k-1,\,k}^{(k)}| > \varepsilon$, и $\tilde{r}_i^{(k)}$ — аппроксимация для $r_i^{(k)}$, определяемого формулой (3.2.2). Допустимое значение главного элемента ε то же, что и в формуле (2.5.10).

Аппроксимация $\tilde{r}_i^{(k)}$ для $r_i^{(k)}$ может быть получена с помощью следующей теоремы (Тьюарсон, 1966).

Теорема 3.2.12. *Если ненулевые элементы в последних* $n-k+1$ *строках и столбцах матрицы* $\hat{A}^{(k)}$ *распределены случайным образом в этих строках и столбцах и* $\hat{r}_i^{(k)}$ *является числом ненулевых элементов в* $(i+k-1)$*-й строке матрицы* $\hat{A}^{(k)}$, *то ожидаемое число ненулевых элементов* $\hat{r}_{i-1}^{(k+1)}$ *соответствующей строки матрицы* $A^{(k+1)}$ *равно для* $i>1$

$$\tilde{r}_{i-1}^{(k+1)}=\hat{r}_i^{(k)}, \qquad \hat{a}_{i+k-1,\,k}^{(k)}=0, \tag{3.2.13}$$

$$\tilde{r}_{i-1}^{(k+1)}=\hat{r}_i^{(k)}+\hat{r}_1^{(k)}-2-\frac{\left(\hat{r}_i^{(k)}-1\right)\left(\hat{r}_1^{(k)}-1\right)}{n-k}, \qquad \hat{a}_{i+k-1,\,k}^{(k)}\neq 0, \tag{3.2.14}$$

где матрица $A^{(k+1)}$ *определяется формулами* (2.5.2), (3.2.10) *и* (2.5.4).

Доказательство. Получим матрицу $\hat{B}_k$ из последних $n-k+1$ строк и столбцов матрицы $\hat{A}^{(k)}$, заменив в них ненулевые элементы единицами. Из определения матриц $\hat{B}_k$ и B_{k+1} следует, что i-я строка матрицы $\hat{B}_k$ соответствует $(i-1)$-й строке матрицы B_{k+1} и $(i+k-1,\ j+k-1)$-й элемент матрицы $\hat{A}^{(k)}$ для всех i и j соответствует $\hat{b}_{ij}^{(k)}$ — $(i,\ j)$-му элементу матрицы $\hat{B}_k$.

Если $\hat{a}_{i+k-1,\,k}^{(k)}=0$, то $\hat{b}_{i1}^{(k)}=0$ и ясно, что i-е строки матриц $\hat{A}^{(k)}$ и $A^{(k+1)}$ совпадают, из чего следует, что $r_{i-1}^{(k+1)}=\tilde{r}_{i-1}^{(k+1)}=\hat{r}_i^{(k)}$, и, таким образом, равенство (3.2.13) доказано.

В противном случае, если $\hat{a}_{i+k-1,\,k}^{(k)}\neq 0$, имеем $\hat{b}_{i1}^{(k)}=1$ и новый ненулевой элемент будет появляться во всех случаях, когда $\hat{b}_{1j}^{(k)}=1$ и $\hat{b}_{ij}^{(k)}=0$. Обозначим через $\Gamma(\sigma)$ вероятность того, что событие σ произойдет. Так как элементы матрицы $\hat{B}_k$ распределены случайным образом (поскольку так распределяются соответствующие элементы матрицы $A^{(k)}$), то

$$\Gamma\left(\hat{b}_{1j}^{(k)}=1 \text{ и } \hat{b}_{ij}^{(k)}=0\right)=\Gamma\left(\hat{b}_{1j}^{(k)}=1\right)\Gamma\left(\hat{b}_{ij}^{(k)}=0\right)=\left(\frac{\hat{r}_1^{(k)}-1}{n-k}\right)\left(1-\frac{\hat{r}_i^{(k)}-1}{n-k}\right). \tag{3.2.15}$$

Здесь использовано то обстоятельство, что если исключить первый столбец матрицы $\hat{B}_k$, то относительное число ненулевых элементов первой и i-й строк для оставшихся $n-k$ столбцов соответственно равно $(\hat{r}_1^{(k)}-1)/(n-k)$ и $(\hat{r}_i^{(k)}-1)/(n-k)$, так как $\hat{b}_{11}^{(k)}=\hat{b}_{i1}^{(k)}=1$. Теперь из формулы (3.2.15) следует, что ожидаемое значение заполнения i-й строки матрицы $\hat{B}_k$ равно

$$(n-k)\left(\frac{\hat{r}_1^{(k)}-1}{n-k}\right)\left(1-\frac{\hat{r}_i^{(k)}-1}{n-k}\right).$$

Прибавляя к этому $\hat{r}_i^{(k)}$ (начальное число ненулевых элементов) и вычитая единицу, поскольку $a_{i+k-1,\,k}^{(k+1)}=0$, имеем

$$\tilde{r}_{i-1}^{(k+1)}=\hat{r}_i^{(k)}+(\hat{r}_1^{(k)}-1)\left(1-\frac{\hat{r}_i^{(k)}-1}{n-k}\right)-1,$$

что после упрощения дает формулу (3.2.14). Этим завершается доказательство теоремы.

Применение установленной теоремы начнем со значения $\tilde{r}_i^{(1)}=r_i^{(1)}$, где $r_i^{(1)}=e_i'B_1V_1$, а матрица B_1 получена из матрицы $A^{(1)}=\tilde{A}$ [см. формулы (3.2.2) и (3.2.9)]. Чтобы сохранить простоту обозначений, будем понимать под $\hat{r}_i^{(k)}$ не только точное число ненулевых элементов $(i+k-1)$-й строки матрицы $\hat{A}^{(k)}$, но и его приближенное значение. Для каждого k, учитывая формулы (3.2.10) и (3.2.11), вычислим

$$\hat{r}^{(k)}=P_k\tilde{r}^{(k)}, \tag{3.2.16}$$

где $\hat{r}^{(k)}$ и $\tilde{r}^{(k)}$ — $(n-k+1)$-мерные векторы, компоненты которых определяют соответственно число ненулевых элементов строк матриц $\hat{A}^{(k)}$ и $A^{(k)}$.

Затем мы используем формулы (3.2.13) и (3.2.14), чтобы по значениям $\hat{r}_i^{(k)}$ получить значения $\tilde{r}_{i-1}^{k+1}$ для всех $1<i\leqslant n-k+1$ на каждом шаге k. Этот метод позволяет избежать вычисления точного значения $r_i^{(k)}$ для каждого k из соответствующих матриц B_k.

Для больших разреженных матриц, хранимых в упакованной форме, это может привести к значительной экономии времени и усилий. Однако значения $\tilde{r}_i^{(k)}$, даваемые формулами (3.2.13) и (3.2.14), были получены на основании вероятностей гипотезы и поэтому являются только аппроксимациями действительных значений всех $r_i^{(k)}$. Для больших разреженных матриц при малых значениях k такая аппроксимация вполне приемлема, но по мере увеличения k она становится все хуже. Поэтому рекомендуется, если можно, периодически через определенные интервалы вычислять точные значения $r_j^{(k)}$, пользуясь соответствующими матрицами B_k.

В некоторых случаях можно точно определять все значения $r_i^{(k)}$, не затрачивая слишком много труда. Например, если $A^{(k)}$ хранится в виде связного списка, описанного в разд. 1.3, то соответствующее значение $r_i^{(k)}$ может быть легко изменено всякий раз, когда в соответствии с формулой (2.5.2) создается новый ненулевой элемент или ненулевой элемент обращается в нуль.

Можно следующим образом суммировать методы этого раздела. Вначале применяем все $c_j^{(k)}$, все $\gamma_j^{(k)}$ или все $\lambda_j^{(k)}$, определенные соответственно формулами (3.2.2), (3.2.4) и (3.2.5), для определения матриц $\tilde{A}$ и Q в формуле (3.2.9). Затем полагаем $A^{(1)} = \tilde{A}$ и используем матрицу B_1, связанную с матрицей $A^{(1)}$, для вычисления $r_i^{(1)}$, согласно формуле (3.2.2). Положим $\tilde{r}_i^{(1)} = r_i^{(1)}$ и преобразуем матрицу $A^{(k)}$ в матрицу $A^{(k+1)}$ для $k = 1, 2, \ldots, n$, пользуясь формулами (3.2.11), (3.2.10), (2.5.2), (2.2.3) и (2.5.4). При этом переход от всех значений $\tilde{r}_i^{(k)}$ ко всем соответствующим значениям $\tilde{r}_i^{(k+1)}$ производится по формулам (3.2.16), (3.2.13) и (3.2.14). Таким путем матрица A преобразуется в некоторую верхнюю треугольную матрицу $\hat{U} = A^{(n+1)}$. Пользуясь формулами (3.2.10) и (2.5.2), можно записать

$$\hat{U} = A^{(n+1)} = L_n P_n \ldots L_1 P_1 A^{(1)}$$

или, с учетом формулы (3.2.9),

$$\hat{U} = L_n P_n \ldots L_1 P_1 A Q = \hat{L} A Q,$$

где

$$\hat{L} = L_n P_n \ldots L_1 P_1.$$

Поэтому

$$A^{-1} = Q \hat{U}^{-1} \hat{L},$$

что дает формулу (2.5.13) разложения матрицы A^{-1} на множители.

В следующем разделе рассматривается, каким образом матрицы P и Q в формуле (3.2.1) могут быть определены априорно так, чтобы матрица $\hat{A}$ имела форму, при которой заполнение или отсутствует, или ограничено только определенными частями матрицы $\hat{A}$.

3.3. Формы, подходящие для гауссова исключения

Одной из наиболее простых форм, исключающих заполнение, когда в качестве главных выбираются диагональные элементы, является полная ленточная форма, которая определяется следующим образом.

Определение. Матрица A, у которой $a_{ij} = 0$ при $|i - j| > \beta$, называется *ленточной матрицей*. Если к тому же $a_{ij} \neq 0$ для всех $|i - j| \leqslant \beta$, то она называется *полной ленточной матрицей*. Величина $2\beta + 1$ называется *шириной ленты* [1]). Отметим, что для симметричной матрицы с переменной локальной шириной ленты, определенной в разд. 1.3, имеем $\beta = \max_i \theta_i$.

Если матрица A — *полная ленточная матрица* и главные элементы выбираются на диагонали начиная с крайнего элемента в верхнем левом углу, тогда из доказательства теоремы 2.5.5 следует, что никакого

[1]) Автор называет шириной ленты (bandwidth) величину β. В русской литературе с понятием ширины ленты связывают величину $2\beta + 1$. — *Прим. перев.*

заполнения не будет, потому что матрица B_k является полной ленточной матрицей, и всякий раз, когда $b_{ik}^{(k)} = b_{kj}^{(k)} = 1$, будет $b_{ij}^{(k)} = 1$. С другой стороны, если матрица A не является полной, некоторые элементы внутри ленты будут нулями и заполнение ограничивается числом таких элементов внутри ленты.

Рассмотрим еще некоторые подходящие формы. Для этого предположим, что путем перестановки строк и столбцов матрицы A по формуле (3.2.1) ее можно привести к матрице $\hat{A}$, имеющей следующую форму:

$$\hat{A} = \begin{vmatrix} A_{11} & A_{12} & \dots & A_{1,p-1} & A_{1p} \\ 0 & A_{22} & \dots & A_{2,p-1} & A_{2p} \\ 0 & 0 & & & \\ \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & & A_{p-1,p-1} & A_{p-1,p} \\ A_{p1} & A_{p2} & & A_{p,p-1} & A_{pp} \end{vmatrix}, \qquad (3.3.1)$$

где диагональные подматрицы A_{ii}, $i = 1, 2, \dots, p$, являются неособенными матрицами. Если главные элементы выбираются из ненулевых элементов диагональных блоков A_{ii}, начиная с блока A_{11}, тогда заполнение может иметь место только в тех блоках матрицы (3.3.1), которые не отмечены нулями. В разд. 3.10 описывается другой порядок выбора главных элементов, который приводит к тому, что отсутствует заполнение даже в матрицах A_{ji}, $j < i$, $i \neq p$.

В свете изложенного желательно было бы определить такие две матрицы перестановок P и Q, чтобы матрица $\hat{A}$ в выражении (3.2.1) имела форму (3.3.1). Если матрица A симметричная, то, вообще говоря, является предпочтительным выразить матрицу $\hat{A}$ также в симметричной форме, так как в этом случае требуется хранить только ненулевые элементы матрицы A, расположенные на главной диагонали и выше нее. Кроме того, в силу теоремы 2.5.19, если диагональные

элементы могут быть выбраны в качестве главных (например, если матрица A положительно определенная), то симметрия сохраняется во время процесса исключения и нижняя треугольная часть матрицы не хранится. В этом случае вместо формулы (3.2.1) имеем

$$PAP' = \hat{A}, \tag{3.3.2}$$

и в матрице (3.3.1) подматрицы $A_{ij} = 0$ для всех $i \neq j$, за исключением $i = p$ или $j = p$.

В следующем разделе мы опишем некоторые методы для определения таких матриц P и Q в формулах (3.2.1) и (3.3.2), которые приводят к различным подходящим формам матрицы $\hat{A}$, частично представленным матрицей (3.3.1). При рассмотрении этих методов нам потребуются некоторые простые понятия теории графов, которые также приведены в следующем разделе. Дополнительные подробности читатель может найти в работах Басейкера и Саати (1965), Харари (1969), Беллмана и др. (1970).

3.4. Матрицы и графы

Пусть b_{ij} обозначает (i, j)-й элемент матрицы B, полученной в результате замещения единицей каждого ненулевого элемента матрицы A. Приведем в соответствие с матрицей B *помеченный граф* Ω, *помеченный направленный граф* Ω_D, *помеченный двудольный граф* Ω_B, *помеченный строчной граф* Ω_R *и помеченный столбцовый граф* Ω_C *согласно следующим определениям.*

Определения

Помеченный граф Ω является набором из n *вершин*, помеченных 1, 2, ..., n, и τ_0 *ребер*. Говорят, что существует *ребро* $[p, q]$, соединяющее вершину p с другой вершиной q, если или b_{pq}, или b_{qp} (или оба) равны единице. Отсюда следует, что τ_0 равно общему числу ненулевых элементов матрицы $B + B'$, расположенных над диагональю.

Помеченный направленный граф Ω_D является набором из n *вершин*, помеченных $1, 2, \ldots, n$, и τ_D дуг. Говорят, что существует *дуга* $[p, q]$ от вершины p к другой вершине q тогда и только тогда, когда $b_{pq}=1$. Таким образом, τ_D равно общему числу недиагональных элементов матрицы B.

Помеченный двудольный граф Ω_B состоит из двух различных наборов вершин R и C, причем каждый набор содержит n элементов, помеченных $1, 2, \ldots, n$,

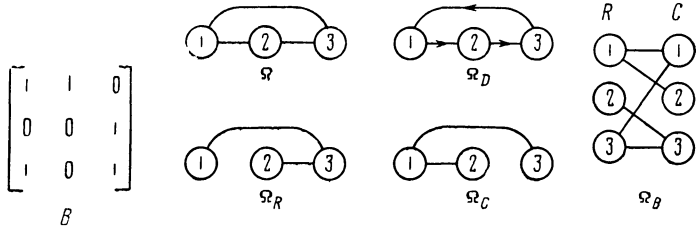

Рис. 3.4.1. Помеченные графы, соответствующие матрице.

и из набора в τ *ребер*, соединяющих вершины R и C. *Ребро* $[p, q]$ с вершиной p из набора R и вершиной q из набора C существует тогда и только тогда, когда $b_{pq} = 1$. Таким образом, τ является общим числом ненулевых элементов матрицы B.

Помеченный строчный граф Ω_R и *помеченный столбцовый граф* Ω_C являются *помеченными графами* соответственно матриц $B*B'$ и $B'*B$, где символ $*$ означает, что при вычислении скалярных произведений векторов в этих матричных произведениях следует применять только булево сложение $\oplus$, т. е. полагать $1 \oplus 1 = 1$. Так как и матрица $B*B'$, и матрица $B'*B$ симметричны, то τ_R (число ребер в Ω_R) и τ_C (число ребер в Ω_C) равны общему числу единиц, расположенных над диагоналями соответственно матриц $B*B'$ и $B'*B$.

На рис. 3.4.1 приведены матрица и соответствующие ей графы Ω, Ω_D, Ω_B, Ω_R и Ω_C.

Если строки и столбцы матрицы B переставлены так, что все ее диагональные элементы остаются на

диагонали, а именно когда

$$PBP' = \tilde{B} \tag{3.4.1}$$

(P — матрица перестановок), то единственное изменение в соответствующих графах Ω и Ω_D будет заключаться в том, что вершины будут помечены по-другому — во всем остальном эти графы остаются без изменений.

Применение различных матриц для перестановки строк и столбцов матрицы B, когда

$$PBQ = \hat{B} \tag{3.4.2}$$

(P и Q матрицы перестановок), оставляет граф Ω_B без изменений, за исключением перенумерации вершин в R и C.

Перестановки строк матрицы B с помощью матрицы перестановок P приводят к перенумерации вершин в графе Ω_R, а перестановка столбцов с помощью матрицы перестановок Q приводит к перенумерации вершин графа Ω_C. Из формулы (3.4.2) имеем

$$PB * B'P' = \hat{B} * \hat{B}' \quad \text{и} \quad Q'B' * BQ = \hat{B}' * \hat{B}, \tag{3.4.3}$$

и, следовательно, исходя из определений Ω_R и Ω_C, можно заключить, что матрицы Q и P в формуле (3.4.2) не оказывают никакого влияния соответственно на графы Ω_R и Ω_C.

Если вершины графов Ω, Ω_D, Ω_B, Ω_R и Ω_C не имеют меток, тогда в их названиях слово *помеченный* опускается и они называются соответственно *граф*, *направленный граф*, *двудольный граф*, *строчный граф* и *столбцовый граф*. Так, матрицам B и $\tilde{B}$ в формуле (3.4.1) соответствуют один и тот же граф, направленный граф, строчный граф и столбцовый граф. Матрицы B и $\hat{B}$ в формуле (3.4.2) имеют одинаковый двудольный граф, строчный граф и столбцовый граф Инвариантность графов, направленных графов, двудольных графов, строчных графов и столбцовых графов к перестановкам строк и столбцов делают их особенно полезными при исследовании заполнения и при

определении подходящих форм для гауссова исключения. Нам потребуются некоторые дополнительные определения для этих целей. Они относятся к графам Ω, Ω_D, Ω_R и Ω_C, но не к графу Ω_B. Дополнительные определения для графа Ω_B будут даны позже.

Определения

Вершины p и q, соединенные ребром в графах Ω, Ω_R или Ω_C или дугой в графе Ω_D, называются *смежными вершинами*. Если существует подмножество различных вершин $v_1, v_2, \ldots, v_\sigma, v_{\sigma+1}$, таких, что для $i = 1, 2, \ldots, \sigma$ вершины v_i и v_{i+1} являются смежными, то говорят, что вершины v_1 и $v_{\sigma+1}$ *связаны путем* $[v_1, v_2, \ldots, v_{\sigma+1}]$ *длиною* σ в графах Ω, Ω_R или Ω_C или *направленным путем длиною* σ в графе Ω_D.

Если в пути $[v_1, v_2, \ldots, v_{\sigma+1}]$ *начальная* вершина v_1 та же, что и *конечная* вершина $v_{\sigma+1}$, то говорят, что путь является *циклом длиною* σ в графах Ω, Ω_R или Ω_C или *направленным циклом* в графе Ω_D *длиною* σ.

Так как в графах Ω, Ω_R, Ω_C и Ω_D общее число вершин равно n, то и максимальное значение, которое σ может иметь, равно n.

Если множество вершин в графе $\Omega(\Omega_D, \Omega_R, \Omega_C)$ может быть разбито на два или более подмножеств, таких, что только вершины внутри подмножества связаны, тогда говорят, что граф $\Omega(\Omega_D, \Omega_R, \Omega_C)$ имеет два или более *несвязных помеченных подграфа.*

Число ребер графа $\Omega(\Omega_R, \Omega_C, \Omega_B)$, на которых данная вершина находится, называется *степенью* вершины. Так, степень вершины i в наборе R графа Ω_B есть число единиц в i-й строке матрицы B. *Степенью* вершины i в графе Ω является число внедиагональных единиц i-й строки матрицы $B \oplus B'$. Число дуг направленного графа Ω_D, начинающихся в данной вершине, есть *степень исхода* вершины. Так, степенью исхода i-й вершины является общее число единиц i-й строки матрицы B.

Степень захода вершины есть число дуг направленного графа Ω_D, оканчивающихся в данной вершине. Так, степенью захода j-й вершины является общее

число недиагональных элементов j-го столбца матрицы B.

Приведем и докажем теперь некоторые теоремы, которые позднее нам потребуются.

Теорема 3.4.4. *Пусть*

$$W = B \oplus B' \oplus I, \tag{3.4.5}$$

$$W^{\sigma+1} = W^{\sigma} * W, \tag{3.4.6}$$

где $\sigma = 1, 2, \ldots, n-1$. *Равенство* $e_i' W^{\sigma} e_j = 1$ *имеет место тогда и только тогда, когда* i*-я и* j*-я вершины помеченного графа, соответствующего* B*, связаны путем, длина которого меньше или равна* σ.

Доказательство проводится методом индукции. Теорема несомненно справедлива для $\sigma = 1$, так как $e_i' W e_j = w_{ij} = 1$, если между i-й и j-й вершинами имеется ребро длиною в единицу. Положим, что теорема верна для некоторого значения σ, и покажем, что она верна также и для $\sigma + 1$. Из формулы (3.4.6) имеем

$$w_{ij}^{\sigma+1} = \sum_{p=1}^{n} w_{ip}^{\sigma} * w_{pj}. \tag{3.4.7}$$

Теперь

$w_{ij}^{\sigma+1} = 1$ тогда и только тогда, когда $w_{ip}^{\sigma} = 1$ и

$w_{pj} = 1$ по крайней мере для одного p. (3.4.8)

Но $w_{ip}^{\sigma} = 1$, если i-я и p-я вершины связаны путем, длина которого равна или меньше σ, и $w_{pj} = 1$, если p-я и j-я вершины связаны ребром, когда $p \neq j$. При $p = j$ на основании формулы (3.4.5) будет $w_{jj} = 1$. Поэтому в любом случае соотношения (3.4.8) имеют место, если i-я и j-я вершины связаны путем, длина которого равна или меньше $\sigma + 1$. Этим завершается доказательство теоремы.

Так как максимальная длина пути равна n (поскольку имеется n вершин), то в последующих разделах матрица W^n используется для определения всех путей и циклов. Для помеченных направленных графов

теорема, подобная теореме 3.4.4, формулируется следующим образом.

Теорема (3.4.9). *Пусть*

$$\hat{W} = B \oplus I, \tag{3.4.10}$$

$$\hat{W}^{\sigma+1} = \hat{W}^{\sigma} * \hat{W}, \tag{3.4.11}$$

где $\sigma = 1, 2, \ldots, n$. *Равенство* $e_i' \hat{W}^{\sigma} e_j = 1$ *верно тогда и только тогда, когда в помеченном направленном графе, соответствующем* B, *имеется направленный путь от* i*-й до* j*-й вершины, длина которого меньше или равна* σ.

Доказательство такое же, как и для теоремы (3.4.4), только матрица W заменяется матрицей $\hat{W}$.

3.5. Диагональная блочная форма

В этом разделе описываются некоторые методы преобразования данной матрицы в *диагональную блочную форму* (BDF). Напомним, что матрица $\hat{A}$, определенная формулой (3.3.1), имеет форму BDF, если для всех $i \neq j$ подматрицы $A_{ij} = 0$ и все диагональные подматрицы A_{ii} являются квадратными матрицами. Если матрица $\hat{A}$ имеет форму BDF, то из формул (3.2.1) и (3.4.2) следует, что и $\hat{B}$ имеет форму BDF. Это означает, что оба графа Ω_R и Ω_C, соответствующие $\hat{B}$, состоят из несвязных помеченных подграфов, и каждый помеченный подграф соответствует отдельному диагональному блоку. Так как матрицам B и $\hat{B}$ соответствуют строчные (и столбцовые) графы, которые отличаются только метками вершин, то графы Ω_R и Ω_C, соответствующие B, могут быть следующим образом использованы для определения матриц P и Q в формуле (3.2.1) (Харари (1962), Тьюарсон (1967 с)).

Теорема 3.5.1. *Если*

$$W = B * B' \tag{3.5.2}$$

и

$$W^{2h+1} = W^{2h} * W^{2h}, \quad h = 0, 1, 2, \ldots, \tag{3.5.3}$$

то существует такое h_1, *что*

$$W^{2^{h_1+1}} \equiv W^{2^{h_1}} = F \tag{3.5.4}$$

и $e_i' F e_j = 1$ *тогда и только тогда, когда* i*-я и* j*-я строки матрицы* B *принадлежат одному и тому же диагональному блоку матрицы* $\bar{B}$.

Доказательство. Если в теореме 3.4.4 вместо матрицы B взять матрицу $B * B'$, то вследствие того, что матрица $B * B'$ симметричная и все ее диагональные элементы ненулевые, уравнение (3.4.5) примет вид

$$W = (B * B') \oplus (B * B')' \oplus I = B * B',$$

что совпадает с формулой (3.5.2). Поэтому из определения строчного графа Ω_R и теоремы 3.4.4 следует, что если $e_i' W^{2^h} e_j = 1$, то существует путь между вершинами i и j. Если ν есть размер наибольшего диагонального блока матрицы $\bar{B}$, тогда все пути в графе Ω_R меньше или равны ν и $\nu \leqslant n$. Поэтому для всех h, таких, что $2^h > \nu$, матрица W^{2^h} остается той же и существует некоторое h_1, для которого формула (3.5.4) верна и $e_i' W^{2^h} e_j = 1$ тогда и только тогда, когда i-я и j-я строки матрицы B принадлежат одному и тому же диагональному блоку матрицы $\bar{B}$. Этим завершается доказательство теоремы.

Чтобы использовать эту теорему, будем применять уравнение (3.5.3) до тех пор, пока не получим $2^h \geqslant n$, так как обычно мы знаем о ν только то, что $\nu \leqslant n$. Пусть $F = W^{2^h}$. Существуют различные практические методы, пригодные вместо формулы (3.5.3) для определения F (Бейкер (1962), Уошелл (1962), Ингермен (1962), Камсток (1964)). Опишем вкратце метод, предложенный Камстоком (1964).

Производится поиск в первой строке матрицы W, пока не встретится нуль, скажем в j-м столбце. Затем этот нуль заменяется выражением

$$\sum_{p=1}^{n} w_{1p} w_{pj},$$

где суммирование производится по правилу алгебры логики. Поиск продолжается в столбцах $j+1$, $j+2$ ... первой строки, пока не найден другой нуль, скажем в q-м столбце. Этот нуль заменяется выражением

$$\sum_{p=1}^{n} w_{1p} w_{pq}.$$

Этот процесс продолжается до тех пор, пока не просмотрена вся строка. Затем подобным образом производится поиск нулей и их замена в других строках матрицы. После обработки всех строк начинают опять с первой строки. Обработка матрицы продолжается до тех пор, пока за полный ее проход в ней не делается никаких изменений. Результирующая матрица и будет матрицей F.

Очевидно, что $f_{ij} = e_i' F e_j = 1$ тогда и только тогда, когда i-я и j-я строки принадлежат одному и тому же диагональному блоку. Таким образом, все строки матрицы B, которые соответствуют ненулевым элементам первого столбца матрицы F, принадлежат первому диагональному блоку. Если все строки матрицы F, для которых $f_{i1} = 1$, учтены, то следующий ненулевой столбец матрицы F может быть использован таким же образом, как и первый столбец этой матрицы, для нахождения строк матрицы B, принадлежащих второму диагональному блоку матрицы B и так далее. Таким путем можно найти диагональные блоки матрицы B, к которым принадлежит каждая строка матрицы B.

Приводимое ниже следствие теоремы 3.5.1 можно применить для определения столбцов матрицы B, принадлежащих различным диагональным блокам матрицы B.

Следствие 3.5.5. *Если матрица F определена как в теореме 3.5.1,*

$$F * B = \bar{F} \tag{3.5.6}$$

и $f_{ij} = e_i' \bar{F} e_j = 1$, то i-я строка и j-й столбец матрицы B принадлежат одному и тому же диагональному блоку матрицы B.

Доказательство. Из уравнения (3.5.6) имеем

$$\bar{f}_{ij} = \sum_{p=1}^{n} f_{ip} b_{pj} \quad \text{(сумма булева)}$$

и $\bar{f}_{ij} = 1$ тогда и только тогда, когда $f_{ip} = b_{pj} = 1$ по крайней мере для одного значения p. Из теоремы 3.5.1 известно, что условие $f_{ip} = 1$ означает, что i-я и p-я строки принадлежат одному и тому же диагональному блоку матрицы $\hat{B}$, и поскольку j-й столбец имеет по крайней мере один ненулевой элемент в p-й строке, постольку j-й столбец должен находиться в том же диагональном блоке, что и i-я и p-я строки. Таким образом, равенство $\bar{f}_{ij} = 1$ означает, что i-я строка и j-й столбец принадлежат одному и тому же диагональному блоку. Этим завершается доказательство следствия.

Чтобы воспользоваться этим следствием, поступаем следующим образом. Для всех столбцов матрицы B, принадлежащих первому диагональному блоку и выбранных согласно теореме 3.5.1, должно быть $\bar{f}_{ij} = 1$. Такие столбцы вычеркиваются или как-то отмечаются в матрице F. Следующая ненулевая строка результирующей матрицы дает множество столбцов матрицы B, которые принадлежат второму диагональному блоку и так далее.

Если в теореме 3.5.1 пользоваться не формулой (3.5.2), а полагать $W = B' * B$, то получим теорему о перестановке столбцов для преобразования матрицы B в матрицу $\hat{B}$. Однако путем небольших изменений можно пользоваться самой теоремой 3.5.1 и таким образом избежать необходимости дублировать работу. Это осуществляется следующим образом.

Теорема 3.5.7. *Если матрица F определена, как в теореме* 3.5.1, *и*

$$e_i'(B' * F * B)\, e_j = 1, \qquad (3.5.8)$$

то i-й и j-й столбцы матрицы B принадлежат одному и тому же диагональному блоку матрицы $\hat{B}$.

Доказательство. Из уравнений (3.5.2), (3.5.3) и (3.5.4) имеем

$$B' * F * B = B' * [(B * B') * (B * B') * \ldots * (B * B')] * B =$$
$$= (B' * B) * (B' * B) * \ldots * (B' * B) =$$
$$= (B' * B)^{\sigma}, \quad \sigma > \nu.$$

Поэтому из определения столбцового графа Ω_C и из тех же рассуждений, которые приводились при доказательстве теоремы 3.5.1, следует, что условие

$$e_i'(B' * F * B)\, e_j = e_i'(B' * B)^{\sigma} e_j = 1$$

предполагает наличие пути, связывающего i-й и j-й столбцы, и поэтому они принадлежат одному и тому же диагональному блоку. Это завершает доказательство теоремы.

Так как формула (3.5.8) содержит одно лишнее матричное умножение по сравнению с формулой (3.5.6), то для перестановки столбцов с целью приведения матрицы к диагональной блочной форме обычно пользуются следствием 3.5.5 вместо теоремы 3.5.7.

Приведем простой пример, показывающий, каким образом можно применить теорему 3.5.1 и следствие 3.5.5 для приведения матрицы B путем перестановок к форме матрицы $\hat{B}$. Пусть

$$B = \begin{vmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{vmatrix}, \quad \text{тогда} \quad W = \begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{vmatrix}$$

и

$$W^2 = \begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{vmatrix}.$$

Так как $W^2 \equiv W$, то

$$F = W \quad \text{и} \quad F * B = W * B = \begin{vmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{vmatrix},$$

и из теоремы 3.5.1 следует, что первая и последняя строки матрицы B принадлежат первому диагональному блоку (так как $f_{11} = w_{11} = f_{41} = w_{41} = 1$), а остальные строки — второму диагональному блоку. Из следствия 3.5.5 заключаем, что второй и третий столбцы матрицы B расположены в первом диагональном блоке, а остальные столбцы — во втором диагональном блоке. Поэтому

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad Q = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

и

$$\hat{B} = PBQ = \begin{bmatrix} \boxed{\begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix}} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \boxed{\begin{matrix} 1 & 1 \\ 1 & 0 \end{matrix}} \end{bmatrix}.$$

3.6. Треугольная блочная форма

Если матрица $\hat{A}$ задана в виде (3.3.1), причем подматрицы $A_{pj} = 0$ для $i \neq p$ и подматрицы A_{ii}, $i = 1, 2, \ldots, p$, являются квадратными матрицами, то говорят, что матрица $\hat{A}$ имеет *треугольную блочную форму* (BTF). В этом параграфе описываются некоторые методы перестановок, которые приводят матрицу B (а значит, и матрицу A) к форме BTF.

Первый метод использует два вида перестановок:

$$\hat{P}B\hat{Q} = \tilde{B} \tag{3.6.1}$$

и

$$\tilde{P}\tilde{B}\tilde{P}' = \hat{B}, \tag{3.6.2}$$

где $\tilde{B}$ — матрица, у которой все диагональные элементы равны единицам, $\hat{B}$ — матрица в форме BTF. Из формул (3.6.1) и (3.6.2) имеем

$$PBQ = \hat{B}, \qquad P = \tilde{P}\hat{P} \qquad \text{и} \qquad Q = \hat{Q}\tilde{P}'. \tag{3.6.3}$$

Задача заключается в определении матриц P и Q по формулам (3.6.3).

Если матрица A неособенная, то должен существовать по меньшей мере один ненулевой член в ее детерминанте. Этот член состоит из произведения n элементов матрицы, взятых по одному в каждой строке и каждом столбце. Таким образом, строки и столбцы матрицы A могут независимо друг от друга переставляться так, чтобы все ненулевые элементы в этом члене были диагональными. Определение матриц $\hat{P}$ и $\hat{Q}$ в формуле (3.6.1), таких, при которых $\tilde{b}_{ii} = 1$ для всех $i = 1, 2, \ldots, n$, производится следующим образом (Стьюард (1962), (1965); Далмейдж и Мендельсон (1963), Кеттлер и Вейль (1969), Харари (1971а), Дафф (1972)).

Алгоритм 3.6.4. Положить

$$B = B_1, \qquad V_1 = \sum_{i=1}^{n} e_i, \qquad U_1 = V_1',$$

$$P_1 = Q_1 = I \qquad \text{и} \qquad k = 1.$$

Шаг 1

Для всех i, j при

$$b_{ij} = 1, \; e_i'V_k = 1 \qquad \text{и} \qquad U_k e_j = 1 \tag{3.6.5}$$

вычислить

$$e_{\alpha_k}'BV_k + U_kBe_{\beta_k} = \min_{i,j}\,(e_i'BV_k + U_kBe_j). \tag{3.6.6}$$

Если нет такой пары (i, j), для которой выполняются условия (3.6.5), то перейти к шагу 2, в противном случае положить равными нулю α_k-й элемент V_k и β_k-й элемент U_k. Переставить k-ю и α_k-ю строки в матрице P_k и k-й и β_k-й столбцы в матрице Q_k.

Заменить k на $k+1$. Если $k=n+1$, перейти к шагу 2, в противном случае вернуться к началу этого шага.

Замечание. Для каждого значения k выбирается (α_k, β_k)-й элемент матрицы B в качестве (k, k)-го элемента матрицы $\tilde{B}$, определяемой формулой (3.6.1). Так как все элементы, лежащие в α_k-й строке или β_k-м столбце матрицы B, не могут быть выбраны позже в качестве диагональных элементов матрицы $\tilde{B}$, то формулы (3.6.5) и (3.6.6) гарантируют, что выбор остальных диагональных элементов матрицы $\tilde{B}$ будет возможен из максимального числа ненулевых элементов матрицы B. Это в свою очередь означает, что к окончанию текущего шага только небольшое число строк и столбцов, вообще говоря, не будет использовано при выборе диагональных элементов матрицы $\tilde{B}$. Если $k=n+1$, то это значит, что все ненулевые диагональные элементы найдены.

Шаг 2.

Вычислить

$$B^{(k)}=P_kBQ_k. \tag{3.6.7}$$

Если $k=n+1$, то $\tilde{B}=B^{(k)}$ и перейти к шагу 3. В противном случае в соответствии с блок-схемой, представленной на рис. 3.6.1, привести матрицу $B^{(k)}$ путем перестановок к матрице $B^{(n+1)}$, все диагональные элементы которой равны единице.

Замечания к рис. 3.6.1. Заметим, что переход к блоку 12 имеет место тогда, когда немаркированные строки маркированных столбцов не содержат ни одной единицы. Так как число маркированных столбцов в этот момент на единицу больше чем число маркированных строк, то матрица является особенной. Следовательно, для неособенных матриц нет перехода к

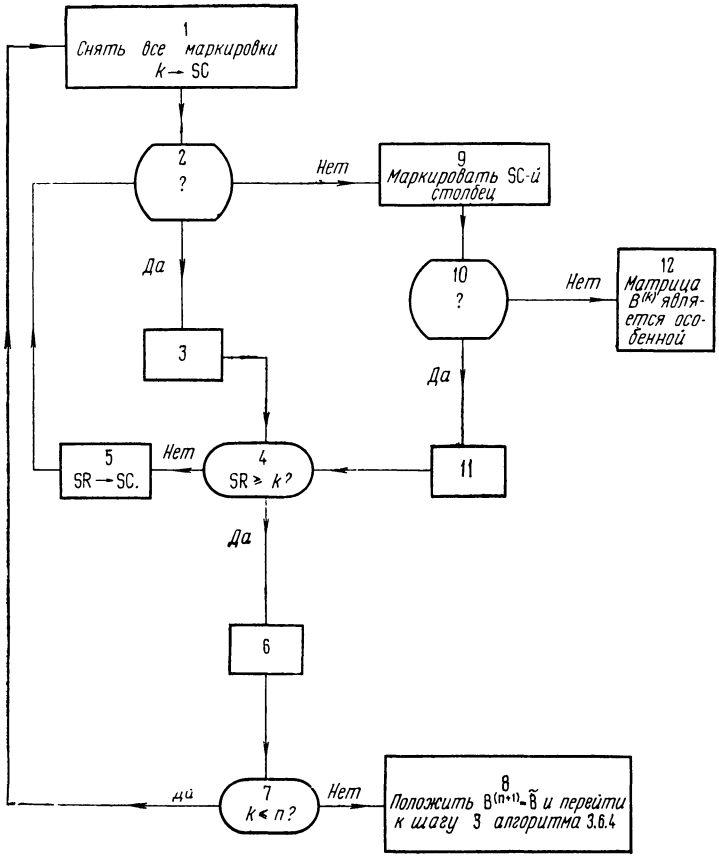

Рис. 3.6.1. Блок-схема для получения диагонали из единиц в матрице $B^{(k)}$.
SC — номер выбранного столбца,
SR — номер выбранной строки,
LICC — список цепочки номеров столбцов.

2. Имеется ли единица в немаркированной строке SC-го столбца?
3. Номер такой строки → SR.
Маркировать SC-й столбец и SR-ю строку.
Поместить SC в LICC.
6. Произвести циклическую перестановку столбцов в матрицах $B^{(k)}$ и Q_k, используя номера столбцов в LICC. Очистить LICC и переставить k-ю и SR-ю строки в матрицах $B^{(k)}$ и P_k. Заменить k на $k+1$.
10. Имеется ли какой-либо маркированный столбец с единицей в немаркированной строке?
11. Поместить номер такого столбца в SC, а номер такой строки в SR. Маркировать SR-ю строку. Исключить из LICC все номера столбцов, следующих в списке за SC.

блоку 12 и всегда возможно получение матрицы $B^{(n+1)}$ со всеми единицами на диагонали. Матрица $B^{(k)}$ содержит все нули в юго-восточном углу, т. е. $e_i' B^{(k)} e_j = 0$ для всех $i, j \geqslant k$. Преобразование матрицы $B^{(k)}$ в матрицу $B^{(k+1)}$ предполагает предварительное определение цепочки от единицы k-го столбца в северо-восточном углу к единице какой-нибудь строки в юго-западном углу матрицы $B^{(k)}$, как показано на рис. 3.6.2. Список столбцов хранится в LICC. Затем осуществляется циклический перенос столбцов $k \to j_1 \to j_2 \to j_3 \to k$, который обеспечивает переход единицы из юго-западного угла в юго-восточный угол. Если в какой-либо момент вычислений в немаркированной строке рассматриваемого столбца единицы нельзя найти, то с помощью блоков 10 и 11 (рис. 3.6.1) производится поиск другой последовательности столбцов, образующих цепочку, до тех пор, пока она не будет включать столбец с единицей в юго-западном углу. Маркировка строк и столбцов матрицы $B^{(k)}$ обеспечивает то, что процесс вычислений, приводящий к столбцу, в котором отсутствуют единицы в немаркированных строках, не может быть снова повторен и цепочка определяется за конечное число шагов. Заметим, что LICC должен быть списком номеров столбцов матрицы $B^{(k)}$, которые образуют цепочку от единицы в k-м столбце к единице в юго-западному углу, и поэтому в блоке 11 из LICC необходимо исключить столбцы, приводящие к тупику в процессе образования всей цепочки.

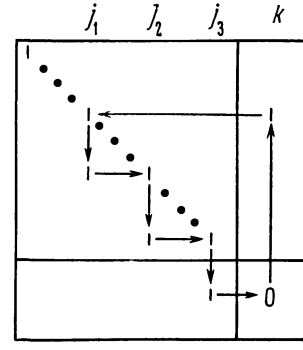

Рис. 3.6.2. Цепь от северо-восточного до юго-западного угла.

Шаг 3.

Положить $P_k = \hat{P}$ и $Q_k = \hat{Q}$. Диагональ матрицы $\hat{B}$ состоит из одних единиц и алгоритм завершен.

Замечание. Определение таких матриц $\tilde{P}$ и $\tilde{Q}$, при которых все диагональные элементы матрицы $\tilde{B}$ ненулевые, эквивалентно такому перенумерованию вершин меченого двудольного графа Ω_B, соответствующего матрице B, при котором любые две вершины наборов R и C с одинаковыми номерами являются смежными (связаны ребром). Максимальная длина ряда единиц расположенных по главной диагонали матрицы, называется *максимальной трансверсалью* (Далмейдж и Мендельсон, 1963). Напомним, что для неособенной матрицы n-го порядка длина максимальной трансверсали равна n.

Определение матрицы $\tilde{P}$ в формуле (3.6.2), составляющее вторую часть первого метода преобразования матрицы B к матрице $\hat{B}$ в форме BTF, основано на следующей теореме.

Теорема 3.6.8. *Если все диагональные элементы матрицы B единицы и*

$$B^{2^{h}+1} = B^{2^{h}} * B^{2^{h}}, \tag{3.6.9}$$

то существует такое h_1, *что*

$$B^{2^{h_1}+1} = B^{2^{h_1}} = F, \tag{3.6.10}$$

и $e_i' F e_j = 1$ *тогда и только тогда, когда существует направленный путь от i-й вершины к j-й вершине графа* Ω_D (*помеченный направленный граф, соответствующий матрице B*).

Доказательство. Так как $B \oplus I = B$, то из теоремы 3.4.9 следует, что $e_i' B^{\nu} e_j = 1$ тогда и только тогда, когда существует направленный путь, длина которого меньше или равна ν, между i-й и j-й вершинами графа Ω_D. Если ν — максимальная длина направленного пути между любыми двумя вершинами графа Ω_D, то $\nu \leqslant n$ и, очевидно, для всех $2^h \geqslant \nu$ матрица B^{2^h} не изменяется. Поэтому существует h_1, удовлетворяющее условию (3.6.10), и, следовательно, $e_i' F e_j = 1$ тогда и только тогда, когда существует направленный путь от i-й вершины к j-й вершине. Это завершает доказательство теоремы.

Если в теореме 3.6.8 вместо матрицы B начать с матрицы $\tilde{B}$, полученной из матрицы B с помощью алгоритма 3.6.4, а затем вычислить матрицу F или в соответствии с формулами (3.6.9) и (3.6.10), или методом Камстока, описанным после доказательства теоремы 3.5.1, то

$$e_i' F e_j = e_j' F e_i = 1$$

означает, что i-я и j-я вершины лежат на одном и том же направленном цикле. Так как вершины графа Ω_D, лежащие на одном и том же направленном цикле, принадлежат одному и тому же диагональному блоку матрицы $\hat{B}$, то можно использовать матрицу F для определения матрицы $\hat{B}$ по полученной матрице $\tilde{B}$ следующим образом (Харари (1962)).

Определим все такие j, что

$$e_1' F e_j = e_j' F e_1 = 1.$$

Тогда все строки и столбцы матрицы $\tilde{B}$ с такими индексами принадлежат одному и тому же диагональному блоку магрицы $\hat{B}$. Исключим (или промаркируем) все такие строки и столбцы в матрице F. Следующая неисключенная (или немаркированная) строка и соответствующий столбец могут теперь быть использованы точно таким же способом, как и первая строка и первый столбец для определения строк и столбцов другого диагонального блока и так далее. Таким путем всем строкам и столбцам матрицы $\tilde{B}$ могут быть поставлены в соответствие диагональные блоки. Заметим, что если $F = F'$, то матрица $\tilde{B}$ может быть преобразована с помощью перестановок к форме BDF (это другой возможный путь решения, хотя и более медленный, чем изложенный в параграфе 3.5). Для того чтобы упорядочить диагональные блоки так, чтобы матрица $\hat{B}$ имела форму BTF, поступаем следующим образом. Каждому диагональному блоку ставим в соответствие строку и столбец, причем каждый элемент этой строки получается логическим суммированием элементов исходной матрицы F, стоящих на пересечении строк, принадлежащих данному

диагональному блоку, и столбцов другого диагонального блока; аналогично вычисляются элементы нового столбца.

Таким путем из матрицы F получается квадратная матрица $\hat{F}$, порядок которой равен числу диагональных блоков. Помеченный направленный граф, соответствующий матрице $\hat{F}$, называется *помеченным направленным графом сгущения* Ω_S матрицы $\tilde{B}$. Его вершинами являются диагональные блоки. Так как граф сгущения не меняется при перестановке номеров вершин, то матрица $\hat{B}$ имеет тот же граф сгущения. Вершина, из которой дуги только исходят, называется *эмиттером*, а вершина, не связанная с какой-либо другой вершиной, называется *изолированной*. Таким образом, направленный граф Ω_S должен содержать эмиттер (матрица $\hat{B}$ не имеет ненулевых блоков ниже диагонали) или изолированную вершину. Назовем диагональный блок, соответствующий этой вершине графа Ω_S, первым диагональным блоком. Не рассматривая строку и столбец матрицы $\hat{F}$, которые соответствуют выбранному первому диагональному блоку, и исключая соответствующие вершину и дуги из направленного графа Ω_S, найдем, что результирующий направленный граф Ω_S (и $\hat{F}$) должен также содержать эмиттер или изолированную вершину и соответствующий диагональный блок берется в качестве второго диагонального блока и т. д. Таким образом определяется порядок диагональных блоков матрицы $\hat{B}$. Все эти перестановки регистрируются, и результирующая матрица перестановок обозначается через $\tilde{P}$. Затем с помощью формулы (3.6.2) определяется верхняя треугольная блочная матрица $\hat{B}$.

Второй метод преобразования путем перестановок матрицы B к форме BTF идентичен первому в том, что касается определения максимальной трансверсали, но отличается своей второй частью. Отличие обусловлено изложенными ниже соображениями.

Можно избежать вычисления матрицы F в соответствии с формулой (3.6.10), если преобразовать матрицу $\tilde{B}$ к треугольной блочной форме $\hat{B}$ следующим образом (Стьюард (1965)).

Заметим, что диагональным блокам второго и более высокого порядка матрицы $\tilde{B}$ отвечают циклы направленного графа, соответствующего матрице. С другой стороны, все столбцы матрицы $\tilde{B}$, в которых нет ненулевых элементов ниже диагонали и которые лежат перед первым диагональным блоком, могут быть легко найдены из матрицы $\tilde{B}$ таким путем.

Шаг 1

Определяем столбец матрицы $\tilde{B}$, содержащий единственную единицу, перемещаем этот столбец и соответствующую строку (в которой, возможно, имеется и несколько единиц) в северо-западный угол и маркируем этот столбец и соответствующую строку. Процесс повторяется до тех пор, пока среди немаркированных столбцов матрицы $\tilde{B}$ больше не будет столбцов с единственной единицей на немаркированной строке.

Шаг 2

Определяем столбцы (и строки), которые лежат в том же диагональном блоке, что и первый немаркированный столбец матрицы $\tilde{B}$. Из шага 1 следует, что первый немаркированный столбец матрицы $\tilde{B}$ имеет по меньшей мере одну недиагональную единицу в немаркированной строке. Столбец, соответствующий строке с первой такой единицей, тоже должен иметь единицу в немаркированной строке, и так далее. В какой-то момент мы сталкиваемся со столбцом, который уже ранее встречался, и этим заканчивается направленный цикл. Заменяем все столбцы направленного цикла одним столбцом, представляющим собой булеву сумму таких столбцов без их диагональных элементов.

Такой процесс называется *стягиванием* столбцов в направленном цикле. *Булева сумма векторов* — это обычное сложение векторов при условии, что $1 \oplus 1 = 1$. Аналогично заменяются все строки направленного цикла одной строкой, являющейся их булевой суммой. Когда маркируется столбец с единственной единицей, то все столбцы, стянутые в этот столбец,

принадлежат одному и тому же блоку. Порядок, в котором столбцы с единственной единицей маркировались, определяет порядок, в котором расположены диагональные блоки. Процесс выполнения шагов 1 и 2 продолжается до тех пор, пока все столбцы и строки матрицы не будут маркированы.

Поясним эти правила на примере матрицы, представленной на рис. 3.6.3, выполняя преобразования шаг за шагом до тех пор, пока не придем к матрице, представленной на рис. 3.6.4. Позиции единиц указаны знаками умножения, нулей — точками, единиц, порожденных стягиванием строк или столбцов, — плюсами.

	1	2	3	4	5	6	7
1	×	×	•	•	•	×	•
2	×	×	•	•	•	•	•
3	•	•	×	•	×	•	×
4	•	•	•	×	•	•	•
5	•	•	•	×	×	•	•
6	•	×	•	•	×	×	•
7	×	•	•	•	•	•	×

Рис. 3.6.3.

1. Столбец 3 имеет единственную единицу. Маркируем столбец 3 и строку 3. Заносим 1 в третью строку колонки «порядок», в которой регистрируется последовательность исключения столбцов.

2. Столбец 7 имеет единственную единицу в немаркированных строках. Маркируем столбец 7 и строку 7 и заносим 2 в колонку «порядок».

3. Нет больше столбцов с единственной единицей в немаркированных строках. Начинаем построение направленного пути со столбца 1, выбирая всегда первую встреченную в столбце единицу. Это дает направленный путь 1, 2, 1, и, следовательно, 1 и 2 находятся в направленном цикле. Стягиваем столбец 2 в столбец 1, для чего дополняем столбец 1 знаком +, чтобы превратить его в булеву сумму прежнего столбца 1 с недиагональными элементами столбца 2. Аналогичным образом стягиваем строку 2 в строку 1 и

маркируем столбец 2 и строку 2. Заносим 1 во вторую строку колонки «стягивание», чтобы указать на то, что столбец 2 был стянут в столбец 1.

4. Начиная со столбца 1, строим направленный путь 1, 6, 1 и стягиваем столбец 6 в столбец 1 и строку 6 в строку 1. Маркируем столбец 6 и строку 6 и заносим 1 в шестую строку колонки «стягивание».

Порядок	Стягивание		1	2	3	4	5	6	7
3		1	X	X	•	•	•	X	•
	1	2	X	X	•	•	•	•	•
1		3	•	•	X	•	X	•	X
5		4	•	•	•	X	•	•	•
4		5	•	•	•	X	X	•	•
	1	6	+	X	•	•	X	X	•
2		7	X	•	•	•	•	•	X

Рис. 3.6.4.

5. Столбец 1 имеет теперь единственную единицу в немаркированных строках. Поэтому заносим 3 в первую строку колонки «порядок» и маркируем строку 1 и столбец 1.

6. Столбец 6 имеет единственную единицу в немаркированных строках, и мы помещаем 4 в колонке «порядок» и маркируем строку 5 и столбец 5.

7. Столбец 4 имеет единственную единицу в немаркированных строках, заносим 5 в колонку «порядок» и маркируем столбец 4 и строку 4.

Теперь уже больше нет немаркированных столбцов и строк и можно по колонкам «порядок» и «стягивание»

установить расположение диагональных блоков. Первый блок содержит элемент (3,3) матрицы $\tilde{B}$, второй — элемент (7,7). Третий блок содержит в качестве диагональных элементы матрицы $\tilde{B}$, стоящие в позициях (1,1), (2,2) и (6,6). Четвертый и пятый блоки

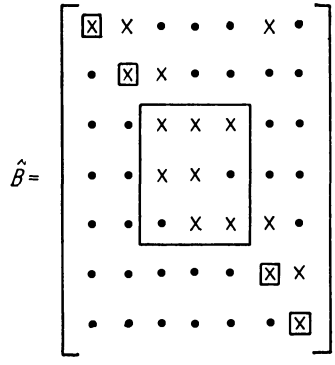

Рис. 3.6.5.

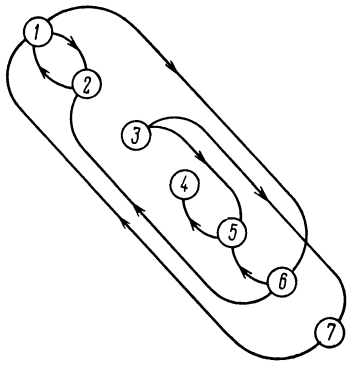

Рис. 3.6.6.

состоят соответственно из элементов (5,5) и (4,4). Таким образом, матрица перестановок $\tilde{P}$ из уравнения (3.6.2) имеет вид

$$\tilde{P} = (e_3, e_7, e_1, e_2, e_6, e_5, e_4),$$

а матрица $\hat{B}$ представлена на рис. 3.6.5.

Если матрица $\tilde{B}$ небольших размеров, то ее направленный граф может быть непосредственно использован для получения $\hat{B}$. Направленный граф, соответствующий матрице $\tilde{B}$ (рис. 3.6.3), представлен на рис. 3.6.6. Ищем вершину, которая является эмиттером (в ней не заканчивается ни одна дуга). Эмиттером будет вершина 3, даем ей номер 1 и исключаем все дуги, выходящие из нее. Теперь эмиттером становится вершина 7, даем ей номер 2 и исключаем все дуги, выходящие из нее. Теперь уже нет других эмиттеров. Вершины 1,2 и 6 расположены в цикле, и мы перенумеровываем их соответственно в 3, 4, 5. Исключаем все дуги, которые выходят из каждой из этих

вершин. Вершина 5 теперь является эмиттером, обозначаем ее через 6 и, наконец, вершине 4 даем номер 7. Таким образом,

$$\tilde{P} = (e_3, e_7, e_1, e_2, e_6, e_5, e_4),$$

как и прежде.

Имеются еще два метода для преобразования матрицы к форме BTF или к форме, подобной ей. Они рассматриваются в разд. 3.7.

3.7. Треугольная ленточная форма

Говорят, что матрица B имеет *треугольную ленточную форму* (BNTF), если существует такое β, $0 < \beta \leqslant n$, что $b_{ij} = 0$, для всех $i - j > \beta$, и значение β не может быть уменьшено перестановками строк и столбцов матрицы B.

Пусть первый ненулевой элемент i-й строки матрицы B находится в q_i-м столбце, тогда определяем

$$\begin{aligned} \beta_i &= i - q_i, & q_i &\leqslant i, \\ \beta_i &= 0, & q_i &> i. \end{aligned} \tag{3.7.1}$$

Очевидно, что если матрица B имеет форму BNTF, то $\max_i \beta_i = \beta$. Заметим, что треугольная блочная форма (BTF), описанная в предыдущем разделе, является специальным случаем формы BNTF ($\beta + 1$ является в этом случае размером наибольшего диагонального блока матрицы B).

В этом разделе будем полагать, что, если это возможно, данная матрица уже преобразована к форме BDF согласно методам, изложенным в разд. 3.5. Если такое преобразование сделано, мы рассмотрим каждый из диагональных блоков в отдельности. Поэтому не будет потери в общности, если мы предположим, что граф матрицы B является связным.

Первый метод преобразования матрицы B к форме BNTF.

Пусть P и Q являются такими матрицами перестановок, что

$$\hat{B} = PBQ, \tag{3.7.2}$$

причем i-я строка и j-й столбец матрицы B становятся соответственно ρ_i-й строкой и μ_j-м столбцом матрицы $\hat{B}$. Нашей задачей является определение таких матриц перестановок P и Q, которые минимизируют

$$\max_i \psi_i,$$

где ψ_i — расстояние от главной диагонали до крайнего левого ненулевого элемента ρ_i-й строки матрицы $\hat{B}$. Заметим, что $\psi_i = 0$ в том случае, когда нет ни одного ненулевого элемента левее диагонали в ρ_i-й строке матрицы $\hat{B}$.

Эта задача может быть решена следующим образом (Чень (1972)). Пусть ненулевые элементы i-й строки матрицы B расположены в $j_{i\alpha}$-ых столбцах, $\alpha = 1, 2, \ldots, r_i$, где r_i — общее число ненулевых элементов i-й строки матрицы B. Если крайний левый ненулевой элемент ρ_i-й строки матрицы $\hat{B}$ лежит в $\mu_{j_{is}}$-м столбце, то

$$\rho_i - \mu_{j_{is}} = \psi_i, \qquad \rho_i \geqslant \mu_{j_{is}}$$

и для всех $\mu_{j_{i\alpha}}$, не лежащих справа от ρ_i-го столбца, выполняется условие $\rho_i + \mu_{j_{i\alpha}} \leqslant \psi_i$. Если мы предпишем, чтобы $\psi_i \geqslant 0$ для всех i, тогда для любого ненулевого элемента вправо от диагонали, скажем лежащего в $\mu_{j_{it}}$-м столбце, будет

$$\rho_i - \mu_{j_{it}} < 0, \quad \text{так как} \quad \rho_i < \mu_{j_{it}}.$$

Поэтому имеем

$$\rho_i - \mu_{j_{i\alpha}} \leqslant \psi_i \quad \text{для всех } \alpha.$$

Учитывая изложенное выше, можно следующим образом формулировать задачу:

Соответственно всем значениям $b_{ij} = 1$ найти такие целые ρ_i, μ_j и ψ_i, которые

$$\text{минимизируют } \max_i \psi_i \qquad (3.7.3)$$

при следующих ограничениях:

$$1 \leqslant \rho_i, \quad \mu_j \leqslant n; \qquad 0 \leqslant \psi_i \leqslant n-1, \tag{3.7.4}$$

$$\rho_i - \mu_{j_{i\alpha}} - \psi_i \leqslant 0; \qquad \alpha = 1, 2, \ldots, r_i \tag{3.7.5}$$

и

$$\rho_i \neq \rho_j, \quad \mu_i \neq \mu_j \quad \text{для} \quad i \neq j. \tag{3.7.6}$$

Это есть задача Чебышева при наличии ограничений, которая может быть выражена как задача целочисленного программирования (Рабинович (1968)) и затем решена обычным способом (Данциг (1963 а)).

Второй метод преобразования матрицы B к форме BNTF.

Этот метод основан на использовании мер для строк и столбцов, полученных из вероятностных соображений. Метод не минимизирует β, но приводит к величине, достаточно мало отличающейся от минимума. Однако он намного проще первого метода, данного уравнениями с (3.7.3) по (3.7.6).

Меры строк и столбцов, используемые в этом методе для преобразования матрицы B к форме BNTF, получаются следующим образом.

Пусть $r_i = r_i^{(1)}$, $c_j = c_j^{(1)}$, где $r_i^{(1)}$ и $c_j^{(1)}$ определены уравнениями (3.2.2), причем $B_1 = B$. Пусть Λ_i множество всех столбцов в i-й строке матрицы B, для которых $b_{ij} = 1$; положим

$$d_i = \sum_{j \in \Lambda_i} c_j. \tag{3.7.7}$$

Для дальнейшего рассмотрения предположим, что матрица B уже приведена к форме BNTF и единицы в области $j \geqslant i - \beta$ распределены случайным образом так, что каждый элемент этой области имеет одинаковую вероятность оказаться ненулевым. Пусть $E(\theta)$ обозначает математическое ожидание θ. Тогда все $E(r_i)$ будут, вообще говоря, уменьшаться с увеличением i от 1 до n. С другой стороны, $E(c_j)$ будут, вообще говоря, возрастать с увеличением j. В соответствии с формулой (3.7.7) для данного i среднее значение всех c_j при $j \in \Lambda_i$ равно d_i/r_i. Нетрудно

обнаружить, что величины $E(d_i/r_i)$, вообще говоря, возрастают с увеличением i. Для заданной i-й строки стандартное отклонение π_i для c_j при $j \in \Lambda_i$ имеет вид

$$\pi_i^2 = \frac{\sum\limits_j c_j^2}{r_i} - \left(\frac{\sum\limits_j c_j}{r_i}\right)^2, \qquad j \in \Lambda_i,$$

что с учетом формулы (3.7.7) дает

$$\pi_i^2 = \frac{\sum\limits_j c_j^2}{r_i} - \left(\frac{d_i}{r_i}\right)^2, \qquad j \in \Lambda_i. \tag{3.7.8}$$

Принимая во внимание допущения относительно характеристик матрицы B, которые были сделаны нами в начале настоящего рассмотрения, нетрудно видеть, что математические ожидания π_i будут, как правило, уменьшаться с увеличением i. Ранее в нашем рассмотрении мы также видели, что с увеличением i все $E(r_i)$ и все $E(r_i/d_i)$, вообще говоря, уменьшаются. Поэтому разумная мера M_i, соответствующая i-й строке, которая учитывает все три величины r_i, π_i и r_i/d_i, может быть взята в виде

$$M_i = r_i + \delta\left(\frac{r_i}{d_i}\right) + \psi\pi_i, \tag{3.7.9}$$

где δ и ψ — некоторые числа, выбранные таким образом, чтобы величины r_i, $\delta(r_i/d_i)$ и $\psi\pi_i$ были одного порядка. Очевидно, все $E(M_i)$ будут, как правило, уменьшаться с увеличением i.

Меры $\tilde{M}_j$ для столбцов могут быть определены по такой же формуле (3.7.9), как и M_i, а именно

$$\tilde{M}_j = \tilde{c}_j + \delta\left(\frac{\tilde{c}_j}{\tilde{d}_j}\right) + \psi\tilde{\pi}_j, \tag{3.7.10}$$

где $\tilde{d}_j$ определяется подобно d_i в формуле (3.7.7) следующим образом:

$$\tilde{d}_j = \sum_i r_i \tag{3.7.11}$$

для всех значений i, при которых $b_{ij} = 1$, а $\tilde{\pi}_j$ даются следующим уравнением, похожим на уравнение

(3.7.8):

$$\tilde{\pi}_j^2 = \frac{\sum_i r_i^2}{c_j} - \left(\frac{d_j}{c_j}\right)^2 \qquad (3.7.12)$$

для всех значений l, при которых $b_{ij} = 1$. Величины $E(\tilde{M}_j)$ будут, вообще говоря, возрастать с увеличением j.

Выше мы видели, что если матрица B имеет достаточно хорошую форму BNTF, то с возрастанием l и j все $E(M_i)$ уменьшаются, в то время как все $E(\tilde{M}_j)$ увеличиваются. Теперь предположим, что матрица B является некоторой разреженной матрицей, не обязательно в форме BNTF, со случайно распределенными ненулевыми элементами. Если вычислить все M_i и все $\tilde{M}_j$ для такой матрицы B и затем упорядочить ее строки по нисходящим значениям всех M_i, а столбцы по возрастающим значениям всех $\tilde{M}_j$, то можно ожидать, что форма преобразованной матрицы B будет достаточно хорошо приближаться к BNTF.

Нам еще предстоит определить значения δ и ψ в (3.7.9) и (3.7.10) в случае произвольной разреженной матрицы B (ненулевые элементы матрицы B случайным образом распределены по всей матрице). На практике нами найдено, что, по-видимому, вполне удовлетворительные результаты получаются, если принять $\delta = \tau^2/n^2$ и $\psi = 2$ (τ — общее число ненулевых элементов матрицы B) (Тьюарсон (1967c)). Эвристическое обоснование такого выбора значений δ и ψ может быть получено, если молчаливо предположить, что для всей матрицы B

$$E(r_i) \approx E(\tilde{c}_j) \approx \tau/n,$$

$$E(d_i) \approx E(\tilde{d}_j) \approx E(r_i)\,E(\tilde{c}_j) \approx \tau^2/n^2,$$

$$E\left(\frac{r_i}{d_i}\right) = E(r_i)/E(d_i)$$

и

$$2\pi_i = E(\tilde{c}_j) - \min \tilde{c}_j$$

для всех значений j, при которых $b_{ij} = 1$. Таким образом,

$$2\pi_i \approx \frac{\tau}{n} \quad \text{и} \quad \frac{\tau^2}{n^2} E\left(\frac{r_l}{d_i}\right) \approx \frac{\tau}{n},$$

и поэтому

$$\psi = 2 \quad \text{и} \quad \delta = \frac{\tau^2}{n^2}.$$

В некоторых случаях, после того как матрица B преобразована к форме BNTF, можно осуществить дополнительные перестановки строк для создания под диагональю *треугольных уголков.*

Треугольные уголки определяются следующим образом. Если в матрице B элементы $b_{ij} = 0$, когда $i \geqslant p_1$ и $j < q_1$ или $i > p_2$ и $q_1 \leqslant j \leqslant q_2$, причем $q_1 \leqslant p_1 < p_2$ и $q_1 < q_2 \leqslant p_2$, то треугольник с вершинами (p_1, q_1), (p_2, q_1) и (p_2, q_2) называется *треугольным уголком* (см. рис. 3.7.1). Если $p_1 = q_1$ и $p_2 = q_2$, то треугольный уголок является половиной диагонального блока.

Вспомним определение β_i, данное формулами (3.7.1), и тот факт, что $\max_i \beta_i \geqslant \beta$ для всех матриц, полученных из B перестановками строк и столбцов. Пусть матрица B имеет форму BNTF и Λ_m обозначает непустое множество индексов i строк, для которых $\beta_i = \beta$. Имеем следующую теорему.

Теорема 3.7.13. *Если матрица B имеет форму* BNTF *и $\beta_{i_2} - \beta_{i_1} > i_2 - i_1$ для некоторых $i_2 > i_1$, то или i_2 не принадлежит множеству Λ_m, или Λ_m включает по меньшей мере еще один (кроме i_2) индекс строки.*

Доказательство. Переставим i_1-ю и i_2-ю строки матрицы B и обозначим новые значения β_{i_1} и β_{i_2} соответственно через $\hat{\beta}_{i_1}$ и $\hat{\beta}_{i_2}$. Тогда

$$\hat{\beta}_{i_1} = \beta_{i_2} - (i_2 - i_1),$$

что, если учесть, что $i_2 > i_1$, означает

$$\hat{\beta}_{i_1} < \beta_{i_2}. \tag{3.7.14}$$

Также

$$\hat{\beta}_{i_2} = \beta_{i_1} + (i_2 - i_1),$$

что, учитывая неравенство $\beta_{i_2} - \beta_{i_1} > i_2 - i_1$, дает

$$\hat{\beta}_{i_2} < \beta_{i_2}. \tag{3.7.15}$$

Если бы i_2 был единственным элементом множества Λ_m, то из неравенств (3.7.14) и (3.7.15) ясно, что мы уменьшили значение β. Это невозможно, поскольку мы предположили, что матрица имеет форму BNTF. Поэтому или i_2 не принадлежит множеству Λ_m, или если оно ему принадлежит, то в множестве Λ_m имеется по меньшей мере один индекс другой строки, чтобы сохранить значение β. Этим завершается доказательство теоремы.

Эта теорема может быть применена для перестановок строк матрицы B до тех пор, пока для всех $i_2 > i_1$

$$\beta_{i_2} - \beta_{i_1} \leqslant i_2 - i_1.$$

К этому моменту вычислений в матрице может появиться некоторое количество треугольных уголков, потому что всякий раз, когда $i_1 < i_2$ и

$$\beta_{i_2} - \beta_{i_1} = i_2 - i_1, \tag{3.7.16}$$

для всех $i_1 \leqslant i \leqslant i_2$ справедливо равенство $\beta_{i_2} - \beta_i = i_2 - i$. Если p_2 и p_1 являются максимумом и минимумом значений соответственно i_2 и i_1, для которых равенство (3.7.16) верно, и крайние левые единицы в p_2-й и $(p_2 + 1)$-й строках находятся соответственно в q_1-м и $(q_2 + 1)$-м столбцах, то треугольник с вершинами (p_1, q_1), (p_2, q_1) и (p_2, q_2) является треугольным уголком (см. рис. 3.7.1). Поэтому мы можем применить методы настоящего раздела, чтобы преобразовать матрицу B к форме BNTF, а затем, переставляя строки, получить под диагональю треугольные уголки в таком количестве, в каком это возможно. Если имеется один или несколько треугольных уголков, для которых $p_1 = q_1$ и $p_2 = q_2$, то матрица $\hat{B}$ имеет форму BTF, как показано на рис. 3.7.2. Все элементы, расположенные на границе заштрихованной области под

диагональю, равны единицам, за исключением тех элементов, которые расположены на горизонтальных участках границы этой области. Все другие элементы заштрихованной области матрицы и элементы, расположенные на горизонтальных участках границы, могут быть и единицами, и нулями. Все элементы,

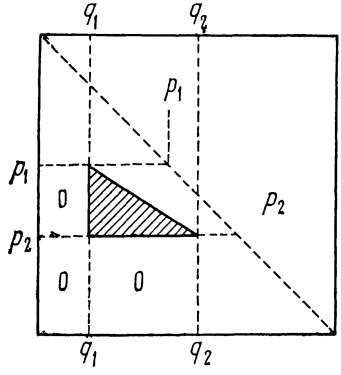

Рис. 3.7.1. Треугольный уголок.

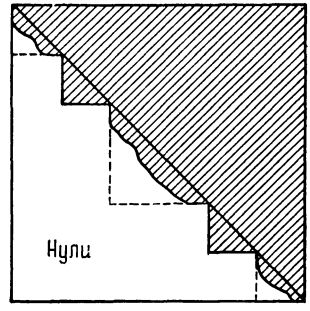

Рис. 3.7.2. Треугольная блочная форма.

лежащие в незаштрихованной области матрицы, являются нулями. Пунктиром указаны диагональные блоки.

Заключим этот раздел замечанием, что упомянутая выше перестановка строк может быть применена даже в тех случаях, когда матрица B имеет форму, только приближающуюся к форме BNTF. В этом случае $\max_i \beta_i$ может, согласно теореме 3.7.13, уменьшаться, и мы не только создаем треугольные уголки, но в определенных случаях ближе приближаемся к форме BNTF.

3.8. Ленточная форма

Во многих приложениях матрица A является симметричной и положительно определенной, и поэтому в целях сохранения симметрии обычно предпочтительней выбирать главные элементы на диаго-

нали. Согласно теореме 2.5.19, это позволит хранить только ненулевые элементы матрицы, расположенные над диагональю. Для преобразования матрицы A к подходящей для гауссова исключения форме в этом случае производятся одинаковые перестановки строк и столбцов. Это можно себе представить как перестановку диагональных элементов матрицы A. Такая перестановка описывается уравнениями (3.3.2) и

$$PBP' = \hat{B}. \tag{3.8.1}$$

Если положить

$$\beta_i = i - q_i, \qquad q_i \leqslant i, \tag{3.8.2}$$

где i и q_i — индексы элемента $\hat{b}_{iq_i}$, являющегося крайней левой единицей i-й строки матрицы $\hat{B}$, то нашей целью будет определение минимальной ширины ленты $2\beta + 1$ и матрицы перестановок P, таких, что

$$\beta = \min_P \max_i \beta_i. \tag{3.8.3}$$

Эта задача может быть выражена как задача Чебышева аналогично тому, что изложено в предыдущем параграфе для первого метода преобразования матрицы B к форме BNTF. Такую постановку задачи мы здесь описывать не будем, так как она обычно не годится для практических целей. Ниже излагаются четыре практических метода перестановок строк и столбцов матрицы, преобразующих ее к *ленточной форме* (BF). Эти методы обеспечивают отсутствие нулей в ленте, если матрица допускает такое преобразование, или приводят к матрице со значением $\max_i \beta_i$, достаточно близким к минимуму β.

Первый метод

Этот метод особенно полезен для задач, в которых исходная нумерация пространственной системы вершин помеченного графа определяет ширину ленты соответствующей матрицы (например, структурные системы конечных элементов, модели тепловых цепей с сосредоточенными параметрами, электрические цепи, системы трубопроводов, конечно-разностные урав-

нения энергетических систем). На рис. 3.8.1 приведен простой пример, иллюстрирующий уменьшение ширины ленты матрицы путем перенумерования вершин соответствующего ей графа. После изменения номеров вершин графа Ω, который соответствует матрице B, получается граф $\hat{\Omega}$, которому соответствует матрица $\hat{B}$. В матрицах B и $\hat{B}$ единицы обозначены знаками умножения, нули — точками. Заметим, что ширина ленты в матрице B равна 9, в то время как в матрице $\hat{B}$ она

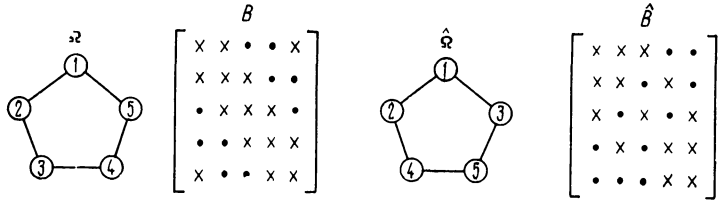

Рис. 3.8.1. Перенумерация вершин для уменьшения ширины ленты.

равна всего 5. Метод перенумерования вершин графа с целью уменьшения ширины ленты в соответствующей матрице может быть в алгоритмической форме описан следующим образом (Розен (1968)).

1. Поместить 1, 2, ..., n в *список вершин* (VL), в котором имеется n ячеек. Определить максимальную ширину ленты в матрице B и две вершины, которые приводят к ней. Если максимум имеет место для нескольких пар вершин, то выбрать любую из них (если $\max_i \beta_i = \beta_p$, то p и $p - \beta_p$ составляют пару вершин). Если вершина с бóльшим номером может быть переставлена с вершиной с меньшим номером для уменьшения ширины ленты (для некоторого $i < p$ имеет место $\beta_i + (p - i) < \beta_p$), то перейти к шагу 7.

2. Если вершина с меньшим номером может быть переставлена с вершиной с бóльшим номером для уменьшения ширины ленты, то перейти к шагу 7.

3. Если вершина с бо́льшим номером может быть переставлена с вершиной с меньшим номером так, что сохраняется ширина ленты, то перейти к шагу 6.

4. Если вершина с меньшим номером может быть переставлена с вершиной с бо́льшим номером так, что сохраняется ширина ленты, то перейти к шагу 6. (Замечание. Целью шагов 3 и 4 является изменить расположение вершин без увеличения ширины ленты так, чтобы могли быть выполнены дополнительные успешные перестановки на первом и втором шагах.)

5. Дальнейшие перестановки невозможны, и теперь матрица B имеет ленточную форму и $e_{\mathrm{VL}(j)}$ будет j-м столбцом матрицы перестановок $P(j = 1, 2, \ldots, n)$, где $\mathrm{VL}(j)$ — целое число j-й ячейки списка вершин VL. Тогда $\hat{A} = PAP'$. Останов.

6. Если на 3-м и 4-м шагах произведено максимальное число непрерывных перестановок или снова выбраны для перестановки две вершины, ранее переставленные друг с другом, то перейти к шагу 5.

7. Произвести указанную перестановку вершин, а именно переставить соответствующие элементы в списке вершин VL и строки и столбцы матрицы B, и вернуться к шагу 1.

Второй метод

Другая схема изменения номеров вершин, приводящая к уменьшению ширины ленты соответствующей матрицы, описывается следующим алгоритмом (Катхилл и Мак-Ки (1969)).

1. Для каждой вершины i графа Ω, соответствующего матрице B, вычислить ее степень ρ_i, равную общему числу недиагональных единиц i-й строки матрицы B. Затем выбрать какую-либо вершину i_1, для которой $\rho_{i_1} = \min_i \rho_i$, и пометить эту вершину первой. На рис. 3.8.2 вершина 1* помечена 1.

2. Присвоить вершинам, смежным с вершиной 1, новые номера, начиная с 2, в порядке возрастания их степеней (если степени некоторых смежных вершин

совпадают, то выбирать любую из них). Эти вершины относят к первому уровню. На рис. 3.8.2 вершинам 2*, 4* и 9* присвоены номера соответственно 2, 3 и 4.

3. Повторить эту процедуру последовательно для каждой из вершин первого уровня — это значит сперва для вершины 2, затем для вершины 3 и т. д. На рис. 3.8.2 еще неперенумерованной вершиной, смеж-

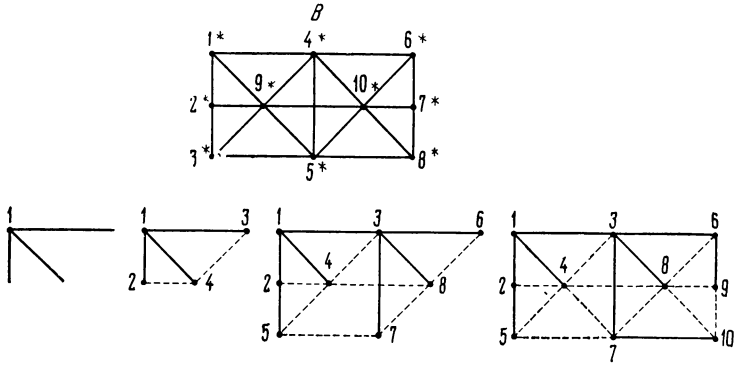

Рис. 3.8.2. Пример схемы перенумерации вершин.

ной с вершиной 2, является вершина 3* — она становится вершиной 5. Смежными с вершиной 3 (которая сначала имела номер 4*) являются вершины 6*, 5* и 10*. Вершина 6* имеет меньшую степень, чем вершина 5*, степень которой в свою очередь меньше степени вершины 10*; поэтому присвоим этим вершинам соответственно номера 6, 7 и 8. Заметим, что вершины 5, 6, 7 и 8 связаны с вершиной 1 путем длины 2 и относятся ко второму уровню.

4. Повторить вышеизложенную процедуру для вершин каждого следующего уровня, пока все n вершин графа Ω не будут перенумерованы. Если Ω состоит из двух или более несвязных подграфов, то процедура заканчивается, как только все вершины в подграфе перенумерованы. В этом случае необходимо выбрать начальную вершину в каждом из несвязных

подграфов и повторить шаги 2, 3, 4 для каждого из них.

5. Наконец, переставить строки и столбцы матрицы B (или A) в соответствии с новыми номерами вершин для получения $\hat{B}$ (или $\hat{A}$). Матрицы B и $\hat{B}$ представлены на рис. 3.8.3. Они соответствуют исход-

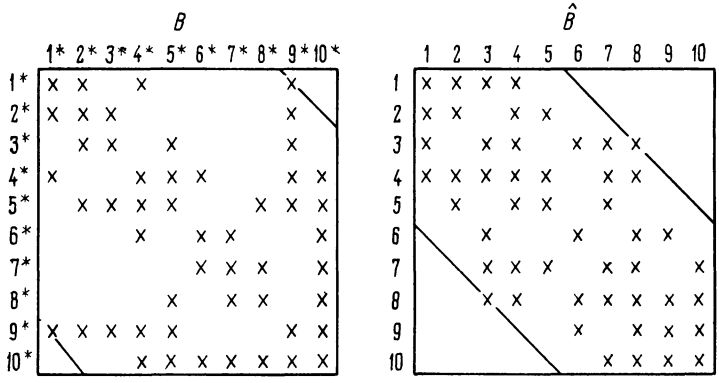

Рис. 3.8.3. Матрица B и ее преобразованная форма $\hat{B}$.

ному графу и перенумерованному графу, представленным на рис. 3.8.2. Заметим, что ширина ленты в матрице B равна 17, а в матрице $\hat{B}$ — 11.

Описанная схема перенумерации не приводит, вообще говоря, к минимальной ширине ленты. Однако можно построить другую последовательность номеров вершин, если в качестве исходной выбрать другую вершину i_1 на шаге 1, такую, для которой

$$\rho_{\min} \leqslant \rho_{i_1} \leqslant \rho_{\min} + \frac{1}{2}\rho_{\max},$$

где $\rho_{\min}$ и $\rho_{\max}$ соответственно минимальное и максимальное значения ρ_i, и если на любом из последующих шагов выбирать другой порядок номеров вершин данного уровня, когда степени вершин совпадают. Для преобразования матрицы B к матрице $\hat{B}$

выбирается такая последовательность номеров вершин, которая приводит к меньшей ширине ленты. Разумеется, имеет смысл построить только небольшое число таких последовательностей, так как в противном случае процесс отнял бы неоправданно много времени.

Третий метод

Итеративный метод, минимизирующий среднюю ширину ленты

$$\bar{\beta}=\frac{1}{n}\sum_{i=1}^{n}\beta_i,$$

где β_i определяется формулой (3.8.2), основан на перестановках строк (столбцов), как и первый метод, и может быть представлен в виде следующего алгоритма (Эйкиуз и Утку (1968)).

1. Переставить две последовательные строки (и соответствующие столбцы) матрицы B, если а) $\bar{\beta}$ уменьшается, б) $\bar{\beta}$ остается без изменения, но строка с меньшим числом единиц внутри ленты расположится в результате перестановки дальше от центральной строки матрицы B. Регистрировать все перестановки в списке перестановок строк RI. (Первоначально список RI содержит целые 1, 2, ..., n, и всякий раз, когда осуществляется перестановка строк, переставляются и элементы списка RI.)

2. Делать перестановки, производя сравнения строк полными циклами. (Полный цикл состоит из $n-1$ шагов. На каждом шаге сравниваются две последовательные строки и производится перестановка, если условия на шаге 1 удовлетворяются. Шаги выполняются в такой последовательности: $(1, 2)$, $(n, n-1)$, $(2, 3)$, $(n-1, n-2)$... и так до центральной строки.)

3. Остановиться, если не сделано ни одной перестановки или не происходит уменьшения $\bar{\beta}$ в течение эмпирически установленного числа циклов, равного

$3 + n/100$. Вектор $e_{\mathrm{RI}(j)}$ является j-м столбцом матрицы перестановок P ($j = 1, 2, \ldots, n$), где $\mathrm{RI}(j)$ — содержимое j-й ячейки списка перестановок строк RI.

Четвертый метод уменьшения ширины ленты

Назовем обычное скалярное произведение (не булево) двух строк данной матрицы B *длиною их пересечения*. Пусть ν_i обозначает сумму длин пересечения i-й строки со всеми строками матрицы B. Если мы предположим, что матрица B имеет ленточную форму

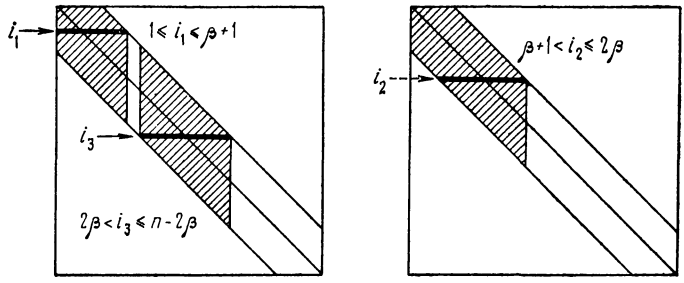

Рис. 3.8.4. Пересечение строк ленточной матрицы.

с шириной ленты $2\beta + 1$, то число единиц в заштрихованных областях на рис. 3.8.4 представляет собой значения ν_i для трех групп значений i. Первая заштрихованная площадь, соответствующая i_1 ($1 \leqslant i_1 \leqslant \beta + 1$), увеличивается вместе с i_1. Это же относится и к заштрихованной площади, соответствующей i_2. Заштрихованная площадь, соответствующая i_3, остается той же при увеличении i_3 от $2\beta + 1$ до $n - 2\beta$. Кроме того, все площади, соответствующие i_1, меньше площадей, соответствующих i_2, которые в свою очередь меньше площадей, соответствующих i_3. Поэтому если молчаливо допустить, что недиагональные единицы внутри ленты распределены случайным образом, то значения всех ν_i будут, вообще говоря, увеличиваться с увеличением i от 1 до $2\beta + 1$ и затем оставаться теми же до значения i, равного $n - 2\beta$. Исходя из симметрии матрицы B, легко заключить,

что для правого нижнего угла матрицы справедливо обратное утверждение. Другими словами, значение ν_i, вообще говоря, уменьшается с увеличением i от $n-2\beta+1$ до n. Заметим также, что любые две строки матрицы B, отстоящие друг от друга более чем на ширину ленты $2\beta+1$, имеют пересечение, равное нулю. Все изложенное выше может быть следующим образом применено для преобразования произвольной разреженной симметричной матрицы к ленточной форме.

Вычислим обычный (не булев) квадрат матрицы B и обозначим его через $\tilde{W}$. Тогда (i, j)-й элемент матрицы $\tilde{W}$ является длиной пересечения i-й строки с j-й строкой и сумма ν_i длин пересечений i-й строки со строками матрицы B равна

$$\nu_i = e_i' \tilde{W} V, \tag{3.8.4}$$

где V есть n-мерный вектор-столбец, все элементы которого единицы. Теперь можно изложить метод в виде следующего алгоритма (Тьюарсон (1971)).

1. Вычислить обычный квадрат матрицы B и обозначить его через $\tilde{W}$ и затем найти ν_i в соответствии с формулой (3.8.4).

2. Вычислить

$$\beta = \max\left(\frac{\tau - n}{2n}, \frac{\rho_{\max} - 1}{2}\right),$$

где τ — общее число единиц в матрице B, $\rho_{\max}$ — максимальное число единиц в одной строке матрицы B. (Замечание. Формула дает оценку снизу для величины β, так как величина $(\tau - n)/2n$ получена из предположения, что лента является полной, а величина $(\rho_{\max} - 1)/2$ — из допущения, что строка с максимальным числом элементов может быть переставлена так, что она войдет в число строк i_3, указанных на рис. 3.8.4, и не имеет ни одного нуля внутри ленты.)

3. Расположить все ν_i в порядке возрастания их величин. Обозначим часть индексов, соответствующих

первым 2β значениям ν_i, через CI, а оставшуюся часть индексов через MI. Разделить группу индексов CI на две подгруппы NW и SE (северо-западный и юго-восточный углы ленточной матрицы) следующим образом (рис. 3.8.5). Определить $\nu_p = \min_i \nu_i$, $i \in$ CI; если имеют место совпадения, то выбирать ν_p с минимальным значением p. Отнести p и все индексы i из CI, для которых $e_p' \tilde{W} e_i \neq 0$, к подгруппе NW. Отнести к подгруппе NW все индексы j из CI, для которых $e_i' \tilde{W} e_j \neq 0$ и $i \in$ NW. Повторить эту процедуру, пока в CI не останется ни одного индекса, который мог бы быть отнесен к подгруппе NW. Другими словами, $e_i' \tilde{W} e_q = 0$ для всех i в NW и $q \notin$ NW, но из CI. Этим определяются все индексы для северо-западного (NW) угла матрицы. Подобным же образом из оставшихся индексов группы CI выделить индексы, которые относятся к юго-восточному SE углу матрицы. Все индексы группы CI, не вошедшие в подгруппы NW и SE, отнести к группе MI. Расположить строки (и столбцы) в подгруппах NW и SE соответственно по возрастающим и убывающим значениям всех ν_i. (Заметим, что всякий раз, когда производятся перестановки строк матрицы B (и A), таким же образом переставляются и столбцы матрицы B (и A), чтобы сохранялась симметрия.) Произвести дополнительные перестановки строк (и столбцов), если требуется, чтобы придать строкам, относящимся к NW и SE, вид заштрихованных областей на рис. 3.8.5. Все эти перестановки строк (столбцов) необходимо регистрировать.

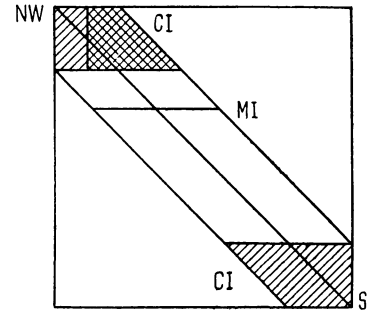

Рис. 3.8.5. Отнесение строки к одному из углов.

4. Отнести каждую из строк группы MI к одной из подгрупп NW или SE следующим образом

(рис. 3.8.5). Пусть V есть n-мерный вектор-столбец, такой, что $e_i'V = 1$, если $i \in$ NW, и $e_i'V = 0$ в противном случае. Тогда вычислить $\hat{v}_p = \max_j (e_j'\tilde{W}V)$ для всех $j \in$ MI. Отнести p к подгруппе NW. Повторить эту процедуру, пока приблизительно половина всех строк группы MI не будет отнесена к подгруппе NW. Записать последовательность, в какой строки из MI отнесены к NW. Таким же способом отнести строки к подгруппе SE и записать порядок, в котором они отнесены к подгруппе SE. (Замечание. Сумма длин пересечений строки $j \in$ MI со строками из подгруппы NW — это дважды заштрихованная область в северо-восточном углу матрицы на рис. 3.8.5. Ясно, что эта область будет максимальна для строки, смежной с последней строкой из NW.)

5. Перестановки строк на шаге 3 и порядок, в котором строки из MI относятся к NW или SE на шаге 4, дают требуемую матрицу перестановок P, так что матрица $\tilde{B} = PBP'$ будет ленточной матрицей.

Существует ряд других подходящих форм для гауссова исключения, но в наши цели не входит описывать в этой главе каждую из них. Многие из этих форм (часть из которых рассматривается в следующем разделе) имеют в своей основе одну из четырех форм BDF, BTF, BNTF и BF, описанных в разд. 3.5, 3.6, 3.7 и 3.8.

3.9. Другие подходящие формы

На рис. 3.9.1 приведены некоторые другие формы, к которым данная матрица может быть приведена (Тьюарсон (1971)). Диагональная блочная форма является частью *односторонне окаймленной диагональной блочной формы* (SBBDF) и *двусторонне окаймленной диагональной блочной формы* (DBBDF). Формы BTF и BNTF являются частями соответственно *окаймленной треугольной блочной формы* (BBTF) и *окаймленной треугольной ленточной формы* (BBNTF). Ленточная матрица связана с *односторонне окаймленной ленточной формой* (SBBF) и *двусторонне окаймленной ленточной формой* (DBBF).

Заметим, что возможны также и комбинации различных форм, представленных на рис. 3.9.1, — две из них приведены на рис. 3.9.2 (Дженнингс (1966), (1968)) Матрицы второго типа являются результатом перенумерации матриц коэффициентов жесткости башенных каркасов.

Если матрица B имеет форму SBBDF, тогда в соответствующем строчном графе вершины, связанные

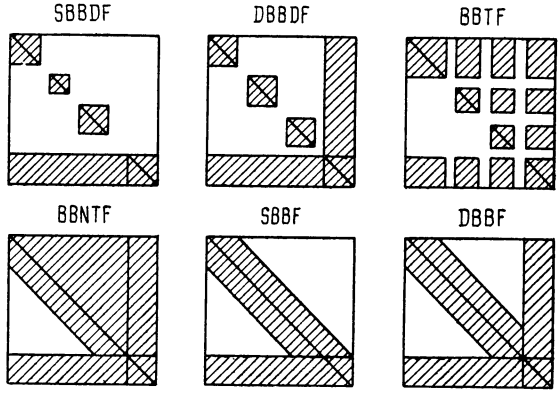

Рис. 3.9.1. Некоторые простые желательные формы.

с окаймляющими строками, имеют, вообще говоря, бóльшую степень, чем другие вершины. Более того, удаление этих вершин разбивает граф на ряд несвязных подграфов, каждый из которых соответствует отдельному диагональному блоку в матрице B. В литературе по теории графов присоединенное множество это такое подмножество вершин, что их удаление (и всех связанных с ними ребер) разбивает граф на два или более несвязных подграфа (Майо (1965)). В энергетических системах трансформаторы связи соответствуют точкам присоединения (Рейд (1971, стр. 125)). Если матрица B имеет форму DBBDF, то вершины, соответствующие окаймляющим строке и столбцу, образуют «присоединенное множество» графа матрицы B. Пусть v_i определено формулой

(3.8.4). Тогда в случаях SBBDF и SBBF мы замечаем, что все окаймляющие строки имеют, вообще говоря, бóльшие ν_i, чем остальные (так как ν_i является суммой длин пересечений i-й строки со всеми строками матрицы B). Этот факт может быть использован для определения окамляющих строк. В случаях DBBDF и DBBF также легко определить строки и столбцы, принадлежащие окаймлению, так как им

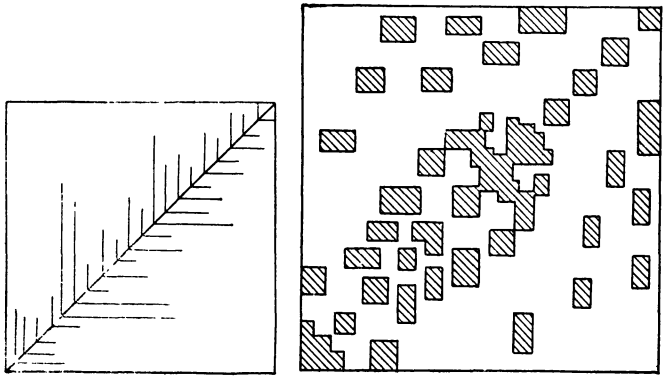

Рис. 3.9.2. Две другие желательные формы.

соответствуют, вообще говоря, бóльшие значения ν_i, чем остальным строкам и столбцам (Огбуобири и др. (1970)).

В случаях BBTF и BBNTF, если переместить окаймляющие строки на самый верх, получим слегка измененные формы BTF и BNTF соответственно (см. рис. 3.9.3). Очевидно, мы не можем пользоваться методами разд. 3.6 для приведения данной матрицы к модифицированной форме BTF, представленной на рис. 3.9.3. Однако второй метод, изложенный в разд. 3.7, может быть применен для преобразования данной матрицы к треугольной ленточной форме, приведенной на рис. 3.9.3. За этим могут последовать (всякий раз, когда это возможно) дальнейшие перестановки в соответствии с теоремой 3.7.13 для получения возможно большего числа уголковых блоков,

так как первая матрица на рис. 3.9.3 может рассматриваться как форма BNTF с уголковыми блоками.

Мы описали некоторые простые практические методы обращения с окаймленными матрицами. Они достаточно хороши для матриц, которые могут быть преобразованы к этим формам так, что ненулевые элементы равномерно распределены в заштрихованных областях. В настоящее время нет определенного алгоритма (подобного тем, что даны в разд. 3.5 и 3.6)

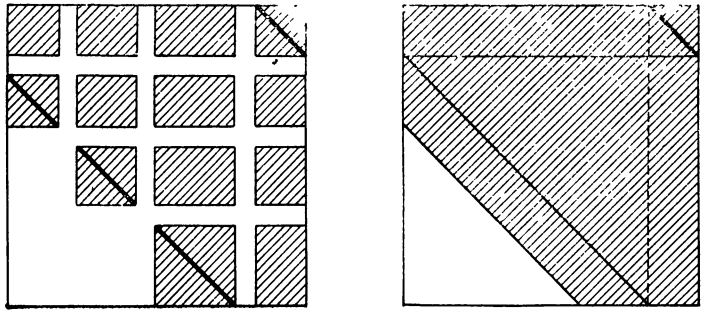

Рис. 3.9.3. Модифицированные формы BTF и BNTF.

для определения строк и (или) столбцов, которые принадлежали бы к окаймлению (Харари 1971 б) в том случае, когда только небольшое число ненулевых элементов составляют это окаймление. Желательно было бы иметь простой практический метод определения наименьшего числа вершин графа (или направленного графа), удаление которых делает граф менее связным.

Опишем теперь метод разрезания Стьюарда (1969) (или удаления дуг направленного графа для разбиения циклов). Он особенно полезен в случаях, когда удаление небольшого числа дуг разбивает большие направленные циклы и соответствующие диагональные блоки в форме BTF становятся меньше (см. разд. 3.6). Этот метод практичен только для разрезания небольших блоков. Мы поясним его на примере. Рассмотрим матрицу и ее граф, приведенные на

рис. 3.9.4. Определим один из наиболее длинных циклов [1, 4, 3, 6, 5, 1]. Два пути между теми же вершинами, направленные в одну сторону, не имеющие

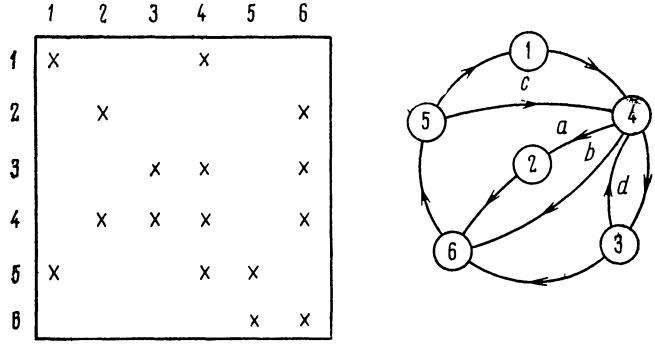

Рис. 3.9.4. Матрица и ее направленный граф.

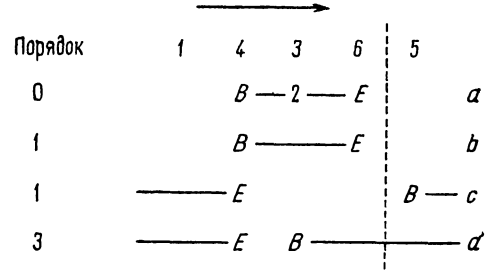

Рис. 3.9.5. Диаграмма шунтирования длинного цикла.

общей дуги и не содержащие циклов, называются *параллельными*. Путь, параллельный дуге в длинном цикле, называется *шунтом*. На рис. 3.9.4 четыре шунта помечены *a*, *b*, *c* и *d*. *Порядок* шунта — это длина шунтированного пути в длинном цикле минус длина шунта. Например, дуга *c* — это [5, 4], а путь длинного цикла, который она укорачивает, что [5, 1, 4], и, следовательно, порядок дуги *c* равен 2 — 1 = 1. Если дуга длинного цикла, имеющая шунт, разрезается,

то остается цикл, проходящий через шунт, и длина всего цикла уменьшится на порядок шунта. Длинный цикл разрезается в тех местах, где нет слишком большого числа шунтов, так как шунты тоже подлежат разрезанию. На рис. 3.9.5 дана диаграмма шунтирования длинного цикла графа, приведенного на рис. 3.9.4. На диаграмме B обозначает начало шунта, а E — его конец. Очевидно, если разрезать дугу [6, 5], то потребуется разрезать только шунт d, и поэтому такой выбор места разреза является наилучшим. Если дугу [6, 5] и шунт [3, 4] удалить (вырезать) из направленного графа на рис. 3.9.4, то вершина 5 становится эмиттером, и она помечается первой вершиной. Теперь, при удалении дуг [5, 1] и [5, 4], становится эмиттером вершина 1 — она помечается второй вершиной и так далее. Матрица, соответствующая перемеченному графу, представлена на рис. 3.9.6. Она имеет форму BTF, если не считать элемент в нижнем левом углу матрицы, подлежащий удалению.

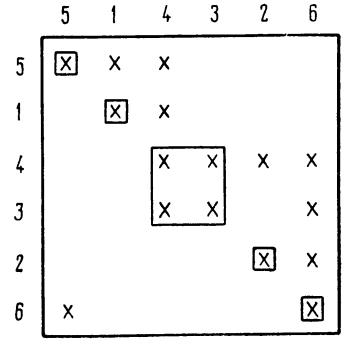

Рис. 3.9.6. Матрица, связанная с перемеченным направленным графом.
В левом нижнем углу находится элемент, подлежащий удалению.

Матрица на рис. 3.9.6 является частным случаем формы BBTF, когда окаймляющая группа блоков состоит всего из единственного недиагонального элемента. Если в процессе разрезания графа нужно удалить ряд элементов, то строки и столбцы матрицы, содержащие эти элементы, могут быть переставлены так, чтобы они стали последними строками и столбцами и матрица приобрела бы форму BBTF. Теперь покажем, каким образом для форм BTF и BBTF могут быть получены элиминативные формы обратных матриц (EFI).

3.10. Обратные матрицы для BTF и BBTF

Рассмотрим форму BBTF, которая дается формулой (3.3.1). Пусть все матрицы A_{ii} ($i = 1, 2, \ldots, p$) неособенные. Метод обращения может быть описан следующим алгоритмом (Тьюарсон (1972); Дафф (1972)).

1. Выполнить следующие шаги для $i = p-1, p-2, \ldots, 2, 1$:

а) Преобразовать матрицу A_{ii} в верхнюю треугольную матрицу U_{ii} с равными единице диагональными элементами и матрицу A_{pi} в нулевую матрицу с помощью прямого *гауссова исключения* (GE). Это приведет к заполнению подматриц A_{ip} и A_{pp}.

б) Применить обратную подстановку гауссова исключения для преобразования матрицы U_{ii} в единичную матрицу I и матриц A_{ji} в нулевые матрицы для всех $j < i$. Это приводит к заполнению всех матриц A_{jp}, $j \leqslant i$.

2. Прямым гауссовым исключением преобразовать матрицу A_{pp} к верхней треугольной матрице U_{pp}, а обратной подстановкой преобразовать матрицу U_{pp} в единичную матрицу I и все модифицированные A_{jp}, $j \neq p$, к нулю.

Очевидно, в этом методе не будет никакого заполнения матриц A_{ji}, $j < i$ и $i \neq p$.

Если задана матрица в форме BTF, то для каждого i ($i = p, p-1, \ldots, 2, 1$) выполняются два шага:

а) преобразовать матрицу A_{ii} в матрицу U_{ii},

б) привести матрицу U_{ii} к единичной матрице I и матрицы A_{ji}, $j < i$, к нулю.

Ясно, что в этом случае не будет заполнения матриц A_{ji}, $j = i$. Обе приведенные выше модификации гауссова исключения легко реализовать, и полученные обратные матрицы в форме EFI являются разреженными. В обоих методах могут быть использованы изложенные в этой и предыдущей главах способы уменьшения заполнения при преобразовании матрицы A_{ii} в матрицу U_{ii}.

3.11. Библиография и комментарии

Интерпретация с помощью теории графов различных подходящих форм матриц, рассмотренных в настоящей главе, дана Харари (1971, а, б). Метод, который аналогичен первому методу разд. 3.7 и минимизирует $\sum_i \psi_i$ вместо $\max_i \psi_i$, предложен Тьюарсоном (1967с). Олвей и Мартин (1965) получили программу, осуществляющую направленный поиск возможных перестановок для определения такой, которая преобразует матрицу к форме BF. Программы для ЭВМ по первому и третьему методам разд. 3.8 разработали Розен (1968) и Эйкиуз и Утку (1968). Методы теории графов для гауссова исключения даны Партером (1961) и Роузом (1970б). Эквивалентность гауссова исключения и исключения узлов показана Огбуобири и др. (1970). Партер (1960) и Меримонт (1969) используют разрезание. Применение методов теории графов для анализа структур и разбиения матриц дано Венке (1964), Ойфингером и др. (1968), Ойфингером (1970). Хорошая схема хранения матриц различных форм, приведенных на рис. 3.9.2, описана Дженнингсом (1966), (1968). Алгоритм решения для симметричных положительно определенных ленточных матриц дается в работе Кантина (1971). Другой алгоритм для симметричной матрицы дается Иенсеном и Парксом (1970). Разбиение матриц и исключение блоков рассматриваются в работах Джорджа (1972) и Роуза и Банча (1972). Для нахождения «точек присоединения» можно пользоваться также теорией рассечения (Бэти и Стьюарт (1971)), которая также называется «теорией декомпозиции» в линейном программировании (Данциг и Вулф (1961); Бендерс (1962)). Она включает в себя определение матрицы близости узлов с помощью обычных (не булевых) степеней матрицы B. Элементы матрицы близости узлов затем используются для определения присоединенного множества.

Глава 4
ПРЯМОЕ ТРЕУГОЛЬНОЕ РАЗЛОЖЕНИЕ

4.1. Введение

В этой главе мы опишем методы представления заданной матрицы A как произведения нижней треугольной матрицы $\tilde{L}$ и верхней треугольной матрицы U вида

$$A = \tilde{L}U. \tag{4.1.1}$$

Нас будет главным образом интересовать вопрос о пригодности этих методов для случая больших разреженных матриц. Эти методы связаны с именами Краута, Дулитла, Холецкого, Банахевича и других (Уэстлейк (1968); Уилкинсон (1965)).

Если разложение (4.1.1) известно, то из него следует, что

$$A^{-1} = U^{-1}\tilde{L}^{-1}. \tag{4.1.2}$$

Разложение матриц U^{-1} и $\tilde{L}^{-1}$ на множители нетрудно определить, так как обе они треугольные матрицы. Поэтому для представления обратной матрицы A^{-1} в факторизованной форме можно применить приведенный выше метод вместо метода, изложенного в гл. 2. Если решение системы уравнений $Ax = b$ желательно иметь только для одной правой части, то нет необходимости вычислять обратную матрицу по формуле (4.1.2). В этом случае решение x может быть получено применением обратной подстановки к системам

$$\tilde{L}y = b \tag{4.1.3}$$

и

$$Ux = y \tag{4.1.4}$$

соответственно для y и x.

Методы, описанные в этой главе, являются по существу модификациями гауссова исключения, изложенного в гл. 2. Они обладают тем преимуществом, что промежуточные редуцированные матрицы $A^{(k)}$, о которых говорилось в гл. 2, не запоминаются, и, кроме того, если скалярные произведения вычисляются с накоплением, эти методы обеспечивают исключительную точность.

4.2. Метод Краута

Обозначим (i, j)-е элементы L и U соответственно через l_{ij} и u_{ij}. Потребуем, чтобы матрица U была верхней треугольной матрицей с единицами на диагонали, т. е. $u_{kk} = 1$, $k = 1, 2, \ldots, n$. Если допустить, что первые $k-1$ строк и столбцов матриц L и U уже определены (см. рис. 4.2.1), тогда, учитывая формулу (4.1.1) и то, что $l_{ip} = 0$ для $p > i$, $u_{pk} = 0$ для $p > k$ и $u_{kk} = 1$, имеем

$$a_{ik} = l_{ik} + \sum_{p=1}^{k-1} l_{ip} u_{pk}, \qquad i \geqslant k,$$

что дает

$$l_{ik} = a_{ik} - \sum_{p=1}^{k-1} l_{ip} u_{pk}, \qquad i \geqslant k. \tag{4.2.1}$$

Таким образом, k-й столбец матрицы L теперь известен. Из формулы (4.1.1) и из того, что $l_{kp} = 0$ для $p > k$, имеем

$$a_{kj} = l_{kk} u_{kj} + \sum_{p=1}^{k-1} l_{kp} u_{pj}, \qquad j > k,$$

что дает

$$u_{kj} = \frac{1}{l_{kk}} \left(a_{kj} - \sum_{p=1}^{k-1} l_{kp} u_{pj} \right), \qquad j > k, \tag{4.2.2}$$

и теперь известна k-я строка U. Таким образом, мы убедились, что если первые $k-1$ строк матрицы U и первые $k-1$ столбцов матрицы L известны, то k-я строка матрицы U и k-й столбец матрицы L могут быть легко вычислены. Первый столбец матрицы L

определяется равенствами

$$l_{i1} = a_{i1}, \qquad i = 1, 2, \ldots, n. \tag{4.2.3}$$

Это следует из формулы (4.1.1) и из того, что первым столбцом матрицы U является вектор e_1. Первую строку матрицы U так же легко найти из формулы (4.1.1)

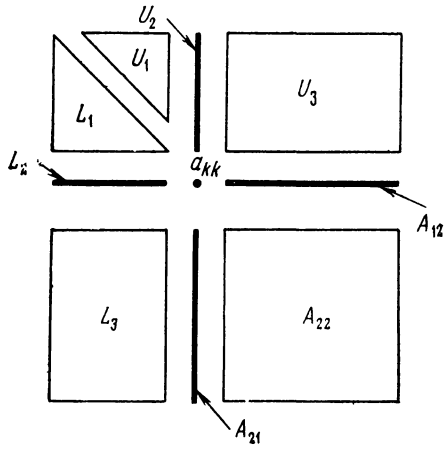

Рис. 4.2.1. Схема хранения для метода Краута.

и на основании того факта, что первая строка матрицы $\tilde{L}$ есть $l_{11}e_1'$, а именно

$$u_{1j} = \frac{a_{1j}}{l_{11}}, \qquad j > 1. \tag{4.2.4}$$

Если допустить, что для $k = 1$ сумма $\sum_{p=1}^{k-1}(\ldots) = 0$, то формулы (4.2.3) и (4.2.4) являются частными случаями соответственно формул (4.2.1) и (4.2.2). Принимая во внимание изложенное выше, все строки и столбцы матриц $\tilde{L}$ и U легко могут быть определены.

Рис. 4.2.1 наглядно показывает, каким образом хранятся различные матрицы к началу k-го шага метода Краута. Первые $k-1$ строк матрицы U — это $[I + U_1, U_2, U_3]$, а первые $k-1$ столбцов матрицы $\tilde{L}$ —

это

$$\begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix}.$$

Заметим, что единичная диагональ матрицы U не хранится. На k-м шаге мы прежде всего используем уравнение (4.2.1) для преобразования элемента a_{kk} и столбца A_{21} в нетривиальные элементы k-го столбца матрицы $\tilde{L}$. Имеем

$$\begin{bmatrix} \hat{a}_{kk} \\ \hat{A}_{21} \end{bmatrix} = \begin{bmatrix} a_{kk} \\ A_{21} \end{bmatrix} - \begin{bmatrix} L_2 \\ L_3 \end{bmatrix} U_2, \qquad (4.2.5)$$

причем $l_{kk} = \hat{a}_{kk}$ и $l_{i+k,\,k} = e_i' \hat{A}_{21}$. Затем следует вычисление нетривиальных элементов k-й строки матрицы U по формуле (4.2.2). Имеем

$$\hat{A}_{12} = (A_{12} - L_2 U_3)/\hat{a}_{kk}, \qquad (4.2.6)$$

причем $u_{k,\,j+k} = \hat{A}_{12} e_j$. Для полного определения матриц $\tilde{L}$ и U необходимо применить следующий порядок вычислений: первый столбец матрицы $\tilde{L}$, первая строка матрицы U, второй столбец матрицы $\tilde{L}$, вторая строка матрицы U и так далее. Как только матрицы $\tilde{L}$ и U становятся известными, легко перейти к формулам (4.1.3) и (4.1.4) для вычисления y и x следующим образом:

$$y_k = \frac{b_k - \sum_{p=1}^{k-1} l_{kp} y_p}{l_{kk}}, \qquad k = 1, 2, \ldots, n \quad (4.2.7)$$

и

$$x_k = y_k - \sum_{p=k+1}^{n} u_{kp} x_p, \qquad k = n, n-1, \ldots, 1, \quad (4.2.8)$$

причем суммы $\sum_{p=1}^{0} (\ldots)$ и $\sum_{p=n+1}^{n} (\ldots)$ полагаются равными нулю.

Так как формулы (4.2.7) и (4.2.2) по существу одинаковы, можно элементы вектора-столбца y получать при вычислении элементов u_{kj} по методу Краута, применяя формулу (4.2.2) к расширенной матрице $(A|b)$ вместо матрицы A и полагая $a_{k,\,n+1}=b_k$ и $u_{p,\,n+1}=y_p$.

Точность метода Краута может быть повышена применением частичного упорядочения при выборе главного элемента следующим образом (Уилкинсон (1965)). Вычисляем l_{ik} согласно формуле (4.2.1), затем определяем

$$|l_{sk}|=\max_i |l_{ik}|, \qquad i\geqslant k, \tag{4.2.9}$$

и переставляем s-ю и k-ю строки матриц L и A. Затем применяем формулы (4.2.1) для $i>k$ (для вычисления остальных элементов k-го столбца матрицы L [1])) и (4.2.2). Каждая такая перестановка строк записывается, и в результате получается матрица перестановок P, такая, что

$$PA=\hat{L}\hat{U}$$

(где L — нижняя треугольная матрица, а U — верхняя треугольная матрица с единичной диагональю). Из последнего уравнения следует, что

$$A^{-1}=\hat{U}^{-1}\hat{L}^{-1}P. \tag{4.2.10}$$

Из сравнения формул (4.2.10) и (4.1.2) очевидно, что, какие бы перестановки строк матрицы A ни производились в процессе треугольного разложения, для получения обратной матрицы A^{-1} требуется произвести такие же перестановки столбцов матрицы $\hat{U}^{-1}\hat{L}^{-1}$. Уилкинсон (1965) доказал, что треугольное разложение с частичным упорядочением является исключительно точным, если скалярные произведения в формулах (4.2.1) и (4.2.2) могут вычисляться с накоплением.

[1]) Они уже вычислены для выбора s по формуле (4.2.9). — *Прим. ред.*

Заметим, что если $l_{kk} = 0$, то вычисления по формуле (4.2.2) производить нельзя. Но если применять формулу (4.2.9), то этого не будет. В этом случае l_{sk} не может быть нулем, так как это означало бы, что k-й столбец нижней треугольной матрицы в треугольном разложении матрицы A равен нулю и матрица A особенная.

Между гауссовым исключением и прямым треугольным разложением Краута существует следующая связь (Уилкинсон (1965)). Для неособенной матрицы A матрицы $\tilde{L}$ и U являются единственными, если они существуют и одна из них является треугольной матрицей с единичной диагональю. Поэтому из уравнений (2.2.7) и (4.1.1) следует, что $\tilde{L} = L^{-1}$ и матрицы U в обоих этих уравнениях идентичны. Если при применении гауссова исключения и метода Краута производятся одни и те же перестановки строк и столбцов, то между этими методами опять же существует вышеупомянутая зависимость.

4.3. Минимизация заполнения для метода Краута

В начале k-го шага метода Краута можно выбрать ненулевой элемент из последних $n - k + 1$ строк и столбцов матрицы A и переместить его в (k, k)-ю позицию так, чтобы заполнение было минимальным. Для определения такого элемента требуется следующее (Тьюарсон (1969а)). Пусть B_k обозначает матрицу, которая получается, если все ненулевые элементы матрицы, представленной на рис. 4.2.1, заменить единицами. Обозначим через $b_{ij}^{(k)}$ (i, j)-й элемент матрицы B_k и пусть S_k, T_k и N_k обозначают подматрицы $b_{ij}^{(k)}$, $i \geqslant k$, $j < k$; $b_{ij}^{(k)}$, $i < k$, $j \geqslant k$ и $b_{ij}^{(k)}$, $i, j \geqslant k$. Определим

$$\Lambda_k = S_k * T_k, \tag{4.3.1}$$

где $*$ означает булево умножение матриц или, другими словами, обычное умножение матриц с дополнительным условием, что $1 + 1 = 1$. Пусть $\bar{\Lambda}_k$ есть матрица, полученная из матрицы Λ_k заменой каждого ее

ненулевого элемента нулем, а нулей — единицами, и пусть

$$\Delta_k = \bar{\Lambda}_k \oplus N_k, \tag{4.3.2}$$

где $\oplus$ означает булево суммирование матриц ($1 \oplus 1 = 1$). Если V есть $(n-k+1)$-мерный вектор-столбец, все элементы которого единицы, то определим

$$\bar{c}^{(k)} = V'\Delta_k \tag{4.3.3}$$

и

$$\bar{r}^{(k)} = \Delta_k V. \tag{4.3.4}$$

Если $\bar{r}_\alpha^{(k)}$ и $\bar{c}_\beta^{(k)}$ обозначают α-й и β-й элементы соответственно векторов $\bar{r}^{(k)}$ и $\bar{c}^{(k)}$, то имеем следующую теорему.

Теорема 4.3.5. *Если* $\bar{r}_s^{(k)} + \bar{c}_t^{(k)} = \max\limits_{\alpha,\beta} (\bar{r}_\alpha^{(k)} + \bar{c}_\beta^{(k)})$ *и пренебречь возможностью полного взаимного уничтожения слагаемых при вычислении скалярных произведений в формулах* (4.2.1) *и* (4.2.2), *то перемещение элемента* $a_{s+k-1,\, t+k-1}$ *в* (k, k)*-ю позицию в начале* k*-го шага метода Краута приводит к наименьшему локальному заполнению.*

Доказательство. Если $e'_\alpha \Delta_k e_\beta = 1$, то из формулы (4.3.2) следует, что $e'_\alpha \bar{\Lambda}_k e_\beta$ и $e'_\alpha N_k e_\beta$ не могут одновременно быть равны нулю. Принимая во внимание, что матрица $\bar{\Lambda}_k$ была получена из матрицы Λ_k заменой ее ненулевых элементов нулями, а нулей — единицами, и учитывая уравнение (4.3.1), находим, что уравнения $e'_\alpha (S_k * T_k) e_\beta = 1$ и $e'_\alpha N_k e_\beta = 0$ не могут быть одновременно верны, если $e'_\alpha \Delta_k e_\beta = 1$. Если перед k-м шагом метода Краута переставить k-й и $(\beta + k - 1)$-й столбцы матрицы, представленной на рис. 4.2.1, то на основании определения S_k, T_k и N_k и того обстоятельства, что мы пренебрегаем возможностью взаимного уничтожения слагаемых при вычислении скалярных произведений, имеем

$$\sum_{p=1}^{k-1} l_{ip} u_{pk} \neq 0 \Leftrightarrow e'_\alpha (S_k * T_k) e_\beta = 1,$$

причем $\alpha + k - 1 = i$ и элемент в (i, k)-й позиции матрицы, в которой произведена перестановка, будет равен нулю тогда и только тогда, когда

$$e'_\alpha N_k e_\beta = 0.$$

Если равенства $e'_\alpha N_k e_\beta = 0$ и $e'_\alpha (S_k * T_k) e_\beta = 1$ одновременно удовлетворяются, то, согласно формуле (4.2.1), имеет место заполнение, так как нуль в (i, k)-й позиции станет ненулевым элементом. Таким образом, мы видим, что если $e'_\alpha \Delta_k e_\beta = 1$ и $(\beta + k - 1)$-й и k-й столбцы переставлены перед k-м шагом метода Краута, то в (i, k)-й позиции заполнения нет. Общее число таких позиций, в которых не может быть заполнения [1]), равно $V' \Delta_k e_\beta$ или $\bar{c}_\beta^{(k)}$, если учесть формулу (4.3.3). Точно таким же способом можно показать, что $\bar{r}_\alpha^{(k)}$ представляет собой общее число позиций [2]), в которых не будет заполнения, если перед k-м шагом метода Краута будут представлены k-я и $(\alpha + k - 1)$-я строки. Поэтому наименьшее локальное заполнение будет иметь место, если перед k-м шагом метода Краута переставить k-ю и $(s + k - 1)$-ю строки и k-й и $(t + k - 1)$-й столбцы, причем $\bar{r}_s^{(k)} + \bar{c}_t^{(k)} = \max_{\alpha, \beta} (\bar{r}_\alpha^{(k)} + \bar{c}_\beta^{(k)})$. Этим завершается доказательство теоремы.

Существенно, чтобы выбранный согласно изложенной теореме элемент $a_{\hat{s}\hat{t}}$, где $\hat{s} = s + k - 1$, $\hat{t} = t + k - 1$, после того как его значение будет изменено по формуле (4.2.1), не стал меньше ε — допустимого значения главного элемента, т. е. чтобы выполнялось условие

$$\left| a_{\hat{s}\hat{t}} - \sum_{p=1}^{k-1} l_{\hat{s}p} u_{p\hat{t}} \right| > \varepsilon. \qquad (4.3.6)$$

---

[1]) Имеется в виду в первоначальном столбце $\beta + k - 1$. — *Прим. ред.*

[2]) Имеется в виду в первоначальной строке $\alpha + k - 1$. — *Прим. ред.*

Выбор ε уже рассматривался в разд. 2.3. Обычно нельзя проверить выполнимость условия (4.3.6) прежде, чем применять теорему 4.3.5, так как это потребовало бы больших вычислений и было бы аналогично «полному упорядочению» в методе Краута, что не рекомендуется (Уилкинсон (1965)). На практике сперва выбирают небольшое число элементов, приводящих к минимуму заполнения или к заполнению, близкому к минимуму, и затем проверяют выполнимость условия (4.3.6), прежде чем выбрать один из этих элементов в качестве *главного* на k-м шаге.

Можно применить теорему 4.3.5 при моделировании метода Краута, если начать с матрицы B_1 (матрицы, полученной из матрицы A путем замены всех ее ненулевых элементов единицами) и использовать булевы умножения и сложение в формулах (4.2.1) и (4.2.2) для регистрации заполнения. Таким образом, для всех шагов могут быть «априорно» выбраны главные элементы и в матрице A выполнены перестановки, в результате которых эти главные элементы располагались бы на главной диагонали, прежде чем действительный метод Краута был бы применен. Это может быть осуществлено, если ни один из вычисленных главных элементов не меньше, чем ε. Если матрица A положительно-определенная, то главные элементы могут быть выбраны указанным выше способом [1]) (Тинни и Уолкер (1967); Густавсон и др. (1970)). Положительно-определенные матрицы и симметричные матрицы, которые могут и не быть положительно-определенными, рассматриваются в разд. 4.5.

4.4. Метод Дулитла (Блэка)

Этот метод является вариантом метода Краута, в котором на k-м шаге вычисляются только k-е строки матриц L и U (Уэстлейк (1968); Тинни и Уокер (1967)). Это является преимуществом, если матрица A хранится по строкам. В этом методе элементы k-й строки матрицы L вычисляются слева направо с по-

[1]) Среди диагональных элементов. — *Прим. ред.*

мощью формулы

$$l_{kj} = a_{kj} - \sum_{p=1}^{j-1} l_{kp} u_{pj}, \qquad j = 1, 2, \ldots, k, \quad (4.4.1)$$

а элементы k-й строки матрицы U вычисляются, как и в методе Краута, по формуле (4.2.2).

Порядок вычислений таков: первая строка матрицы L, первая строка матрицы U, вторая строка матрицы L, вторая строка матрицы U и так далее.

В этом методе для минимизации локального заполнения нельзя пользоваться теоремой 4.3.5, так как к началу k-го шага известна только первая строка подматрицы S_k [1]). Возможно, лучше всего моделировать сначала метод Краута, а затем априорно выбрать главные элементы, как об этом упоминалось в предыдущем параграфе.

Можно также получать матрицы L и U не по строкам, а по столбцам, вычислив

$$u_{ik} = \frac{a_{ik} - \sum_{p=1}^{i-1} l_{ip} u_{pk}}{l_{ii}}, \qquad i = 1, 2, \ldots, k-1, \quad (4.4.2)$$

и затем применив формулу (4.2.1) для определения l_{ik}, $i \geqslant k$. Этот вариант метода Краута полезен тогда, когда ненулевые элементы матрицы A хранятся по столбцам.

В следующем разделе рассмотрим LU-разложение для симметричных матриц.

4.5. Метод Холецкого (квадратных корней, Банахевича)

Хорошо известно, что если матрица A неособенная и симметричная, то она может быть представлена в виде $A = LL'$ или $A = U'U$. При этом матрица A должна быть упорядочена так, чтобы ни одна из ее северо-западных главных подматриц не была особенной. Матрица L — это единственная нижняя треуголь-

[1]) Так же, как и подматрицы N_k. — *Прим. ред.*

ная матрица[1]), а матрица U — это единственная верхняя треугольная матрица. Если $A = U'U$, то

$$\sum_{p=1}^{k} u_{pk}u_{pj} = a_{kj}, \qquad k \leqslant j. \tag{4.5.1}$$

Поэтому для элементов k-й строки матрицы U имеем

$$u_{kk} = \left(a_{kk} - \sum_{p=1}^{k-1} u_{pk}^2\right)^{\frac{1}{2}} \tag{4.5.2}$$

и

$$u_{kj} = \frac{a_{kj} - \sum_{p=1}^{k-1} u_{pk}u_{pj}}{u_{kk}}, \qquad j > k. \tag{4.5.3}$$

В этих формулах принято, что $\sum_{p=1}^{k-1}(\ldots) = 0$ для $k = 1$. Для вычисления строк матрицы U формулы применяются поочередно. Если матрица A не положительно-определенная, то диагональные элементы u_{kk} могут быть комплексными числами.

Если A — положительно-определенная симметричная матрица, то метод Холецкого является наилучшим для треугольного разложения (Уилкинсон (1965)). Для того чтобы ошибки округления были малыми, не требуется никаких перестановок строк или столбцов. С другой стороны, если матрица A симметричная, но не положительно-определенная, то для удержания ошибок округления в разумных пределах необходимо производить выбор главных элементов, и это нарушает симметрию. Если на каждом шаге брать наибольший диагональный элемент, то это сохранит симметрию, но не гарантирует устойчивость с точки зрения ошибок округления.

В свете сказанного разложение Холецкого применимо к симметричным разреженным матрицам, если произвольный порядок разложения не влияет отрицательно на точность вычислений. К счастью, во многих

[1]) Дающая факторизацию данной матрицы A. — *Прим. ред.*

практических приложениях это имеет место (например, см. Тинни и Уолкер (1967)).

Для минимизации заполнения в методе Холецкого можно использовать теорему 4.3.5. Однако в этом случае

$$S_k = T'_k, \tag{4.5.4}$$

что с учетом формул (4.3.1) и (4.3.2), очевидно, означает, что Δ_k симметричная. Поэтому из равенств (4.3.3) и (4.3.4) можно заключить, что

$$\bar{c}^{(k)} = (\bar{r}^{(k)})'. \tag{4.5.5}$$

Имея в виду вышеизложенное, можно вместо теоремы 4.3.5 использовать следующее следствие.

Следствие 4.5.6. *Если* $\bar{r}_s^{(k)} = \max\limits_{\alpha}(\bar{r}_\alpha^{(k)})$ *и мы пренебрегаем возможностью взаимного уничтожения слагаемых в скалярных произведениях формулы* (4.5.3), *то перемещение элемента* $a_{s+k-1,\, s+k-1}$ *в* (k, k)*-ю позицию в начале* k*-го шага метода Холецкого приводит к наименьшему заполнению.*

Доказательство. На любом шаге метода Холецкого только диагональные элементы матрицы A могут быть переставлены друг с другом, в противном случае будет нарушена симметрия. Поэтому в теореме 4.3.5 необходимо брать $\alpha = \beta$. Учитывая равенство (4.5.5), отсюда получаем

$$\bar{r}_s^{(k)} + \bar{c}_s^{(k)} = \max_{\alpha}(\bar{r}_\alpha^{(k)} + \bar{c}_\alpha^{(k)})$$

или

$$\bar{r}_s^{(k)} = \max_{\alpha} \bar{r}_\alpha^{(k)}.$$

Начиная с этого места, доказательство следствия такое же, как и доказательство теоремы 4.3.5 (с l_{kp}, замененным на u_{pk}).

Изложенное выше следствие можно применить при моделировании метода Холецкого, если исходить из верхней треугольной части матрицы B_1 (полученной из матрицы A путем замены в ней ненулевых элементов единицами) и произвести булевы умножения и суммирования в формуле (4.5.3) для регистрации заполне-

ния. Заметим, что нет нужды моделировать формулу (4.5.2). Кроме того, не нужно производить деления в формуле (4.5.3) при булевом моделировании, так как знаменатель всегда равен единице. Поэтому может быть определена матрица перестановок P и матрица A упорядочена с помощью уравнения $\hat{A} = PAP$. Действительный метод Холецкого применяется затем к матрице $\hat{A}$.

4.6. Подходящие формы для треугольного разложения

В главе 3 были приведены некоторые элементарные формы, к которым может быть априорно преобразована данная матрица так, что заполнение, если оно имеется, ограничивается определенными областями этих форм.

Теперь покажем, что такие же формы желательны и для треугольного разложения. В конце разд. 4.2 мы указали на то, что $\tilde{L} = L^{-1}$, где $\tilde{L}$ и L — нижние треугольные матрицы, полученные соответственно в методе Краута и в гауссовом исключении, при условии, что порядок и выбор главных элементов в обоих случаях одинаковы. Теперь, учитывая формулу (2.2.6), имеем

$$L_n \ldots L_2 L_1 \tilde{L} = I_n,$$

и из формул (2.2.3) и (2.2.4) является очевидным, что матрица L_k преобразует k-й столбец матрицы $L_{k-1} \ldots \ldots L_2 L_1 \tilde{L}$ к вектору e_k, а все остальные столбцы остаются без изменений. Таким образом, можно заключить, что

$$\eta_i^{(k)} = 0, \qquad i < k,$$
$$\eta^{(k)} = \frac{1}{l_{kk}} \quad \text{и} \quad \eta_i^{(k)} = -\frac{l_{ik}}{l_{kk}}, \qquad i > k. \tag{4.6.1}$$

Из всего сказанного с очевидностью следует, что элементы матрицы $\tilde{L}$ и нетривиальные элементы множителей в разложении L имеют одну и ту же структуру распределения ненулевых элементов и могут поэтому храниться в одинаковых ячейках. Таким образом, в

обоих методах, и в методе Краута, и в гауссовом исключении, заполнение одно и то же, и поэтому для обоих методов желательны одни и те же формы.

Если A является симметричной матрицей или матрицей с симметричной структурой распределения ненулевых внедиагональных элементов, то ленточная форма (BF), двусторонне окаймленная ленточная форма (DBBF), диагональная блочная форма (BDF), двусторонне окаймленная диагональная блочная форма (DBBDF) или комбинации этих форм являются желательными для метода Краута (для метода Холецкого, если $A = A'$).

Если матрица A несимметричная, она может быть преобразована перед применением метода Краута к односторонне окаймленной ленточной форме (SBBF), односторонне окаймленной диагональной блочной форме (SBBDF), треугольной ленточной форме (BNTF), окаймленной треугольной ленточной форме (BBNTF), треугольной блочной форме (BTF), окаймленной треугольной блочной форме (BBTF) или к комбинации этих форм.

4.7. Библиография и комментарии

Задачу оптимального упорядочения для треугольного разложения матриц, встречающихся в некоторых практических приложениях, рассматривали Карпентьер (1963), Сато и Тинни (1963), Эдельман (1963, 1968), Чан (1969), Мак-Кормик (1969), Тинни (1969), Ашкенази (1971) и Дженнингс и Тафф (1971). Во многих задачах электрических цепей треугольное разложение является особенно полезным (Тинни и Уокер (1967); Эрисман (1972)). Некоторые теоретические результаты сравнения метода Краута с другими методами даны Брейтоном и др. (1969).

Метод Краута может быть обобщен в том смысле, что для любого заданного k мы полагаем или $u_{kk} = 1$, или $l_{kk} = 1$. В этом случае число делений часто может быть сокращено, если группа, имеющая меньшее число ненулевых элементов, нормализована (Густавсон и др. (1970)).

Глава 5
ИСКЛЮЧЕНИЕ ГАУССА — ЖОРДАНА

5.1. Введение

Если на каждом шаге гауссова исключения (см. разд. 2.2) исключаются не только ненулевые элементы под диагональю, но и те ненулевые элементы, которые находятся над диагональю, то процесс называется *исключением Гаусса — Жордана*. Таким образом, заданная матрица коэффициентов A приводится непосредственно к единичной матрице в противоположность методу Гаусса, когда матрица A первоначально преобразуется к верхней треугольной (с единичной диагональю) матрице U, которая затем приводится к единичной матрице.

В этой главе мы опишем метод исключения Гаусса — Жордана и покажем, каким образом используются матрицы, связанные с различными шагами процесса исключения, для представления обратной матрицы A^{-1} в форме разложения на множители, которая называется *мультипликативной формой обратной матрицы* (PFI). Мы также покажем, каким образом для данной разреженной матрицы может быть получена разреженная форма PFI.

5.2. Основной метод

При исключении Гаусса — Жордана (GJE) к данной матрице A применяется последовательность элементарных преобразований для приведения ее к единичной матрице I_n. Та же последовательность преобразований, примененная к вектору b в правой части системы уравнений $Ax = b$, дает решение (Фаддеев и Фаддеева (1960); Фокс (1965)).

Пусть $A^{(k)}$ обозначает матрицу в начале k-го шага исключения, причем $k = 1, 2, \ldots, n$ и $A^{(1)} \equiv A$, а

$A^{(n+1)} = I_n$. Обозначим (i, j)-й элемент матрицы $A^{(k}$ через $a_{ij}^{(k)}$. Матрица $A^{(k)}$ в своих первых $k-1$ столбцах совпадает с единичной матрицей I_n. На k-м шаге k-й столбец матрицы $A^{(k)}$ преобразуется в вектор e_k с помощью элементарных преобразований строк. Имеем

$$A^{(k+1)} = T_k A^{(k)}, \tag{5.2.1}$$

где

$$T_k = I_n + (\zeta^{(k)} - e_k) e_k' \tag{5.2.2}$$

и элементы вектора-столбца $\zeta^{(k)}$ даются в виде

$$\zeta_i^{(k)} = -\frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}, \qquad i \neq k, \quad \text{и} \quad \zeta_k^{(k)} = \frac{1}{a_{kk}^{(k)}}. \tag{5.2.3}$$

Теперь из формулы (5.2.1) и из условий $A^{(1)} \equiv A$ и $A^{(n+1)} = I_n$ имеем

$$T_n \ldots T_2 T_1 A = I_n,$$

что дает нам *мультипликативную форму обратной матрицы* (PFI) в виде

$$A^{-1} = T_n \ldots T_2 T_1. \tag{5.2.4}$$

Если к концу k-го шага исключения Гаусса — Жордана (GJE) вектор-столбец $\zeta^{(k)}$ хранится на месте k-го столбца матрицы $A^{(k+1)}$ (который в дальнейшем не потребуется), тогда нетривиальные элементы формы PFI будут замещать матрицу A при завершении процесса исключения.

Мультипликативная форма представления обратной матрицы является существенной частью большинства программ линейного программирования, где вопрос минимизации заполнения играет значительную роль (Данциг и Орчард-Хейс (1954); Ларсон (1962); Смит и Орчард-Хейс (1963); Вулф и Катлер (1963); Тьюарсон (1966), (1967а); Орчард-Хейс (1968); Данциг и др. (1969); Брейтон и др., (1969)). В разд. 5.4 мы рассмотрим минимизацию заполнения ненулевыми элементами в процессе исключения Гаусса — Жордана.

В случае когда решение системы линейных уравнений $Ax = b$ требуется только для небольшого числа правых частей, нет необходимости хранить PFI, так как каждая матрица преобразования T_k может быть применена также и к правым частям, когда она применяется к матрице $A^{(k)}$ в соответствии с формулой (5.2.1). Даже в этом случае для больших разреженных матриц очень полезно минимизировать заполнение ненулевыми элементами при преобразовании матрицы $A^{(k)}$ в матрицу $A^{(k+1)}$ в соответствии с формулой (5.2.1).

В следующем разделе мы рассмотрим соотношение между элиминативной формой обратной матрицы (EFI), которая была определена в разд. 2.4, и мультипликативной формой (PFI).

5.3. Связь между формами PFI и EFI

Вспомним из разд. 2.2, что в процессе обратной подстановки метода Гаусса (GE) для матрицы, обратной к верхней треугольной матрице U с единичной диагональю, разложение на множители получается путем выбора в качестве главных последовательно расположенных элементов диагонали, начиная с нижнего правого угла матрицы. В этом случае заполнения ненулевыми элементами не может происходить и нетривиальные элементы множителей в разложении обратной матрицы U^{-1} получаются только изменением знаков у тех элементов матрицы U, которые лежат над диагональю. Однако для вычисления матрицы U^{-1} таким способом необходимо, чтобы все строки матрицы U были известны. Другими словами, необходимо ждать завершения процесса прямого гауссова исключения.

Другим путем нахождения матрицы U^{-1} является выбор диагональных элементов в качестве главных, начиная с левого верхнего угла матрицы и двигаясь вниз по диагонали. В этом случае некоторое заполнение ненулевыми элементами, вообще говоря, будет происходить. Однако на k-м шаге при любом частном значении k требуются только первые k строк матрицы

U. Так как к моменту завершения k-го шага прямого гауссова исключения первые k строк матрицы U известны, то матрица U^{-1} может быть вычислена в процессе прямого гауссова исключения. Это и есть в точности то, что делается при методе исключения Гаусса — Жордана (GJE). Вычисление матрицы U^{-1} в форме разложения на множители (которое, как мы покажем, является тем же, что и вычисление матрицы U^{-1} в явном виде) сочетается с прямым гауссовым исключением. Мы увидим, что в методе GJE исключение элементов над диагональю эквивалентно вычислению матрицы U^{-1} в соответствии с приведенным выше вторым методом. Этот метод для вычисления матрицы, обратной к верхней треугольной матрице U с единичной диагональю, может быть в математической форме описан следующим образом.

Пусть

$$U^{(k+1)} = \tilde{U}_k U^{(k)}, \qquad k = 1, 2, \ldots, n, \tag{5.3.1}$$

причем

$$U^{(1)} = U, \qquad U^{(n+1)} = I_n \tag{5.3.2}$$

и

$$\tilde{U}_k = I_n + \tilde{\xi}^{(k)} e_k', \tag{5.3.3}$$

где элементы вектора-столбца $\tilde{\xi}^{(k)}$ даются в виде

$$\tilde{\xi}_i^{(k)} = 0, \qquad i \geqslant k, \quad \text{и} \quad \tilde{\xi}_i^{(k)} = -u_{ik}^{(k)}, \qquad i < k \tag{5.3.4}$$

($u_{ij}^{(k)}$ — (i, j)-й элемент матрицы $U^{(k)}$).

Принимая во внимание формулы (5.3.3) и (5.3.4), имеем $\tilde{U}_1 = I_n$, и поэтому из формул (5.3.1) и (5.3.2) следует, что

$$U^{-1} = \tilde{U}_n \ldots \tilde{U}_3 \tilde{U}_2. \tag{5.3.5}$$

Для установления связи между методами GE и GJE нам потребуются следующие результаты (Брейтон и др. (1969)). Пусть матрицы L_k, T_k и $\tilde{U}_k$ определяются соответственно формулами (2.2.3), (5.2.2) и (5.3.1). Тогда имеем следующие леммы.

Лемма 5.3.6. *Последние* $n-k+1$ *строк матриц* $L_k \ldots L_2L_1A$ *и* $T_k \ldots T_2T_1A$ *идентичны при* $k=1, 2, \ldots, n$.

Доказательство. Лемма, несомненно, верна для $k=1$, так как в каждом из двух методов GE и GJE первый столбец матрицы A приводится к вектору e_1 идентичными преобразованиями строк, при которых в качестве главного взят элемент $a_{11}^{(1)}$. Допустим, что лемма справедлива для некоторого k. Тогда на $(k+1)$-м шаге, в каждом из двух методов GE и GJE, в $(k+1)$-м столбце $(k+1)$-й элемент делается равным единице, а все элементы, расположенные ниже, обращаются в нули. Другими словами, над последними $n-k$ строками производятся идентичные преобразования. Поэтому последние $n-k$ строк матриц $L_{k+1} \ldots L_2L_1A$ и $T_{k+1} \ldots T_2T_1A$ тоже идентичны. Доказательство леммы завершается по индукции.

Лемма 5.3.7. *Последние* $n-k+1$ *строк матриц* L_k *и* T_k *идентичны при* $k=1, 2, \ldots, n$.

Доказательство. Принимая во внимание формулы (2.2.4) и (5.2.3), мы находим, что последние $n-k+1$ строк обеих матриц L_k и T_k образуются из последних $n-k+1$ элементов k-го столбца соответственно матриц $L_{k-1} \ldots L_2L_1A$ и $T_{k-1} \ldots T_2T_1A$. Поэтому, учитывая лемму 5.3.6, последние $n-k+1$ строк матриц L_k и T_k идентичны.

Лемма 5.3.8. *Если матрицы* $U^{(k)}$ *и* $A^{(k)}$ *определяются соответственно формулами* (5.3.1) *и* (5.2.1), *то первые* $k-1$ *строк обеих матриц совпадают при* $k=2, 3, \ldots, n$.

Доказательство. Так как $e_1'L_1A = e_1'U$, то, учитывая лемму 5.3.6 и равенство $\tilde{U}_1 = I_n$, имеем $e_1'A^{(2)} = e_1'T_1A = e_1'L_1A = e_1'U = e_1'\tilde{U}_1U = e_1'U^{(2)}$. Таким образом, лемма справедлива при $k=2$. Предположим, что лемма справедлива при некотором k. Тогда из формулы (5.2.1) на основании леммы 5.3.6 и учитывая, что в конце k-го шага прямого гауссова исключения

$e_k' L_k \dots L_2 L_1 A = e_k' U$, имеем

$$e_k' A^{(k+1)} = e_k' T_k \dots T_2 T_1 A = e_k' L_k \dots L_2 L_1 A = \\ = e_k' U = e_k' U^{(k+1)}.$$

Теперь из формул (5.2.1), (5.2.2), (5.2.3), (5.3.1), (5.3.3) и (5.3.4) следует, что на k-ом шаге над первыми $k-1$ строками обеих матриц $A^{(k)}$ и $U^{(k)}$ производятся идентичные преобразования. Поэтому не только k-е строки, но также и первые $k-1$ строк матриц $A^{(k+1)}$ и $U^{(k+1)}$ являются идентичными. Методом индукции по k доказательство леммы завершается.

Лемма 5.3.9. *Первые $k-1$ строк матриц $\tilde{U}_k$ и T_k совпадают.*

Доказательство. Принимая во внимание формулы (5.3.3), (5.3.4), (5.2.2) и (5.2.3), находим, что нетривиальные элементы первых $k-1$ строк в обеих матрицах $\tilde{U}_k$ и T_k образуются из первых $k-1$ элементов k-х столбцов соответственно матриц $U^{(k)}$ и $A^{(k)}$. Поэтому из леммы 5.3.8 заключаем, что первые $k-1$ строк матриц $\tilde{U}_k$ и T_k идентичны.

Лемма 5.3.10. *Если матрицы U_k и U^{-1} даются соответственно формулами* (5.3.3) *и* (5.3.5), *то k-е столбцы этих матриц идентичны.*

Доказательство. Из формул (5.3.3) и (5.3.4) имеем

$$\tilde{U}_p e_q = (I_n + \tilde{\xi}^{(p)} e_p') e_q = e_q, \qquad p \neq q,$$
$$\tilde{U}_p e_p = e_p + \tilde{\xi}^{(p)}$$

и

$$\tilde{U}_q \tilde{\xi}^{(p)} = (I_n + \tilde{\xi}^{(q)} e_q') \tilde{\xi}^{(p)} = \tilde{\xi}^{(p)}, \qquad q > p,$$

так как $e_q' \tilde{\xi}^{(p)} = 0$.
Поэтому имеем

$$U^{-1} e_k = \tilde{U}_n \dots \tilde{U}_2 e_k = \tilde{U}_n \dots \tilde{U}_{k+1} \tilde{U}_k e_k = \\ = \tilde{U}_n \dots \tilde{U}_{k+1} (e_k + \tilde{\xi}^{(k)}) = e_k + \tilde{\xi}^{(k)} = \tilde{U}_k e_k,$$

что завершает доказательство.

На основании лемм 5.3.7, 5.3.9 и 5.3.10 имеем следующие два уравнения, которые показывают связь между формами PFI и EFI:

$$e_i' T_k e_k = e_i' \tilde{U}_k e_k = e_i' U^{-1} e_k, \qquad i < k, \tag{5.3.11}$$

и

$$e_i' T_k e_k = e_i' L_k e_k, \qquad i \geqslant k. \tag{5.3.12}$$

Из формулы (5.3.11) ясно, что нетривиальные элементы PFI, расположенные над ведущей диагональю, являются элементами матрицы U^{-1}, выраженной в явной форме, где U является верхней треугольной матрицей с единичной диагональю, полученной в конце процесса прямого гауссова исключения. С другой стороны, из формулы (5.3.12) следует, что нетривиальные элементы в обоих формах PFI и EFI, расположенные на диагонали и под ней, являются идентичными. Таким образом, структура распределения нулей и ненулевых элементов в обеих формах PFI и EFI одинакова для элементов, расположенных на диагонали и под ней. Над диагональю структура распределения нулей и ненулевых элементов у формы PFI такая же, как и у матрицы U^{-1}, в то время как у формы EFI она такая, как у матрицы U. Вообще говоря, матрица U^{-1} содержит больше ненулевых элементов, чем матрица U, и поэтому форма PFI обычно не так разрежена, как форма EFI.

5.4. Минимизация общего числа ненулевых элементов в форме PFI

Ввиду тесной связи, существующей между формами PFI и EFI, все сказанное в разд. 2.3 относительно ошибок округления и выбора главного элемента остается в силе и в случае формы PFI. Методы минимизации общего числа ненулевых элементов в форме PFI по существу подобны тем, что приведены в разд. 2.5 (Маркович (1957); Ларсон (1962); Смит и Орчард-Хейс (1963); Вулф и Катлер (1963); Диксон (1965); Тьюарсон (1966), (1967, b); Орчард-

Хейс (1968)). Мы сейчас вкратце рассмотрим те изменения, которые нужно внести в методы, приведенные в разд. 2.5, с тем чтобы они могли быть использованы в случае формы PFI.

На k-м шаге метода GJE k-я строка матрицы $A^{(k)}$ умножается на различные коэффициенты и складывается со всеми другими строками. Вследствие этого имеет место заполнение ненулевыми элементами не только области ниже диагонали (как в методе GE), но также и области, расположенной над диагональю. Для минимизации этого заполнения нам потребуются следующие результаты.

Если к началу k-го шага метода GJE переставить s-ю и k-ю строки и t-й и k-й столбцы, то вместо элемента $a_{kk}^{(k)}$ в качестве главного элемента будет взят элемент $a_{st}^{(k)}$. Конечно, $|a_{st}^{(k)}|$ должно быть больше допустимого значения главного элемента ε. При таком выборе главного элемента мы получим следующее. Если P_k и Q_k есть матрицы, полученные перестановкой соответственно s-й и k-й строк и t-го и k-го столбцов единичной матрицы I_n, то матрица

$$\hat{A}^{(k)} = P_k A^{(k)} Q_k \tag{5.4.1}$$

имеет элемент $a_{st}^{(k)}$ в (k, k)-й позиции. Тогда вместо формул (5.2.1) и (5.2.3) имеем

$$A^{(k+1)} = T_k \hat{A}^{(k)}, \qquad k = 1, 2, \ldots, n, \tag{5.4.2}$$

и

$$\zeta_i^{(k)} = -\frac{\hat{a}_{ik}^{(k)}}{\hat{a}_{kk}^{(k)}}, \qquad i \neq k, \quad \text{и} \quad \zeta_k^{(k)} = \frac{1}{\hat{a}_{kk}^{(k)}} \tag{5.4.3}$$

и из формул (5.4.1) и (5.4.2) имеем

$$A^{-1} = Q_1 Q_2 \ldots Q_n T_n P_n \ldots T_2 P_2 T_1 P_1. \tag{5.4.4}$$

Пусть B_k есть матрица, полученная из последних $n - k + 1$ столбцов матрицы $A^{(k)}$ путем замены в них ненулевых элементов единицами. (Заметим, что матрицы $A^{(k)}$ и B_k не идентичны соответствующим матрицам разд. 2.5.) Теперь имеем следующую теорему (Тьюарсон (1976б)), которая подобна теореме 2.5.5.

Теорема 5.4.5. *Если элемент* $a^{(k)}_{i,j+k-1}$, *где* $i \geqslant k$, *выбран в качестве главного элемента на* k*-м шаге метода* GJE, *то локальное заполнение дается* (i,j)*-м элементом матрицы* G_k, *равной*

$$G_k = B_k \bar{B}'_k B_k, \tag{5.4.6}$$

где $\bar{B}_k$ — *матрица, полученная из матрицы* B_k *путем замены всех ее нулей единицами, а всех единиц — нулями.*

Доказательство. Если в матрице $A^{(k)}$ $(p, q+k-1)$-й элемент равен нулю, но $(i, q+k-1)$-й и $(p, j+k-1)$-й элементы — оба ненулевые, то из формул (5.4.1), (5.4.2), (5.2.2) и (5.4.3) следует, что $(p, q+k-1)$-й элемент матрицы $A^{(k+1)}$ будет ненулевым. Начиная с этого места, ход доказательства настоящей теоремы совпадает с доказательством теоремы 2.5.5 при условии, что GE заменено GJE и матрица M является матрицей размеров $n \times (n-k+1)$, все элементы которой равны единице. Эту часть доказательства поэтому мы здесь не приводим.

Приводимое ниже следствие непосредственно вытекает из теоремы 5.4.5, и на этом основании его доказательство опущено.

Следствие 5.4.7. *Если на* k*-м шаге метода* GJE *элемент* $a^{(k)}_{st}$ *выбран в качестве главного элемента, причем* $s \geqslant k$, $t = \beta + k - 1$, *а* s *и* β *определяются формулой*

$$g^{(k)}_{s\beta} = \min_{i,j} e'_i G_k \tilde{e}_j \tag{5.4.8}$$

для всех $|a^{(k)}_{i,j+k-1}| > \varepsilon$, $i \geqslant k$ (ε — *некоторое подходящим образом выбранное допустимое значение главного элемента, а* $\tilde{e}_j$ — j*-й столбец матрицы* I_{n-k+1}), *то локальное заполнение будет минимальным.*

Так как элементы всех матриц T_k вычисляются по элементам всех матриц $A^{(k)}$, то минимизация локального заполнения всех матриц $A^{(k)}$ будет минимизировать число ненулевых элементов формы PFI при

условии, что обеспечение локальных минимумов приводит к глобальному минимуму. Это может быть верным для некоторых матриц, но не является таковым для произвольных матриц.

Более простой, хотя и менее точный метод выбора главного элемента, при котором локальное заполнение мало, основан на следующей теореме (Маркович (1957)).

Теорема 5.4.9. *Если элемент* $a^{(k)}_{i,j+k-1}$, *где* $i \geqslant k$, *выбран в качестве главного элемента на* k*-м шаге метода GJE, то максимально возможное заполнение (не обязательно совпадающее с реальным) дается формулой*

$$\hat{g}^{(k)}_{ij} = (r^{(k)}_i - 1)(c^{(k)}_j - 1), \tag{5.4.10}$$

где

$$r^{(k)}_i = e'_i B_k V_k, \qquad c^{(k)}_j = V' B_k \tilde{e}_j, \tag{5.4.11}$$

V_k *и* V *— соответственно* $(n-k+1)$*-мерный и* n*-мерный векторы-столбцы, все элементы которых единицы, а* $\tilde{e}_j$ *—* j*-й столбец матрицы* I_{n-k+1}.

Доказательство. Из формулы (5.4.11) видно, что $r^{(k)}_i$ и $c^{(k)}_j$ обозначают общее число ненулевых элементов соответственно в i-й строке и в $(j+k-1)$-м столбце матрицы $A^{(k)}$. Поэтому максимум возможного заполнения, который может иметь место, когда $(i, j+k-1)$-й элемент матрицы $A^{(k)}$ выбран в качестве главного элемента, будет равен $(r^{(k)}_i - 1)(c^{(k)}_j - 1)$, что завершает доказательство.

Легко показать, что $\hat{g}^{(k)}_{ij}$ является (i, j)-м элементом матрицы $\hat{G}_k$: из формул (5.4.10), (5.4.11) и на основании равенства $e'_i V = V'_k \tilde{e}_j = 1$ имеем

$$\begin{aligned}\hat{g}^{(k)}_{ij} &= (e'_i B_k V_k - 1)(V' B_k \tilde{e}_j - 1) = \\ &= e'_i (B_k V_k - V)(V' B_k - V'_k)\tilde{e}_j,\end{aligned}$$

поэтому

$$\hat{G}_k = (B_k V_k - V)(V' B_k - V'_k). \tag{5.4.12}$$

Для того чтобы воспользоваться теоремой 5.4.9 вместо формулы (5.4.8), будем производить выбор

главного элемента на k-м шаге с помощью уравнения

$$\hat{g}_{s\beta}^{(k)} = \min e_i' \hat{G}_k \tilde{e}_j \qquad (5.4.13)$$

для всех $|a_{i,j+k-1}^{(k)}| > \varepsilon$.

Заметим, что главный элемент $a_{s,\beta+k-1}^{(k)}$, выбранный в соответствии с формулой (5.4.13), не обязательно приводит к наименьшему локальному заполнению.

Приведенные в разд. 3.2 методы минимизации объема памяти для хранения формы EFI, основанные на априорной перестановке столбцов, могут быть также применены и для формы PFI. Такая возможность вытекает из следующих соображений. При $k = 1$ уравнения (3.2.2) и (5.4.11) одинаковы, и поэтому все $c_j^{(1)}$, $\gamma_j^{(1)}$ и $\lambda_j^{(1)}$ будут идентичными для методов GE и GJE. Более того, в обоих методах GE и GJE на $(k+1)$-м шаге главный элемент может быть выбран только из последних $n-k$ строк и столбцов. Поэтому только последние $n-k$ из величин $r^{(k)}$, связанных с методом GJE, должны быть изменены на k-м шаге. Это вместе с тем обстоятельством что B_k — это матрица размеров $n \times (n-k+1)$ (а не размеров $(n-k+1) \times (n-k+1)$, как в методе GE), позволяет нам иначе сформулировать теорему 3.2.12.

Теорема 5.4.14. *Если ненулевые элементы последних $n-k+1$ строк и столбцов матрицы $\hat{A}^{(k)}$ распределены в этих строках и столбцах случайным образом и $\hat{r}_i^{(k)}$ есть число ненулевых элементов в i-й строке матрицы $\hat{A}^{(k)}$ для $i \geqslant k$, тогда $\tilde{r}_i^{(k+1)}$ — ожидаемое число ненулевых элементов i-й строки матрицы $A^{(k+1)}$ для $i > k$ — дается формулами*

$$\tilde{r}_i^{(k+1)} = \hat{r}_i^{(k)}, \qquad \hat{a}_{ik}^{(k)} = 0, \qquad (5.4.15)$$

$$\tilde{r}_i^{(k+1)} = \hat{r}_i^{(k)} + \hat{r}_k^{(k)} - 2 - \frac{(\hat{r}_i^{(k)} - 1)(\hat{r}_k^{(k)} - 1)}{n-k}, \qquad \hat{a}_{ik}^{(k)} \neq 0. \qquad (5.4.16)$$

Доказательство. Такое же, что и для теоремы 3.2.12.

Итак, мы видим, что столбцы матрицы A могут быть априорно упорядочены в соответствии с возрастающими значениями всех $c_j^{(1)}$, $\gamma_j^{(1)}$ или $\lambda_j^{(1)}$ и главные элементы выбираются согласно формуле (3.2.11), но значение $\hat{r}^{(k)}$ изменяется с помощью формул (5.4.15) и (5.4.16) вместо формул (3.2.13) и (3.2.14) (Тьюарсон (1966), (1967а); Орчард-Хейс (1968)).

Как и в разд. 3.3, можно преобразовать матрицу A с помощью матриц перестановок строк (P) и столбцов (Q) так, чтобы матрица $\hat{A} = PAQ$ имела одну из форм, желательных для метода GJE. Мы перейдем к рассмотрению этого вопроса.

5.5. Подходящие формы для метода GJE

Если главные элементы последовательно берутся на диагонали матрицы A, начиная с верхнего левого угла, следующие формы, описанные в гл. 3, являются также желательными и для метода GJE: BDF, SBBDF, DBBDF, транспонированная форма BTF (нижняя треугольная блочная форма), транспонированная форма BBTF (окаймленная нижняя треугольная блочная форма). В гл. 3 уже были приведены методы, позволяющие преобразовать матрицу A к одной из этих форм.

Подходящая форма для матрицы $\hat{A}$, которая включена в некоторые программы для ЭВМ (Орчард-Хейс, 1968), определяется матрицами перестановок P и Q, такими, что

$$PAQ = \hat{A} = \begin{bmatrix} I & A_{12} & A_{13} & A_{14} \\ 0 & A_{22} & 0 & 0 \\ 0 & A_{32} & A_{33} & 0 \\ 0 & A_{42} & A_{43} & A_{44} \end{bmatrix},$$

где A_{22} и A_{44} — нижние треугольные матрицы, а A_{33} — квадратная матрица. Главные элементы выбираются последовательно из матриц I, A_{22}, A_{33} и A_{44}. За исключением матрицы A_{33}, главные элементы выби-

раются на диагонали. Заполнение имеет место только в областях A_{13}, A_{33} и A_{43}, когда главные элементы выбираются из матрицы A_{33}. Это заполнение может быть минимизировано любым из методов, изложенных в разд. 5.4.

5.6. Библиография и комментарии

Метод GJE описан во многих руководствах по численному анализу, например Хильдебранда (1956), Фаддеева и Фаддеевой (1960), Фокса (1965), Рэлстона (1965), Уэстлейка (1968). Форма PFI описана у Данцига и Орчард-Хейса (1954), Гасса (1958), Хедли (1962) и Данцига (1963а).

Примеры вычислений с применением формы PFI и способов сохранения разреженности даны Ларсоном (1962), Смитом и Орчард-Хейсом (1963), Вулфом и Катлером (1963), Диксоном (1965), Бауманном (1965), Тьюарсоном (1966, 1967а), Орчард-Хейсом (1968), Данцигом и др. (1969), Брейтоном и др. (1969), Билом (1971) и Де Бюше (1971).

Сравнение форм EFI и PFI с точки зрения заполнения для разреженных матриц со случайным распределением ненулевых элементов дано Брейтоном и др. (1969).

Глава 6

МЕТОДЫ ОРТОГОНАЛИЗАЦИИ

6.1. Введение

Во многих практических приложениях требуется преобразовать заданную разреженную матрицу в другую матрицу с ортонормированными столбцами. Применение программ ортонормирования хорошо известно (Дэвис (1962)). В этой главе мы рассмотрим задачу определения оптимального порядка, в котором столбцы заданной разреженной матрицы должны быть ортонормированы, чтобы результирующаяся матрица была как можно более разреженной. Нас будут интересовать методы ортонормирования *Грама — Шмидта*, *Хаусхолдера* и *Гивенса* (Тьюарсон (1968а), (1970а)).

6.2. Метод Грама — Шмидта

Пусть A обозначает матрицу размеров $m \times n$, где $m \geqslant n$, ранг которой равен n. *Метод Грама — Шмидта* заключается в определении такой верхней треугольной матрицы U, что столбцы матрицы AU ортонормированы (Дэвис (1962); Райс (1966)). Если матрица A разреженная, то, вообще говоря, предпочтительней найти такую матрицу перестановок Q, чтобы и матрица AQU, и матрица U были разреженными. Методы для осуществления этого будут рассмотрены в следующем разделе. В настоящем разделе мы рассмотрим слегка измененный вариант метода Грама — Шмидта, более точный по сравнению с обычным методом с точки зрения ошибок округления (Райс (1966); Тьюарсон (1968а)). Он называется *модифицированным методом Грама — Шмидта* (RGS) и состоит из n шагов. Если $A^{(k)}$ обозначает матрицу к на-

чалу k-го шага, $k = 1, 2, \ldots, n$, а $A^{(1)} \equiv A$, то после n шагов все столбцы матрицы $A^{(n+1)}$ будут ортонормированными. В матрице $A^{(k)}$ первые $k-1$ столбцов ортонормированы, а k-й столбец ортогонален к ним. На k-м шаге k-й столбец матрицы A нормируется, а последние $n-k$ столбцов делаются ортогональными к нему. Результирующая матрица обозначается через $A^{(k+1)}$. Если $a_j^{(k)}$ и $a_{ij}^{(k)}$ обозначают соответственно j-й

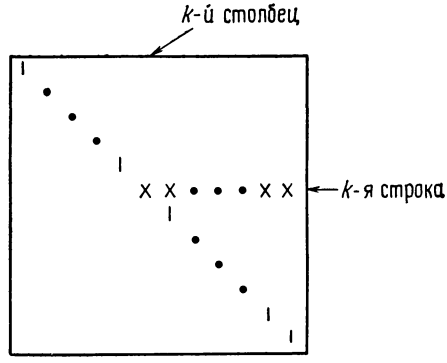

Рис. 6.2.1. Элементарная матрица $\hat{U}_k$ на k-м шаге.

столбец и (i, j)-й элемент матрицы $A^{(k)}$, то метод может быть в математической форме описан следующим образом:

$$A^{(k+1)} = A^{(k)}\hat{U}_k, \qquad k = 1, 2, \ldots, n, \tag{6.2.1}$$

где элементарная верхняя треугольная матрица U_k (см. рис. 6.2.1) дается формулой

$$\hat{U}_k = I_n + e_k(\hat{\xi}^{(k)} - e_k'), \tag{6.2.2}$$

причем элементы вектора-строки $\hat{\xi}^{(k)}$ определяются так:

$$\hat{\xi}_j^{(k)} = 0, \qquad j < k;$$

$$\hat{\xi}_k^{(k)} = \frac{1}{\left(a_k^{(k)\prime} a_k^{(k)}\right)^{\frac{1}{2}}} \quad \text{и} \quad \hat{\xi}_j^{(k)} = -\frac{a_k^{(k)\prime} a_j^{(k)}}{a_k^{(k)\prime} a_k^{(k)}}, \qquad j > k. \tag{6.2.3}$$

Матрица U_k является единичной матрицей, в которой k-я строка заменена вектором-строкой $\hat{\xi}^{(k)}$. Из формул (6.2.1), (6.2.2) и (6.2.3) следует, что

$$a_j^{(k+1)} = \begin{cases} a_j^{(k)}, & j < k, \\ \dfrac{a_k^{(k)}}{\left(a_c^{(k)\prime} a_k^{(k)}\right)^{\frac{1}{2}}}, & j = k, \\ a_j^{(k)} - \dfrac{a_j^{(k)\prime} a_c^{(k)}}{a_c^{(k)\prime} a_k^{(k)}} a_k^{(k)}, & j > k. \end{cases} \tag{6.2.4}$$

В такой форме метод RGS обычно приводится в учебниках.

Если A — квадратная матрица n-го порядка ($m = n$), то из формулы (6.2.1) и из того факта, что обратная к матрице с ортонормированными столбцами равна транспонированной к ней матрице, следует, что

$$A^{-1} = \hat{U}_1 \hat{U}_2 \ldots \hat{U}_n A^{(n+1)\prime}. \tag{6.2.5}$$

Таким образом, метод RGS может быть использован для вычисления матрицы A^{-1}.

6.3. Минимизация ненулевых элементов в методе RGS

Из формулы (6.2.4) видно, что во всех столбцах, для которых существует отличное от нуля скалярное произведение со столбцом $a_k^{(k)}$, может иметь место некоторое заполнение. Поэтому для минимизации локального заполнения из последних $n - k + 1$ столбцов выбирается столбец $a^{(k)}$, который привел бы к наименьшему заполнению, если перед k-м шагом метода RGS сделать его k-м столбцом. Это осуществляется перестановкой k-го и s-го столбцов матрицы $A^{(k)}$:

$$\hat{A}^{(k)} = A^{(k)} Q_k, \tag{6.3.1}$$

где Q_k получается перестановкой k-го и s-го столбцов единичной матрицы I_n. Тогда вместо формулы (6.2.1)

применяем

$$A^{(k+1)} = \hat{A}^{(k)}\hat{U}_k, \qquad k = 1, 2, \ldots, n. \tag{6.3.2}$$

Если $\hat{a}_j^{(k)}$ и $\hat{a}_{ij}^{(k)}$ обозначают соответственно j-й столбец и (i, j)-элемент матрицы $\hat{A}^{(k)}$, то матрица $\hat{U}_k$ дается формулой (6.2.2), но элементы векторов $\hat{\xi}^{(k)}$ даются формулами

$$\hat{\xi}_j^{(k)} = 0, \qquad j < k,$$

$$\hat{\xi}_k^{(k)} = \frac{1}{\left(\hat{a}_k^{(k)'}\hat{a}_k^{(k)}\right)^{\frac{1}{2}}} \quad \text{и} \quad \hat{\xi}_j^{(k)} = -\frac{\hat{a}_k^{(k)'}\hat{a}_j^{(k)}}{\hat{a}_k^{(k)'}\hat{a}_k^{(k)}}, \qquad j > k, \tag{6.3.3}$$

вместо формул (6.2.3).

Для определения столбца, который приводит к наименьшему локальному заполнению, можно воспользоваться следующей теоремой, в которой через $b_j^{(k)}$ обозначен j-й столбец матрицы B_k, полученной из последних $n - k + 1$ столбцов матрицы $A^{(k)}$ путем замены всех ненулевых элементов единицами.

Теорема 6.3.4. *Если $(t + k - 1)$-й и k-й столбцы матрицы $A^{(k)}$ переставлены и затем выполняется k-й шаг метода RGS, то максимальное заполнение дается t-м диагональным элементом матрицы G_k, где*

$$G_k = \left(B_k' * B_k\right)\bar{B}_k' B_k, \tag{6.3.5}$$

причем $*$ обозначает булево умножение, а матрица $\bar{B}_k$ получается из матрицы B_k путем замены всех ее нулей единицами, а всех единиц нулями.

Доказательство. Если $(t + k - 1)$-й столбец матрицы $A^{(k)}$ сделан ортогональным к ее $(j + k - 1)$-му столбцу, то максимальное число дополнительных ненулевых элементов, созданных в j-м столбце, равно $\bar{b}_j^{(k)'} b_t^{(k)}$, где $\bar{b}_j^{(k)}$ — j-й столбец матрицы $\bar{B}_k$. С другой стороны, если $b_t^{(k)'} * b_j^{(k)} = 0$, то ни один ненулевой элемент не создается. Таким образом, в обоих случаях общее заполнение для j-го столбца может быть дано в виде

$$\left(b_t^{(k)'} * b_j^{(k)}\right)\bar{b}_j^{(k)'} b_t^{(k)}.$$

Принимая во внимание то обстоятельство, что $\bar{b}_t^{(k)'} b_t^{(k)} = 0$, общее заполнение для всех столбцов имеет вид

$$g_{tt}^{(k)} = \sum_{j=1}^{n-k+1} \left(b_t^{(k)'} * b_j^{(k)}\right) \bar{b}_j^{(k)'} b_t^{(k)}, \qquad (6.3.6)$$

или

$$g_{tt}^{(k)} = \sum_{j=1}^{n-k+1} \left(\tilde{e}_t' B_k' * B_k \tilde{e}_j\right) \tilde{e}_j' \bar{B}_k' B_k \tilde{e}_t,$$

где $\tilde{e}_j$ является j-м столбцом матрицы I_{n-k+1}, или

$$g_{tt}^{(k)} = \tilde{e}_t' \left(B_k' * B_k\right) \sum_{j=1}^{n-k+1} \tilde{e}_j \tilde{e}_j' \bar{B}_k' B_k \tilde{e}_t =$$

$$= \tilde{e}_t' \left(B_k' * B_k\right) \bar{B}_k' B_k \tilde{e}_t = \tilde{e}_t' G_k \tilde{e}_t,$$

так как $\sum_{j=1}^{n-k+1} \tilde{e}_j \tilde{e}_j' = I_{n-k+1}$, что и завершает доказательство теоремы.

Заметим, что условие $b_p^{(k)'} * b_q^{(k)} = 1$ не обязательно означает, что скалярное произведение соответствующих столбцов матрицы $A^{(k)}$ не равно нулю. Также заполнение в столбце $b_j^{(k)}$ может быть меньше, чем $\bar{b}_j^{(k)'} b_t^{(k)}$, так как в действительном методе RGS может иметь место взаимное уничтожение слагаемых. Этим объясняется, почему $g_{tt}^{(k)}$ дает максимальное, а не действительное заполнение. Наш опыт учит, что такие случаи [1]) редки и когда они имеют место, то составляют лишь небольшой процент от общего объема вычислений. Поэтому действительное заполнение очень близко к $g_{tt}^{(k)}$. Как бы там ни было, из приведенной выше теоремы следует, что для того, чтобы минимизировать заполнение, мы определяем

$$g_{ss}^{(k)} = \min_t g_{tt}^{(k)} = \min_t \left[\tilde{e}_t' G_k \tilde{e}_t\right] \qquad (6.3.7)$$

в начале k-го шага метода RGS, затем полагаем $s = \hat{s} + k - 1$ и применяем формулы (6.3.1), (6.3.2),

[1]) Имеется в виду — взаимного уничтожения слагаемых. — *Прим. ред.*

(6.2.2) и (6.3.3.) для вычисления матрицы $A^{(k+1)}$ с помощью матрицы $A^{(k)}$. Заметим, что для определения $\hat{s}$ на каждом шаге k должны вычисляться только диагональные элементы произведения матриц $B_k' * B_k$ и $\bar{B}_k' B_k$. Даже это требует больших затрат труда на каждом шаге. Поэтому, прежде чем применить метод RGS, мы можем априорно упорядочить столбцы матрицы A по возрастающим значениям диагональных элементов матрицы $(B_1' * B_1)\,\bar{B}_1' B_1$. Можно также, используя теорему 6.3.4, производить изменения порядка столбцов матрицы $A^{(k)}$ через определенные интервалы, а не на каждом шаге k. Это, вообще говоря, приводит к дальнейшему уменьшению заполнения. Для определения столбцов, которые приводят к нулевому заполнению в других столбцах, полезно использовать приводимые ниже следствия из теоремы 6.3.4.

Следствие 6.3.8. *Если всякий раз, когда* $b_t^{(k)'} * b_j^{(k)} = 1$, *имеем* $b_t^{(k)'} b_t^{(k)} = b_j^{(k)'} b_t^{(k)}$, *то* $g_{tt}^{(k)} = 0$.

Доказательство. Пусть V является m-мерным вектором, все элементы которого единицы. Тогда

$$0 = b_t^{(k)'} b_t^{(k)} - b_j^{(k)'} b_t^{(k)} = V' b_t^{(k)} - b_j^{(k)'} b_t^{(k)} =$$
$$= \left(V' - b_j^{(k)'}\right) b_t^{(k)} =$$
$$= \bar{b}_j^{(k)'} b_t^{(k)}$$

и следствие вытекает из формулы (6.3.6).

Следствие 6.3.9. *Если* $b_t^{(k)'} b_t^{(k)} = 1$, *то* $g_{tt}^{(k)} = 0$.

Доказательство. Если $b_t^{(k)'} b_t^{(k)} = 1$, то $b_t^{(k)'} * b_j^{(k)} = 1$ означает, что $b_t^{(k)'} b_j^{(k)} = 1$, и на основании следствия 6.3.8 имеем $g_{tt}^{(k)} = 0$.

Таким образом, из этого следствия видно, что все столбцы, содержащие всего один ненулевой элемент, должны быть ортонормированы в первую очередь. Следующая теорема показывает, что в действительности столбец с единственным ненулевым элементом приводит к уменьшению общего числа ненулевых элементов в столбцах, с которыми он взаимодействует.

Теорема 6.3.10. *Если* $a_{pk}^{(k)}=1$, *но* $a_{ik}^{(k)}=0$ *для всех* $i\neq p$, *то* $a_{pj}^{(k+1)}=0$ *для всех* $j>k$.

Доказательство. Из формулы (6.2.4) для $j>k$ имеем

$$a_{pj}^{(k+1)}=a_{pj}^{(k)}-\left(a_{pk}^{(k)}a_{pj}^{(k)}/\left(a_{pk}^{(k)}\right)^2\right)a_{pk}^{(k)}=$$
$$=a_{pj}^{(k)}-a_{pj}^{(k)}=0,$$

что и завершает доказательство теоремы.

Пусть V и V_k — соответственно m-мерный и $(n-k+1)$-мерный векторы-столбцы, все элементы которых единицы, а e_i и $\tilde{e}_i$ обозначают i-е столбцы матриц I_n и I_{n-k+1} соответственно. Тогда, подобно формулам (5.4.11), определим

$$r_i^{(k)}=e_i'B_kV_k \quad \text{и} \quad c_j^{(k)}=V'B_k\tilde{e}_j. \tag{6.3.11}$$

Теперь можем описать алгоритм априорного упорядочения столбцов матрицы A для минимизации заполнения.

Алгоритм 6.3.12. Определить матрицу B_1 по матрице A. Пусть R_1 обозначает вектор-строку, составленную из n натуральных чисел так, что ее j-й элемент равен j. Положить $k=1$.

1. Вычислить $c_j^{(k)}$ по формуле (6.3.11). Найти $c_t^{(k)}=\min_j c_j^{(k)}$. В случае совпадения значений минимума для нескольких j выбрать $c_t^{(k)}$ с наименьшим индексом. Если $c_t^{(k)}>1$, перейти к шагу 2. В противном случае положить все $b_{pj}^{(k)}=0$, если $b_{pt}^{(k)}=1$, заменить t-й столбец матрицы B_k ее первым столбцом и исключить первый столбец из дальнейшего рассмотрения. Переставить $(t+k-1)$-й и k-й элементы вектора R_k. Положить k равным $k+1$. Если $k=n$, то перейти к шагу 3, в противном случае вернуться к началу настоящего шага.

2. Вычислить G_k согласно формуле (6.3.5). Упорядочить последние $n-k+1$ элементов R_k соответственно возрастающим значениям диагональных элементов матрицы G_k и обозначить результирующий вектор через R_{n+1}.

3. Упорядочить столбцы матрицы A согласно значениям элементов вектора R_{n+1} так, чтобы новая позиция j-го столбца матрицы A соответствовала значению j-го элемента вектора R_{n+1}.

Замечания. На первом шаге изложенного алгоритма мы определим все столбцы матрицы A, которые содержат всего один ненулевой элемент. После того как такие столбцы делаются ортогональными к остальным столбцам матрицы, могут появиться еще столбцы с одним ненулевым элементом (согласно теореме 6.3.10), и они также определяются на первом шаге. На втором шаге оставшиеся столбцы матрицы A упорядочиваются соответственно величинам заполнения, которое каждый столбец создавал бы, если бы его выбрали в конце первого шага алгоритма. Заметим, что в существующей практике выполнения алгоритма перестановка столбцов матрицы $A^{(k)}$ только запоминается, а не производится фактически. Информация о перестановках используется в дальнейшем при осуществлении ортогонализации по методу RGS.

Матрицы перестановок P и Q, при которых матрица $\hat{A} = PAQ$ имеет форму, подходящую для метода RGS, могут быть найдены способами, описанными в гл. 3. Если матрица $\hat{A}$ имеет одну из форм: BDF, BTF, BNTF, SBBDF, DBBDF, BBTF или BBNTF, то область, в которой может иметь место заполнение, ограничивается заштрихованными частями этих форм.

В следующем разделе мы рассмотрим метод ортогональной триангуляризации Хаусхолдера (1959). В нем используются элементарные эрмитовы матрицы для преобразования матрицы A к верхней треугольной форме.

6.4. Метод триангуляризации Хаусхолдера

Этот метод состоит из n шагов. Для k-го шага

$$A^{(k+1)} = H_k A^{(k)}, \qquad k = 1, 2, \ldots, n, \tag{6.4.1}$$

где

$$H_k = I_m - \alpha_k^{-1} \hat{\eta}^{(k)} \hat{\eta}^{(k)'}, \tag{6.4.2}$$

а элементы вектора-столбца $\hat{\eta}^{(k)}$ даются формулами

$$\hat{\eta}_i^{(k)} = 0, \qquad i < k, \tag{6.4.3}$$

$$\hat{\eta}_k^{(k)} = a_{kk}^{(k)} \pm \beta_k, \qquad \hat{\eta}_i^{(k)} = a_{ik}^{(k)}, \qquad i > k,$$

причем

$$\beta_k^2 = \sum_{i=k}^{m} (a_{ik}^{(k)})^2, \quad \alpha_k = \beta_k^2 \pm \beta_k a_{kk}^{(k)} \tag{6.4.4}$$

и для устойчивости β_k берется с тем же знаком, что и $a_{kk}^{(k)}$. Мы начинаем с матрицы $A^{(1)} \equiv A$. После n шагов метода *триангуляризации Хаусхолдера* (HT) первые n строк матрицы $A^{(n+1)}$ представляют собой верхнюю треугольную матрицу, которую обозначим через $\bar{U}$, а последние $m - n$ строк матрицы $A^{(n+1)}$ содержат одни нули (напомним, что $m \geqslant n$). Заметим, что только $n - 1$ шагов требуется для метода HT, если $m = n$. Пусть

$$H_n H_{n-1} \ldots H_1 = H, \tag{6.4.5}$$

а матрицу, составленную из первых n строк матрицы H, обозначим через $\hat{H}$. Тогда из формулы (6.4.1) и из того, что H — ортогональная матрица, следует

$$A^{(n+1)} = HA$$

и

$$A = \hat{H}'\bar{U}$$

и, наконец,

$$A\bar{U}^{-1} = \hat{H}'. \tag{6.4.6}$$

Так как $\bar{U}^{-1}$ — верхняя треугольная матрица и столбцы матрицы $\hat{H}'$ ортонормированы, то из формулы (6.4.6) следует, что метод HT является другим возможным путем ортонормирования столбцов матрицы A. Высокая точность метода HT делает его привлекательным для вычислений (Уилкинсон (1965)). Конечно, он требует бóльших затрат труда, чем метод RGS. Матрица H хранится в факторизованной форме (6.4.5); фактически требуется хранить только ненулевые элементы векторов $\eta^{(k)}$ и все значения α_k (Тьюарсон (1968а)).

Для рассмотрения вопроса о заполнении для метода НТ нам потребуется следующая лемма.

Лемма 6.4.7. *Если k-й шаг метода НТ задан формулами от* (6.4.1) *до* (6.4.4), *то*

а) *первые $k-1$ строк и столбцов матриц $A^{(k+1)}$ и $A^{(k)}$ одинаковы,*

б) $a_{kk}^{(k+1)} = \mp \beta_k$ и $a_{ik}^{(k+1)} = 0$, $i > k$.

Доказательство. Из формул (6.4.2) и (6.4.3) видно, что первые $k-1$ строк и столбцов матрицы H_k такие же, как и соответствующие столбцы и строки единичной матрицы I_m. Отсюда, принимая во внимание формулу (6.4.1), следует справедливость первой части леммы. Теперь из формул (6.4.3) и (6.4.4) имеем

$$\hat{\eta}^{(k)'} a_k^{(k)} = \left(a_{kk}^{(k)} \pm \beta_k\right) a_{kk}^{(k)} + \sum_{i=k+1}^{m} \left(a_{ik}^{(k)}\right)^2 =$$

$$= \pm \beta_k a_{kk}^{(k)} + \beta_k^2 = \alpha_k,$$

откуда, учитывая формулы (6.4.1) и (6.4.2), вытекает, что

$$a_k^{(k+1)} = a_k^{(k)} - \alpha_k^{-1} \left(\hat{\eta}^{(k)'} a_k^{(k)}\right) \eta^{(k)} =$$

$$= a_k^{(k)} - \hat{\eta}^{(k)}$$

или

$$a_{kk}^{(k+1)} = \mp \beta_k$$

и

$$a_{ik}^{(k+1)} = 0, \quad i > k.$$

Этим завершается доказательство леммы.

Из приведенной леммы ясно, что единственно возможным на k-м шаге метода НТ является заполнение в последних $n-k+1$ строках и $n-k$ столбцах матрицы $A^{(k)}$. Для минимизации этого заполнения воспользуемся приведенными ниже теоремами (6.4.8) и (6.4.9). Пусть B_k — матрица, полученная путем замены ненулевых элементов единицами в последних $n-k+1$ строках и столбцах матрицы $A^{(k)}$. Обозначим j-й столбец этой матрицы через $b_j^{(k)}$, а (i, j)-й эле-

мент — через $b_{ij}^{(k)}$, так что

$$b_j^{(k)} = B_k \tilde{e}_j, \quad b_{ij}^{(k)} = \tilde{e}_i' B_k \tilde{e}_j,$$

где $\tilde{e}_j$ — j-й столбец единичной матрицы I_{n-k+1}. Тогда имеем следующую теорему.

Теорема 6.4.8. *Если* $b_{11}^{(k)} = 1$, *то максимальное значение заполнения на* k*-м шаге метода* НТ *дается первым диагональным элементом матрицы* G_k, *определенной формулой* (6.3.5).

Доказательство. Из формул (6.4.1), (6.4.2), (6.4.3) и (6.4.4) имеем

$$a_q^{(k+1)} = a_q^{(k)} - \alpha_k^{-1} \left(\hat{\eta}^{(k)\prime} a_q^{(k)}\right) \hat{\eta}^{(k)}, \quad q > k.$$

Но

$$\eta^{(k)\prime} a_q^{(k)} = \sum_{i=k}^{m} a_{ik}^{(k)} a_{iq}^{(k)} \pm \beta_k a_{kq}^{(k)},$$

и так как равенство $b_{11}^{(k)} = 1$ гарантирует выполнение неравенства $a_{kk}^{(k)} \neq 0$, то условие

$$b_1^{(k)\prime} * b_j^{(k)} = 0,$$

где $j = q - k + 1$, влечет за собой $\sum_{i=k}^{m} a_{ik}^{(k)} a_{iq}^{(k)} = 0$ и $\hat{\eta}_{A^{(k)}}^{(k)\prime} a_q^{(k)} = 0$. Следовательно, в q-м столбце матрицы $A^{(k)}$ не будет заполнения, если равенство $b_1^{(k)\prime} * b_j^{(k)} = 1$ не имеет места. Если $b_1^{(k)\prime} * b_j^{(k)} = 1$, то заполнение не может превышать величины $\left(b_1^{(k)\prime} * b_j^{(k)}\right) \bar{b}_j^{(k)\prime} b_1^{(k)}$, где вектор $\bar{b}_j^{(k)}$ получен из вектора $b_j^{(k)}$ путем замены в нем всех единиц нулями, а всех нулей — единицами. (Заметим, что равенство $b_1^{(k)\prime} * b_j^{(k)} = 1$ не означает, что $\hat{\eta}^{(k)\prime} a_q^{(k)} \neq 0$; так как может иметь место взаимное уничтожение слагаемых скалярного произведения.) Учитывая равенство $\bar{b}_t^{(k)\prime} b_t^{(k)} = 0$, можно выразить максимум заполнения всех столбцов матрицы $A^{(k)}$ в виде

$$g_{11}^{(k)} = \sum_{j=1}^{n-k+1} \left(b_1^{(k)\prime} * b_j^{(c)}\right) \bar{b}_j^{(c)\prime} b_1^{(k)},$$

что соответствует формуле (6.3.6) при $t = 1$. Подобным же образом мы заключаем, что

$$g_{11}^{(k)} = \tilde{e}_1' G_k \tilde{e}_1.$$

Этим завершается доказательство теоремы.

Заметим, что если $a_{kk}^{(k)} = 0$, то может существовать столбец ν, для которого $\hat{\eta}^{(k)'} a_\nu^{(k)} \neq 0$, но $a_k^{(k)'} a_\nu^{(k)} = 0$, поэтому будет иметь место излишнее заполнение в ν-м столбце. Этого можно избежать перестановкой k-й и s-й строк матрицы $A^{(k)}$ перед k-м шагом метода НТ, причем s определяется из условия $a_{sk}^{(k)} \neq 0$. Для учета перестановок строк и столбцов имеется следующая теорема.

Теорема 6.4.9. *При перестановке двух столбцов матрицы B_k должны быть переставлены также соответствующие диагональные элсменты матрицы G_k. Перестановка же двух строк матрицы B_k никакого влияния на матрицу G_k не оказывает.*

Доказательство. Пусть матрицы P_k и Q_k получены из единичной матрицы I_{n-k+1} путем перестановки двух строк и двух столбцов соответственно. Теперь если мы замещаем матрицу B_k матрицей $P_k B_k Q_k$, то правая часть уравнения (6.3.5) будет равна

$$(Q_k' B_k' P_k' * P_k B_k Q_k) Q_k' \bar{B}_k' P_k' P_k B_k Q_k =$$

$$= Q_k' (B_k' * B_k) \bar{B}_k' B_k Q_k = Q_k' G_k Q_k,$$

так как

$$P_k' * P_k = Q_k Q_k' = Q_k * Q_k' = P_k' P_k = I_{n-k+1}.$$

Этим завершается доказательство.

Из теорем (6.4.8) и (6.4.9) следует, что для минимизации заполнения в методе НТ мы определяем к началу k-го шага индекс $\hat{s}$ согласно формуле (6.3.7) и другой индекс $s \geqslant k$, такой, что $a_{s,\,\hat{s}+k-1}^{(k)} \neq 0$. Затем переставляем k-й и $(\hat{s} + k - 1)$-й столбцы и k-ую и s-ую строки матрицы $A^{(k)}$. Очевидно, столбцы матрицы B_k, содержащие всего один ненулевой элемент, не приводят к какому-либо заполнению и должны рассмат-

риваться первыми. Это эквивалентно предварительной перестановке строк и столбцов матрицы A, которая позволила бы получить наибольшую верхнюю треугольную матрицу в верхнем левом углу, прежде чем применить метод HT к оставшейся матрице.

6.5. Сопоставление заполнений в методах RGS и HT

В предыдущем разделе мы отметили, что метод HT может быть применен к заданной матрице A для получения матрицы $\hat{A}'$ с ортонормированными столбцами. Соотношение между матрицами A и $\hat{A}'$ дается формулой (6.4.6). Если мы вспомним, что $\hat{A}'$ обозначает матрицу, составленную из первых n столбцов матрицы $H_1'H_2' \ldots H_n'$, и будем ее хранить в факторизованной форме, то потребуются только ненулевые элементы всех векторов $\eta^{(k)}$ и все α_k. Хранение матрицы $\hat{A}'$ в факторизованной форме не влечет за собой никаких осложнений, так как матрица $\hat{A}'$ в дальнейшем обычно используется для умножения на вектор (или матрицу). Мы сейчас покажем, что матрица $\hat{A}'$ в факторизованной форме является более разреженной, чем матрица с ортонормированными столбцами, полученная в методе RGS.

Принимая во внимание теоремы 6.3.4, 6.4.8 и 6.4.9, можно утверждать, что максимально возможное заполнение на k-м шаге и в методе RGS, и в методе HT дается минимальными диагональными элементами соответствующих матриц G_k. Из уравнения (6.3.5) и из того, что матрица B_k имеет размеры $m \times (n-k+1)$ для метода RGS и $(n-k+1) \times (n-k+1)$ для метода HT, ясно, что на k-м шаге в методе RGS создается, вообще говоря, большее число новых ненулевых элементов, чем на соответствующем шаге в методе HT. Более того, из формулы (6.4.3) следует, что только последние $n-k+1$ ненулевых элементов k-го столбца матрицы $A^{(k)}$ хранятся для вектора $\hat{\eta}^{(k)}$ в методе HT в противоположность эквивалентному хранению ненулевых элементов m-мерного вектора, k-го столбца матрицы $A^{(k)}$, в методе RGS. Поэтому является оче-

видным, что факторизованная форма матрицы A', вообще говоря, значительно более разрежена, чем ортонормированные столбцы, полученные в методе RGS.

6.6. Метод Якоби

Вращения Якоби, являющиеся элементарными ортогональными преобразованиями (Уилкинсон (1965)), могут также использоваться для преобразования заданной матрицы в матрицу $A^{(n+1)}$, которая аналогична матрице, полученной в методе HT. Метод Якоби (JM) состоит из $n-1$ основных шагов, каждый из которых в свою очередь состоит из нескольких промежуточных шагов. Если $A^{(k)}$ обозначает матрицу в начале k-го основного шага, то первые $k-1$ столбцов матрицы $A^{(k)}$ уже имеют форму верхней треугольной матрицы. Если (i, j)-й элемент матрицы $A^{(k)}$ обозначить через $a_{ij}^{(k)}$, то в течение k-го основного шага все $a_{ik}^{(k)} \neq 0$, $i > k$, обращаются в нули. На каждом промежуточном шаге с помощью плоского вращения преобразуется в нуль один из элементов $a_{ik}^{(k)} \neq 0$, $i > k$. Таким образом, общее число промежуточных шагов на k-м основном шаге равно числу ненулевых элементов $a_{ik}^{(k)}$, $i > k$. Рассмотрим первый промежуточный шаг k-го основного шага. Если $a_{pk}^{(k)}$ является первым ненулевым элементом k-го столбца, который лежит под диагональю матрицы $A^{(k)}$, то мы определяем ортогональную матрицу R_{pk} следующим образом:

$$R_{pk} = I_n + (\tau - 1)(e_k e_k' + e_p e_p') + \omega (e_k e_p' - e_p e_k'), \qquad (6.6.1)$$

где

$$\tau = \frac{a_{kk}^{(k)}}{\left(a_{kk}^{(k)^2} + a_{pk}^{(k)^2}\right)^{1/2}} \qquad (6.6.2)$$

и

$$\omega = \frac{a_{pk}^{(k)}}{\left(a_{pk}^{(k)^2} + a_{pk}^{(k)^2}\right)^{1/2}}.$$

Таким образом, матрица R_{pk} получена из единичной матрицы путем замены ее элементов (k, k), (k, p),

(p, k) и (p, p) соответственно величинами τ, ω, $-\omega$ и τ. Мы сейчас покажем, что все строки матриц $A^{(k)}$ и $R_{pk}A^{(k)}$ одинаковы, за исключением k-й и p-й строк, которые взаимодействуют друг с другом, причем (p, k)-й элемент матрицы $R_{pk}A^{(k)}$ обращается в нуль.

Для i, не равного k или p, из формулы (6.6.1) и из того, что $e_i'e_j = 0$, $i \neq j$, имеем

$$e_i'R_{pk}A^{(k)} = e_i'A^{(k)}. \tag{6.6.3}$$

С другой стороны,

$$\begin{aligned} e_k'R_{pk}A^{(k)} &= (e_k' + (\tau - 1)\, e_k' + \omega e_p')\, A^{(k)} = \\ &= \tau e_k'A^{(k)} + \omega e_p'A^{(k)}. \end{aligned} \tag{6.6.4}$$

Аналогично

$$\begin{aligned} e_p'R_{pk}A^{(k)} &= (e_p' + (\tau - 1)\, e_p' - \omega e_k')\, A^{(k)} = \\ &= \tau e_p'A^{(k)} - \omega e_k'A^{(k)}. \end{aligned} \tag{6.6.5}$$

Таким образом, p-я и k-я строки матрицы $R_{pk}A^{(k)}$ являются линейными комбинациями соответствующих строк матрицы $A^{(k)}$. Наконец, из формул (6.6.5) и (6.6.2) следует, что

$$e_p'R_{pk}A^{(k)}e_k = \tau a_{pk}^{(k)} - \omega a_{kk}^{(k)} = 0. \tag{6.6.6}$$

Из формул (6.6.4) и (6.6.5) видно, что заполнение имеет место не только в p-й строке, но также и в k-й строке матрицы $A^{(k)}$. Так как k-я строка матрицы $R_{pk}A^{(k)}$ используется снова на следующем промежуточном шаге для преобразования некоторого элемента $a_{qk}^{(k)}$, $q > p$, в нуль, то ненулевые элементы, созданные в k-й строке на первом шаге, могут также создавать ненулевые элементы в q-й строке. Это случится всякий раз, когда $a_{kj}^{(k)} = 0$, $a_{pj}^{(k)} \neq 0$ и $a_{qj}^{(k)} = 0$. Мы назовем это *взаимодействием второго порядка между* p-й и q-й строками. Аналогично имеем *взаимодействие третьего и высшего порядка* между строками. Таким образом, важно минимизировать заполнение k-й строки на каждом промежуточном шаге. Так же, как и в методах HT и RGS, мы рассмотрим матрицу B_k, полученную из последних $n - k + 1$ строк и столбцов матрицы $A^{(k)}$ путем замены каждого ненулевого элемента едини-

цей. Если $(s+k-1,\ t+k-1)$-й элемент матрицы $A^{(k)}$ перемещен в $(k,\ k)$-ю позицию в начале k-го основного шага метода Якоби, то заполнение может быть определено с помощью матрицы B_k, если пренебречь взаимным уничтожением слагаемых в процессе вычислений. Во всяком случае, если принять в расчет взаимное уничтожение слагаемых при вычислениях, то фактическое заполнение будет меньшим, чем заполнение, вычисленное с помощью матрицы B_k. Из формул (6.6.4) и (6.6.5) видно, что общее заполнение на k-м шаге зависит не только от s-й строки матрицы B_k, но также и от каждой i-ой строки матрицы B_k, для которой $b_{it}^{(k)}=1$, $b_{it}^{(k)}=e_i'B_ke_t$. Заполнение будет тем меньше, чем меньше строк, для которых $b_{it}^{(k)}=1$, и чем меньше в этих строках содержится ненулевых элементов.

Общее число единиц во всех строках, для которых $b_{it}^{(k)}=1$, дается формулой (3.2.6):

$$d_t^{(k)}=\sum_i r_i^{(k)},$$

где суммирование производится для всех i, для которых $b_{it}^{(k)}=1$. Используя формулу (3.2.2), имеем

$$d_t^{(k)}=\sum_i b_{it}^{(k)}e_i'B_kV_k=\sum_i e_t'B_k'e_ie_i'B_kV_k$$

или

$$d_t^{(k)}=e_t'B_k'B_kV_k. \tag{6.6.7}$$

Таким образом, для минимизации заполнения мы выбираем столбец t следующим образом. Определяем

$$d_t^{(k)}=\min_j d_j^{(k)}. \tag{6.6.8}$$

и затем, чтобы минимизировать заполнение, вызванное взаимодействием второго и высшего порядков, располагаем строки, для которых $b_{it}^{(k)}=1$, в порядке возрастания значений всех $r_i^{(k)}$ и выбираем s из условия

$$r_s^{(k)}=\min_i r_i^{(k)}, \tag{6.6.9}$$

где минимум ищется по всем i, для которых $b_{it}^{(k)}=1$.

Можно выбрать s и t, используя общее заполнение k-й строки на соответствующем основном шаге. Кроме того, может быть определено заполнение и для других строк, но это требует затраты слишком большого труда, и потому не удобно для практического использования.

В матрице B_k (s, j)-й элемент будет ненулевым в конце k-го шага в том случае, если

$$e_t' B_k' * B_k e_j = 1.$$

Поэтому общее число новых ненулевых элементов в строке s будет равно

$$\gamma_{st}^{(k)} = \sum_j e_t' B_k' * B_k e_j - r_s^{(k)}$$

или, если воспользоваться формулой (3.2.2),

$$\gamma_{st}^{(k)} = e_t' (B_k' * B_k) V_k - e_s' B_k V_k. \qquad (6.6.10)$$

Таким образом, мы можем выбрать s и t из условия

$$\gamma_{st}^{(k)} = \min_{[i,\, j]} \gamma_{ij}^{(k)}, \qquad (6.6.1)$$

где минимум берется по всем значениям i и j, при которых $b_{ij}^{(k)} = 1$.

6.7. Библиография и комментарии

Метод триангуляризации Хаусхолдера и метод Якоби с анализом ошибок округления изложены у Уилкинсона (1965). На основании экспериментальных вычислений Райс (1966) показал, что модифицированный метод Грама — Шмидта приводит к лучшим результатам, чем обычный метод Грама — Шмидта. Анализ ошибок округления для метода RGS дается Бьёрком (1967). Применение программ ортонормирования в численном анализе рассматривается Дэвисом (1962).

Глава 7

СОБСТВЕННЫЕ ЗНАЧЕНИЯ И СОБСТВЕННЫЕ ВЕКТОРЫ

7.1. Введение

Имеются два хорошо известных прямых метода для вычисления собственных значений и собственных векторов симметричных матриц: *метод Гивенса* (GM) и *метод Хаусхолдера* (HM). По существу они эквивалентны методам триангуляризации Якоби и Хаусхолдера, описанным в гл. 6. В обоих методах применяется ряд ортогональных преобразований подобия для приведения заданной матрицы к трехдиагональной форме, так как собственные значения и собственные векторы трехдиагональной матрицы легко определяются (Уилкинсон (1965); Фокс (1965)). В следующих двух разделах мы дадим краткое описание этих методов и обрисуем некоторые приемы минимизации заполнения в случае, когда заданная матрица преобразуется к трехдиагональной форме (Тьюарсон (1970a)).

В случае несимметричных матриц применяется модификация гауссова исключения для приведения заданной матрицы к форме Хессенберга, в которой все $a_{ij} = 0$ при $i > j + 1$ (Уилкинсон (1965)). Собственные значения матрицы Хессенберга легко находятся (Фокс (1965)). В разд. 7.4 вслед за кратким описанием этого метода приводятся способы минимизации заполнения (Тьюарсон (1970c)).

Если заданная матрица A симметричная, то во многих случаях можно произвести такую перестановку, при которой верхний левый угол результирующей матрицы будет иметь трехдиагональную форму. Это можно осуществить, если удастся найти строку, в которой не больше одного внедиагонального элемента. Эту строку (и соответствующий столбец) перемещаем так, чтобы она стала первой строкой (первым столбцом).

Затем мы исключаем первую строку и первый столбец из дальнейшего рассмотрения и повторяем описанную выше процедуру для оставшихся строк и столбцов. Если на каком-либо шаге нельзя будет найти строку с одним внедиагональным элементом, то процесс прекращается. Ясно, что на этом шаге левый верхний угол преобразованной матрицы будет иметь трехдиагональную форму. Теперь нам остается только преобразовать квадратную матрицу в нижнем правом углу к трехдиагональной форме, применив один из методов GM или HM. Поэтому в остальной части настоящей главы эту подматрицу без потери общности будем обозначать через A.

7.2. Метод Гивенса

Этот метод приводит матрицу A к трехдиагональной форме с помощью вращений Якоби (Уилкинсон (1965)). К началу k-го основного шага первые $k-1$ строк и столбцов матрицы $A^{(k)}$ имеют трехдиагональную форму. Основной k-й шаг состоит не больше чем из $n-k-1$ промежуточных шагов, в процессе которых последовательно вводятся нули в позиции $k+2$, $k+3$, ..., n k-й строки и k-го столбца. Применяя обозначения, подобные тем, что были в разд. 6.6, определим

$$R_{pk} = I_n + (\tau - 1)\left(e_{k+1}e'_{k+1} + e_p e'_p\right) + \\ + \omega\left(e_{k+1}e'_p - e_p e'_{k+1}\right), \qquad (7.2.1)$$

где $a_{pk}^{(k)}$ — первый ненулевой элемент после $(k+1)$-й строки в k-м столбце. Тогда первый промежуточный шаг в основном k-м шаге может быть представлен в виде

$$A_1^{(k)} = R_{pk}A^{(k)}R'_{pk}. \qquad (7.2.2)$$

Теперь $e'_i R_{pk} = e'_i$, если $i \neq k+1$, p. Поэтому все строки и столбцы матриц $A_1^{(k)}$ и $A^{(k)}$, имеющие индексы, отличные от $k+1$ и p, будут одинаковыми и

$$e'_{k+1}R_{pk}A^{(k)} = \left(e'_{k+1} + (\tau - 1)\,e'_{k+1} + \omega e'_p\right)A^{(k)} = \\ = \tau e'_{k+1}A^{(k)} + \omega e'_p A^{(k)}. \qquad (7.2.3)$$

Аналогично имеем

$$e'_p R_{pk} A^{(k)} = \tau e'_p A^{(k)} - \omega e'_{k+1} A^{(k)}. \qquad (7.2.4)$$

Таким образом, $(k+1)$-я и p-я строки матрицы $R_{pk}A^{(k)}$ являются линейными комбинациями соответствующих строк матрицы $A^{(k)}$.

Подобным же образом можно убедиться, что в матрице $R_{pk}A^{(k)}R'_{pk}$ $(k+1)$-й и p-й столбцы являются линейными комбинациями соответствующих столбцов матрицы $R_{pk}A^{(k)}$. Более того, если положить

$$\tau = \frac{a^{(k)}_{k+1,\,k}}{\left((a^{(k)}_{pk})^2 + (a^{(k)}_{k+1,\,k})^2\right)^{\frac{1}{2}}}$$

и

$$\omega = \frac{a^{(k)}_{pk}}{\left((a^{(k)}_{pk})^2 + (a^{(k)}_{k+1,\,k})^2\right)^{\frac{1}{2}}},$$

то

$$\begin{aligned} e'_p A^{(k)}_1 e_k &= e'_p R_{pk} A^{(k)} R'_{pk} e_k = \\ &= \left(\tau e'_p A^{(k)} - \omega e'_{k+1} A^{(k)}\right) e_k = \\ &= \tau a^{(k)}_{pk} - \omega a^{(k)}_{k+1,\,k} = 0. \end{aligned}$$

Так же можно показать, что $e'_k A^{(k)}_1 e_p = 0$. Таким образом, (p, k)-й и (k, p)-й элементы матрицы $A^{(k)}_1$ равны нулю. Повторным применением формулы (7.2.2) преобразуются в нуль все элементы, расположенные в позициях $k+2$, $k+3$, ..., n k-й строки и k-го столбца матрицы $A^{(k)}$. Результирующая матрица обозначается через $A^{(k+1)}$ и k-й основной шаг завершен. Нам было бы желательно определить такой элемент $a^{(k)}_{s+k,\,t+k-1} \neq 0$, $s \neq t-1$, что если с помощью симметричной перестановки строк и столбцов его сделать $(k+1, k)$-м элементом, то заполнение будет минимальным [1]). Это

[1]) Автор не учитывает, что выбор главного элемента в столбце $t+k-1$, $t \geqslant 2$, и связанная с этим перестановка k-го и $(t+k-1)$-го столбцов влекут за собой симметричную перестановку k-й и $(t+k-1)$-го строк. В результате ненулевой эле-

трудная комбинаторная задача, и ее решение не могло бы быть полезным для практики из-за чрезмерно большого объема вычислений, которого оно потребовало бы. Поэтому обычно предпочитают методы, близкие к оптимальным, не связанные с большой затратой труда. Анализ, приведенный в разд. 6.6, с минимальными изменениями может быть применен также и здесь. Отличие заключается в том, что в разд. 6.6 k-я строка взаимодействует с другими строками, в то время как здесь $(k+1)$-я строка взаимодействует с другими строками, а затем $(k+1)$-й столбец взаимодействует с другими столбцами. Так как матрицы $A^{(k)}$, $A_1^{(k)}$ и $A^{(k+1)}$ все симметричные, то операции для достижения минимума заполнения строк приведут, вообще говоря, также и к минимизации заполнения столбцов. Если B_k — матрица, полученная из последних $n-k$ строк и $n-k+1$ столбцов матрицы $A^{(k)}$ путем замены каждого ненулевого элемента единицей, то или с помощью формул (6.6.8) и (6.6.9), или применяя формулу (6.6.11) можно определить s и t, причем $s \neq t-1$ (так как диагональный элемент матрицы $A^{(k)}$ не может быть сделан $(k+1, k)$-м элементом симметричными перестановками строк и столбцов).

Интересная модификация метода GM для ленточных матриц дана Шварцем (1968). В этом методе для приведения симметричной ленточной матрицы к трехдиагональной форме применяется соответствующая последовательность вращений, аналогичных вращению, представленному формулой (7.2.1). В течение всего процесса преобразований сохраняется свойство ленточности заданной матрицы (см. рис. 7.2.1). Исключение двух симметричных элементов внутри ленты создает, вообще говоря, два одинаковых ненулевых элемента в симметричных позициях вне ленты. Эти элементы исключаются последовательностью вращений, которые сдвигают их вниз на λ строк и λ столбцов $(2\lambda+1-$

мент, стоящий в позиции $(k, k-1)$, займет позицию $(t+k-1, k-1)$, где ранее стоял нуль. Таким образом, первые $k-1$ строк и столбцов утратят тот вид, который был предположен в начале раздела. — *Прим. ред.*

ширина ленты заданной матрицы) и в конечном счете за границы матрицы. На рис. 7.2.1 показан весь процесс исключения элемента a_{41} при $n = 10$ и $\lambda = 3$, причем ввиду симметрии матрицы A рассматриваются только элементы, лежащие на главной диагонали и

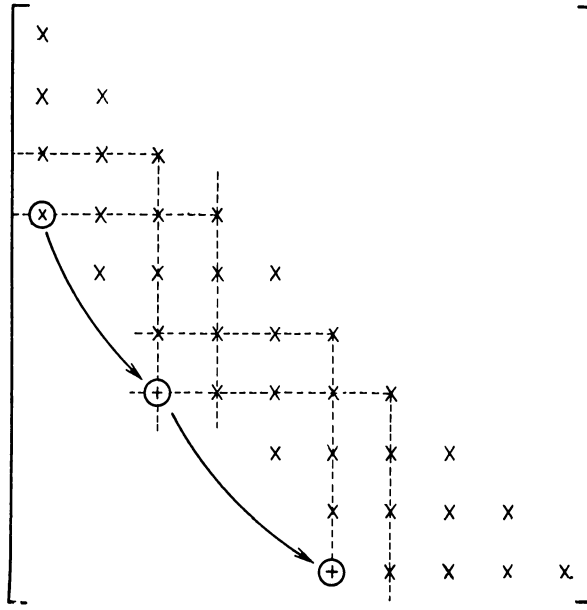

Рис. 7.2.1. Вращения для ленточной матрицы.

под ней. Пусть вращение R_{pq} определяется так же, как в формуле (6.6.1), и τ и ω выбраны из условия преобразования данного элемента в нуль. Например, на рис. 7.2.1 сначала применяется вращение R_{34} для того, чтобы сделать a_{41} равным нулю. Это создает ненулевой элемент в (7, 3)-й позиции и используется R_{67} для обращения его в нуль. В свою очередь это создает (10, 6)-й ненулевой элемент. Тогда с помощью $R_{9,10}$ обращается в нуль (10, 6)-й элемент.

Подобным же образом с помощью вращения R_{23} (которое оставляет (4, 1)-й элемент без изменения)

исключается (3, 1)-й элемент. Два дополнительных вращения сдвинут ненулевые элементы, созданные в шестой строке и втором столбце, вниз и за границу матрицы. Продолжение этого процесса для других строк и столбцов преобразует матрицу A к трехдиагональной форме.

7.3. Метод Хаусхолдера

Этот метод приводит матрицу A к трехдиагональной форме с помощью элементарных эрмитовых ортогональных матриц (Уилкинсон (1965)). В начале k-го шага матрица $A^{(k)}$ имеет трехдиагональную форму для своих первых $k-1$ строк и столбцов. На k-м шаге вводятся нули в k-ю строку и k-й столбец, причем сохраняются нули, введенные на предыдущих шагах. Другими словами, все элементы в позициях $k+2$, $k+3$, ..., n k-й строки и k-го столбца обращаются в нуль. Метод подобен тому, что описан в разд. 6.4. Он состоит из $n-2$ шагов, таких, что

$$A^{(k+1)} = H_k A^{(k)} H_k, \quad k = 1, 2, \ldots, n-2 \qquad (7.3.1)$$

($A^{(1)} \equiv A$ и $A^{(n-1)}$ — трехдиагональная матрица), где

$$H_k = I_n - \alpha_k^{-1} \hat{\eta}^{(k)} \hat{\eta}^{(k)'} \qquad (7.3.2)$$

и элементы вектора-столбца $\hat{\eta}^{(k)}$ заданы соотношениями

$$\hat{\eta}_i^{(k)} = 0, \quad i \leqslant k, \qquad (7.3.3)$$

$$\hat{\eta}_{k+1}^{(k)} = a_{k+1,\,k}^{(k)} \pm \beta_k, \quad \hat{\eta}_i^{(k)} = a_{ik}^{(k)}, \quad i > k+1,$$

причем

$$\beta_k^2 = \sum_{i=k+1}^{n} (a_{ik}^{(k)})^2, \quad \alpha_k = \beta_k^2 \pm \beta_k a_{k+1,\,k}^{(k)}. \qquad (7.3.4)$$

Для устойчивости знак β_k должен быть выбран таким же, как и знак $a_{k+1,\,k}^{(k)}$.

Пусть B_k — матрица, полученная из последних $n-k$ строк и $n-k+1$ столбцов матрицы $A^{(k)}$ путем замены каждого ненулевого элемента единицей. Тогда аналогично теореме 6.4.8 имеем следующую теорему.

Теорема 7.3.5. *Если* $b_{11}^{(k)} = 1$ *и*

$$\hat{A}^{(k)} = H_k A^{(k)}, \tag{7.3.6}$$

где матрица H_k *определяется соотношениями* (7.3.2), (7.3.3) *и* (7.3.4), *то максимальное значение заполнения дается* (1, 1)-*м элементом матрицы* G_k, *определенной по формуле* (6.3.5).

Доказательство. Если мы сравним формулы (7.3.6), (7.3.2), (7.3.3) и (7.3.4) соответственно с формулами (6.4.1), (6.4.2), (6.4.3) и (6.4.4), то увидим, что отличие между ними заключается в том, что в первой группе формул не используется k-я строка матрицы $A^{(k)}$. Это отличие учитывается в этом разделе тем, что матрица B_k составляется из последних $n - k$ строк, а не из последних $n - k + 1$ строк, как в теореме 6.4.8. Исходя из сказанного, дальнейший ход доказательства настоящей теоремы такой же, как и для теоремы 6.4.8, и мы его здесь повторять не будем.

Из определения G_k видно, что теорема 6.4.9 применима также и к матрице B_k, определенной в этом параграфе. Поэтому если мы воспользуемся формулой (6.3.7) для выбора $\hat{s}$ и затем возьмем такое $s = \hat{s} - 1$, при котором $b_{s\hat{s}}^{(k)} = 1$, и переместим $(s + k, \hat{s} + k - 1)$-й элемент матрицы $A^{(k)}$ на $(k + 1, k)$-ю позицию перед применением преобразования (7.3.6), то заполнение будет минимизировано[1]). Так как в формуле (7.3.1) и матрица $A^{(k+1)}$, и матрица $A^{(k)}$ симметричные, то описанный выше выбор s и $\hat{s}$ для минимизации заполнения в случае, когда матрица $A^{(k)}$ умножается слева на матрицу H_k, будет, вообще говоря, минимизировать также заполнение в случае, когда матрица $H_k A^{(k)}$ умножается справа на матрицу H_k. Можно вычислить действительное заполнение для преобразования (7.3.1), но так как это связано со слишком большими вычислениями, то никакого преимущества для практики такое вычисление не дает.

---

[1]) См. примечание на стр. 154.

В свете изложенного, прежде чем выполнить k-й шаг метода НМ, мы перемещаем $(s+k, \hat{s}+k-1)$-й элемент матрицы $A^{(k)}$ в $(k+1, k)$-ю позицию симметричными перестановками строк и столбцов. Этим достигается то, что заполнение будет небольшим.

Более простым путем нахождения $\hat{s}$ является выбор столбца матрицы B_k, который взаимодействует с небольшим числом других столбцов. Другими словами, такой столбец дает минимиальное число ненулевых булевых скалярных произведений с другими столбцами:

$$e_{\hat{s}}'(B_k' * B_k)V = \min_j e_j'(B_k' * B_k)V. \qquad (7.3.7)$$

Конечно, это является, вообще говоря, менее точной оценкой заполнения, чем использование формулы (6.3.7).

До сих пор в этой главе мы рассмотрели метод Гивенса и метод Хаусхолдера для симметричных разреженных матриц. В следующем разделе мы покажем, как с помощью элементарных преобразований подобия можно привести несимметричную матрицу к верхней форме Хессенберга.

7.4. Приведение к форме Хессенберга

Пусть $A^{(k)}$ — матрица, у которой первые $k-1$ столбцов имеют форму Хессенберга, т. е. $a_{ij}^{(k)}=0$ для всех $i>j+1$ и $j<k$. Тогда на k-м шаге применяются элементарные преобразования подобия (которые очень похожи на матрицы L_k разд. 2.2) для обращения в нуль элементов k-го столбца в позициях $k+2$, $k+3$, ..., n. Это осуществляется для $k=1, 2, \ldots, n-2$, в результате чего матрица $A^{(n-1)}$ имеет форму Хессенберга. Таким образом, имеем

$$A^{(k+1)} = L_{k+1}A^{(k)}L_{k+1}^{-1}, \qquad k=1, 2, \ldots, n-2, \quad (7.4.1)$$

где

$$L_{k+1} = I_n + \eta^{(k+1)}e_{k+1}', \qquad (7.4.2)$$

причем элементы вектора-столбца $\eta^{(k)}$ даются в виде

$$\eta_i^{(k+1)} = 0, \qquad i \leqslant k+1, \tag{7.4.3}$$

$$\eta_i^{(k+1)} = -\frac{a_{ik}^{(k)}}{a_{k+1,k}^{(k)}}, \qquad i > k+1.$$

Из формул (7.4.2) и (7.4.3) следует, что

$$L_{k+1}^{-1} = I_n - \eta^{(k+1)} e_{k+1}'. \tag{7.4.4}$$

Приведенные выше уравнения означают, что для получения матрицы $A^{(k+1)}$ из матрицы $A^{(k)}$ необходимо вначале добавить ко всем строкам с элементами $a_{ik}^{(k)} \neq 0$, $i > k+1$, умноженную на соответствующий коэффициент $(k+1)$-ю строку и затем добавить к $(k+1)$-му столбцу умноженные на соответствующие коэффициенты столбцы результирующей матрицы, отвечающие элементам $a_{jk}^{(k)} \neq 0$, $j > k+1$. Таким образом, заполнение имеет место во всем $(k+1)$-м столбце и в последних $n-k-1$ строках и столбцах матрицы $A^{(k)}$. В последних $n-k-1$ компонентах $(k+1)$-го столбца заполнение имеет место дважды: один раз при умножении слева на матрицу L_{k+1} и один раз при умножении справа на матрицу L_{k+1}^{-1}. Конечно, в первых $k+1$ компонентах $(k+1)$-го столбца заполнение имеет место только в результате умножения справа на матрицу L_{k+1}^{-1}. С другой стороны, для последних $n-k-1$ строк и столбцов заполнение имеет место только при умножении слева на L_{k+1}. Пусть B_k обозначает матрицу, полученную из последних $n-k$ строк и $n-k+1$ столбцов матрицы $A^{(k)}$ путем замены каждого ненулевого элемента единицей. Тогда имеем следующую теорему.

Теорема 7.4.5. *Если* $(s+k, t+k-1)$*-й элемент матрицы* $A^{(k)}$ *перемещается в* $(k+1, k)$*-ю позицию симметричными перестановками строк и столбцов*[1]), *причем* $s \neq t-1$, *и результирующая матрица умножается слева на матрицу* L_{k+1}, *то максимальное за-*

[1]) См. примечание на стр. 154.

полнение дается (s, t)*-м элементом матрицы* G_k*, определенной формулой* (2.5.6).

Доказательство. Ограничение $s \neq t-1$ требуется потому, что диагональный элемент матрицы $A^{(k)}$ не может быть перемещен в $(k+1, k)$-ю позицию с помощью симметричных перестановок строк и столбцов. С учетом того, что (i, j)-й элемент матрицы B_k соответствует $(i+k, j+k-1)$-му элементу матрицы $A^{(k)}$, доказательство такое же, как и для теоремы 2.5.5, и нет нужды его здесь повторять.

Так как матрица L_{k+1}^{-1} изменяет только $(k+1)$-й столбец матрицы $A^{(k)}$, то, если пренебречь заполнением, которое при этом создается, можно использовать теорему 7.4.5, чтобы выбрать главный элемент для минимизации заполнения на каждом шаге. Можно вычислить заполнение для $(k+1)$-го столбца, вызванное умножением на матрицу L_{k+1}^{-1}, но это слишком трудоемкий процесс, чтобы его использовать на практике.

Теперь мы опишем, каким образом заполнение $(k+1)$-го столбца может быть минимизировано при умножении матрицы $A^{(k)}$ только справа на матрицу L_{k+1}^{-1}. Пусть N_k обозначает матрицу, полученную из последних $n-k+1$ столбцов матрицы $A^{(k)}$ путем замены каждого ненулевого элемента единицей. Пусть $\tilde{B}_k$ обозначает матрицу, состоящую только из последних $n-k+1$ строк матрицы N_k. Кроме того, пусть $I^{(q)}$ обозначает матрицу, полученную из единичной матрицы $(n-k+1)$-го порядка путем замены ее q-го диагонального элемента нулем. Тогда имеем следующую теорему.

Теорема 7.4.6. *Если* $(p+k-1, q+k-1)$*-й элемент матрицы* $A^{(k)}$ *перемещается в* $(k+1, k)$*-ю позицию путем симметричных перестановок строк и столбцов* [1]) *и результирующая матрица умножается* справа *на матрицу* L_{k+1}^{-1}*, то максимальное заполнение* $(k+1)$*-го столбца дается в виде*

$$\gamma_{pq}^{(k)} = e_p' \bar{N}_k' \left(N_k * I^{(q)} \tilde{B}_k\right) e_q. \qquad (7.4.7)$$

[1]) См. примечание на стр. 154.

Доказательство. Из соотношений (7.4.3) и (7.4.4), из определений N_k и $\tilde{B}_k$ и из условия, что в матрице $A^{(k)}$ $(q+k-1)$-й и $(p+k-1)$-й столбцы становятся соответственно k-м и $(k+1)$-м столбцами, имеем

$$\gamma_{pq}^{(k)} = (\bar{N}_k e_p)' \left(\sum_{i \neq q}^{*} \tilde{b}_{iq}^{(k)} N_k e_i \right),$$

где $\tilde{b}_{iq}^{(k)}$ — (i, q)-й элемент матрицы $\tilde{B}_k$ и $\sum^*$ означает булеву сумму столбцов. Теперь

$$\gamma_{pq}^{(k)} = e_p' \bar{N}_k' \left(N_k * \sum_{i \neq q} e_i e_i' \tilde{B}_k e_q \right) = e_p' \bar{N}_k' \left(N_k * I^{(q)} \tilde{B}_k \right) e_q,$$

что завершает доказательство теоремы.

Принимая во внимание теоремы 7.4.5 и 7.4.6 и то, что (i, j)-й элемент матрицы B_k такой же, как и $(i+1, j)$-й элемент матрицы $\tilde{B}_k$, мы можем, очевидно, выбрать главный элемент $a_{s+k,\, t+k-1}^{(k)}$ следующим образом:

$$g_{st}^{(k)} + \gamma_{s+1,\, t}^{(k)} = \min_{i,\, j} \left(g_{ij}^{(k)} + \gamma_{i+1,\, j}^{(k)} \right), \qquad i \neq j+1, \quad (7.4.8)$$

где $g_{ij}^{(k)}$ — (i, j)-й элемент матрицы G_k [1]). Это должно, вообще говоря, минимизировать заполнение.

Более простой, хотя и менее точный, путь для выбора главного элемента, основан на зависимости заполнения от общего числа ненулевых элементов в s-й строке и t-м столбце. Поэтому мы выбираем s и t следующим образом. Пусть B_k определено так, как это сделано для теоремы 7.4.5, а V_k так, как для формулы (3.2.2), но V_k' — $(n-k)$-мерный вектор-строка, все элементы которого равны единице. Тогда с учетом формул (3.2.2) и (3.2.3) уравнение

$$\hat{g}_{st}^{(k)} = \min_{i,\, j} \hat{g}_{ij}^{(k)}, \qquad i \neq j-1, \qquad (7.4.9)$$

может быть использовано для определения s и t.

[1]) См. примечание на стр. 154.

7.5. Собственные векторы

Собственный вектор x, соответствующий известному собственному значению λ, может быть легко получен потому, что уравнение $Ax = \lambda x$ означает, что

$$(A - \lambda I)\, x = 0. \qquad (7.5.1)$$

Заметим, что $A - \lambda I$ — особенная матрица, так как $x \neq 0$, и поэтому мы могли бы опустить одно из уравнений системы (7.5.1) [1]) и решить оставшуюся систему неоднородных уравнений для $n - 1$ отношений компонент вектора x. Ошибки округления и другие вопросы вычислений для этого случая упоминаются у Фокса (1965). При решении неоднородной системы уравнений могут быть использованы различные способы минимизации заполнения и (или) вычислительных затрат, приведенные в предыдущих главах.

7.6. Библиография и комментарии

Различные прямые методы вычисления собственных значений и собственных векторов полных матриц и анализ ошибок даны у Фокса (1965) и Уилкинсона (1965). Вращения Якоби для ленточных матриц описаны у Рутисхаузера (1963) и Шварца (1968). Заполнение для методов Гивенса и Хаусхолдера рассматривается Тьюарсоном (1970a). Также у Тьаюрсона (1970c) излагаются некоторые способы минимизации заполнения при приведении заданной матрицы к форме Хессенберга.

---

[1]) Здесь автор ошибается: можно опустить лишь то из уравнений (7.5.1), которое является линейной комбинацией прочих! — *Прим. ред.*

Глава 8
ИЗМЕНЕНИЕ БАЗИСА И РАЗНЫЕ ВОПРОСЫ

8.1. Введение

В некоторых приложениях необходимо вносить некоторые изменения в матрицу A после того, как обратная к ней матрица (в форме PFI или EFI) определена. Это имеет место, например, в линейном программировании, где на каждом шаге симплексного метода один столбец «базиса» заменяется «небазисным» столбцом, и обратная матрица базиса изменяется так, чтобы стать обратной к измененному базису (Орчард-Хейс (1968)). Другим примером, из области электрических цепей, является метод, известный под названием метода разбиения Крона (Крон (1963)), где изменение в матрице A задается матрицей малого ранга. В этой главе мы рассмотрим некоторые методы включения результатов изменения матрицы A в обратную к ней матрицу (Данциг (1963б); Бартельс и Голуб (1969); Брейтон и др. (1969); Форрест и Томлин (1972)). В разд. 8.2 мы опишем различные методы изменения форм EFI и PFI, если изменен один столбец в заданной матрице (изменения в строках матрицы A могут быть учтены путем рассмотрения изменений соответствующих столбцов матрицы A'). Метод разбиения Крона для разреженных матриц описан в разд. 8.3. Разложение на множители матрицы, обратной к A, дается в разд. 8.4. Оно получается, если обращение матрицы U производится способом, отличным от того, который приведен в разд. 2.2.

8.2. Изменение обратной матрицы A^{-1} при изменениях в столбце матрицы A

Предположим, что имеется EFI, PFI или одна из других форм обратной к A матрицы. Формы EFI и PFI заданы соответственно формулами (2.4.1) и

(5.2.4). Пусть $\hat{A}$ обозначает матрицу, полученную из матрицы A путем замены ее q-го столбца новым столбцом, скажем $\hat{a}_q$. Мы опишем несколько возможных путей решения вопроса построения матрицы $\hat{A}^{-1}$ по известной матрице A^{-1}.

Первый метод

Если A^{-1} обозначает какую-либо форму обратной к A матрицы, то каждый столбец матрицы $A^{-1}\hat{A}$ совпадает с соответствующим столбцом единичной матрицы I_n, за исключением q-го столбца. Действительно, имеем

$$A^{-1}\hat{A} = I_n + (A^{-1}\hat{a}_q - e_q)\, e_q' = I_n + (\hat{a}_q^{(n+1)} - e_q)\, e_q', \quad (8.2.1)$$

где $\hat{a}_q^{(n+1)} = A^{-1}\hat{a}_q$. Поэтому

$$\hat{A}^{-1} = [I_n + (\hat{a}_q^{(n+1)} - e_q)\, e_q']^{-1} A^{-1} = \hat{T}_q A^{-1}, \quad (8.2.2)$$

где, принимая во внимание разд. 5.2,

$$\hat{T}_q = I_n + (\hat{\xi}^{(q)} - e_q)\, e_q', \quad (8.2.3)$$

причем

$$\hat{\xi}_i^{(q)} = -\frac{\hat{a}_{iq}^{(n+1)}}{\hat{a}_{qq}^{(n+1)}}, \quad i \neq q, \quad \text{и} \quad \hat{\xi}_q^{(q)} = \frac{1}{\hat{a}_{qq}^{(n+1)}}. \quad (8.2.4)$$

Таким образом, $\hat{A}^{-1}$ имеет на один множитель больше, чем разложение на множители матрицы A^{-1}. Напомним, что только ненулевые элементы вектора $\xi^{(q)}$ должны храниться для вычисления матрицы $\hat{T}_q$.

Другие столбцы матрицы A могут быть заменены подобным же способом. Конечно, каждая такая замена добавляет еще один множитель (подобный матрице $\hat{T}_q$) к обратной матрице. Если требуется изменить только небольшое число столбцов, то разумно пользоваться изложенным здесь методом. С другой стороны, если заменяется ряд столбцов матрицы A, как в линейном программировании, то было бы желательным избавиться от тех множителей матрицы A^{-1}, которые соответствуют заменяемым столбцам исходной матрицы A. Каждый из этих столбцов можно представить

себе удаленным из «базиса», а новый столбец (которым заменяется исходный) — вставленным на его место. В следующих двух методах, пригодных только для формы EFI, исключается множитель, соответствующий удаляемому из базиса столбцу.

Второй метод

Мы покажем, что замена a_q на $\hat{a}_q$ приводит к замещению матрицы U_q в формуле (2.4.1) матрицей T_q, определяемой формулой (8.2.3), причем вектор $\xi^{(q)}$ дается соотношениями

$$\hat{\xi}_i^{(q)} = -\frac{\hat{a}_{iq}^{(t)}}{\hat{a}_{qq}^{(t)}}, \quad i \neq q, \quad \text{и} \quad \hat{\xi}_q^{(q)} = \frac{1}{\hat{a}_{qq}^{(t)}}, \tag{8.2.5}$$

где

$$\hat{a}_q^{(t)} = U_{q+1} \ \ldots \ U_n L_n \ \ldots \ L_1 \hat{a}_q. \tag{8.2.6}$$

Очевидно, матрица $L_n \ \ldots \ L_1 \hat{A}$, за исключением q-го столбца, совпадает с верхней треугольной матрицей $A^{(n+1)}$, определяемой формулой (2.2.5). Обозначим q-й столбец матрицы $L_n \ \ldots \ L_1 \hat{A}$ через $\hat{a}_q^{(n+1)}$. Как и в разд. 2.2, разложение на множители матрицы, обратной к матрице $L_n \ \ldots \ L_1 \hat{A}$, получится, если главные элементы брать на ведущей диагонали. Так как матрицы $L_n \ \ldots \ L_1 \hat{A}$ и $L_n \ \ldots \ L_1 A$ отличаются только q-м столбцом, то ввиду соотношений (2.2.9), (2.2.10) и (2.2.11) матрицы $U_{q+1} \ \ldots \ U_n L_n \ \ldots \ L_1 \hat{A}$ и $U_{q+1} \ \ldots \ U_n L_n \ \ldots \ L_1 A$ будут совпадать, за исключением q-го столбца. Столбец q для первой из этих двух матриц выражается формулой (8.2.6). Из формул (8.2.3), (8.2.5) и (8.2.6) видно, что матрица $\hat{T}_q$ преобразует q-й столбец матрицы $U_{q+1} \ \ldots U_n L_n \ \ldots \ L_1 \hat{A}$ в вектор-столбец e_q и не повлияет на другие столбцы. Другими словами, матрица $\hat{T}_q U_{q+1} \ \ldots \ U_n L_n \ \ldots \ L_1 \hat{A}$ и матрица $U_q U_{q+1} \ \ldots \ U_n L_n \ \ldots \ L_1 A$ совпадают. Поэтому

$$I_n = U_2 \ \ldots \ U_n L_n \ \ldots \ L_1 A \equiv$$

$$\equiv U_2 \ \ldots \ U_{q-1} \hat{T}_q U_{q+1} \ \ldots \ U_n L_n \ \ldots \ L_1 \hat{A}$$

и

$$\hat{A}^{-1} = U_2 \ \ldots \ U_{q-1} \hat{T}_q U_{q+1} \ \ldots \ U_n L_n \ \ldots \ L_1. \tag{8.2.7}$$

Чтобы изменить другой, q_1-й столбец, после того как q-й столбец был заменен, необходимо учесть две возможности:

1. Если $q_1 \leqslant q$, то $\hat{T}_{q_1}$ замещает U_{q_1} и

$$\hat{a}_{q_1}^{(t)} = U_{q_1+1} \ \ldots \ U_{q-1}\hat{T}_q U_{q+1} \ \ldots \ U_n L_n \ \ldots \ L_1 \hat{a}_{q_1}.$$

2. Если $q_1 > q$, то $\hat{T}_{q_1}$ вставляется непосредственно после $\hat{T}_q$, причем

$$\hat{a}_{q_1}^{(t)} = \hat{T}_q U_{q+1} \ \ldots \ U_n L_n \ \ldots \ L_1 \hat{a}_{q_1}$$

и матрица U_{q_1} не замещается.

Из приведенных двух случаев ясно, что столбцы следует, если можно, замещать в порядке уменьшения их индексов (Брейтон и др. (1969)).

Третий метод

Этот метод особенно подходит для программ линейного программирования (Брейтон и др. (1969); Форрест и Томлин (1972)). Как и ранее, пусть a_q, q-й столбец матрицы A, замещается столбцом $\hat{a}_q$ и $\hat{A}$ обозначает измененную матрицу. Если $\hat{A}^{(n+1)} = L_n \ \ldots \ L_1 \hat{A}$ и $U = L_n \ \ldots \ L_1 A$, то только q-е столбцы матриц $\hat{A}^{(n+1)}$ и U различны. Теперь определяются элементарные матрицы $\hat{U}_q$ и $\hat{T}_q$, такие, что последние $n - q$ элементов q-й строки матрицы $\hat{U}_q \hat{A}^{(n+1)}$ равны нулю и e_q является q-м столбцом матрицы $U^{(q)} = \hat{T}_q \hat{U}_q \hat{A}^{(n+1)}$. Очевидно, матрица $U^{(q)}$ легко обращается, так как она получается из матрицы U путем замены ее q-й строки и q-го столбца соответственно на e_q' и e_q. Таким образом, принимая во внимание формулы (2.2.9), (2.2.10) и (2.2.11), $U^{(q)-1}$ получается из матрицы $U_2 \ \ldots \ U_n$, если исключить матрицу U_q и положить в формуле (2.2.11) каждое $\xi_q^{(k)} = 0$, $k > q$. Ясно, что

$$\hat{A}^{-1} = U^{(q)-1} \hat{T}_q \hat{U}_q L_n \ \ldots \ L_1. \qquad (8.2.8)$$

Чтобы воспользоваться этой формулой для вычисления $\hat{A}^{-1}$, нам нужно определить матрицы $\hat{U}_q$ и $\hat{T}_q$. Это может быть осуществлено следующим образом. Если

$$\hat{U}_q = I_n + e_q \hat{\xi}^{(q)}, \tag{8.2.9}$$

причем

$$e'_q + \hat{\xi}^{(q)} = e'_q U_{q+1} \dots U_n,$$

то, принимая во внимание условие $\hat{A}^{n+1} e_j = U e_j$, $j \neq q$, имеем

$$\begin{aligned} e'_q \hat{U}_q \hat{A}^{(n+1)} e_j &= (e'_q + \hat{\xi}^{(q)}) U e_j = \\ &= e'_q U_{q+1} \dots U_n U e_j = \\ &= e'_q e_j = 0, \quad j \neq q, \end{aligned}$$

так как $U_{q+1} \dots U_n$ преобразует последние $n - q$ столбцов матрицы U в соответствующие столбцы единичной матрицы I_n (см. разд. 2.2). Кроме того, так как $e'_j \hat{U}_q = e'_j$, $j \neq q$, матрица $\hat{U}_q$, заданная формулой (8.2.9), является желаемой матрицей, которая приводит все недиагональные элементы q-й строки матрицы $\hat{A}^{(n+1)}$ к нулю, а остальные строки сохраняет без изменений.

Пусть

$$\hat{a}_q^{(t)} = \hat{U}_q L_n \dots L_1 \hat{a}_q. \tag{8.2.10}$$

Так как $e'_q \hat{U}_q \hat{A}^{(n+1)} = \hat{a}_{qq}^{(t)} e'_q$, исключение Гаусса — Жордана, произведенное для q-го столбца матрицы $\hat{U}_q \hat{A}^{(n+1)}$, не окажет влияния на другие столбцы, и матрица $\hat{T}_q$ будет определяться той же формулой (8.2.3), в которой $\hat{\zeta}^{(q)}$ даются соотношениями (8.2.5) и $\hat{a}_q^{(t)}$ — формулой (8.2.10).

Рассмотрим, каким образом столбец матрицы $\hat{A}$, скажем столбец $\hat{a}_{q_1}$, заменить столбцом $\hat{a}_{q_1}$. Легко проверить, что

$$\hat{\hat{A}}^{-1} = U^{(q, q_1)^{-1}} \hat{T}_{q_1} \hat{U}_q \hat{T}_q \hat{U}_q L_n \dots L_1, \tag{8.2.11}$$

где матрица $U^{(q, q_1)}$ получена из матрицы $U^{(q)}$ путем замены ее q_1-й строки и q_1-го столбца соответственно

на e'_{q_1} и e_{q_1}, а матрицы $\hat{T}_{q_1}$ и $\hat{U}_{q_1}$ получены из матрицы $\hat{T}_q\hat{U}_qL_n \ldots L_1\hat{\hat{A}}$ тем же способом, каким матрицы $\hat{T}_q$ и $\hat{U}_q$ получаются из $L_n \ldots L_1\hat{A}$. Нет необходимости преобразовывать q-й столбец матрицы $\hat{U}_q\hat{A}^{(n+1)}$ к единичному вектору e_q, если запомнить, что при обращении матрицы $\hat{U}_q\hat{A}^{(n+1)}$ ее q-й столбец должен быть сперва преобразован к единичному вектору (Форрест и Томлин (1972)). В этом случае матрицы $\hat{T}_q$ и $\hat{T}_{q_1}$ в формуле (8.2.11) отсутствуют. Однако при обращении матрицы $U^{(q,\,q_1)}$ q-й и q_1-й столбцы (вначале столбец с меньшим индексом) должны быть приведены к единичным векторам ранее, чем другие столбцы матрицы.

8.3. Метод разбиения Крона

Пусть K, E и C обозначают матрицы соответственно размеров $n \times r$, $r \times r$ и $r \times n$ и

$$\hat{A} = A + KEC.$$

Тогда легко проверить непосредственным умножением, что

$$\hat{A}^{-1} = \left[I_n - A^{-1}KE\left(I_r + CA^{-1}KE\right)^{-1}C\right]A^{-1}. \quad (8.3.1)$$

Если матрица A^{-1} может быть представлена в виде разложения на множители, то матрица $\hat{A}^{-1}$ может быть вычислена следующим образом:

1. Вычислить матрицу $Y = A^{-1}KE$ размеров $n \times r$.
2. Решить уравнение $(I_r + CY)'Z' = Y'$ относительно матрицы Z' размеров $r \times n$.
3. Вычислить $\hat{A}^{-1} = (I_n - ZC)A^{-1}$.

Для хранения $\hat{A}^{-1}$ нам требуется хранить только матрицы Z, C и A^{-1}. Мы теперь покажем, что первый метод предыдущего параграфа является частным случаем метода Крона. Если $K = \hat{a}_q - a_q$, $E = 1$ и $C = e'_q$, то матрица, которая умножается справа на матрицу A^{-1} в уравнении (8.3.1), если принять во внимание

соотношения (8.2.3) и (8.2.4), равна

$$I_n - A^{-1}(\hat{a}_q - a_q)\left[1 + e_q' A^{-1}(\hat{a}_q - a_q)\right]^{-1} e_q' =$$
$$= I_n - \left(\hat{a}_q^{(n+1)} - e_q\right)\left[1 + e_q'\left(\hat{a}_q^{(n+1)} - e_q\right)\right]^{-1} e_q' =$$
$$= I_n - \left(\hat{a}_q^{(n+1)} - e_q\right)\left[\hat{a}_{qq}^{(n+1)}\right]^{-1} e_q' = \hat{T}_q.$$

Таким образом, уравнение (8.3.1) становится таким же, как и уравнение (8.2.2).

В следующем разделе мы покажем, каким образом матрицы типа (8.2.9) могут быть использованы для представления матрицы A^{-1} в факторизованной форме (Цолленкопф (1971)).

8.4. Бифакторизация

Матрица U, полученная в конце прямого гауссова исключения в разд. 2.2, может быть преобразована в единичную матрицу I_n с помощью элементарных операций над столбцами, таких, что

$$U\hat{U}_1\hat{U}_2 \ldots \hat{U}_{n-1} = I_n. \tag{8.4.1}$$

В этой формуле матрица $\hat{U}_k$ для $k = 1, 2, \ldots, n-1$ преобразует k-ю строку матрицы $U\hat{U}_1 \ldots \hat{U}_{k-1}$ в e_k' путем вычитания умноженного на различные коэффициенты k-го столбца матрицы $U\hat{U}_1 \ldots \hat{U}_{k-1}$, который равен e_k, из последующих столбцов. Очевидно, все остальные строки матрицы $U\hat{U}_1 \ldots \hat{U}_{k-1}$ остаются без изменений и последние $n-k$ строк такие же, как и у матрицы U. Поэтому матрица $\hat{U}_k$ имеет вид

$$\hat{U}_k = I_n + e_k\hat{\xi}^{(k)}, \tag{8.4.2}$$

где

$$\hat{\xi}_i^{(k)} = 0, \quad j \leqslant k, \quad \text{и} \quad \hat{\xi}_j^{(k)} = -u_{kj}, \quad j > k. \tag{8.4.3}$$

Из формулы (8.4.1) имеем $(\hat{U}_1 \ldots \hat{U}_{n-1})U = I_n$, откуда, принимая во внимание формулы (2.2.6) и (2.2.7), получаем, что

$$\hat{U}_1 \ldots \hat{U}_{n-1}L_n \ldots L_1A = I_n$$

и, следовательно,

$$A^{-1} = \hat{U}_1 \ \ldots \ \hat{U}_{n-1} L_n \ \ldots \ L_1. \qquad (8.4.4)$$

Так как матрица $A^{(k+1)}$ определяется формулами (2.2.2), (2.2.3) и (2.2.4), а матрица U — формулами (2.2.6) и (2.2.7), то первые k строк обеих матриц совпадают. Из этого условия и из соотношений (8.4.2) и (8.4.3) следует, что матрица U_k может быть вычислена, как только будет известна матрица $A^{(k+1)}$. Другими словами, все матрицы L_k и все матрицы U_k могут быть вычислены в следующем порядке:

$$L_1, \ \hat{U}_1, \ L_2, \ \hat{U}_2, \ \ldots, L_{n-1}\hat{U}_{n-1}, \ L_n.$$

Обратная подстановка, которая рассматривалась в разд. 2.2, здесь отсутствует.

8.5. Библиография и комментарии

Первый метод, описанный в разд. 8.2, хорошо известен в линейном программировании, если «базисная» обратная матрица хранится в мультипликативной форме (Данциг (1963а)). Схемы упаковки для хранения мультипликативной формы обратной матрицы (PFI) даются Смитом (1969) и Де-Бюше (1971). Элиминативная форма обратной матрицы (EFI), которая требуется для второго и третьего методов, впервые рекомендовалась Марковичем (1957) и позднее Данцигом (1963б) для специальных структур «лестничных» матриц. Данциг и др. (1969) показали превосходство формы EFI над формой PFI для матриц общего типа в линейном программировании с точки зрения скорости, точности вычисления обратной матрицы и ее разреженности. Данциг (1963б) для своего «лестничного» алгоритма исследовал изменение формы EFI, при котором сохранялась особая структура множителей. Бартельс и Голуб (1969) предложили треугольную схему изменений, которая имеет определенные желаемые свойства. Однако Форрест и Томлин (1972) отметили, что возникают некоторые практические затруднения при использовании этой схемы

для разреженных матриц больших размеров. Третий метод разд. 8.2, предложенный Брейтоном и др. (1969) и примененный и развитый Томлиным (1970) и Форрестом и Томлиным (1972) в линейном программировании, признан, по-видимому, наиболее пригодным для решения задач этой области с разреженными матрицами больших размеров. Этот метод используется в настоящее время для решения реальных практических задач линейного программирования (Форрест и Томлин (1972)).

Если сравнить первый и второй методы разд. 8.2, то, принимая во внимание соотношения (8.2.1)—(8.2.4) и (8.2.5)—(8.2.7), мы можем заключить, что не только вычисление матрицы $\hat{T}_q$ в первом методе более трудоемко, но и сама матрица имеет тенденцию быть более плотной, чем матрица $\hat{T}_q$ во втором методе. Кроме того, во втором методе матрица U_q замещается матрицей $\hat{T}_q$ в отличие от первого метода, в котором матрица $\hat{T}_q$ становится дополнительным множителем в разложении A^{-1}. Таким образом, второй метод, вообще говоря, лучше первого.

В работах Шуберта (1970) и Бройдена (1971) рассматривается выбор коррекций ранга 1 для разреженных матриц при решении нелинейных разреженных систем методами квазиньютоновского типа, при котором результирующая матрица тоже разрежена, но представляет собой лучшее приближение для якобиана.

Метод Крона (Крон (1963)) описан также Ротом (1959) и Спиллерсом (1968).

СПИСОК ЛИТЕРАТУРЫ *)

Айронс (Irons B. M.), 1970, A frontal solution program for finite element analysis, *Int. J. Numer. Methods Eng.* **2**, 5—32.

Ашкенази (Ashkenazi V.), 1971, Geodetic normal equations, In «Large Sparse Sets of Linear Equations» (J. K. Redi, ed.), pp. 57—74. Academic Press, New York.

Банч (Bunch J. R.), 1969, On direct methods for solving symmetric systems of linear equations, Ph. D. thesis. Univ. of California, Berkeley, California.

Бартельс, Голуб (Bartels R. H. and Golub G. H), 1969, The simplex method of linear programming using LU decomposition, *Comm. ACM*, **12**, 266—268.

Басейкер, Саати (Busacker R. C. and Saaty T. L.), 1965, Finite Graphs and Networks, McGraw-Hill, New York.

Бауман (Baumann R.), 1965, Some new aspects of load flow calculation, *IEEE Trans. Power Apparatus and Systems*, **85**, 1164—1176.

Бауман (Baumann R), 1971, Sparseness in power systems equations, In «Large Sparse Sets of Linear Equations» (J. K. Reid, ed.), pp. 105—126, Academic Press, New York.

Бауэр (Bauer F. L.), 1963, Optimally scaled matrices. *Numer. Math.*, **5**, 73—87.

Бейкер (Baker J. M.), 1962, A note on multiplying Boolean matrices, *Comm. ACM*, **5**, 102.

Беллман, Кук, Локетт (Bellman R., Cooke K. L. and Lockett J. A.), 1970, Algorithms. Graphs and Computers, Academic Press, New York.

Бендерс (Benders J. F.), 1962, Partitioning procedures for solving mixed-variable programming problems, *Numer. Math.*, **4**, 238—252.

Берри (Berry R. D.), 1971, An optimal ordering of electronic circuit equations for a sparse matrix solution, *IEEE Trans. Circuit Theory* CT-**18**, 40—50.

Бертеле, Бриоши (Bertele U. and Brioschi F.), 1971, On the theory of the elimination process, *J. Math. Anal Appl.*, **35**, 48—57.

Бил (Beale F. M. L.), 1971, Sparseness in linear programming. In «Large Sparse Sets of Linear Equations» (J. K. Reid. ed.), pp. 1—16. Academic Press, New York.

*) Не на все работы имеются ссылки в книге.

Брейнин (Branin F. H. Jr.), 1959, The relation between Kron's method and the classical methods of network analysis, WESCON Convention Record (Part 2), 1—29.

Брейнин (Branin F. H. Jr.), 1967, Computer methods of network analysis, *Proc. IEEE*, **55**, 1787—1801.

Брейтон, Густавсон, Уиллогби (Brayton R. K., Gustavson F. G. and Willoughby R. A.), 1969, Some results on sparse matrices, Rep. No RC2332, IBM, Yorktown Heights, New York. (A shorter version of this appeared in 1970 in *Math. Comput*, **24**, 937—954).

Бри (Bree D. Jr.), 1965, Some remarks on the application of graph theory to the solution of sparse systems of linear equations, Ph. D. thesis. Math. Dept., Princeton Univ., Princeton, New Jersey.

Бройден (Broyden C. G.), 1971, The convergence of an algorithm for solving sparse non-linear systems, *Math. Comput.*, **25**, 285—294.

Бьёрк (Björck A.), 1967, Solving linear least squares problems by Gram Schmidt orthogonalization, *Nordish Tidskr. Informations-Behandling* (BIT) **7**, 1—21.

Бэти, Стьюарт (Baty J. P. and Stewart K. L.), 1971, Organization of network equations using disection theory. In «Large Sparse Sets of Linear Equations» (J. K. Reid, ed), pp. 169—190, Academic Press, New York.

Бюше (Buchet J. de), 1971, How to take into account the low density of matrices to design a mathematical programming package. In «Large Sparse Sets of Linear Equations» J. K. Reid. ed.), pp. 211—218. Academic Press, New York.

Ван-дер-Слюиз (van der Sluis A.), 1969, Condition numbers and equilibration of matrices. *Numer. Math.*, **14**, 14—23.

Варга (Varga R. A.), 1962, Matrix Iterative Analysis, Prentice-Hall; Englewood Cliffs, New Jersey.

Вейл (Weil R. L. Jr.), 1968, The decomposition of economic production systems, *Econometrica*, **36**, 260—278.

Венке (Wenke V. K.), 1964, Praktische Anwendung linearer Wirtschafts-modelle, Unternehmenforshung, **8**, 33—46.

Вулф (Wolfe P.), 1965, Error in solution of linear programming problems. In «Error in Digital Computation» (L. B. Rall. ed.), vol 2, pp. 271—284. Wiley, New York.

Вулф (Wolfe P.), 1969, Trends in linear programming computations, In «Sparse Matrix Proceedings» (R. A. Willoughby. ed.), Rep. No RA1 (11707), pp. 107—112, IBM, Yorktown Heights, New York.

Вулф, Катлер (Wolfe P. and Cutler L.), 1963, Experiments in linear programming (R. L. Graves and P. Wolfe. eds.), pp 211—218. McGraw-Hill, New York.

Гасс (Gass S.), 1958, Linear Programming: Methods and Applications, McGraw-Hill, New York.

Гиббс (Gibbs N. E.), 1969, The bandwidth of graphs, Ph. D. thesis, Purdue Univ

Гимон, Кинг (Guymon G. L. and King I. P.), 1972, Application of the finite element method to regional transport phenomena. In «Sparse Matrices and Their Applications» (D. J. Rose and R. A. Willoughby, eds.), pp. 115—120, Plenum Press, New York.

Гир (Gear C. W.), 1971, Simultaneous numerical solution of differential-algebraic equations, *IEEE Trans. Circuit Theory*, **CT-18**, 89—95.

Глейзер (Glaser G. H.), 1972, Automatic bandwidth reduction techniques, Rep. 72—260. DBA. Systems Inc., Melbourne, Florida.

Глейзер, Салиба (Glaser G. H. and Saliba M. S.), 1972, Application of sparse matrices to analytical photogrammetry. In «Sparse Matrices and Their Applications» (D. J. Rose and R. A. Willoughby, eds.), pp. 135—146, Plenum Press, New York.

Густавсон (Gustavson F. G.), 1972, Some basic techniques for solving sparse systems of linear equations. In «Sparse Matrices and Their Applications» (D. J. Rose and R. A. Willoughby, eds.), pp. 41—52, Plenum Press, New York.

Густавсон, Лайниджер, Уиллогби (Gustavson F. G., Liniger W. and Willoughby R. A.), 1970, Symbolic generation of an optimal Crout algorithm for sparse systems of equations, *ACM J.*, **17**, 87—109.

Далмейдж, Мендельсон (Dulmage A. L. and Mendelsohn N. S.), 1962, On the inversion of sparse matrices, *Math. Comput.*, **16**, 494—496.

Далмейдж, Мендельсон (Dulmage A. L. and Mendelsohn N. S.), 1963, Two algorithms for dipartite graphs, *SIAM J. Appl. Math.*, **11**, 183—194.

Далмейдж, Мендельсон (Dulmage A. L. and Mendelsohn N. S.), 1967, Graphs and matrices, In «Graph Theory and Theoretical Physics» (F. Harary, ed.), pp. 167—277, Academic Press, New York.

Данциг (Dantzig G. B.), 1963a, Linear Programming and Extensions, Princeton Univ. Press, Princeton, New Jersey.

Данциг (Dantzig G. B.), 1963b, Compact basis triangularization for the simplex method, In «Recent Advances in Mathematical Programming» (R. L. Graves, P. Wolfe. eds.), pp. 125—132, McGraw-Hill, New York.

Данциг, Харвей, Мак-Найт, Смит (Dantzig G. B., Harvey R. P., McKnight R. D., and Smith S. S.), 1969, Sparse matrix techniques in two mathematical programming codes, In «Sparse Matrix Proceedings» (R. A. Willoughby. ed.), Rep. No. RA I (≠ 11707), pp. 85—99. IBM, Yorktown Heights, New York.

Данциг, Орчард-Хейс (Dantzig G. B. and Orchard-Hays W.), 1954, The product form of inverse in the simplex method, *Math. Comput*, 8, 64—67.

Данциг, Вулф (Dantzig G. B. and Wolfe P.), 1961, The decomposition algorithm for linear programs, *Econometrica*, **29**, 767—778.

Дафф (Duff I.), 1972, On a factored form of the inverse for sparse matrices, D. Phil thesis, Oxford University.

Дженнингс (Jennings A), 1966, A compact storage scheme for the solution of symmetric linear simultaneous equations, *Compt. J.*, **9**, 281—285.

Дженнингс (Jennings A.), 1968, A sparse matrix scheme for the computer analysis of structures, *Internat. J. Comput. Math.*, **2**, 1—21.

Дженнингс, Тафф (Jennings A. and Tuff A. D.), 1971, A direct method for the solution of large sparse symmetric simultaneous equations, In «Large Sparse Sets of Linear Equations» (J. K. Reid. ed), pp. 97—104, Academic Press, New York.

Джордж (George J. A.), 1971, Computer implementation of the finite element method, Ph. D. thesis, Comput. Sci. Dept., Stanford Univ., Stanford, California.

Джордж (George J. A.), 1972, Block elimination of finite element systems of equations, In «Sparse Matrices and Their Applications» (D. J. Rose and R. A. Willoughby, eds.), pp. 101—114, Plenum Press, New York.

Диксон (Dickson J. C.), 1965, Finding permutation operations to produce a large triangular submatrix, 28th Nat. Meeting of OR Society of America, Houston, Texas.

Дуглас (Douglas A.), 1971, Examples concerning efficient strategies for Gaussian elimination, *Computing.*, **8**, 382—394.

Дэвис (Davis P. J.), 1962, Orthonormalizing codes in numerical analysis, In «Survey of Numerical Analysis» (J. Todd. ed.), pp. 347—379, McGraw-Hill, New York.

Зевкевич (Ziewkiewicz O. C.), 1967, The Finite Element Method in Structural and Continuum Mechanics, McGraw-Hill, New York.

Ивенс (Evans D. J.), 1972, New iterative procedures for the solution sparse systems of linear difference equations, In «Sparse Matrices and Their Applications» (D. J. Rose and R. A. Willoughby, eds), pp. 89—100, Plenum Press, New York.

Иенсен (Jensen H. G), 1967, Efficient matrix techniques applied to transmission tower design, *Proc. IEEE*, **55**, 1997—2000.

Иенсен, Паркс (Jensen H. G. and Parks G. A.), 1970. Efficient solutions for linear matrix equations, *J. Struct. Div. Proc. Amer. Soc. Civil Eng.*, **96**, 40—64.

Ингерман (Ingerman P. Z.), 1962, Path matrix, *Comm. ACM*, **5**, 556.

Камсток (Comstock D. R), 1964, A note on multiplying Boolean matrices II, *Comm ACM*, **7**, 13.

Кантин (Cantin G.), 1971, An equation solver of very large capacity, *Internt. J. Numer. Methods Eng*, **3**, 379—388.

Карпентьер (Carpentier J), 1963, Ordered eliminations, Proc. Power Systems Comput Conf, London

Карпентьер (Carpentier J), 1965, Éliminations ordonnées — un processus diminuant le volume des calculs dans la résolutions des systèmes linéaires à matrice creuse, «Troisième Congr. de Calcul et de Traitement de l'Information AFCALTI», pp. 63—71. Dunod, Paris.

Каррé (Carré B. A.), 1971, An elimination method for minimal-cost network flow problems, In «Large Sparse Sets of Linear Equations» (J. K. Reid. ed.), pp. 191—210, Academic Press, New York.

Каррé (Carré B. A), 1966, The partitioning of network equations for block iterations, *Comput. J.*, **9**, 84—97.

Катхилл (Cuthill E H.), 1971, Several strategies fo reducing the bandwidth of matrices. Tech. note CMD-42-71, Naval Ship. Res. and Develop. Center, Bethesda, Maryland.

Катхилл, Мак-Ки (Cuthill E. H. and McKee J.), 1969, Reducing the bandwidth of sparse symmetric matrices, Tech. note AML-40-69, Appl. Math. Lab. Naval Ship. Res. and Develop. Center, Washington, D. C.

Кертис, Рейд (Curtis A. R. and Reid J. K.), 1971a, Fortran subroutines for the solution of sparse sets of linear equations, Rep. R 6844, Atomic Energy Res. Establishment, Harwell, England.

Кертис, Рейд (Curtis A. R. and Reid J. K.), 1971c, The solution of large sparse systems of linear equations, Proceedings of IFIP Rep. TR 450. Atomic Energy Res. Establishment, Harwell, England.

Кертис, Пауэлл, Рейд (Curtis A. R., Powell M. J. D. and Reid J. K.), 1972, On the estimation of sparse Jacobian matrices, Rep. TP 476, Atomic Energy Res. Estiblishment, Harwell, England.

Кеттлер, Вейл (Kettler P. C. and Weil R. L.), 1969, An algorithm to provide structure for decomposition, In «Sparse Matrix Proceedings» (R. A. Willoughby, ed.), Rep. No. RAI (≠11707), pp. 11—24, IBM, Yorktown Heights, New York.

Кинг (King I. P.), 1970, An automatic reordering scheme for simultaneous equations derived from network analysis, *Internat. J. Numer. Methods Eng.*, **2**, 523—533.

Клейсен (Clasen R. J.), 1966, Techniques for automatic tolerance control in linear programming, *Comm. ACM*, **9**, 802.

Клюев, Коковкин-Щербак (Klyuyev V. V. and Kokovkin-Shcherbak N. I.), 1965, On the minimization of the number of arithmetic operations for the solution of linear algebraic systems of equations (translated by G. J. Tee.), Rep. CS-24. Comput. Sci. Dept., Stanford University, Stanford, California.

Крон (Kron G.), 1963, Diakoptics, McDonald, London.

Лайвсли (Livesley R. K.), 1960—1961, The analysis of large structural systems, *Comput. J.*, **3**, 34—39.

Лайнигер, Уиллогби (Liniger W. and Willoughby R. A.), 1969, Efficient numerical integration of stiff systems of differential equations, *SIAM J. Numer. Anal*, **7**, 47—66.

Ларсон (Larson L. J.), 1962, A modified inversion procedure for product form of inverse in linear programming codes, *Comm. ACM*, **5**, 382—383.

Леви (Levy R.), 1971, Resequencing of the structural stiffness matrix to improve computational efficiency. *JPL Quart. Tech. Rev.* **1**, 61—70.

Ли (Lee H. B.), 1969, An implementation of Gaussian elimination for sparse systems of linear equations, In «Sparse Matrix Proceedings» (R. A. Willoughby, ed.), Rep. No. RA 1 (≠ 11707), pp. 75—84, IBM, Yorktown Heights, New York.

Либль, Седлачек (Liebl P. and Sedláček J.), 1966, Umformung von Quadratmatrizen auf quasitrianguläre Form mit Mitteln der Graphentheorie, *App. Mat.*, **11**, 1—9.

Луксан (Luksan L), 1972, A collection of programs for operations involving sparse matrices, Res. Rep. Z-483. Inst. of Radio Eng. and Electronics, CSAV, Prague, Czechoslovakia.

Майо (Mayoh B. H.), 1965, A graph technique for inverting certain matrices, *Math. Comput.*, **19**, 644—645.

Мак-Кормик (McCormick C. W.), 1969, Application of partially banded matrix methods to structural analysis, In «Sparse Matrix Proceedings» (R. A. Willoughby, ed.), Report No. RAI (≠ 11707), pp. 155—158, IBM, Yorktown Heights, New York.

Мак-Нами (McNamee J. M.), 1971, A sparse matrix package, Algorithm 408, *Comm. ACM*, **14**, 265—273.

Маркович (Markowitz H. M.), 1957, The elimination form of the inverse and its application to linear programming, *Management Sci*, **3**, 255—269.

Маурзер (Maurser W. D.), 1968, Programming, Introduction to Computer Languages and Techniques, Holden-Day, San Francisco, California.

Меримонт (Marimont R. B.), 1959, A new method for checking the consistency of precedence matrices, *J. Assoc. Comput. Mach.*, **6**, 164—171.

Меримонт (Marimont R. B.), 1960, Application of graphs and Boolean matrices to computer programming, *SIAM Rep.*, **2**, 259—268.

Меримонт (Marimont R. B.), 1969, System connectivity and matrix properties, *Bull. Math. Biophys.*, **31**, 255—274.

Мюллер-Мербах (Mueller-Merbach H.), 1964, On round-off errors in Linear Programming, Res. Rep., Operations Res. Center. Univ. oh California, Berkeley, California.

Надинг, Калерт-Уормболд (Nuding E. and Kahlert-Warmbold I.), 1970, A computer oriented representation of matrices, *Computing*, **6**, 1—8.

Натан, Ивен (Nathan A. and Even R. K.), 1967—1968, The inversion of sparse matrices by a strategy derived from their graphs. *Comput. J.*, **10**, 190—194.

Норин, Поттл (Norin R. S. and Pottle C.), 1971, Effective ordering of sparse matrices arising from nonlinear electrical networks, *IEEE Trans. Circuit Theory*, **CT-18**, 139—145.

Огбуобири (Ogbuobiri E. C.), 1970, Dynamic storage and retrieval in sparsity programming. *IEEE Trans. Power Apparatus Systems* **PAS 89**, 150—155.

Огбуобири (Ogbuobiri E. C.), 1971, Sparsity techniques in power-system grid-expansion planning. In «Large Sparse Sets of Linear Equations» (J. K. Reid, ed.), pp. 219—230. Academic Press, New York.

Огбуобири, Тинни, Уокер (Ogbuobiri E. C., Tinney W. F. and Walker J. W.), 1970, Sparsity-directed decomposition for Gaussian elimination on matrices, *IEEE Trans. Power. Apparatus Systems*, **PAS 89**, 141—155.

Ойфингер (Eufinger J.), 1970, Eine Untersuchung zur Auflösung magerer Gleichungssysteme, *J. Reine Angewandte Math.*, **245**, 208—220.

Ойфингер, Егер, Венке (Eufinger J., Jaeger A. and Wenke V. K.), 1968, An algorithm for the partitioning of a large system of sparse linear equations using graph theoretical methods, In «Methods of Operations Research» (R. Henn, H. P. Kunzi, H. Schubert, eds.), pp. 118—128, Verlag Anton Hain, Meisenheim, Germany.

Олвей, Мартин (Alway G. G. and Martin D. W.), 1965, An algorithm for reducing the bandwidth of a matrix of symmetrical configuration, *Comput. J.*, 8, 264—272.

Олвуд (Allwood R. J.), 1971, Matrix methods of structural analysis, In «Large Sparse Sets of Linear Equations» (J. K. Reid, ed.), pp. 17—24, Academic Press, New York.

Орчард-Хейс (Orchard-Hays W.), 1968, Advanced Linear Programming Computing Techniques, McGraw-Hill, New York.

Орчардс-Хейс (Orchard-Hays W.), 1969, MP systems technology for large sparse matrices, In Sparse Matrix Proceedings (R. A. Willoughby, ed.), Rep. No. RAl (11707), pp. 59—64, IBM, Yorktown Heights, New York.

Партер (Parter S.), 1960, On the eigenvalues and eigenvectors of a class of matrices, *J. SIAM*, 8, 376—388.

Партер (Parter S.), 1961, The use of linear graphs in Gauss elimination, *SIAM Rev.*, **3**, 119—130.

Патон (Paton K.), 1971, An algorithm for the blocks and cutnodes of a graph, *Comm. ACM*, **14**, 468—445.

Поп, Хенсон (Pope A. J. And Hanson R. H.), 1972, An algorithm for the pseudoinverse of sparse matrices, NOAA, Geodetic Research Lab., Rockville, Maryland, (Paper presented at spring A. G. U. meeting, Washington, D. C.).

Пэлекол (Palacol E. L.), 1969, The finite element method of structural analysis, In «Sparse Matrix Proceedings» (R. A. Willoughby, ed.). Report No. RAl (11707), pp. 101—106, IBM, Yorktown Heights, New York.

Рабинович (Rabinowitz P.), 1968, Applications of linear programming to numerical analysis, *SIAM Rev.*, **10**, 121—159.

Райс (Rice J. H.), 1966, Experiments on Gram-Schmidt orthogonalization, *Math. Comput.*, **20**, 325—328.

Рейд (Reid J. K.), 1971, Large Sparse Sets of Linear Equations, Proc. Oxford Conf. Inst. Math. Appl., April 1970. Academic Press, New York.

Робертс (Roberts E. J.), 1970, The fully indecomposable matrix and its associated dipartite graph an investigation of combinatorial and structural properties, Rep. TM X-58037, NASA Manned Spracecraft Center, Houston, Texas.

Роджерс (Rogers A.), 1971, Matrix Methods in Urban and Regional Analysis, Holden-Day, San Francisco, California.

Розен (Rosen R.), 1968, Matrix bandwidth minimization, *Proc. 23rd Nat. Conf. ACM Publ.* **P-68**, pp. 585—595. Brandon Systems Press, Princeton, New Jersey.

Росс, Харари (Ross I. C. and Harary F.), 1959, A description of strengthening and weakening members of a group, *Sociometry*, **22**, 139—147.

Рот (Roth J. P.), 1959, An application of algebraic topology: Kron's method of tearing, *Quart. Appl. Math.*, **17**, 1—24.

Роуз (Rose D. J), 1970a, Symmetric elimination on sparse positive definite systems and the potential flow network problem, Ph. D. thesis, Harvard Univ.

Роуз (Rose D. J.), 1970b, Triangulated graphs and the elimination process, *J. Math. Anal. Appl.*, **32**, 597—609.

Роуз (Rose D. J.), 1972, A graph-theoretic study of the numerical solution of sparse positive definite systems of linear equations, Math. Dept. Rep. University of Denver, Denver, Colorado.

Роуз, Банч (Rose D. J. and Bunch J. R.), 1972, The role of partitioning in the numerical solution of sparse systems, In «Sparse Matrices and Their Applications» (D. J. Rose and R. A. Willoughby, eds), pp. 177—190, Plenum Press, New York.

Роуз, Уиллогби (Rose D. J. and Willoughby R. A.), 1972, «Sparse Matrices and Their Applications», Proc. IBM Conf. Sept. 1970, Plenum Press, New York.

Рубинштейн (Rubinstein M. F), 1967, Combined analysis by substructures and recursion, *Proc. J. Struct. Div. ASCE*, **93**, 231—235.

Рутисхаузер (Rutishauser H.), 1963, On Jacobi Rotation patterns. *Proc. Symposia Appl. Math.* **15**, pp. 219—240, Amer. Math. Soc. Providence, Rhode Island.

Рэлстон (Ralston A.), 1965, A first Course in Numerical Analysis. McGraw-Hill, New York.

Сато, Тинни (Sato N. and Tinney W. F.), 1963, Techniques for exploiting the sparsity of the network admittance matrix, *IEEE Trans. Power Apparatus Systems*, **PAS-82**, 944—950.

Сегетова (Segethova J.), 1970, Elimination procedures for sparse symmetric linear algebraic systems of a special structure. Rep. No. 70121, Comput. Sci. Center. Univ. of Maryland.

Смит (Smith D. M.), 1969, Data logistics for matrix inversion, In «Sparse Matrix Proceedings» (R. A. Willoughby, ed.), Report No. RA 1 (11707), pp. 127—137, IBM, Yorktown Heights, New York.

Смит, Орчард-Хейс (Smith D. M. and Orchard-Hays W.), 1963, Computational efficiency in product form LP codes, In «Recent Advances in Mathematical Programming» (R. L. Graves and P. Wolfe, eds.), pp. 211—218, McGraw-Hill, New York.

Спиллерс (Spillers W. R.), 1968, Analysis of large structures: Kron's method and more recent work, *J. Struct. Dir. ASCE*, 94, **ST-11**, 2521—2534.

Спиллерс, Хикерсон (Spillers W. R. and Hickerson N.), 1968, Optimal elimination for sparse symmetric systems as a graph problem, *Quart. Appl. Math.*, **26**, 425—432.

Стаг, Эль-Абиад (Stagg G. W. and El-Abiad A H), 1968, Computer Methods in Power System Analysis, McGraw-Hill, New York.

Стьюард (Steward D. V.), 1962, On an approach to techniques for the analysis of the structure of large systems of equations, *SIAM Rev.*, **4**, 321—342.

Стьюард (Steward D. V.), 1965, Partitioning and tearing systems of equations, *SIAM J Numer Anal.*, **2**, 345—365.

Стьюард (Steward D. V.), 1969, Tearing analysis of the structure of disorderly sparse matrices, in «Sparse Matrix Proceedings» (R. A. Willoughby, ed.), Rep. No. RA1 (11707), pp. 65—74. IBM, Yorktown Heights, New York.

Тинни (Tinney W. F.), 1969, Comments on sparsity techniques for power system problems, In «Sparse Matrix Proceedings» (R. A. Willoughby, ed.), Report No. RA1 (11707), pp. 25—34, IBM, Yorktown Heights, New York.

Тинни, Огбуобири (Tinney W. F. and Ogbuobiri E. C.), 1970, Sparsity techniques: theory and practice, March 1970 Rep. Bonnville Power Administration, Portland, Oregon.

Тинни, Уокер (Tinney W. F. and Walker J. W.), 1967, Direct solutions of sparse network equations by optimally ordered triangular factorization, *Proc. IEEE*, **55**, 1801—1809.

Томлин (Tomlin J. A.), 1970, Maintaining sparse inverse in the simplex method, Rep. 70—15, Operations Res. Dept. Stanford University, Stanford, California.

Томлин (Tomlin J. A.), 1972a, Modifying triangular factors of the basis in the simplex method, In «Sparse Matrices and Their Applications» (D. J. Rose and R. A. Willoughby, eds.), pp 77—85. Plenum Press. New York

Томлин (Tomlin J. A.), 1972b, Pivoting for size and sparsity in linear programming inversion routines. *IMA J* (в печати).

Точер (Tocher J. L.), 1966, Selective inversion of stiffness matrices. *Proc. Struct. Div. ISCE*, **92**, 75—88.

Тьюарсон (Tewarson R. P.), 1966, On the product form of inverses of sparse matrices, *SIAM Rev*, **8**, 336—342.

Тьюарсон (Tewarson R. P.), 1967a, On the product form of inverses of sparse matrices and graph theory, *SIAM Rev.*, **9**, 91—99.

Тьюарсон (Tewarson R. P), 1967b, Solution of a system of simultaneous linear equations with a sparse coefficient matrix by eliminatoin methods, *Nordisk. Tidskr. Informations Behandling (BIT)*, **7**, 226—239.

Тьюарсон (Tewarson R. P), 1967c, Row column permutation of sparse matrices, *Comput, J.*, **10**, 300—305.

Тьюарсон (Tewarson R. P.), 1968a, On the orthonormalization of sparse vectors, *Computing* (Arch. Elektron. Rechnen), **3**, 268—279.

Тьюарсон (Tewarson R. P.), 1968b, Solution of linear equations with coefficient matrix in band form, *Nordisk. Tidskr. Informations Behandling (BIT)*, **8**, 53—58.

Тьюарсон (Tewarson R. P.), 1969a, The Crout reduction for sparse matrices, *Comput. J.*, **12**, 158—159.

Тьюарсон (Tewarson R. P.), 1969b, The Gaussian elimination and sparse systems. In «Sparse Matrix Proceedings» (R. A. Willoughby, ed.), Report No. RA1 (11707), pp. 35—42, IBM, Yorktown Heights, New York.

Тьюарсон (Tewarson R. P.), 1970a, On the transformation of symmetric sparse matrices to the triple diagonal form, *Internal. J. Comput. Math.*, **2**, 247—258.

Тьюарсон (Tewarson R. P.), 1970b, Computations with sparse matrices, *SIAM Rev.*, **12**, 527—544.

Тьюарсон (Tewarson R. P.), 1970c, On the reduction of a sparse matrix to Hessenberg form, *Internat. J. Comput. Math.*, **2**, 283—295.

Тьюарсон (Tewarson R. P.), 1971, Sorting and ordering sparse linear systems, In «Large Sparse Sets of Linear Equations» (J. K. Reid, ed.), pp. 151—168, Academic Press, New York.

Тьюарсон (Tewarson R. P.), 1972, On the Gaussian elimination for inverting sparse matrices, *Computing (Arch. Elektron. Rechnen)*, **9**, 1—7.

Тьюарсон, Чен (Tewarson R. P. and Cheng K. Y.), 1972, A desirable form for sparse matrices when computing their inverses in factored forms, *Computing*, **9** в печати.

Уивер (Weaver W. Jr.), 1967, Computer Programs for Structural Analysis», Van Nostrand Reinhold, Princeton, New Jersey.

Уилкинсон (Wilkinson J. H.), 1965, The Algebraic Eigenvalue Problem, Oxford Univ. Press, London and New York. (Русский перевод: Уилкинсон Дж. Х., Алгебраическая проблема собственных значений, «Наука», М., 1970.)

Уиллогби (Willoughby R. A.), ed., 1969, Sparse Matrix Proceedings, Rep. NO RA1 (11707). IBM, Yorktown Heights, New York.

Уоршелл (Warshall S.), 1962, A theorem on Boolean matrices, *J. ACM*, **9**, 11—12.

Уэстлейк (Westlake J. R.), 1968, A Handbook of Numerical Matrix Inversion and Solution of Linear Equations, Wiley, New York.

Фаддеев Д. К., Фаддеева В. Н. 1960, Вычислительные методы линейной алгебры, Физматгиз, М.

Фалкерсон, Гросс (Fulkerson D. R. and Gross O. A.), 1965, Incidence matrices and interval graphs, *Pacific J. Math.*, **15**, 835—855.

Фалкерсон, Вулф (Fulkerson D. R. and Wolfe P.), 1962, An algorithm for scaling matrices, *SIAM Rev.*, **4**, 142—146

Фокс (Fox L.), 1965, Introduction to Numerical Linear Algebra, Oxford Univ. Press (Clarendon), London and New York.

Форрест, Томлин (Forrest J. J. H. and Tomlin J. A.), 1972, Updating triangular factors of the basis to maintain sparsity in the product form simplex method, *Math. Programming*, **2**, 263—268.

Форсайт (Forsythe G. E.), 1967, Today's computational methods of linear algebra, *SIAM Rev.*, **9**, 489—515.

Форсайт, Молер (Forsythe G. E. and Moler C. B.), 1967, Computer Solution of Linear Algebraic Systems, Prentice-Hall. Englewood Cliffs, New Jersey.

Харари (Harary F.), 1959, A graph theoretic method for complete reduction of a matrix with a view toward finding its eigenvalues, *J. Math. Phys.*, **38**, 104—111.

Харари (Harary F.), 1960, On the consistency of precedence matrices, *J. Assoc. Comput. Mach.*, **7**, 255—259.

Харари (Harary F.), 1962, A graph theoretic approach to matrix inversion by partitioning, *Numer. Math.*, **4**, 128—135.

Харари (Harary F.), 1967, Graphs and Matrices, *SIAM Rev.*, **9**, 83—90.

Харари (Harary F.), 1969, «Graph Theory». Addison-Wesley. Reading, Massachusetts. [Русский перевод. Харари Ф., Теория графов, «Мир», М., 1973]

Харари (Harary F.), 1971a, Sparse matrices and graph theory, In «Large Sparse Sets of Linear Equations» (J. K. Reid, ed.) pp. 139—150, Academic Press, New York.

Харари (Harary F.), 1971b, Sparse digraphs: Classification and algorithms, Paper presented at IFIP Conf., Ljubljana, Yugoslavia.

Хаусхолдер (Householder A. S.), 1958, Unitary triangularization of a nonsymmetric matrix, *J. ACM*, **5**, 339—342.

Хедли (Hadley G.), 1962, Linear Programming, Addison-Wesley Reading, Massachusetts.

Хечтел, Брейтон, Густавсон (Hachtel G., Brayton R. and Gustavson F.), 1971, The sparse tableau approach to network analysis and design, *IEEE Trans. Circuit Theory* **CT-18**, 101—113.

Хетчел, Густавсон, Брейтон и Грейпс (Hachtel G., Gustavson F., Brayton R. and Grapes T.), 1969, A sparse matrix approach to network analysis, Proc. Cornell Conf. Computerized Electron.

Хильдебранд (Hildebrand F. B), 1956, Introduction to Numerical Analysis, McGraw-Hill, New York.

Хименес (Jimenez A. J.), 1969, Computer handling of sparse matrices, Rep. No. TR 00.1873. IBM, Yorktown Heights, New York.

Хип (Heap B. R.), 1966, Random matrices and graphs, *Numer. Math.*, **8**, 114—122.

Хси, Гаузи (Hsieh H. Y. and Ghausi M. S.), 1971a, On sparse matrices and optimal pivoting algorithms, Tech. Rep. 400-213, Electrical Eng. Dept., New York. Univ., New York.

Хси, Гаузи (Hsieh H Y. and Ghausi M. S), 1971b, A probabilistic approach to optimal pivoting and prediction of fill-in for random sparse matrices, Tech. Rep. 400-214. Electrical Eng. Dept., New York Univ., New York.

Цолленкопф (Zollenkopf K.), 1971, Bi-Factorization: Basic computational algorithm and programming techniques, In «Large Sparse Sets of Linear Equations» (J. K. Reid, ed.), pp. 75—96, Academic Press, New York.

Чан (Chang A.), 1969, Application of sparse matrix methods in electric power system analysis, In «Sparse Matrix Proceedings» (R A. Willoughby, ed), Rep. No. RA1 (≠ 11707), pp. 113—121, IBM, Yorktown Heights, New York.

Чень (Chen W. K.), 1967, On directed graph solution of linear algebraic equations, *SIAM Rev*, **9**, 692—707.

Чень (Chen Y. T.), 1972, Permutation of irreducible sparse matrices to upper triangular form, *IMA J.*, **10**, 15—18.

Чень, Тьюарсон (Chen Y. T. and Tewarson R. P.), 1972a, On the fill-in when sparse vectors are orthonormalized, *Computing (Arch. Elektron. Rechnen)*, **9**, 53—56.

Чень, Тьюарсон (Chen Y. T. and Tewarson R. P.), 1972b, On the optimal choice of pivots for the Gaussian elimination, *Computing*, **9** (в печати).

Черчилл (Churchill M. E.), 1971, A sparse matrix procedure for power system analysis programs, In «Large Sparse Sets of Linear Equations» (J. K. Reid, ed.), pp. 127—138. Academic Press, New York.

Шварц (Schwarz H. R), 1968, Tridiagonalization of a symmetric band matrix, *Numer. Math.*, **12**, 231—241.

Шуберт (Schubert L. K.), 1970, Modification of a quasi-Newton method for non-linear equations with sparse Jacobian, *Math. Comput.*, **25**, 27—30.

Эдельман (Edelmann H), 1963, Ordered triangular factorization of matrices, Proc. Power Systems Comput. Conf. London.

Эдельман (Edelmann H.), 1968, Massnahmen zur Reduktion des Rechenaufwands bei der Berechnung grosser elektrischer Netze. *Elektron Rechenanlagen* **10**, 118—123.

Эйкиуз, Утку (Akyuz F. A. and Utku S), 1968, An automatic relabeling scheme for bandwidth minimization of stiffness matrices, *AIAA J.*, **6**, 728—730.

Эйрени, Смит, Шода (Arany I., Smyth W. F. and Szoda L.), 1971, An improved method for reducing the bandwidth of sparse, symmetric matrices, IFIP Conf., Ljubljana, Yugoslavia.

Эрисман (Erisman A. M), 1972, Sparse matrix approach to the frequency domain analysis of linear passive electrical networks, In «Sparse Matrices and Their Application» (D. J. Rose and R. A. Willoughby, eds.), pp. 31—40, Plenum Press, New York.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

ОГЛАВЛЕНИЕ

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., д. 2, издательство «Мир».

ИБ № 475

Р. Тьюарсон

РАЗРЕЖЕННЫЕ МАТРИЦЫ

Редактор А. Бряндинская
Художник К. Сиротов
Художественный редактор В. Шаповалов
Технический редактор Л. Чуркина

Сдано в набор 21/VI 1976 г. Подписано к печати 13/I 1977 г. Бумага тип. № 3
$84\times108^{1}/_{32}=3$ бум. л. Усл. печ. л. 10,08. Уч.-изд. л. 8,99. Изд. № 1/8811.
Цена 68 коп. Зак. 255.

Издательство «Мир»
Москва, 1-й Рижский пер., 2

Ордена Трудового Красного Знамени Ленинградская типография № 2
имени Евгении Соколовой Союзполиграфпрома
при Государственном комитете Совета министров СССР
по делам издательств, полиграфии и книжной торговли.
198052, Ленинград, Л-52, Измайловский проспект, 29