СМБ

В. В. ВОЕВОДИН
Ю. А. КУЗНЕЦОВ

МАТРИЦЫ И ВЫЧИСЛЕНИЯ

22.19
В 63
УДК 519.6

Матрицы и вычисления. Воеводин В. В., Кузнецов Ю. А.— М.: Наука. Главная редакция физико-математической литературы, 1984.— 320 с.

Книга представляет собой справочное пособие по линейной алгебре. Это пособие охватывает как основные теоретические вопросы линейной алгебры, так и ее численные методы. Описание ведется с учетом особенностей реализации методов на ЭВМ. Отличительной чертой данного справочного пособия является отсутствие доказательств и теоретических обоснований приводимых фактов и методов.

Для студентов, аспирантов и научных сотрудников, деятельность которых связана с решением задач алгебры на ЭВМ.

В $\frac{1702070000-038}{053(02)-84}$ 64-83

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Линейная алгебра со своим математическим аппаратом и численными методами исключительно активно проникает в самые различные области, в особенности связанные с вычислениями. Построение процесса решения сложной проблемы на основе сведения к последовательности решений задач линейной алгебры является в настоящее время одним из важнейших направлений развития вычислительной математики.

Как правило, в прикладных задачах линейная алгебра играет вспомогательную роль. С точки зрения специалиста, решающего основную проблему, единственное, что требуется от линейной алгебры,— это предложить подходящий способ решения его частной задачи. Довольно часто такой способ можно выбрать из уже имеющегося обширного багажа теоретических и практических знаний в области линейной алгебры. Однако не менее часто это сделать не удается. Оказывается, что существующие знания не учитывают специфику основной проблемы, и поэтому невозможно решить задачу даже с использованием ЭВМ с нужной степенью эффективности из-за недостаточной скорости, точности или нехватки ресурсов вычислительной техники. После этого начинается утомительный поиск необходимых фактов с целью конструирования подходящего численного метода.

Настоящая книга предназначена для оказания помощи читателю в осуществлении такого поиска. Она представляет собой монографию справочного характера. Главная ее отличительная черта заключается в полном отсутствии каких-либо доказательств. Изложены только факты, определения и вычислительные схемы. В необходимых случаях приведены краткие пояснения. Материал систематизирован и описан таким образом, что каждый новый факт является, в основном, следствием предыдущих. Поэтому данную книгу можно рассматривать и как конспективное изложение основ теории матриц и численных методов. Порядок расположения определений и фактов в целом соответствует порядку их изложения в курсах линейной алгебры и численных методов, читаемых в высших учебных заведениях.

Весь материал книги разбит на две главы. Первая глава посвящена основам теории, вторая — численным методам. Содержание отдельных параграфов приведено в оглавлении после их названия. Для указания расположения утверждений принята двойная нумерация: левые цифры означают номер параграфа, правые — порядковый номер утверждения в данном параграфе. Все необходимые пояснения выделены петитом и не отмечены нумерацией. В случае необходимости ссылок на факты из текста пояснений они даются по ближайшему утверждению с номером.

Обилие теоретического и вычислительного материала в области линейной алгебры не позволило дать достаточно полное изложение всех ее разделов в одной книге. Поэтому авторы ограничились описанием лишь тех из них, знание которых особенно необходимо широкому кругу читателей. В случае имеющейся неоднозначности в определениях, терминологии и классификации, как правило, излагается наиболее распространенная точка зрения.

В этой книге §§ 1—17, 20—33 написаны В. В. Воеводиным, §§ 18, 19, 34—40 — Ю. А. Кузнецовым.

В. В. Воеводин, Ю. А. Кузнецов

ГЛАВА 1

ОСНОВЫ ТЕОРИИ

§ 1. Множества, элементы, операции

В этой книге мы будем рассматривать те или иные совокупности объектов, объединенных некоторым общим признаком. Число различных видов объектов будет невелико. В основном это матрицы и векторы. Однако большое число признаков, связывающих объекты, порождает огромное разнообразие их совокупностей.

Совокупность объектов, объединенных общим признаком, принято называть *множеством*, а сами объекты — *элементами* множества. Как правило, мы будем обозначать множества прописными латинскими буквами: A, B, ..., а их элементы — малыми: a, b, ... Мы будем писать $x \in A$, если элемент x принадлежит множеству A, и $x \notin A$, если элемент x не принадлежит множеству A. Иногда для удобства мы будем вводить в рассмотрение так называемое *пустое* множество (т. е. множество, которое не содержит ни одного элемента) и обозначать его $\varnothing$.

Среди всевозможных множеств особый интерес вызывают те, над элементами которых допускается выполнение некоторых операций. Пусть задано некоторое множество A, содержащее хотя бы один элемент.

1.1. Будем говорить, что в множестве A определена *алгебраическая операция*, если указан закон, по которому любой паре элементов a, b, взятых из этого множества в определенном порядке, однозначным образом ставится в соответствие некоторый третий элемент c, также принадлежащий этому множеству.

Эта операция может быть названа *сложением*, и тогда c будет называться *суммой* элементов a и b и обозначаться символом $c = a + b$; эта операция может быть названа *умножением*, и тогда c будет называться *произведением* элементов a и b и обозначаться символом $c = ab$.

Вообще терминология и символика для операции, определенной в множестве A, не будет играть в дальнейшем какой-либо существенной роли. Как правило, мы будем пользоваться символикой суммы и произведения независимо от того, каким образом определена операция в действительности. Если же нам потребуется подчеркнуть некоторые общие свойства алгебраической операции, то будем обозначать операцию символом $*$.

1.2. Алгебраическая операция называется *коммутативной*, если результат ее применения не зависит от порядка выбора элементов, т. е. для любых двух элементов a и b из заданного множества имеет место равенство $a * b = b * a$.

1.3. Алгебраическая операция называется *ассоциативной*, если для любых трех элементов a, b, c исходного множества $a * (b * c) = (a * b) * c$.

Ассоциативность операции позволяет говорить об однозначно определенном результате применения алгебраической операции к трем элементам a, b, c, понимая под ним любое из выражений $a * (b * c)$. $(a * b) * c$, и писать $a * b * c$ без скобок.

1.4. Результат вычисления выражения $a_1 * a_2 * \ldots * a_n$ для неассоциативной операции в общем случае при $n > 2$ зависит от расстановки скобок.

1.5. Результат вычисления выражения $a_1 * a_2 * \ldots * a_n$ для ассоциативной операции не зависит от расстановки скобок при всех n.

1.6. Результат вычисления выражения $a_1 * a_2 * \ldots * a_n$ для ассоциативной и коммутативной операции при всех n не зависит ни от расстановки скобок, ни от порядка элементов.

При обсуждении свойств алгебраической операции мы неявно предполагали возможность проверки любых двух элементов множества на их совпадение или несовпадение между собой. Мы нигде не предполагали, что совпадающие элементы действительно представляют собой один элемент, а не являются различными объектами. По существу же мы лишь использовали то, что некоторая группа элементов, которые мы называли равными, в определенных ситуациях проявляет себя одинаково.

В самых различных вопросах мы будем сталкиваться с необходимостью разбиения того или иного множества на группы элементов, объединенных по некоторому признаку. Если при этом ни один элемент не принадлежит двум различным группам, то будем говорить о разбиении множества на непересекающиеся группы, или на *классы*.

Пусть задан некоторый признак. Будем считать, что в отношении любой пары элементов a, b из множества A можно сказать, что либо элемент a связан с элементом b данным признаком, либо не связан с ним. Если элемент a связан с b, то будем писать $a \sim b$ и говорить, что *a эквивалентен b*.

1.7. Признак называется *отношением эквивалентности*, если он удовлетворяет следующим условиям:

— *рефлексивность*: $a \sim a$ для всех $a \in A$;

— *симметричность*: если $a \sim b$, то $b \sim a$;

— *транзитивность*: если $a \sim b$ и $b \sim c$, то $a \sim c$.

1.8. Множество можно разбить на классы по любому отношению эквивалентности.

Любые два элемента с точки зрения рассматриваемого признака могут быть либо эквивалентными, либо не эквивалентными. Ничто не изменится, если эквивалентные элементы мы назовем *равными* (в отношении данного признака), а не эквивалентные — *не равными* (в отношении того же признака). Введение *отношения равенства* позволяет разбить все множество на классы элементов, которые по тем или иным причинам мы решили считать равными. Это означает, что различие между элементами, входящими в один класс, не имеет для нас никакого значения. Следовательно, во всех ситуациях, которые мы будем в дальнейшем рассматривать, элементы, названые равными, должны проявлять себя одинаково.

Вернемся снова к изучению операций. Пусть в множестве A определена алгебраическая операция. Предположим, что она обладает тем свойством, что уравнения

$$a * x = b, \quad y * a = b$$

имеют, и притом единственные, решения при любых a, b. Тогда каждой упорядоченной паре элементов a, b из A мы можем поставить в соответствие однозначно определенные элементы x, y из A, т. е. ввести две алгебраические операции.

1.9. Описанные операции называются *правой* и *левой обратными* операциями по отношению к основной операции.

1.10. Если существует правая и левая обратные операции, то говорят, что основная операция имеет *обратную* операцию.

1.11. Не каждая алгебраическая операция, даже коммутативная и ассоциативная, имеет правую и левую обратные операции.

1.12. Если алгебраическая операция коммутативна и обратная операция для нее существует, то правая и левая обратные операции совпадают.

1.13. Из ассоциативности или коммутативности алгебраической операции не обязательно следует ассоциативность или коммутативность обратной операции, даже если обратная операция существует.

1.14. Если правая и левая обратные операции совпадают, то исходная алгебраическая операция коммутативна.

1.15. Если алгебраическая операция имеет обратную, то правая и левая обратные операции также имеют обратные.

1.16. Операция, обратная к левой или правой обратной, не обязательно совпадает с исходной алгебраической операцией.

Множества с одной алгебраической операцией в некотором смысле являются самыми простыми, и поэтому естественно начать наши исследования именно с таких множеств. Мы будем считать свойства операции аксиомами и затем выводить из них следствия. Это позволит в дальнейшем сразу применить результаты исследований ко всем множествам, в которых операции имеют аналогичные свойства, независимо от конкретных особенностей.

1.17. *Группой* называется множество G с одной алгебраической операцией, ассоциативной (хотя не обязательно коммутативной), причем для этой операции должна существовать обратная операция. Группа, состоящая из конечного числа элементов, называется *конечной* группой.

Заметим, что обратную операцию нельзя считать второй независимой операцией в группе, так как она определяется через основную. Назовем, как это принято в теории групп, операцию, заданную в G, *умножением* и условимся употреблять соответствующую символику.

1.18. Во всякой группе G существует, и притом единственный, элемент e, удовлетворяющий равенствам $ae = ea = a$ для всех a из G. Этот элемент называется *единицей* группы G.

1.19. Во всякой группе G любой элемент a обладает единственным *обратным* элементом a^{-1}, для которого $aa^{-1} = a^{-1}a = e$.

1.20. Имеет место соотношение $e^{-1} = e$.

1.21. Для любых элементов $a_1, a_2, \ldots, a_n$ из группы G имеет место соотношение $(a_1 a_2 \ldots a_n)^{-1} = a_n^{-1} a_{n-1}^{-1} \ldots a_1^{-1}$.

Проверка того факта, является ли группой множество с одной ассоциативной операцией, облегчается тем, что в определении группы требование о выполнимости обратной операции можно заменить предположением о существовании единицы и обратных элементов, причем лишь с одной стороны (например, правой) и без предположения об их единственности.

1.22. Множество G с одной ассоциативной операцией будет группой, если в G существует хотя бы один элемент e, обладающий свойством $ae = a$ для всех a из G, и по отношению к нему всякий элемент a из G обладает хотя бы одним правым обратным элементом a^{-1}, т. е. $aa^{-1} = e$.

1.23. Группа называется *коммутативной* или *абелевой*, если групповая операция коммутативна.

В этом случае операцию, как правило, называют *сложением* и вместо символа произведения ab пишут символ суммы $a + b$. Единицу абелевой группы называют *нулевым* элементом и обозначают символом 0. Обратную операцию называют *вычитанием*, а обратный элемент — *противоположным*. Обозначают его символом $-a$. Мы будем считать, что по определению символ разности $a - b$ означает сумму $a + (-b)$.

Если все же по каким-либо причинам операцию в коммутативной группе мы будем называть *умножением*, то обратную операцию будем считать *делением*. Равные в этом случае произведения $a^{-1}b$ и ba^{-1} будем обозначать через b/a и называть *частным* от деления b на a.

Рассмотрим множество K, в котором введены две операции. Назовем одну из них сложением, а другую — умножением и будем использовать соответствующую символику. Будем предполагать, что обе операции связаны *законом дистрибутивности*, т. е. для любых трех элементов a, b, c из K имеют место соотношения

$$(a + b)c = ac + bc, \quad a(b + c) = ab + ac.$$

1.24. Множество K называется *кольцом*, если в нем определены две операции — сложение и умножение, обе ассоциативные, а также связанные законом дистрибутивности, причем сложение коммутативно и обладает обратной операцией.

1.25. Кольцо называется *коммутативным*, если умножение коммутативно, и некоммутативным — в противном случае.

1.26. Любое кольцо является абелевой группой по сложению.

1.27. В любом кольце существует единственный нулевой элемент 0. При этом для всякого элемента a из кольца имеют место равенства

$$0 + a = a + 0 = a, \quad 0 \cdot a = a \cdot 0 = 0.$$

1.28. Ненулевые элементы кольца, произведение которых есть нулевой элемент, называются *делителями нуля*.

1.29. Коммутативное кольцо, в котором есть единичный элемент и каждый ненулевой элемент имеет обратный, называется *полем*.

1.30. Поле не имеет делителей нуля.

Используя запись частного a/b в виде произведения ab^{-1}, легко показать, что во всяком поле сохраняются все обычные правила обращения с дробями с точки зрения операций сложения, вычитания, умножения и деления, а именно:

$$\frac{a}{b} \pm \frac{c}{d} = \frac{ad \pm bc}{bd}, \quad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}, \quad \frac{-a}{b} = -\frac{a}{b}.$$

Кроме того, $a/b = c/d$ тогда и только тогда, когда $ad = bc$, если, конечно, $b \neq 0$ и $d \neq 0$. Поэтому все поля с точки зрения обычных правил обращения с дробями неотличимы от множества чисел. По этой причине элементы поля часто называются *числами*.

Рассмотрим теперь множество K и поле P произвольной природы. Предположим, что для всех элементов из K определены операции сложения и умножения на числа из P. Будем называть элементы из K *векторами*, независимо от их конкретной природы.

1.31. Множество K называется *линейным* или *векторным пространством* над полем P, если для всех векторов из K определены операции сложения и умножения на числа из P, причем выполнены следующие аксиомы:

A. Каждой паре векторов x, y отвечает вектор $x + y$, называемый *суммой* x и y, причем:

— сложение коммутативно: $x + y = y + x$;

— сложение ассоциативно: $x + (y + z) = (x + y) + z$;

— существует единственный нулевой вектор 0 такой, что $x + 0 = x$ для любого вектора x;

— для каждого вектора x существует единственный противоположный вектор $-x$ такой, что $x + (-x) = 0$.

B. Каждой паре α, x, где α — число, а x — вектор, отвечает вектор αx, называемый *произведением* α и x, причем:

— умножение на число ассоциативно: $\alpha(\beta x) = (\alpha\beta)x$;

— $1 \cdot x = x$ для любого вектора x.

C. Операции сложения и умножения связаны между собой следующими соотношениями:

— умножение на число дистрибутивно относительно сложения векторов: $\alpha(x + y) = \alpha x + \alpha y$;

— умножение на вектор дистрибутивно относительно сложения чисел: $(\alpha + \beta)x = \alpha x + \beta x$.

Перечисленные аксиомы не претендуют на логическую независимость. Свойства A описывают множество векторов с точки зрения операции сложения и говорят о том, что оно по отношению к этой операции является абелевой группой. Свойства B описывают множество векторов с точки зрения операции умножения вектора на число. Свойства C описывают связь двух операций между собой.

1.32. В любом линейном пространстве для каждого вектора x имеет место равенство $0 \cdot x = 0$, где в правой части 0 означает нулевой вектор, а в левой 0 — число нуль.

1.33. В любом линейном пространстве для любого вектора x справедливо соотношение $-x = (-1)x$.

1.34. В любом линейном пространстве имеет место равенство $\alpha \cdot 0 = 0$ для любого числа α.

С точки зрения операций умножения, сложения и вычитания формально имеют место все правила эквивалентных преобразований алгебраических выражений в любом линейном пространстве. В дальнейшем эти правила мы уже не будем оговаривать особо.

1.35. Множество L линейного пространства K называется его *линейным подпространством*, если при тех же операциях, что и в пространстве K, оно само является линейным пространством.

1.36. Множество, состоящее из одного нулевого вектора, является линейным подпространством. Это подпространство называется *нулевым*.

Наименьшим подпространством является нулевое, наибольшим — само исходное линейное пространство. Эти два пространства называются *тривиальными*, остальные — *нетривиальными*.

1.37. Для того чтобы множество векторов линейного пространства было его подпространством, необходимо и достаточно, чтобы это множество вместе с каждой парой элементов x, y содержало и все их линейные комбинации $\alpha x + \beta y$.

Мы будем называть линейное пространство *рациональным*, *вещественным* или *комплексным* в зависимости от того, является ли поле P полем рациональных, вещественных или комплексных чисел, и обозначать соответственно буквами **P**, **R** и **C**. Тот факт, что в названии и в обозначении отсутствует какая-либо ссылка на элементы самого пространства, имеет глубокий смысл, но об этом мы будем говорить позднее.

Важнейшим моментом в создании основ математического анализа является введение понятия функции. Это понятие можно распространить на произвольные множества. Правило, по которому каждому элементу x некоторого непустого множества X ставится в соответствие единственный элемент y некоторого множества Y, называется *оператором*. Результат применения оператора A к элементу x обозначают $y = A(x)$ или $y = Ax$ и говорят, что оператор A *действует* из X в Y, или *отображает* X в Y.

Множество X называется *областью определения* оператора A. Элемент y называется *образом* элемента x, а x — *прообразом* элемента y. Совокупность всех образов называется *областью значений* оператора A. В том случае, когда каждый элемент $y \in Y$ имеет, и притом только один, прообраз, оператор называется *взаимно однозначным*. Оператор называется также *отображением*, *преобразованием* или *операцией*.

В дальнейшем мы будем рассматривать лишь так называемые *линейные* операторы. Их отличительные особенности заключаются в следующем. Во-первых, областью определения линейного оператора всегда является некоторое линейное пространство или его подпространство. Во-вторых, свойства линейного оператора тесно связаны с операциями над векторами линейного пространства. В общей теории операторов линейные операторы играют столь же важную роль, как и линейные функции в математическом анализе.

1.38. Пусть заданы линейные пространства X, Y над одним и тем же полем P. Рассмотрим оператор A, областью определения которого является пространство X, а областью значений — некоторое множество из Y. Оператор A называется *линейным*, если

$$A(\alpha u+\beta v)=\alpha Au+\beta Av$$

для любых векторов $u,\ v\in X$ и любых чисел $\alpha,\ \beta\in P$.

1.39. Область значений любого линейного оператора есть подпространство пространства Y.

Оператор, который каждому вектору $x\in X$ ставит в соответствие нулевой вектор из Y, является, очевидно, линейным. Он называется *нулевым* оператором и обозначается символом 0. Оператор B, построенный по предписанию $Bx=-Ax$, где A есть линейный оператор из X в Y, также является линейным оператором из X в Y. Он называется оператором, *противоположным* оператору A. Зафиксируем произвольное число α и каждому вектору $x\in X$ поставим в соответствие вектор $\alpha x\in X$. Построенный таким способом оператор будет линейным оператором, действующим из X в X. Он называется *скалярным*. При $\alpha=0$ мы получаем нулевой оператор, при $\alpha=1$ — так называемый *тождественный* оператор. Тождественный оператор обозначается символом E или I. По определению всегда $Ex=x$ ($Ix=x$).

1.40. Два оператора A, B, действующие из X в Y, называются *равными*, если $Ax=Bx$ для всех $x\in X$. Равенство операторов обозначается $A=B$.

1.41. Оператор C называется *суммой* операторов A, B, действующих из X в Y, если $Cx=Ax+Bx$ для всех $x\in X$. Сумму операторов обозначают $C=A+B$.

1.42. Оператор C называется *произведением* оператора A, действующего из X в Y, на число λ из поля P, если выполняется равенство $Cx=\lambda\cdot Ax$ для всех $x\in X$. Это произведение обозначают $C=\lambda A$.

1.43. Для линейных операторов A, B операторы $A+B$ и λA являются линейными.

1.44. Операция сложения линейных операторов является алгебраической, ассоциативной и коммутативной.

1.45. Множество всех линейных операторов, действующих из линейного пространства X в линейное пространство Y, есть линейное пространство.

Рассмотрим три линейных пространства X, Y, Z над одним и тем же полем P. Пусть A — оператор, действующий из X в Y, B — оператор, действующий из Y в Z.

1.46. Оператор C, действующий из X в Z, называется *произведением* оператора B на оператор A, если $Cx=B(Ax)$ для всех $x\in X$. Произведение операторов B и A обозначают $C=BA$.

1.47. Произведение линейных операторов есть снова линейный оператор.

1.48. Произведение операторов обладает следующими свойствами:

$$(AB)C=A(BC),\qquad \lambda(BA)=(\lambda B)A=B(\lambda A),$$

$$(A+B)C=AC+BC,\quad A(B+C)=AB+AC.$$

1.49. Множество всех линейных операторов, действующих в линейном пространстве X, вообще говоря, есть некоммутативное кольцо.

§ 2. Системы векторов

Рассмотрим множество всех линейных пространств, заданных над одним и тем же полем P. Естественно спросить, чем же похожи и чем различаются между собой все эти пространства.

Каждое линейное пространство в своем описании содержит две существенно различные части. Во-первых, *линейное пространство* есть совокупность конкретных объектов, называемых *векторами*. Во-вторых, над этими конкретными объектами определены операции сложения и умножения на число. Поэтому можно интересоваться либо природой векторов и их свойствами, либо свойствами указанных операций независимо от природы элементов.

Во всех практически интересных случаях построение и исследование линейных пространств осуществляется в два этапа: сначала, учитывая природу векторов, определяют операции сложения и умножения на число, а затем на основе свойств этих операций изучают сами пространства. Поэтому два пространства, устроенные одинаково по отношению к операциям сложения и умножения на число, можно считать обладающими одинаковыми свойствами.

2.1. Два линейных пространства, заданных над одним и тем же полем, называются *изоморфными*, если между их векторами можно установить такое взаимно однозначное соответствие, при котором сумме любых двух векторов первого пространства будет отвечать сумма соответствующих векторов второго пространства, а произведению какого-либо числа на вектор первого пространства будет отвечать произведение того же числа на соответствующий вектор второго пространства.

2.2. С точки зрения всех следствий, вытекающих из аксиом линейного пространства, изоморфные пространства неотличимы друг от друга.

2.3. Отношение изоморфизма есть отношение эквивалентности.

Согласно утверждению 1.8 множество всех линейных пространств, заданных над одним и тем же полем, может быть разбито на классы. При этом в каждый класс будут входить изоморфные пространства, и только они. Так как изоморфные пространства неотличимы друг от друга, то вместо изучения всех линейных пространств, заданных над одним полем, можно изучать лишь представителей из каждого класса. К описанию этих представителей мы и переходим.

2.4. *Вектором* размерности n (или *n-мерным вектором*) называется упорядоченная совокупность из n чисел поля P. Если x — вектор, определяемый числами $\alpha_1, \alpha_2, \ldots, \alpha_n$, то будем писать

$$x = (\alpha_1, \alpha_2, \ldots, \alpha_n).$$

2.5. Числа $\alpha_1, \ldots, \alpha_n$ называются *координатами* вектора x.

2.6. Если векторы x, y размерности n заданы своими координатами: $x = (\alpha_1, \ldots, \alpha_n)$, $y = (\beta_1, \ldots, \beta_n)$, то *суммой* этих векторов называется вектор

$$x + y = (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_n + \beta_n).$$

2.7. Если вектор x задан своими координатами 2.4, то *произведением вектора на число* λ из поля P называется вектор

$$\lambda x = (\lambda\alpha_1, \lambda\alpha_2, \ldots, \lambda\alpha_n).$$

2.8. *Нулевым* вектором называется вектор, все координаты которого равны нулю, т. е. $0 = (0, \ldots, 0)$.

2.9. Вектором, *противоположным* вектору 2.4, называется вектор

$$-x = (-\alpha_1, \ldots, -\alpha_n).$$

2.10. Множество векторов 2.4 с операциями 2.6, 2.7 есть линейное пространство. Это пространство называется *арифметическим* пространством размерности n и обозначается P_n. Размерность пространства P_n обозначается $\dim P_n$.

2.11. Линейное пространство, изоморфное пространству P_n, называется *конечномерным*.

2.12. Линейное пространство, не изоморфное никакому пространству P_n, называется *бесконечномерным*.

Таким образом, все линейные пространства, заданные над одним и тем же полем P, разбиваются на два класса — конечномерные и бесконечномерные. Типичным примером бесконечномерного пространства является множество всех вещественных функций, заданных на одном и том же отрезке, с естественными операциями сложения функций и умножения функции на вещественное число. Различные пространства получаются при наложении на допустимые функции различных условий гладкости. Вообще говоря, бесконечномерные пространства в большинстве случаев оказываются связанными с пространствами типа P_∞, которые определяются аналогично 2.4—2.7, но с бесконечным числом координат каждого вектора. Однако мы не будем изучать такие пространства.

Мы будем иметь дело в дальнейшем только с вещественными и комплексными арифметическими пространствами $\mathbf{R}_n$, $\mathbf{C}_n$ и с изоморфными им пространствами. При этом под размерностью произвольного конечномерного пространства будем понимать размерность изоморфного ему арифметического пространства. Если при формулировке какого-либо определения или утверждения не подчеркивается, о каком пространстве идет речь, то это означает, как правило, что рассматриваются сразу и вещественное и комплексное пространства, причем слово «арифметическое» чаще всего будет опускаться.

2.13. Вектор e называется *единичным*, если одна его координата равна единице, а остальные равны нулю.

2.14. В n-мерном арифметическом пространстве совокупность единичных векторов

$$p_1 = (1, 0, 0, \ldots, 0), \quad p_2 = (0, 1, 0, \ldots, 0), \quad \ldots, \quad p_n = (0, 0, 0, \ldots, 1)$$

называется *естественным базисом* пространства.

2.15. Для любого вектора $x=(\alpha_1, \alpha_2, \ldots, \alpha_n)$ справедливо следующее представление:

$$x=\alpha_1 p_1+\alpha_2 p_2+\ldots+\alpha_n p_n.$$

2.16. Пусть заданы система векторов $e_1, e_2, \ldots, e_n$ и вектор x. Если при некоторых числах $\alpha_1, \alpha_2, \ldots, \alpha_n$ выполняется равенство

$$x=\alpha_1 e_1+\alpha_2 e_2+\ldots+\alpha_n e_n,$$

то говорят, что вектор x линейно выражается через векторы $e_1, \ldots, e_n$, или представлен в виде *разложения* по этим векторам.

2.17. Правая часть равенства 2.16 называется *линейной комбинацией* векторов $e_1, \ldots, e_n$, числа $\alpha_1, \ldots, \alpha_n$ — *коэффициентами* линейной комбинации.

2.18. Пусть задана система векторов $e_1, \ldots, e_n$. Множество всех линейных комбинаций векторов $e_1, \ldots, e_n$ называется *линейной оболочкой* этой системы векторов и обозначается $L(e_1, \ldots, e_n)$.

2.19. Линейная оболочка любой системы векторов из любого линейного пространства является линейным подпространством.

2.20. Линейная оболочка векторов $e_1, \ldots, e_n$ есть «наименьшее» линейное подпространство, содержащее эти векторы, т. е. любое линейное подпространство, содержащее какую-либо систему векторов, содержит и ее линейную оболочку.

Интерес к линейным оболочкам определяется несколькими обстоятельствами. Во-первых, любая линейная оболочка устроена очень просто — это совокупность всех линейных комбинаций векторов заданной системы. Во-вторых, любая линейная оболочка является линейным пространством. И наконец, согласно представлению 2.15 любое арифметическое пространство есть линейная оболочка векторов своего естественного базиса. В общем случае каждое линейное пространство содержит в себе бесчисленное множество других линейных пространств — линейных оболочек всевозможных своих систем векторов. В основе исследования связи линейных пространств, линейных оболочек и порождающих их систем векторов лежит одно из самых фундаментальных понятий — понятие линейной зависимости системы векторов.

2.21. Система векторов $e_1, e_2, \ldots, e_n$ называется *линейно зависимой*, если один из векторов e_i линейно выражается через остальные векторы системы или эта система состоит из одного нулевого вектора.

2.22. Система векторов, не являющаяся линейно зависимой, называется *линейно независимой.*

2.23. Система векторов $e_1, e_2, \ldots, e_n$ линейно независима тогда и только тогда, когда из равенства

$$\alpha_1 e_1+\alpha_2 e_2+\ldots+\alpha_n e_n=0$$

следует равенство нулю всех коэффициентов линейной комбинации.

2.24. Система, содержащая нулевой вектор, линейно зависима.

2.25. Система, состоящая из одного ненулевого вектора, линейно независима.

2.26. Если в системе векторов некоторая подсистема линейно зависима, то и вся система линейно зависима.

2.27. Система векторов линейно независима тогда и только тогда, когда линейно независима любая ее подсистема.

2.28. Пусть система векторов $e_1, \ldots, e_n$ линейно независима, а система $e_1, \ldots, e_n, x$ линейно зависима. Тогда вектор x линейно выражается через векторы $e_1, \ldots, e_n$.

2.29. Система векторов $e_1, e_2, \ldots, e_n$ линейно зависима тогда и только тогда, когда либо $e_1 = 0$, либо некоторый вектор e_k, $2 \leqslant k \leqslant n$, линейно выражается через предшествующие векторы.

2.30. Система векторов $e_1, \ldots, e_n$ линейно зависима тогда и только тогда, когда либо $e_n = 0$, либо некоторый вектор e_k, $1 \leqslant k \leqslant n-1$, линейно выражается через последующие векторы.

2.31. Если какой-либо вектор линейного пространства единственным образом представляется в виде линейной комбинации векторов $e_1, \ldots, e_n$, то эта система векторов линейно независима.

2.32. Если система векторов линейно независима, то любой вектор ее линейной оболочки единственным образом представляется в виде линейной комбинации векторов системы.

2.33. Если система векторов линейно зависима, то для любого вектора ее линейной оболочки существует бесконечно много разложений по векторам системы.

2.34. Если каждый из векторов линейно независимой системы $e_1, \ldots, e_n$ линейно выражается через векторы $y_1, \ldots, y_m$, то $n \leqslant m$.

Линейная оболочка может порождаться различными системами векторов — как зависимыми, так и независимыми. Естественно предположить, что все эти системы должны обладать некоторыми общими свойствами.

2.35. Две системы векторов $e_1, \ldots, e_n$ и $f_1, \ldots, f_m$ называются *эквивалентными*, если каждый из векторов одной системы линейно выражается через векторы другой системы.

2.36. Эквивалентность систем векторов есть отношение эквивалентности (см. 1.7).

2.37. Две системы векторов эквивалентны тогда и только тогда, когда их линейные оболочки совпадают.

2.38. Эквивалентные линейно независимые системы векторов содержат одинаковое число векторов.

2.39. Любые $n+1$ векторов из линейной оболочки системы, содержащей n векторов, линейно зависимы.

2.40. Если не все векторы системы нулевые, то в этой системе можно выбрать эквивалентную ей линейно независимую подсистему. Всякая такая подсистема называется *базой* данной системы.

2.41. Все базы одной и той же системы векторов состоят из одинакового числа векторов. Это число называется *рангом* системы. Если все векторы системы нулевые, то ее ранг считается равным нулю.

2.42. Пусть ранг системы векторов равен r. Если некоторая ее подсистема содержит более r векторов, то она линейно зависима.

2.43. Любая линейно независимая подсистема, содержащая r векторов, является базой системы.

2.44. Любую линейно независимую подсистему данной системы векторов можно достроить до базы этой системы.

2.45. Если векторы одной системы линейно выражаются через векторы другой системы, то ранг первой системы не больше ранга второй.

2.46. Если две системы векторов имеют одинаковый ранг и векторы одной системы линейно выражаются через векторы другой, то эти системы эквивалентны.

Эквивалентные системы векторов позволяют дать полное описание строения линейных пространств как линейных оболочек некоторых систем векторов и указать свойства этих систем.

2.47. Система из векторов естественного базиса арифметического пространства линейно независима.

2.48. В арифметическом пространстве размерности n не может существовать линейно независимая система, содержащая более n векторов.

2.49. В арифметическом пространстве размерности n любая линейно независимая система, содержащая n векторов, эквивалентна естественному базису.

2.50. Любое арифметическое пространство размерности n является линейной оболочкой любой своей линейно независимой системы из n векторов.

2.51. Никакое арифметическое пространство размерности n не может являться линейной оболочкой системы, содержащей менее n векторов.

2.52. Линейно независимая система из n векторов арифметического пространства размерности n называется *базисом* пространства.

Базис имеет огромное значение при изучении линейных пространств и постоянно используется в самых различных исследованиях. Именно существование базисов в произвольных конечномерных пространствах позволяет создавать конструктивные методы изучения как самих пространств, так и функций в этих пространствах. Базис позволяет легко описать строение любого линейного пространства, причем простота описания в значительной мере определяется конкретным выбором базиса. Конечно, естественный базис является самым простым и удобным, однако им можно воспользоваться очень редко, да и то только в арифметическом пространстве. Например, любая линейная оболочка является линейным пространством. Но среди векторов, образующих ее базисы и рассматриваемых как векторы основного арифметического пространства, может просто не быть ни одного единичного вектора.

В произвольном конечномерном линейном пространстве, вообще говоря, все базисы равноправны. Их неравноправие начинается лишь тогда, когда мы рассматриваем базисы по отношению к каким-либо другим объектам линейного пространства, например по отношению к какому-нибудь фиксированному базису или системе векторов или по отношению к каким-либо функциям, введенным в линейном пространстве, и т. п. Как уже отмечалось, любой вектор может быть представлен своим разложением по любому базису.

2.53. Если имеет место равенство 2.16, где $e_1, \ldots, e_n$ являются векторами базиса, то оно называется *разложением вектора x по базису*, а числа $\alpha_1, \ldots, \alpha_n$ называются *координатами* вектора относительно этого базиса.

2.54. Разложение любого вектора по любому базису единственно.

2.55. Все координаты нулевого вектора относительно любого базиса равны нулю.

2.56. При сложении двух векторов их координаты относительно любого базиса складываются.

2.57. При умножении вектора на число его координаты относительно любого базиса умножаются на это число.

Аналогичные утверждения можно было бы продолжить, однако в этом нет никакой необходимости. Предположим, что задано некоторое конечномерное линейное пространство K размерности n. В частности, это может быть и арифметическое пространство. Выберем в K какой-либо базис $e_1, \ldots, e_n$ и каждому вектору x из K поставим в соответствие вектор x_e, составленный из координат вектора x в выбранном базисе. Это соответствие не только отображает пространство K взаимно однозначно на арифметическое пространство, но и отображает изоморфно. Поэтому при таком отображении любая линейно зависимая система переходит в линейно зависимую, ранг системы не изменяется, базис переходит в базис и т. д.

Установленная связь между произвольными конечномерными и арифметическими пространствами позволяет ограничиться исследованиями, связанными только с арифметическими пространствами. Но чтобы не потерять общности исследований, мы время от времени будем обращаться к произвольным линейным пространствам.

§ 3. Матрицы и операторы

Как уже отмечалось ранее, множество всех линейных операторов, действующих из одного линейного пространства в другое, есть линейное пространство (см. 1.45). Исследование этого пространства с помощью базиса является эффективным и приводит к возникновению исключительно важного матричного аппарата.

3.1. Для любого базиса $e_1, \ldots, e_m$ m-мерного пространства X и любой системы векторов $f_1, \ldots, f_m$ n-мерного пространства Y существует такой линейный оператор A, что $f_k = Ae_k$ для всех k. Если $x = \xi_1 e_1 + \ldots + \xi_m e_m$, то

$$Ax = \sum_{k=1}^{m} \xi_k f_k.$$

3.2. Линейный оператор, действующий из пространства X в пространство Y, однозначно определяется совокупностью образов какого-либо базиса пространства X.

3.3. Линейное пространство операторов, действующих из m-мерного пространства X в n-мерное пространство Y, есть конечномерное пространство размерности mn.

Если в пространстве линейных операторов выбрать некоторый базис, то тем самым устанавливается изоморфное соответствие пространства операторов арифметическому пространству векторов. При этом операция произведения операторов порождает соответствующую операцию над векторами. Однако прямой подход в данном случае оказывается не самым лучшим. Значительно более интересные результаты можно получить, если базис в пространстве операторов выбрать связанным определенным образом с базисами тех пространств, в которых заданы сами операторы. К тому же оказывается, что вектор арифметического пространства, соответствующий оператору, удобнее записывать не в виде строки 2.4, а в виде некоторой прямоугольной таблицы.

3.4. *Матрицей* называется совокупность вещественных или комплексных чисел a_{ij}, расположенных в виде прямоугольной таблицы:

$$\begin{bmatrix} a_{11} & a_{12} & \cdot & a_{1m} \\ a_{21} & a_{22} & \cdot & a_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdot & a_{nm} \end{bmatrix}.$$

3.5. Числа a_{ij} называются *элементами* матрицы. Индексы i, j означают, что элемент a_{ij} расположен на пересечении i-й строки и j-го столбца матрицы. Если m или n равно единице, то соответствующий индекс может отсутствовать.

3.6. Если матрица имеет n строк и m столбцов, то она называется матрицей *размера* $n \times m$.

3.7. Матрица называется *квадратной* матрицей порядка n, если $n = m$. В общем случае матрица называется *прямоугольной*.

3.8. Прямоугольная матрица размера $n \times 1$, состоящая из одного столбца

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdot\ \cdot \\ \alpha_n \end{bmatrix},$$

называется *столбцевой* и обозначается $(\alpha_1, \alpha_2, \ldots, \alpha_n)$.

3.9. Прямоугольная матрица размера $1 \times m$, состоящая из одной строки $[\beta_1\, \beta_2 \ldots \beta_m]$, называется *строчной* и обозначается $(\beta_1, \beta_2, \ldots, \beta_m)$.

3.10. Столбцевая матрица называется также *вектор-столбцом*, строчная матрица называется *вектор-строкой*. Если не возникает каких-либо недоразумений, то оба вида матриц называются просто *векторами*.

Для обозначения матрицы используются различные символы, например: A, $A(n \times m)$, $A(a_{ij})$, $A = (a_{ij})$ и т. п., или же она указывается явно в виде таблицы, в зависимости от того, какие характеристики матрицы нужно отметить. В случае необходимости указать элемент, расположенный на пересечении i-й строки и j-го столбца матрицы, или k-й элемент вектора, заданные выражением $\{\cdot\}$, используются символы соответственно $\{\cdot\}_{ij}$ или $\{\cdot\}_k$. В некоторых случаях один или оба индекса ставятся вверху. По отношению к векторам, заданным своим разложением по какому-либо базису, символ $\{\cdot\}_k$ означает не что иное, как k-ю координату этого разложения.

3.11. Пусть в m-мерном пространстве X задан базис $e_1, \ldots, e_m$, в n-мерном пространстве Y — базис $q_1, \ldots, q_n$. Линейный оператор A, действующий из X в Y, однозначно определяется набором чисел $\{Ae_i\}_j$ для $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$.

3.12. При фиксированных базисах $e_1, \ldots, e_m$ и $q_1, \ldots, q_n$ в пространствах X, Y матрица размера $n \times m$ с элементами a_{ij} называется *матрицей оператора* A в выбранных базисах, если $a_{ij} = \{Ae_i\}_j$ для всех i, j.

3.13. При введенных выше обозначениях выполняются соотношения

$$Ae_j = \sum_{s=1}^{n} a_{sj} q_s.$$

3.14. Пусть $y = Ax$, где A есть линейный оператор, действующий из пространства X в пространство Y. Если векторы x, y заданы своими разложениями

$$x = \sum_{j=1}^{m} \xi_j e_j, \quad y = \sum_{i=1}^{n} \eta_i q_i$$

по базисам, то при введенных выше обозначениях выполняются соотношения

$$\begin{aligned} \eta_1 &= a_{11}\xi_1 + a_{12}\xi_2 + \ldots + a_{1m}\xi_m, \\ \eta_2 &= a_{21}\xi_1 + a_{22}\xi_2 + \ldots + a_{2m}\xi_m, \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \eta_n &= a_{n1}\xi_1 + a_{n2}\xi_2 + \ldots + a_{nm}\xi_m. \end{aligned}$$

Между линейными операторами, действующими из m-мерного пространства X в n-мерное пространство Y над общим полем P, и прямоугольными матрицами размера $n \times m$ с элементами из того же поля P установлено взаимно однозначное соответствие. По существу, матрица оператора есть вектор того арифметического пространства размерности mn, которому изоморфно пространство линейных операторов. Все различие заключается лишь в том, что этот вектор удобно записывать не строкой своих координат, а матрицей.

Базис пространства линейных операторов, с помощью которого устанавливается отображение данного пространства на арифметическое, устроен очень просто. Это есть совокупность из mn линейных операторов E_{ij}, у которых при выбранных базисах в пространствах X, Y все элементы матриц, за исключением одного, являются нулевыми. Ненулевой элемент находится в позиции (i, j) и равен единице. Координаты образа, прообраза и линейного оператора в своих базисах связаны между собой соотношениями 3.14.

Чтобы говорить об изоморфизме множества линейных операторов и матриц, необходимо определить для матриц понятие равенства и ввести соответствующие операции сложения, произведения и умножения на число.

3.15. Матрицы A и B называются *равными*, если они имеют одинаковые размеры и $\{A\}_{ij} = \{B\}_{ij}$ для всех i, j. Равенство матриц A и B обозначается $A = B$.

3.16. Отношение равенства матриц есть отношение эквивалентности.

3.17. *Суммой* матриц A и B размеров $m \times n$ называется матрица C размера $m \times n$, если $\{C\}_{ij} = \{A\}_{ij} + \{B\}_{ij}$ для всех i, j. Эта операция обозначается $C = A + B$.

3.18. *Произведением* матрицы A размера $m \times n$ на число α называется матрица C размера $m \times n$, если $\{C\}_{ij} = \alpha\{A\}_{ij}$ для всех i, j. Эта операция обозначается $C = \alpha A$.

3.19. *Разностью* матриц A и B размеров $m \times n$ называется матрица C размера $m \times n$, если $\{C\}_{ij} = \{A\}_{ij} - \{B\}_{ij}$ для всех i, j. Эта операция обозначается $C = A - B$.

3.20. *Произведением* матрицы A размера $m \times n$ и матрицы B размера $n \times p$ называется матрица C размера $m \times p$, если

$$\{C\}_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

для всех i, j. Эта операция обозначается $C = AB$.

3.21. Операция сложения матриц коммутативна и ассоциативна, т. е.

$$A + B = B + A, \quad (A + B) + C = A + (B + C).$$

3.22. Всегда выполняется соотношение $A - B = A + (-1)B$.

3.23. Для выполнимости операции произведения двух матриц необходимо и достаточно, чтобы число столбцов левого сомножителя равнялось числу строк правого сомножителя.

3.24. Операция произведения матриц не является в общем случае коммутативной.

3.25. Если для каких-либо матриц A и B выполняется равенство $AB = BA$, то матрицы называются *коммутирующими* или *перестановочными*.

3.26. Если операция произведения матриц выполнима, то она ассоциативна, т. е. $(AB)C = A(BC)$.

3.27. Для выполнимости операции произведения нескольких матриц, заданных в определенном порядке, необходимо и достаточно, чтобы число столбцов каждой матрицы равнялось числу строк соседней матрицы справа. В этом случае произведение матриц определяется однозначно и может быть вычислено при произвольном порядке расстановки скобок.

3.28. Если операция произведения матриц выполнима, то она дистрибутивна по отношению к операции сложения, т. е.

$$(A+B)C = AC + BC, \quad D(A+B) = DA + DB.$$

3.29. В терминах введенных операций соотношения 3.14 могут быть записаны следующим образом:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \dots \\ \eta_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \dots \\ \xi_m \end{bmatrix}.$$

Правая часть равенства есть произведение матрицы на вектор.

3.30. Множества линейных операторов и матриц с соответствующими операциями изоморфны.

3.31. Множество прямоугольных матриц одинаковых размеров $m \times n$ образует линейное пространство размерности mn.

3.32. Множество квадратных матриц одного порядка, вообще говоря, образует некоммутативное кольцо.

Важно подчеркнуть, что все операторные и соответствующие матричные равенства будут выглядеть совершенно одинаково, если символ Ax понимать как произведение оператора A на вектор x. Так как символика и свойства операций над матрицами и операторами в этом случае полностью совпадают, то любое преобразование операторного равенства приводит к аналогичному преобразованию матричного равенства. Поэтому с формальной точки зрения безразлично, иметь ли дело с матричными или операторными соотношениями. Использование матричного аппарата позволяет создавать конструктивные методы исследования, и в этом заключается основное его достоинство. Операторный аппарат оказывается более удобным, например, в тех случаях, когда нужно подчеркнуть общие свойства или когда по каким-либо причинам приходится отказываться от рассмотрения конкретных базисов.

В дальнейшем мы не будем делать различия между операторными и матричными соотношениями. Более того, все новые понятия и факты, имеющие место в отношении операторов (матриц), мы, как правило, без особой оговорки будем распространять и на матрицы (операторы).

3.33. Матрица A' размера $n \times m$ называется *транспонированной* по отношению к матрице A размера $m \times n$, если $\{A'\}_{ij} = \{A\}_{ji}$ для всех i, j.

3.34. Имеют место соотношения

$$(\alpha A)' = \alpha A', \quad (A+B)' = A' + B', \quad (AB)' = B'A', \quad (A')' = A.$$

3.35. Матрица $\bar{A}$ размера $m \times n$ называется *комплексно сопряженной* по отношению к матрице A размера $m \times n$, если $\{\bar{A}\}_{ij} = \overline{\{A\}_{ij}}$ для всех i, j. Здесь черта означает комплексное сопряжение.

3.36. Имеют место соотношения

$$(\overline{\alpha A}) = \bar{\alpha}\bar{A}, \quad (\overline{A+B}) = \bar{A} + \bar{B},$$

$$(\overline{AB}) = \bar{A}\cdot\bar{B}, \quad (\bar{\bar{A}}) = A.$$

3.37. Матрица A^* размера $n \times m$ называется *сопряженной* по отношению к матрице A размера $m \times n$, если $\{A^*\}_{ji} = \overline{\{A\}_{ij}}$ для всех i, j.

3.38. Имеют место соотношения

$$(\alpha A)^* = \bar{\alpha}A^*, \qquad (A+B)^* = A^* + B^*,$$

$$(AB)^* = B^*A^*, \qquad (A^*)^* = A, \quad (\bar{A})' = (\overline{A'}).$$

3.39. Совокупность элементов a_{ij} матрицы, для которых $i = j$, называется *главной диагональю*; соответствующие элементы a_{ij} называются *диагональными*. Все остальные элементы называются *внедиагональными*.

3.40. Сумма диагональных элементов матрицы A называется *следом* матрицы A и обозначается $\operatorname{tr} A$.

3.41. Имеют место соотношения

$$\operatorname{tr} A = \operatorname{tr} A' \quad \operatorname{tr} A^* = \overline{\operatorname{tr} A}, \quad \operatorname{tr}(\alpha A) = \alpha \operatorname{tr} A,$$

$$\operatorname{tr}(A+B) = \operatorname{tr} A + \operatorname{tr} B, \quad \operatorname{tr}(BA) = \operatorname{tr}(AB).$$

3.42. Все элементы матрицы A равны нулю тогда и только тогда, когда $\operatorname{tr}(AA^*) = 0$.

3.43. Матрица, все элементы которой равны нулю, называется *нулевой* и обозначается символом 0.

3.44. Для любой матрицы A и нулевых матриц соответствующих размеров справедливы равенства

$$A + 0 = 0 + A = A, \quad 0\cdot A = A\cdot 0 = 0.$$

3.45. Матрица A называется *диагональной*, если все ее внедиагональные элементы равны нулю. Диагональная матрица с элементами $a_{11}, \ldots, a_{mm}$ обозначается $\operatorname{diag}(a_{11}, a_{22}, \ldots, a_{mm})$.

3.46. Диагональная квадратная матрица с равными диагональными элементами называется *скалярной*.

3.47. Диагональная квадратная матрица с диагональными элементами, равными единице, называется матрицей *тождественного преобразования* или *единичной* и обозначается E или I.

3.48. При умножении прямоугольной матрицы A справа (слева) на диагональную матрицу $\operatorname{diag}(d_1, d_2, \ldots)$ все столбцы (строки) матрицы A умножаются как векторы на числа $d_1, d_2, \ldots$

3.49. Если квадратная матрица A перестановочна с квадратной диагональной матрицей D, имеющей попарно различные диагональные элементы, то матрица A — диагональная.

3.50. Если квадратная матрица A перестановочна со всеми квадратными матрицами того же порядка, то матрица A — скалярная.

3.51. Множество матриц, перестановочных с любой фиксированной матрицей A, является подпространством.

3.52. Множество матриц, перестановочных с любой фиксированной матрицей A, является кольцом.

3.53. Если матрицы A, B перестановочные, то имеют место соотношения

$$(A+B)^2 = A^2 + 2AB + B^2,$$
$$A^2 - B^2 = (A+B)(A-B),$$
$$(A+B)^n = A^n + C_n^1 A^{n-1}B + \ldots + C_n^{n-1}AB^{n-1} + B^n,$$
$$A^n - B^n = (A-B)(A^{n-1} + A^{n-2}B + \ldots + AB^{n-2} + B^{n-1}).$$

3.54. Матрицей *перестановок* называется квадратная матрица, у которой в каждой строке и в каждом столбце только один элемент отличен от нуля и равен единице.

3.55. При умножении прямоугольной матрицы A справа (слева) на матрицу перестановок переставляются столбцы (строки) матрицы A.

3.56. Для любой матрицы перестановок P всегда выполняются соотношения $P'P = PP' = E$.

3.57. Произведение матриц перестановок одного порядка есть снова матрица перестановок.

3.58. Множество всех матриц перестановок одного порядка есть конечная группа по умножению. При этом единичным элементом группы является единичная матрица, а элементом, обратным для матрицы перестановок P,— матрица P'.

3.59. Матрица A с элементами a_{ij} называется *правой* или *верхней* (*левой* или *нижней*) *треугольной*, если $a_{ij} = 0$ при $i > j$ ($i < j$).

3.60. Матрица A с элементами a_{ij} называется *строго правой* или *строго верхней* (*строго левой* или *строго нижней*) *треугольной*, если $a_{ij} = 0$ при $i \geqslant j$ ($i \leqslant j$).

3.61. Сумма и произведение любых двух треугольных матриц одного наименования есть треугольная матрица того же наименования.

3.62. Множество треугольных матриц одного размера и одного наименования есть линейное пространство.

3.63. Множество квадратных треугольных матриц одного порядка и одного наименования есть кольцо.

3.64. Множество квадратных треугольных матриц с ненулевыми диагональными элементами одного порядка и одного наименования есть группа по умножению.

§ 4. Определители, миноры, дополнения

4.1. Совокупность чисел $j_1, j_2, \ldots, j_n$, среди которых нет равных и каждое из которых есть одно из чисел $1, 2, \ldots, n$, называется *перестановкой* этих чисел. Перестановка $1, 2, \ldots, n$ называется *нормальной*.

4.2. В множестве из n чисел общее количество перестановок равно $n!$.

4.3. Говорят, что в данной перестановке числа i, j образуют *инверсию*, если $i > j$, но i стоит в перестановке раньше j.

4.4. Перестановка называется *четной*, если ее числа составляют четное количество инверсий, и *нечетной* — в противном случае.

4.5. *Транспозицией* называется преобразование перестановки, при котором меняются местами какие-либо два числа, не обязательно стоящие рядом.

4.6. Всякая транспозиция меняет четность перестановки.

4.7. Все $n!$ перестановок из n чисел можно расположить в таком порядке, что каждая следующая будет получаться из предыдущей при помощи одной транспозиции, причем начинать можно с любой перестановки.

4.8. Число четных перестановок равно числу нечетных и равно $\frac{1}{2}n!$.

4.9. *Определителем* n-го порядка, соответствующим квадратной матрице порядка n, называется алгебраическая сумма $n!$ членов, составленная следующим образом. Членами определителя служат всевозможные произведения по n элементов матрицы, взятых по одному в каждой строке и каждом столбце. Член берется со знаком плюс, если индексы столбцов его элементов образуют четную перестановку при условии, что сами элементы расположены в порядке возрастания номеров строк, и со знаком минус — в противном случае.

4.10. Если необходимо указать явный вид элементов матрицы A, то для ее определителя употребляется обозначение

$$\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix}.$$

Если же явный вид не нужен, то используется обозначение $\det A$ или $|A|$.

4.11. Определитель не меняется при транспонировании матрицы. Это утверждение означает, что все свойства определителя, которые имеют место по отношению к строкам матрицы, будут справедливы также и по отношению к столбцам. Поэтому в дальнейшем мы будем указывать только свойства, касающиеся строк.

4.12. Определитель равен нулю тогда и только тогда, когда строки матрицы образуют линейно зависимую систему.

4.13. Если какую-либо строку матрицы умножить на число α, то и определитель умножится на число α.

4.14. Если все элементы матрицы порядка n умножить на число α, то определитель умножится на α^n.

4.15. Определитель меняет знак, если любые две различные строки матрицы поменять местами.

4.16. Определитель не меняется, если к какой-нибудь строке матрицы прибавить любую линейную комбинацию остальных строк.

4.17. Если все элементы i-й строки представлены в виде суммы $a_{ij} = a'_{ij} + a''_{ij}$, то определитель равен сумме двух определителей, у которых все строки, кроме i-й, такие же, как в исходном определителе, а i-я строка в одном из слагаемых состоит из элементов a'_{ij}, в другом — из элементов a''_{ij}.

4.18. Определитель p-го порядка, составленный из элементов матрицы A, стоящих на пересечении строк с номерами $i_1, i_2, \ldots$ $\ldots, i_p$ и столбцов с номерами $j_1, j_2, \ldots, j_p$, называется *минором* p-го порядка (матрицы или определителя) и обозначается

$$A\begin{pmatrix} i_1 & i_2 & \ldots & i_n \\ j_1 & j_2 & \ldots & j_n \end{pmatrix}.$$

При этом совпадение каких-либо индексов в верхней (нижней) строке в обозначении минора означает, что совпадают между собой соответствующие строки (столбцы) самого минора.

4.19. Минор, расположенный в первых p строках и первых p столбцах, называется *ведущим* или *угловым* минором. Минор, расположенный в столбцах и строках с одинаковыми номерами, называется *главным*.

4.20. (*Формула Бине — Коши.*) Пусть квадратная матрица C порядка n равна произведению двух прямоугольных матриц A и B соответственно размеров $n \times m$ и $m \times n$, причем $m \geqslant n$. Тогда

$$C\begin{pmatrix} 1 & 2 & & n \\ 1 & 2 & & n \end{pmatrix} = \sum_{1 \leqslant k_1 < \ldots < k_n \leqslant m} A\begin{pmatrix} 1 & 2 & & n \\ k_1 & k_2 & \ldots & k_n \end{pmatrix} B\begin{pmatrix} k_1 & k_2 & \ldots & k_n \\ 1 & 2 & \ldots & n \end{pmatrix}.$$

4.21. Пусть квадратная матрица C порядка n равна произведению двух прямоугольных матриц A и B соответственно размеров $n \times m$ и $m \times n$, причем $m < n$. Тогда $\det C = 0$.

4.22. Определитель произведения двух квадратных матриц равен произведению определителей сомножителей.

4.23. Пусть в строках $i_1, \ldots, i_k$ и столбцах $j_1, \ldots, j_k$ определителя расположен минор M порядка k. Минор N порядка $n-k$, расположенный в строках и столбцах матрицы, оставшихся после вычеркивания строк $i_1, \ldots, i_k$ и столбцов $j_1, \ldots, j_k$, называется *дополнительным* минором для минора M. Число

$$(-1)^{\sum\limits_{p=1}^{k}(i_p+j_p)} N$$

называется *алгебраическим дополнением* минора M. Алгебраическое дополнение элемента a_{ij} (минор 1-го порядка) матрицы A обозначается A_{ij}.

4.24. (*Теорема Лапласа.*) Пусть в определителе d порядка n произвольно выбраны k строк (столбцов), где $1 \leqslant k \leqslant n-1$. Тогда сумма произведений всех миноров k-го порядка, содержащихся в выбранных строках (столбцах), на их алгебраические дополнения равна определителю d.

4.25. Пусть в определителе d порядка n произвольно выбраны две различные группы по k строк (столбцов), где $1 \leqslant k \leqslant n-1$. Тогда сумма произведений всех миноров k-го порядка, содержащихся в строках (столбцах) одной группы, на алгебраические дополнения соответствующих миноров другой группы равна нулю.

4.26. Для любой квадратной матрицы $A(a_{ij})$ порядка n всегда выполняются следующие соотношения:

$$\sum_{s=1}^{n} a_{ks}A_{is} = \begin{cases} \det A, & k = i, \\ 0, & k \neq i; \end{cases} \qquad \sum_{s=1}^{n} a_{sk}A_{si} = \begin{cases} \det A, & k = i, \\ 0, & k \neq i. \end{cases}$$

4.27. Определитель диагональной матрицы равен произведению диагональных элементов.

4.28. Определитель единичной матрицы равен единице.

4.29. Определитель треугольной матрицы равен произведению диагональных элементов.

Прежде чем переходить к дальнейшим исследованиям различных аспектов, связанных с определителем, заметим следующее. Существует много эквивалентных определений этого понятия. Мы рассмотрели одно из них, самое распространенное, но, пожалуй, и самое формальное. За этим определением трудно «разглядеть», что же представляет собой эта функция, если отвлечься от использования конкретных базисов в линейных пространствах.

Иногда определитель вводится как некоторая рекуррентная (по порядку матрицы) функция, вычисляемая на основе теоремы Лапласа. Наверное, наиболее наглядно определитель можно было бы ввести как ориентированный объем системы векторов, о чем речь пойдет дальше. Однако, чем проще выглядит определение определителя, тем сложнее доказать свойство 4.11. Для установления этого свойства неизбежно приходится в той или иной мере переходить к определению 4.9.

В действительности определитель есть некоторая скалярная функция, являющаяся не чем иным, как числовой мерой степени линейной зависимости или независимости системы векторов. Оказывается, что при вполне естественных ограничениях существует только одна функция, обладающая свойствами определителя.

4.30. На множестве квадратных матриц существует единственная функция Φ, обладающая следующими свойствами:

— Φ есть линейная функция по каждой вектор-строке матрицы;

— Φ равна нулю на всех матрицах с линейно зависимыми вектор-строками;

— Φ равна единице на единичной матрице.

4.31. Функция Φ совпадает с определителем.

4.32. Наивысший порядок отличных от нуля миноров прямоугольной матрицы A называется ее *рангом* и обозначается символом $\operatorname{rank} A$. По определению $\operatorname{rank} 0 = 0$.

4.33. Любой отличный от нуля минор порядка $\operatorname{rank} A$ называется *базисным минором* матрицы A.

4.34. Строки и столбцы, на которых расположен базисный минор, называются *базисными*.

4.35. Любые базисные строки (столбцы) матрицы образуют базу ее вектор-строк (вектор-столбцов).

4.36. Для любой прямоугольной матрицы ранги ее систем вектор-строк и вектор-столбцов совпадают и равны рангу матрицы.

4.37. На пересечении любых базисных строк и базисных столбцов матрицы находится базисный минор.

4.38. В любых r линейно независимых строках (столбцах) матрицы найдется ненулевой минор порядка r.

4.39. Перестановками строк и столбцов прямоугольной матрицы можно добиться того, что все ведущие миноры, порядок которых не превосходит ранга матрицы, будут отличны от нуля.

4.40. Перестановками строк (столбцов) квадратной матрицы с ненулевым определителем можно добиться того, что все ведущие миноры будут отличны от нуля.

4.41. Прямоугольная матрица A размера $m \times n$ называется матрицей *полного ранга*, если $\operatorname{rank} A = \min\{m, n\}$.

4.42. Для любой квадратной матрицы A полного ранга все ведущие миноры матриц A^*A и AA^* неотрицательны, а ведущие миноры, порядок которых не превосходит $\operatorname{rank} A$, строго положительны.

4.43. Матрица размера $n \times m$ имеет ранг 1 тогда и только тогда, когда она может быть представлена в виде произведения двух ненулевых матриц размеров $n \times 1$ и $1 \times m$.

4.44. Любую матрицу ранга r можно представить в виде суммы r матриц ранга 1 и нельзя представить в виде суммы меньшего числа матриц ранга 1.

4.45. (*Неравенство Фробениуса.*) Если A, B, C — прямоугольные матрицы и произведение ABC существует, то

$$\operatorname{rank} AB + \operatorname{rank} BC \leqslant \operatorname{rank} B + \operatorname{rank} ABC.$$

4.46. (*Неравенство Сильвестра.*) Если прямоугольные матрицы A и B имеют соответственно n столбцов и строк, то

$$\operatorname{rank} A + \operatorname{rank} B - n \leqslant \operatorname{rank} AB \leqslant \min\{\operatorname{rank} A, \operatorname{rank} B\}.$$

4.47. Имеют место соотношения

$$\operatorname{rank}(\alpha A) = \operatorname{rank} A, \ \alpha \neq 0,$$
$$\operatorname{rank} A = \operatorname{rank} AA^* = \operatorname{rank} A^*A,$$
$$\operatorname{rank} A = \operatorname{rank} A' = \operatorname{rank} A^*,$$
$$\operatorname{rank} AB \leqslant \min \{\operatorname{rank} A, \operatorname{rank} B\},$$
$$\operatorname{rank} (A + B) \leqslant \operatorname{rank} A + \operatorname{rank} B.$$

4.48. От умножения слева или справа на матрицы, определитель которых отличен от нуля, ранг не меняется.

4.49. Матрица называется *невырожденной* (*неособенной*), если ее определитель не равен нулю, и *вырожденной* (*особенной*) — в противном случае.

4.50. Матрица является невырожденной тогда и только тогда, когда ее строки (столбцы) линейно независимы.

4.51. Произведение невырожденных матриц одного порядка есть невырожденная матрица.

4.52. Произведение любой квадратной матрицы и вырожденной матрицы есть вырожденная матрица.

4.53. Матрица $\widehat{A}$ с элементами $\widehat{a}_{ij}$ называется *присоединенной* для квадратной матрицы A с элементами a_{ij}, если $\widehat{a}_{ij} = A_{ji}$, где A_{ji} есть алгебраическое дополнение a_{ji}.

4.54. Согласно 4.26 матрица и присоединенная к ней связаны между собой соотношением

$$\widehat{A}A = A\widehat{A} = \det A \cdot E.$$

4.55. Определитель матрицы, присоединенной для квадратной матрицы порядка n, равен $(\det A)^{n-1}$.

4.56. Для любой невырожденной матрицы A существует единственная матрица A^{-1} такая, что $AA^{-1} = A^{-1}A = E$. Матрица A^{-1} называется *обратной* к матрице A.

4.57. Обратная матрица связана с присоединенной соотношением $A^{-1} = (\det A)^{-1}\widehat{A}$, и ее элементы таковы:

$$A^{-1} = \left\| \begin{array}{cccc} \frac{A_{11}}{\det A} & \frac{A_{21}}{\det A} & & \frac{A_{n1}}{\det A} \\ \frac{A_{12}}{\det A} & \frac{A_{22}}{\det A} & & \frac{A_{n2}}{\det A} \\ \cdot & & & \cdot\ \cdot \\ \frac{A_{1n}}{\det A} & \frac{A_{2n}}{\det A} & \cdots & \frac{A_{nn}}{\det A} \end{array} \right\|.$$

4.58. Множество всех невырожденных матриц одного порядка есть группа по умножению.

4.59. Имеют место соотношения

$$(\bar{A})^{-1} = \overline{(A^{-1})}, \quad (A')^{-1} = (A^{-1})', \quad (A^*)^{-1} = (A^{-1})^*,$$
$$(AB)^{-1} = B^{-1}A^{-1}.$$

4.60. Всегда выполняется равенство $\det A^{-1} = (\det A)^{-1}$.

4.61. Пусть матрица A — невырожденная и матрица B ранга 1 представлена в виде $B = xy'$, где x, y — столбцевые матрицы. Тогда

$$(A + B)^{-1} = A^{-1} - \beta A^{-1}BA^{-1},$$

где $\beta = (1 + \alpha)^{-1}$, $\alpha = y'A^{-1}x$.

4.62. Если к матрице A прибавить матрицу ранга 1, то и к обратной матрице A^{-1} прибавляется матрица ранга 1.

4.63. Квадратная матрица A называется *вполне неотрицательной* (*вполне положительной*), если все миноры любых порядков этой матрицы неотрицательны (положительны).

4.64. Произведение вполне неотрицательных (вполне положительных) матриц есть матрица вполне неотрицательная (вполне положительная).

4.65. Пусть A — квадратная матрица порядка n. Для заданного числа p, $1 \leqslant p \leqslant n$, упорядочим все $N = C_n^p$ сочетаний из n чисел $1, 2, \ldots, n$ по p чисел $k_1 < k_2 < \ldots < k_p$ в лексикографическом порядке. Это означает, что сочетание $k_1 < k_2 < \ldots < k_p$ предшествует сочетанию $k_1' < k_2' < \cdots < k_p'$, если $k_1 = k_1', \ldots$ $\ldots, k_{l-1} = k_{l-1}'$, но $k_l < k_l'$, $1 \leqslant l \leqslant p$. Составим квадратную матрицу A_p с элементами

$$\alpha_{ij} = A\begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ j_1 & j_2 & \ldots & j_p \end{pmatrix},$$

если номер сочетания $i_1 < i_2 < \ldots < i_p$ равен i, а номер сочетания $j_1 < j_2 < \ldots < j_p$ равен j. Полученная матрица A_p называется p-й *ассоциированной* с A матрицей.

4.66. Ассоциированная матрица для единичной есть единичная.

4.67. Ассоциированная матрица для диагональной есть диагональная.

4.68. Ассоциированная матрица для треугольной есть треугольная того же наименования.

4.69. Имеют место соотношения

$$A_1 = A, \quad A_n = \det A, \quad (AB)_p = A_pB_p, \quad (A^{-1})_p = (A_p)^{-1}.$$

4.70. Миноры матрицы и обратной к ней связаны между собой соотношениями

$$A^{-1}\begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ k_1 & k_2 & \ldots & k_p \end{pmatrix} = \frac{1}{\det A} \cdot (-1)^{\sum\limits_{s=1}^{p}(i_s+k_s)} A\begin{pmatrix} k_1' & k_2' & \ldots & k_{n-p}' \\ i_1' & i_2' & \ldots & i_{n-p}' \end{pmatrix}.$$

Здесь $i_1 < i_2 < \ldots < i_p$ вместе с $i'_1 < i'_2 < \cdots < i'_{n-p}$ и $k_1 < k_2 < \ldots$ $\cdots < k_p$ вместе с $k'_1 < k'_2 < \cdots < k'_{n-p}$ составляют полную систему индексов $1, 2, \ldots, n$.

Среди других числовых функций от матриц отметим *перманент матрицы*. Перманент отличается от определителя тем, что в определении 4.9 каждый член всегда берется со знаком плюс. Для этой функции справедливы многие свойства определителя, например 4.11, 4.13, 4.17, имеют место аналоги теоремы Лапласа 4.24, формулы Бине — Коши 4.20 и т. п. Однако перманент не меняется при перестановке строк матрицы. Он не может служить мерой линейной зависимости или независимости системы векторов. Легко построить примеры, показывающие, что перманент может быть отличен от нуля, даже если строки матрицы линейно зависимы, и равен нулю для матрицы с линейно независимыми строками. Мы не будем более детально касаться свойств перманента.

§ 5. Скалярное произведение

Абстрактные линейные пространства в некотором смысле беднее своими понятиями и свойствами, чем множества направленных отрезков на плоскости и в пространстве. Беднее прежде всего потому, что в них не нашли отражения важнейшие факты, связанные с измерениями длин, углов, площадей, объемов и т. д. Распространять метрические понятия на абстрактные линейные пространства можно различным образом. Однако самым эффективным способом задания возможности измерений является аксиоматическое введение скалярного произведения векторов.

5.1. Вещественное линейное пространство E называется *евклидовым*, если каждой паре векторов x, y из E поставлено в соответствие вещественное число (x, y), называемое *скалярным* или *внутренним* произведением, причем выполнены следующие аксиомы:

$$(x, y) = (y, x),$$
$$(\lambda x, y) = \lambda(x, y),$$
$$(x + y, z) = (x, z) + (y, z),$$
$$(x, x) > 0 \text{ при } x \neq 0, \quad (0, 0) = 0$$

для произвольных векторов x, y, z из E и произвольного вещественного числа λ.

5.2. Комплексное линейное пространство U называется *унитарным*, если каждой паре векторов x, y из U поставлено в соответствие комплексное число (x, y), называемое *скалярным* или *внутренним* произведением, причем выполнены следующие аксиомы:

$$(x, y) = \overline{(y, x)},$$
$$(\lambda x, y) = \lambda(x, y),$$
$$(x + y, z) = (x, z) + (y, z),$$
$$(x, x) > 0 \text{ при } x \neq 0, \quad (0, 0) = 0$$

для произвольных векторов x, y, z и произвольного комплексного числа λ. Черта в первой аксиоме означает комплексное сопряжение.

Сравнение аксиом евклидова и унитарного пространств показывает, что они различаются «всего лишь» комплексным сопряжением в первой аксиоме. Однако это небольшое отличие исключительно важно, и его нельзя забывать. Если бы в комплексном пространстве аксиомы скалярного произведения полностью совпадали с аксиомами евклидова пространства, то такое пространство имело бы совсем другие свойства, чем унитарное.

При указанных выше аксиомах большинство фактов евклидова и унитарного пространств оказываются либо одинаковыми, либо очень похожими. Если какие-либо определения или факты в обоих типах пространств совпадают полностью или с точностью до замены комплексных чисел вещественными, мы будем приводить соответствующие формулировки только для унитарных пространств без дополнительной оговорки.

Естественно, что любое подпространство евклидова (унитарного) пространства само становится евклидовым (унитарным) пространством, если в нем сохранить скалярное произведение.

5.3. Со скалярным произведением можно выполнять формальные алгебраические преобразования, т. е.

$$\left(\sum_{i=1}^{r}\alpha_i x_i,\ \sum_{j=1}^{s}\beta_j y_j\right)=\sum_{i=1}^{r}\sum_{j=1}^{s}\alpha_i\bar{\beta}_j(x_i,\ y_j)$$

для любых векторов x_i, y_j, чисел α_i, β_j и любого числа r, s слагаемых.

5.4. Пусть $e_1, \ldots, e_n$ — базис линейного пространства и для произвольных векторов x, y имеют место разложения

$$x=\xi_1e_1+\xi_2e_2+\ldots+\xi_ne_n,$$
$$y=\eta_1e_1+\eta_2e_2+\ldots+\eta_ne_n.$$

Скалярное произведение всегда можно ввести следующим образом:

$$(x,\ y)=\xi_1\bar{\eta}_1+\xi_2\bar{\eta}_2+\ldots+\xi_n\bar{\eta}_n.$$

5.5. (*Неравенство Коши — Буняковского.*) Для любых векторов x, y справедливо неравенство

$$|(x,\ y)|^2\leqslant(x,\ x)(y,\ y).$$

Это неравенство иногда называют *неравенством Шварца.*

Мы не будем сейчас обращать внимание на исследование арифметических пространств со скалярным произведением. Наличие естественного базиса в этих пространствах делает соответствующие иллюстрации почти очевидными. Так, если векторы заданы своими координатами, то формула 5.4 представляет собой наиболее распространенный способ задания скалярного произведения через координаты векторов, неравенство Коши — Буняковского превращается в известное неравенство

$$\left|\sum_{i=1}^{n}\xi_i\eta_i\right|^2\leqslant\left(\sum_{i=1}^{n}|\xi_i|^2\right)\left(\sum_{j=1}^{n}|\eta_j|^2\right)$$

для двух последовательностей комплексных чисел и т. д.

Однако в арифметическом пространстве скалярное произведение далеко не всегда вводится через естественный базис, и к тому же в форме 5.4. Чтобы не потерять в дальнейшем общности, мы будем использовать в наших исследованиях только сформулированные выше аксиоматические свойства скалярного произведения. Естественно, что при этом все утверждения остаются в силе и в том случае, когда скалярное произведение задано в виде 5.4.

5.6. Неравенство Коши — Буняковского обращается в равенство тогда и только тогда, когда векторы x, y *коллинеарны*, т. е. либо $x=\alpha y$, либо $y=\alpha x$ для некоторого числа α.

5.7. Вектор x называется *нормированным*, если $(x, x)=1$.

5.8. Система векторов называется *нормированной*, если нормированы все ее векторы.

5.9. Любой ненулевой вектор y можно нормировать, если умножить его на число $\lambda=(y, y)^{-1/2}$.

5.10. Векторы x, y называются *ортогональными*, если $(x, y)=0$.

5.11. Система векторов называется *ортогональной*, если либо она состоит из одного вектора, либо ее векторы попарно ортогональны.

5.12. Нормированная ортогональная система векторов называется *ортонормированной*.

5.13. Ортогональная система ненулевых векторов всегда линейно независима.

5.14. Единственный вектор, который ортогонален ко всем векторам пространства, есть нулевой вектор.

5.15. Если $e_1, \ldots, e_n$ — ортонормированная система векторов и для векторов x, y имеют место разложения

$$x=\alpha_1 e_1+\alpha_2 e_2+\ldots+\alpha_n e_n,$$
$$y=\beta_1 e_1+\beta_2 e_2+\ldots+\beta_n e_n,$$

то справедливы равенства

$$\alpha_i=(x, e_i), \quad \beta_i=(y, e_i),$$
$$(x, y)=\alpha_1\bar{\beta}_1+\alpha_2\bar{\beta}_2+\ldots+\alpha_n\bar{\beta}_n,$$
$$(x, x)=|\alpha_1|^2+|\alpha_2|^2+\ldots+|\alpha_n|^2.$$

5.16. В любом конечномерном пространстве со скалярным произведением существует ортонормированный базис.

5.17. Если для какого-либо базиса одно из равенств 5.15 выполняется для всех векторов, то этот базис по отношению к данному скалярному произведению является ортонормированным.

5.18. Любую ортонормированную систему векторов можно дополнить до ортонормированного базиса.

5.19. Два множества векторов называются *ортогональными*, если каждый вектор одного множества ортогонален каждому вектору другого множества. Ортогональность множеств F и G обозначается $F\perp G$.

5.20. Для того чтобы вектор x был ортогонален к подпространству L, необходимо и достаточно, чтобы он был ортогонален ко всем векторам какого-либо базиса подпространства L.

5.21. Для того чтобы два подпространства были ортогональны, необходимо и достаточно, чтобы каждый вектор какого-либо базиса одного подпространства был ортогонален ко всем векторам какого-либо базиса другого подпространства.

5.22. Совокупность всех векторов, ортогональных множеству F, называется *ортогональным дополнением* множества F и обозначается $F^{\perp}$.

5.23. Ортогональное дополнение любого множества есть подпространство.

5.24. Множество K векторов некоторого линейного пространства называется *суммой* множеств $L_1, L_2, \ldots, L_m$ этого пространства и обозначается $K = L_1 + L_2 + \ldots + L_m$, если оно состоит из всех векторов вида $x = x_1 + x_2 + \ldots + x_m$, где $x_i \in L_i$ для всех i.

5.25. Сумма подпространств есть подпространство.

5.26. Говорят, что линейное пространство K есть *прямая сумма* своих подпространств $L_1, L_2, \ldots, L_m$, и обозначают

$$K = L_1 \dotplus L_2 \dotplus \ldots \dotplus L_m,$$

если любой вектор $x \in K$ единственным образом представляется в виде суммы $x = x_1 + x_2 + \ldots + x_m$, где $x_i \in L_i$ для всех i. Если $m = 2$, то вектор x_1 (x_2) называется *проекцией вектора* x на подпространство L_1 (L_2) вдоль подпространства L_2 (L_1).

5.27. Для того чтобы пространство K было прямой суммой своих подпространств, необходимо и достаточно, чтобы объединение базисов этих подпространств составляло базис всего пространства.

5.28. Для того чтобы сумма подпространств была прямой, необходимо и достаточно, чтобы нулевой вектор единственным образом был разложим по векторам базисов подпространств.

5.29. Множество F некоторого линейного пространства называется *пересечением* множеств L_1, L_2 этого пространства и обозначается $F = L_1 \cap L_2$, если оно состоит из векторов, одновременно принадлежащих как L_1, так и L_2.

5.30. Пересечение подпространств есть подпространство.

5.31. Для любых подпространств L_1, L_2 имеет место равенство

$$\dim (L_1 \cap L_2) + \dim (L_1 + L_2) = \dim L_1 + \dim L_2.$$

5.32. Размерность суммы любого числа подпространств не меньше, чем максимальная из размерностей этих подпространств.

5.33. Для того чтобы сумма подпространств была прямой, необходимо и достаточно, чтобы пересечение каждого из подпространств с суммой остальных содержало лишь нулевой вектор.

5.34. Сумма K подпространств L_1, L_2, ..., L_m называется *ортогональной* и обозначается

$$K = L_1 \oplus L_2 \oplus \ldots \oplus L_m,$$

если подпространства попарно ортогональны.

5.35. Ортогональная сумма ненулевых подпространств всегда является прямой суммой.

5.36. Евклидово или унитарное пространство есть ортогональная сумма любого своего линейного подпространства и его ортогонального дополнения.

5.37. Для того чтобы ранг какой-либо системы векторов равнялся размерности пространства, необходимо и достаточно, чтобы единственным вектором пространства, ортогональным всем векторам системы, был нулевой вектор.

5.38. Пусть $K = L_1 \dotplus \ldots \dotplus L_m$ и для векторов x, y имеют место разложения $x = x_1 + \ldots + x_m$, $y = y_1 + \ldots + y_m$. Для того чтобы сумма K была ортогональной, необходимо и достаточно, чтобы для всех x, y выполнялось равенство $(x, y) = (x_1, y_1) + \ldots + (x_m, y_m)$.

5.39. Для любых двух подпространств L, M евклидова или унитарного пространства K справедливы соотношения

$$\dim L + \dim L^{\perp} = \dim K;$$

$$(L^{\perp})^{\perp} = L, \ K^{\perp} = 0, \ 0^{\perp} = K;$$

$$(L + M)^{\perp} = L^{\perp} \cap M^{\perp}, \ (L \cap M)^{\perp} = L^{\perp} + M^{\perp};$$

$$\text{если } L \subset M, \text{ то } M^{\perp} \subset L^{\perp}.$$

5.40. *Длиной* $|x|$ вектора x называется величина $+(x, x)^{1/2}$.

5.41. Для любого вектора x и числа λ выполняется равенство $|\lambda x| = |\lambda|\,|x|$.

5.42. Если векторы $x_1, \ldots, x_s$ попарно ортогональны, то $|x_1 + \ldots + x_s|^2 = |x_1|^2 + \ldots + |x_s|^2$.

5.43. Для произвольных векторов x, y имеют место соотношения $|\,|x| - |y|\,| \leqslant |x - y| \leqslant |x| + |y|$.

5.44. Для любых векторов x, y имеет место *тождество параллелограмма* $|x - y|^2 + |x + y|^2 = 2(|x|^2 + |y|^2)$.

5.45. *Углом* $\{x, y\}$ между ненулевыми векторами x, y евклидова пространства называется угол, определяемый соотношениями

$$\cos\{x, y\} = \frac{(x, y)}{|x|\,|y|}, \quad 0 \leqslant \{x, y\} \leqslant \pi.$$

Если среди векторов x, y есть хотя бы один нулевой, то угол между такими векторами считается не определенным. В унитарном пространстве понятие угла между векторами обычно не вводится.

5.46. Угол между ненулевыми векторами равен 0 или π тогда и только тогда, когда векторы коллинеарны.

5.47. Пусть $e_1, \ldots, e_n$ — ортонормированный базис евклидова пространства, x — произвольный вектор. Всегда выполняется равенство

$$\cos^2\{x, e_1\} + \cos^2\{x, e_2\} + \ldots + \cos^2\{x, e_n\} = 1.$$

5.48. *Расстоянием* $\rho(x, y)$ между векторами x, y называется величина $|x-y|$.

5.49. Расстояние между векторами удовлетворяет следующим свойствам:

$$\rho(x, y) = \rho(y, x);$$

$$\rho(x, y) > 0, \text{ если } x \neq y, \quad \rho(x, y) = 0, \text{ если } x = y;$$

$$\rho(x, y) \leqslant \rho(x, z) + \rho(z, y) \text{ (\textit{неравенство треугольника}).}$$

5.50. *Расстоянием* $\rho(A, B)$ между множествами A, B векторов одного и того же пространства называется величина

$$\rho(A, B) = \inf \rho(x, y), \quad x \in A, \quad y \in B.$$

5.51. *Углом* $\{x, L\}$ между ненулевыми вектором x и подпространством L евклидова пространства называется наименьший из углов, которые x образует с векторами из L.

5.52. Если евклидово пространство разложено в ортогональную сумму подпространств $L_1, \ldots, L_p$, то

$$\cos^2\{x, L_1\} + \cos^2\{x, L_2\} + \ldots + \cos^2\{x, L_p\} = 1.$$

5.53. Каковы бы ни были вектор f и подпространство L, всегда существует, и притом единственное, разложение $f = g + h$, где $g \in L$, $h \perp L$. Вектор g называется *ортогональной проекцией* вектора f на подпространство L, вектор h — *перпендикуляром*, опущенным из f на L. Проекция обозначается $\mathrm{pr}_L f$, перпендикуляр — $\mathrm{ort}_L f$.

5.54. Для любых векторов x, y и числа λ всегда выполняются соотношения

$$\mathrm{pr}_L x = \mathrm{ort}_{L^\perp} x, \qquad \mathrm{ort}_L x = \mathrm{pr}_{L^\perp} x,$$

$$\mathrm{pr}_L(x+y) = \mathrm{pr}_L x + \mathrm{pr}_L y, \qquad \mathrm{pr}_L(\lambda x) = \lambda\, \mathrm{pr}_L x,$$

$$\mathrm{ort}_L(x+y) = \mathrm{ort}_L x + \mathrm{ort}_L y, \qquad \mathrm{ort}_L(\lambda x) = \lambda\, \mathrm{ort}_L x,$$

$$\mathrm{pr}_L(\mathrm{pr}_L x) = \mathrm{pr}_L x, \qquad \mathrm{ort}_L(\mathrm{ort}_L x) = \mathrm{ort}_L x.$$

5.55. Для того чтобы сумма $L_1 + L_2$ подпространств L_1, L_2 была ортогональной, необходимо и достаточно, чтобы для любого вектора x выполнялось равенство

$$\mathrm{pr}_{L_1+L_2} x = \mathrm{pr}_{L_1} x + \mathrm{pr}_{L_2} x.$$

5.56. Если подпространства $L_1, L_2, \ldots, L_m$ попарно ортогональны, то для любого вектора x справедливо неравенство

$$|x|^2 \geqslant \sum_{i=1}^{m} |\mathrm{pr}_{L_i} x|^2,$$

причем равенство достигается тогда и только тогда, когда x принадлежит ортогональной сумме этих подпространств.

5.57. Расстояние между вектором и подпространством равно длине перпендикуляра, опущенного из вектора на подпространство.

5.58. Ближайшим вектором подпространства к заданному вектору является его проекция на это подпространство.

5.59. Угол между вектором и подпространством совпадает с углом между этим вектором и его проекцией на подпространство.

5.60. Пусть $e_1, \ldots, e_s$ — ортонормированный базис подпространства L. Для любого вектора x имеют место соотношения

$$\mathrm{pr}_L x = \sum_{i=1}^{s} (x, e_i) e_i, \qquad \mathrm{ort}_L x = x - \sum_{i=1}^{s} (x, e_i) e_i,$$

$$|\mathrm{pr}_L x|^2 = \sum_{i=1}^{s} |(x, e_i)|^2, \qquad |\mathrm{ort}_L x| = |x|^2 - \sum_{i=1}^{s} |(x, e_i)|^2.$$

5.61. Две системы векторов $x_1, \ldots, x_k$ и $y_1, \ldots, y_k$ называются *биортогональными* или *двойственными*, если

$$(x_i, y_i) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

5.62. Каждая из биортогональных систем линейно независима.

5.63. Для любого базиса существует единственный биортогональный базис.

5.64. Пусть $e_1, \ldots, e_n$ и $f_1, \ldots, f_n$ — пара биортогональных базисов. Для любого k, $1 \leqslant k < n$, ортогональное дополнение к линейной оболочке векторов $e_1, \ldots, e_k$ совпадает с линейной оболочкой векторов $f_{k+1}, \ldots, f_n$.

5.65. Пусть в пространстве со скалярным произведением задана система векторов $x_1, \ldots, x_n$. Обозначим через L_0 нулевое подпространство, через L_i — линейную оболочку векторов $x_1, \ldots, x_i$. *Объемом* $V(x_1, \ldots, x_n)$ системы векторов $x_1, \ldots, x_n$ называется величина

$$V(x_1, \ldots, x_n) = \prod_{i=0}^{n-1} |\mathrm{ort}_{L_i} x_{i+1}|.$$

5.66. Для любой системы векторов $x_1, \ldots, x_n$ справедливо *неравенство Адамара*

$$V(x_1, \ldots, x_n) \leqslant \prod_{i=0}^{n-1} |x_{i+1}|,$$

причем равенство имеет место тогда и только тогда, когда система $x_1, \ldots, x_n$ ортогональна или содержит нулевой вектор.

5.67. Если система $x_1, \ldots, x_n$ нормирована и ее объем равен единице, то она ортонормирована.

5.68. Среди нормированных систем ортонормированная система имеет максимальный объем.

5.69. Для любых двух ортогональных множеств векторов $x_1, \ldots, x_p$ и $y_1, \ldots, y_r$ справедливо равенство

$$V(x_1, \ldots, x_p, y_1, \ldots, y_r) = V(x_1, \ldots, x_p)V(y_1, \ldots, y_r).$$

5.70. Разложим векторы $x_1, \ldots, x_n$ по какому-нибудь ортонормированному базису $e_1, \ldots, e_n$ и из координат этого разложения, расположенных по строкам или столбцам, составим квадратную матрицу. Модуль определителя данной матрицы всегда совпадает с объемом системы векторов $x_1, \ldots, x_n$.

5.71. Пусть $e_1, \ldots, e_n$ — ортонормированная система векторов. Если для системы $\varepsilon_1, \ldots, \varepsilon_n$ выполняется неравенство

$$\sum_{i=1}^{n} |\varepsilon_i|^2 < 1,$$

то система векторов $e_1 + \varepsilon_1, \ldots, e_n + \varepsilon_n$ линейно независима.

5.72. Евклидовы (унитарные) пространства K и K' называются *евклидово* (*унитарно*) *изоморфными*, если они изоморфны как вещественные (комплексные) линейные пространства и, кроме того, для любой пары векторов x, y из K и соответствующих векторов x', y' из K' выполняется равенство $(x, y) = (x', y')$.

5.73. Для того чтобы два евклидовых (унитарных) пространства были евклидово (унитарно) изоморфны, необходимо и достаточно, чтобы были равны их размерности.

5.74. Пусть даны два евклидовых (унитарных) пространства одной размерности. Выберем в каждом пространстве по какому-нибудь ортонормированному базису и установим соответствие между теми векторами, которые имеют одинаковые координаты соответственно в своих базисах. Это соответствие есть евклидов (унитарный) изоморфизм.

Установление изоморфизма между двумя пространствами со скалярным произведением позволяет в теоретическом плане ограничиться исследованием арифметических пространств со скалярным произведением 5.4 по отношению к естественному базису. В реальных ситуациях не всегда можно указать ортонормированный базис, позволяющий установить изоморфное соответствие арифметическому пространству. Именно этим, в основном, и объясняется наше стремление сохранить бо́льшую общность. Кроме того, не всегда за определенными фактами в арифметических пространствах легко увидеть существенные моменты. В качестве примера приведем следующее утверждение.

5.75. Пусть в арифметических пространствах скалярные произведения введены согласно 5.4 по отношению к естественным базисам. Для любой прямоугольной матрицы A и любых векторов x, y соответствующих размеров всегда выполняется равенство

$$(Ax, y) = (x, A^*y).$$

Рассматривая это равенство с точки зрения формальных матричных определений и естественного скалярного произведения, трудно увидеть что-нибудь большее, чем то, что написано.

В заключение мы хотели бы обратить внимание на утверждение 5.71. Из него вытекает, что при малых изменениях ортонормированных систем сохраняется их линейная независимость. В других системах это свойство уже не проявляется в столь четком виде, так как малость изменения зависит в общем случае от меры линейной независимости исходной системы. Эта особенность ортонормированных систем определила их широкое использование при построении самых различных вычислительных алгоритмов.

§ 6. Системы линейных алгебраических уравнений

6.1. *Системой линейных алгебраических уравнений* относительно неизвестных $x_1, x_2, \ldots, x_m$ называется совокупность уравнений вида

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m &= y_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m &= y_2, \\ &\cdot \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m &= y_n. \end{aligned}$$

Числа a_{ij} называются *коэффициентами* системы, числа y_i — ее *правыми частями.*

6.2. Упорядоченная совокупность неизвестных, удовлетворяющая каждому из уравнений, называется *решением* системы.

6.3. Если система имеет хотя бы одно решение, то она называется *совместной*; в противном случае система называется *несовместной.*

6.4. Если система совместна, то каждое ее решение называется *частным*; совокупность всех частных решений называется *общим* решением.

6.5. Система называется *неоднородной*, если среди ее правых частей есть хотя бы одна отличная от нуля; в противном случае система называется *однородной.*

6.6. Система, полученная путем замены всех правых частей нулями, называется *приведенной* однородной системой.

6.7. Однородная система всегда совместна, так как одним из ее частных решений является нулевое решение.

6.8. Две совместные системы линейных алгебраических уравнений относительно одних и тех же неизвестных называются *эквивалентными*, если каждое решение одной системы является решением другой системы.

6.9. В терминах матричных операций система линейных алгебраических уравнений может быть записана следующим образом:

$$\begin{bmatrix} a_{11} & a_{12} & & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ & & & \\ a_{n1} & a_{n2} & \ldots & a_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_m \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix}.$$

6.10. Матрица A с элементами a_{ij} называется *матрицей системы*, вектор x с неизвестными $x_1, \ldots, x_m$ — *вектором неизвестных*, вектор y, составленный из чисел $y_1, \ldots, y_n$,— *вектором правых частей* системы или просто *правой частью*.

В этих обозначениях система линейных алгебраических уравнений выглядит так: $Ax = y$.

Решение систем линейных алгебраических уравнений является одной из самых важных задач линейной алгебры. Как вытекает из соотношений 3.14, к ней сводится нахождение по заданному образу при линейном отображении одного пространства в другое всех или части его прообразов. К решению линейных систем сводится также определение координат векторов при их разложении по заданному базису, установление принадлежности вектора заданной линейной оболочке и многое другое. Все эти задачи особенно наглядны в арифметических пространствах, если к тому же под линейным преобразованием понимать умножение матрицы на вектор.

Совпадение символической записи для линейного преобразования и системы линейных алгебраических уравнений, конечно, не случайно. если принять во внимание изоморфизм линейных операторов и матриц. Задача решения систем, по существу, полностью совпадает с задачей исследования множества прообразов по заданному образу при фиксированных базисах в пространствах. Поэтому снова, вводя новые понятия и получая новые факты в отношении матриц, мы будем распространять их, как правило, и на операторы без специальных оговорок.

6.11. В терминах векторных операций (точнее, операций над столбцевыми матрицами) система линейных алгебраических уравнений может быть записана следующим образом:

$$x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \ldots \\ a_{n1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \ldots \\ a_{n2} \end{bmatrix} + \cdots + x_m \begin{bmatrix} a_{1m} \\ a_{2m} \\ \ldots \\ a_{nm} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix}.$$

6.12. Система линейных алгебраических уравнений совместна тогда и только тогда, когда вектор правой части принадлежит линейной оболочке вектор-столбцов матрицы системы.

6.13. Матрица, полученная приписыванием справа к матрице системы столбца правых частей, называется *расширенной* матрицей системы.

6.14. (*Теорема Кронекера — Капелли.*) Для того чтобы система линейных алгебраических уравнений была совместна, необходимо и достаточно, чтобы ранг расширенной матрицы системы равнялся рангу матрицы системы.

6.15. Общее решение приведенной однородной системы образует в m-мерном арифметическом пространстве подпространство размерности $m-r$, где r — ранг матрицы системы.

6.16. Любой базис подпространства решений приведенной однородной системы называется *фундаментальной системой* решений.

6.17. Общее решение неоднородной системы получается путем прибавления к общему решению приведенной однородной системы любого частного решения неоднородной системы.

6.18. Разность любых двух частных решений неоднородной системы есть частное решение приведенной однородной системы.

6.19. Для того чтобы совместная система имела единственное решение, необходимо и достаточно, чтобы ранг матрицы системы был равен числу неизвестных.

6.20. Для того чтобы однородная система имела ненулевое решение, необходимо и достаточно, чтобы ранг матрицы системы был меньше числа неизвестных.

6.21. Однородная система с квадратной матрицей имеет ненулевое решение тогда и только тогда, когда матрица системы вырожденная.

6.22. Система линейных алгебраических уравнений $Ax=y$ с квадратной матрицей A имеет единственное решение тогда и только тогда, когда матрица A невырожденная. В этом случае решение задается формулой $x=A^{-1}y$.

6.23. (*Формулы Крамера.*) Рассмотрим систему линейных алгебраических уравнений с невырожденной матрицей. Обозначим через d определитель матрицы системы, через d_j — определитель, отличающийся от d лишь тем, что в нем j-й столбец заменен столбцом правых частей. Тогда единственное решение системы может быть вычислено по формулам $x_j = d_j/d$ для всех j.

Несмотря на то, что формулы Крамера дают ясное представление решения системы с невырожденной матрицей, они редко используются в практических расчетах. Основные причины этого заключаются в большой трудоемкости процесса вычисления всех определителей и в численной его неустойчивости. Однако в теоретическом отношении формулы Крамера нередко оказываются полезными, так как позволяют исследовать зависимость решения от элементов матрицы и правой части.

Теорема Кронекера — Капелли формулирует необходимое и достаточное условие разрешимости системы в терминах ранга матрицы. Это не очень удобно, так как не позволяет заметить той глубокой связи, которая существует между линейными системами и уравнениями других типов. В дальнейших исследованиях существенно используется соотношение 5.75, поэтому будем считать, что скалярное произведение введено согласно 5.4, через естественный базис.

6.24. Пусть X, Y — арифметические пространства размерности соответственно m, n и A — матрица размера $n\times m$. Множество векторов $x\in X$, для которых $Ax=0$, называется *ядром* матрицы A и обозначается $\ker A$. Размерность ядра матрицы A иногда называется ее *дефектом*. Множество векторов $y\in Y$, для которых

$y = Ax$ хотя бы для одного вектора $x \in X$, называется *образом* матрицы A и обозначается $\operatorname{im} A$.

6.25. Образ и ядро любой матрицы суть подпространства, при этом всегда имеют место равенства

$$\dim (\operatorname{im} A) = \operatorname{rank} A, \quad \dim (\ker A) = m - \operatorname{rank} A.$$

6.26. Для любой матрицы A и сопряженной к ней матрицы A^* всегда выполняются соотношения

$$\operatorname{rank} A = \operatorname{rank} A^*,$$
$$\dim (\ker A) = \dim (\ker A^*),$$
$$\ker A^* = (\operatorname{im} A)^{\perp}, \quad \ker A = (\operatorname{im} A^*)^{\perp},$$
$$\ker (A^*A) = \ker A, \quad \operatorname{im} (A^*A) = \operatorname{im} A^*,$$
$$\ker (AA^*) = \ker A^*, \operatorname{im} (AA^*) = \operatorname{im} A.$$

6.27. Для арифметических пространств X, Y со скалярным произведением при любой матрице A справедливы разложения

$$X = \ker A \oplus \operatorname{im} A^* = \ker (A^*A) \oplus \operatorname{im} (A^*A),$$
$$Y = \ker A^* \oplus \operatorname{im} A = \ker (AA^*) \oplus \operatorname{im} (AA^*).$$

6.28. Или неоднородная система $Ax = y$ имеет решение при любой правой части, или сопряженная однородная система $A^*u = 0$ имеет по крайней мере одно ненулевое решение.

6.29. (*Альтернатива Фредгольма.*) Или неоднородная система с квадратной матрицей всегда имеет, и притом единственное, решение при любой правой части, или сопряженная однородная система имеет по крайней мере одно ненулевое решение.

6.30. (*Теорема Фредгольма.*) Для того чтобы неоднородная система с квадратной матрицей была разрешима, необходимо и достаточно, чтобы правая часть была ортогональна ко всем решениям сопряженной однородной системы.

6.31. Если система $Ax = y$ совместна, то среди ее решений существует только одно, принадлежащее $\operatorname{im} A^*$. Это решение называется *нормальным*.

6.32. Среди всех решений системы $Ax = y$ нормальное решение является единственным, которое ортогонально ядру матрицы A.

6.33. Среди всех решений системы $Ax = y$ нормальное решение имеет минимальную длину $|x|$.

6.34. Для того чтобы система линейных алгебраических уравнений $Ax = y$ была совместна, необходимо и достаточно, чтобы правая часть y была ортогональна ядру матрицы A^*.

6.35. Система $A^*Ax = A^*y$ совместна при любой матрице A и любой правой части y.

6.36. Если система $Ax = y$ совместна, то она эквивалентна системе $A^*Ax = A^*y$.

6.37. *Псевдорешением* или *обобщенным решением* системы $Ax = y$ называется решение системы $A^*Ax = A^*y$.

6.38. Среди всех векторов пространства X псевдорешения, и только они, обеспечивают:

— ортогональность вектора *невязки* $Ax - y$ образу матрицы A;

— минимальность длины вектора невязки $|Ax - y|$.

6.39. Пусть задана система $Ax = y$. Рассмотрим вектор $\hat{y}$, равный проекции вектора y на образ матрицы A. Система $Ax = \hat{y}$ всегда совместна, и множество ее решений совпадает с множеством псевдорешений исходной системы.

6.40. Для любой системы линейных алгебраических уравнений всегда существует, и притом единственное, нормальное псевдорешение.

6.41. Среди всех псевдорешений системы $Ax = y$ нормальное псевдорешение, и только оно, ортогонально ядру матрицы A.

Таким образом, если разрешимость $Ax = y$ не гарантируется, то мы всегда можем заменить решение этой системы решением системы $A^*Ax = A^*y$. При этом обеспечивается минимизация длины невязки $Ax - y$.

В случае системы с невырожденной матрицей ее решение находится с помощью обратной матрицы. Обратная матрица играет существенную роль при выполнении многих исследований. Однако она была определена лишь для невырожденных матриц, и пока мы не имеем соответствующего аналога для вырожденных и тем более для прямоугольных матриц. Этот аналог может быть построен на основе псевдорешений.

Пусть задана произвольная прямоугольная матрица A. Каждому вектору $y \in Y$ поставим в соответствие вектор $x_0 \in X$, являющийся нормальным псевдорешением системы $Ax = y$. Это соответствие порождает отображение $x_0 = A^+y$ пространства Y в пространство X с помощью некоторого оператора A^+. Принимая во внимание, что псевдорешения линейно зависят от правой части, а нормальное псевдорешение к тому же ортогонально ядру матрицы A, легко показать, что оператор A^+ — линейный. Матрица оператора A^+ и есть аналог обратной матрицы.

6.42. Пусть A — прямоугольная матрица размера $n \times m$ и ранга $r > 0$. Существуют матрицы B, C размеров соответственно $n \times r$, $r \times m$ и ранга r такие, что

$$A = BC = \begin{bmatrix} b_{11} & \dots & b_{1r} \\ b_{21} & & b_{2r} \\ \dots & \dots & \dots \\ b_{n1} & \dots & b_{nr} \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ \dots & \dots & \dots & \dots \\ c_{r1} & c_{r2} & \dots & c_{rm} \end{bmatrix}.$$

Это разложение матрицы A называется *скелетным*.

6.43. В скелетном разложении в качестве столбцов матрицы B можно взять любые базисные столбцы матрицы A. Тогда столбцы матрицы C состоят из коэффициентов линейных комбинаций, с помощью которых выражаются все столбцы матрицы A через базисные.

6.44. Матрица A^+ размера $m \times n$ называется *псевдообратной* или *обобщенной обратной* для матрицы A размера $n \times m$, если выполняются условия

$$AA^+A = A, \quad A^+ = UA^* = A^*V,$$

где U, V — некоторые матрицы.

6.45. Для любой матрицы A не может существовать двух различных псевдообратных матриц.

6.46. Если матрица A размера $n \times m$ есть матрица полного ранга, то матрица A^+ имеет вид

$$A^+ = \begin{cases} (A^*A)^{-1}A^*, & n \geqslant m, \\ A^*(AA^*)^{-1}, & n \leqslant m. \end{cases}$$

6.47. Если матрица A представлена своим скелетным разложением 6.42, то матрица A^+ имеет вид

$$A^+ = C^+B^+ = C^*(CC^*)^{-1}(B^*B)^{-1}B^*.$$

6.48. Матрица, псевдообратная для нулевой, также является нулевой.

6.49. Матрицы A, A^+, A^* связаны между собой соотношениями

$$(A^*)^+ = (A^+)^*, \quad (A^+)^+ = A,$$
$$(AA^+)^* = AA^+, \quad (AA^+)^2 = AA^+,$$
$$(A^+A)^* = A^+A, \quad (A^+A)^2 = A^+A,$$
$$A^+AA^+ = A^+.$$

6.50. (*Уравнения Пенроуза.*) Псевдообратная матрица единственным образом определяется уравнениями

$$AA^+A = A, \qquad A^+AA^+ = A^+,$$
$$(A^+A)^* = A^+A, \quad (AA^+)^* = AA^+.$$

6.51. Для псевдообратной матрицы A^+ и сопряженной матрицы A^* совпадают ядро и образ.

6.52. Всегда имеют место соотношения

$$\mathrm{pr}_{\mathrm{im}A}\, y = AA^+y, \quad \mathrm{pr}_{\mathrm{im}A^*}\, x = A^+Ax.$$

6.53. Псевдорешения системы $Ax = y$, и только они, являются решениями системы $Ax = AA^+y$.

6.54. Для нормального псевдорешения x_0 любой системы линейных алгебраических уравнений $Ax = y$ справедливо равенство $x_0 = A^+y$.

6.55. Рассмотрим систему $Ax = y$ и сопряженную систему $A^*u = v$. Нормальные решения x_0, u_0 этих систем связаны между собой соотношением $(x_0, v) = (y, u_0)$.

Наряду с системой $Ax = y$ в формулировках приведенных фактов неоднократно встречалась и сопряженная система $A^*u = v$, т. е. система с матрицей A^*. Это не является случайным, о чем свидетельствует утверждение 6.55.

Мы хотим обратить внимание на следующее обстоятельство. В практических задачах довольно часто приходится решать системы с одной и той же матрицей A, но с различными правыми частями y. При этом нередко интерес представляют не сами решения, а некоторые линейные функционалы от них. Известно, что линейный функционал может быть представлен как скалярное произведение с фиксированным вектором. Поэтому вместо того, чтобы каждый раз решать систему и затем вычислять линейный функционал, согласно 6.55 можно один раз решить сопряженную систему, а каждый раз вычислять только линейный функционал. Принимая во внимание, что решение системы является исключительно трудоемкой задачей, нетрудно понять преимущество второго подхода.

Остановимся теперь на геометрической интерпретации системы линейных алгебраических уравнений и ее решения. Выражения, которыми задаются отдельные уравнения системы, представляют в декартовых координатах на плоскости и в пространстве соответственно уравнения прямой линии и плоскости. Поэтому решение системы можно трактовать как нахождение общих точек пересечения прямых и плоскостей. Заметим, что как уравнение прямой на плоскости, так и уравнение плоскости в пространстве описываются через скалярные произведения. Все эти факты имеют прямую аналогию и в случае систем общего вида.

6.56. Пусть L — некоторое подпространство линейного пространства K и x_0 — некоторый вектор из K. Множество H векторов $x = x_0 + y$, где y — любой вектор из L, называется *плоскостью* в линейном пространстве. Вектор x_0 называется вектором *сдвига*, подпространство L — *направляющим* подпространством. *Размерностью* плоскости называется размерность ее направляющего подпространства.

6.57. Каждая плоскость однозначно определяет свое направляющее подпространство и неоднозначно — вектор сдвига.

6.58. Для каждой плоскости существует единственный вектор сдвига, ортогональный направляющему подпространству.

6.59. Если пересечение плоскостей содержит вектор x_0, то оно представляет собой плоскость, образованную сдвигом пересечения направляющих подпространств на вектор x_0.

6.60. В линейном пространстве размерности m любая плоскость размерности $m-1$ называется *гиперплоскостью*.

6.61. В пространстве со скалярным произведением любая гиперплоскость может быть задана уравнением $(x, a) = y$, где a — ненулевой вектор, называемый *нормальным*, y — число.

6.62. Систему линейных алгебраических уравнений 6.1 можно рассматривать как систему гиперплоскостей

$$(x, a_1) = y_1, \quad (x, a_2) = y_2, \; \ldots, \; (x, a_n) = y_n$$

в арифметическом пространстве, а решение системы — как пересечение этих гиперплоскостей.

6.63. Общее решение системы линейных алгебраических уравнений представляет собой либо пустое множество, либо плоскость. Размерность этой плоскости равна разности между размерностью пространства и рангом линейной оболочки векторов $a_1, \ldots, a_n$.

6.64. Расстояние между вектором v и гиперплоскостью H, заданной уравнением $(x, a) = y$, равно

$$\rho(v, H) = |(v, a) - y|/|a|.$$

6.65. Пусть система уравнений 6.1 задана системой гиперплоскостей 6.62, причем векторы $a_1, \ldots, a_n$ нормированные. Тогда множество псевдорешений системы 6.1 совпадает с множеством векторов пространства, для которых сумма квадратов расстояний до всех гиперплоскостей 6.62 является минимальной.

6.66. Любой вектор f пространства может быть единственным образом представлен в виде суммы $f = g + h$, где g принадлежит гиперплоскости H, а h ортогонален направляющему подпространству. Вектор g называется *проекцией вектора* f на гиперплоскость H и обозначается $\mathrm{pr}_H f$.

6.67. Если гиперплоскость H задана уравнением $(x, a) = y$, то имеет место равенство

$$\mathrm{pr}_H f = f - \frac{(f, a) - y}{|a|^2} a.$$

6.68. Если совместная система уравнений 6.1 задана системой гиперплоскостей 6.62, то проекция любого вектора f, не являющегося решением, на любую из гиперплоскостей ближе к любому решению, чем вектор f, в смысле расстояния 5.48.

В связи с геометрической интерпретацией системы линейных алгебраических уравнений полезным является введение понятия прямой линии и исследование ее взаимного расположения с гиперплоскостью.

6.69. Плоскость размерности 1 называется *прямой линией.*

6.70. Любая прямая линия есть множество векторов вида $z = x_0 + tq$, где t — числа, x_0, q — заданные векторы. Вектор q называется *направляющим вектором* прямой.

6.71. Пусть заданы прямая линия и гиперплоскость. Возможны три ситуации:

— прямая имеет одну общую точку с гиперплоскостью;

— прямая целиком принадлежит гиперплоскости;

— прямая не имеет общих точек с гиперплоскостью.

В первом случае говорят, что прямая *пересекает* гиперплоскость, во втором — *принадлежит* гиперплоскости, в третьем — *параллельна* гиперплоскости.

В заключение отметим, что проведенные исследования систем линейных алгебраических уравнений позволяют дать конструктивное решение задачи проектирования вектора на подпространство.

6.72. Пусть заданы вектор b и подпространство L. Выберем в подпространстве L любую систему векторов $a_1, a_2, \ldots, a_m$, линейная оболочка которых совпадает с подпространством L.

Предположим, что векторы b и a_i заданы своими координатами в каком-либо ортонормированном базисе. Обозначим через A матрицу, столбцами которой являются векторы a_i. Тогда ортогональная проекция вектора b на подпространство L задается формулой

$$\mathrm{pr}_L\, b = AA^+b.$$

Матрица AA^+ называется *матрицей ортогонального проектирования* или *проектором* на подпространство L.

6.73. Если векторы $a_1, \ldots, a_m$ в 6.72 образуют базис подпространства L, то матрица $A(A^*A)^{-1}A^*$ есть проектор на подпространство L.

Нет необходимости подробно обсуждать проекторы, так как о них, по существу, уже говорилось выше.

§ 7. Матрицы простой структуры

Выбор базисов в линейных пространствах однозначно определяет матрицу линейного оператора. Это позволяет свести исследование различных свойств оператора и операторных соотношений к исследованию аналогичных матричных свойств. Очевидно, что это исследование осуществляется тем эффективнее, чем проще вид матрицы оператора. В общем случае матрица оператора зависит от базисов.

7.1. Пусть $e_1, \ldots, e_m$ и $f_1, \ldots, f_m$ — два базиса одного и того же пространства, причем

$$\begin{aligned} f_1 &= p_{11}e_1 + p_{21}e_2 + \ldots + p_{m1}e_m,\\ f_2 &= p_{12}e_1 + p_{22}e_2 + \ldots + p_{m2}e_m,\\ &\cdot\quad\cdot\quad\cdot\quad\cdot\\ f_m &= p_{1m}e_1 + p_{2m}e_2 + \ldots + p_{mm}e_m. \end{aligned}$$

Матрицей преобразования координат при переходе от базиса $e_1, \ldots, e_m$ к базису $f_1, \ldots, f_m$ называется матрица

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m}\\ p_{21} & p_{22} & \cdots & p_{2m}\\ \cdot & \cdot & \cdot & \cdot\\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}.$$

7.2. Пусть в обозначениях 7.1 x_e и x_f означают вектор-столбцы, составленные из координат вектора x в соответствующих базисах. Всегда выполняется равенство $x_e = Px_f$.

7.3. Для того чтобы матрица P была матрицей некоторого преобразования координат, необходимо и достаточно, чтобы она была невырожденной.

7.4. Если P есть матрица преобразования координат при переходе от базиса $e_1, \ldots, e_m$ к базису $f_1, \ldots, f_m$, то P^{-1} есть матрица преобразования координат при переходе от базиса $f_1, \ldots$ $\ldots, f_m$ к базису $e_1, \ldots, e_m$.

7.5. Пусть P, R, S — матрицы преобразования координат при переходе от базисов $\{e_1, \ldots, e_m\}$, $\{f_1, \ldots, f_m\}$, $\{e_1, \ldots, e_m\}$ соответственно к базисам $\{f_1, \ldots, f_m\}$, $\{r_1, \ldots, r_m\}$, $\{r_1, \ldots, r_m\}$. Всегда выполняется равенство $S = PR$.

7.6. Два базиса одного вещественного пространства называются *одноименными*, если определитель их матрицы преобразования координат положительный.

7.7. Все базисы вещественного пространства можно разбить на два класса одноименных базисов. Один из классов называется *левым*, второй — *правым*.

7.8. Пусть линейный оператор действует из пространства X в пространство Y. Выберем в пространстве X базис $e_1, \ldots, e_m$, в пространстве Y — базис $q_1, \ldots, q_n$ и обозначим через A матрицу оператора в этих базисах. Предположим, что в пространствах X, Y выбраны также другие базисы $f_1, \ldots, f_m$ и $t_1, \ldots, t_n$ соответственно и B есть матрица оператора в новых базисах. Пусть P и Q — матрицы преобразования координат при переходе от $e_1, \ldots, e_m$ к $f_1, \ldots, f_m$ и от $q_1, \ldots, q_n$ к $t_1, \ldots, t_n$. Всегда выполняется равенство $B = Q^{-1}AP$.

Описанная связь матриц оператора указывает направление поиска тех базисов, в которых матрица оператора имеет простейший вид. Решение этой задачи существенно зависит от того, совпадают или не совпадают пространства, задающие область определения и область значений оператора. Если пространства, в которых действует оператор, различны, то простейший вид матрицы оператора находится очень легко. Так же легко он находится и в тех случаях, когда пространства совпадают, но по тем или иным причинам предоставляется возможность выбирать различные базисы для образов и прообразов оператора.

7.9. Две прямоугольные матрицы A и B одинаковых размеров называются *эквивалентными*, если существуют такие две невырожденные матрицы P и Q, что $B = Q^{-1}AP$.

7.10. Признак эквивалентности матриц есть отношение эквивалентности.

7.11. Две матрицы эквивалентны тогда и только тогда, когда они соответствуют одному и тому же линейному оператору в подходящим образом выбранных базисах.

7.12. Для того чтобы две прямоугольные матрицы одинаковых размеров были эквивалентны, необходимо и достаточно, чтобы они имели один и тот же ранг.

7.13. Все эквивалентные матрицы ранга r эквивалентны матрице вида

$$E_r = \begin{bmatrix} 1 & 0 & & 0 & 0 & & 0 \\ 0 & 1 & & 0 & 0 & \ldots & 0 \\ \cdot & \cdot & & \cdot & & & \cdot \\ 0 & 0 & & 1 & 0 & & 0 \\ 0 & 0 & \ldots & 0 & 0 & & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & & 0 & 0 & \cdot & 0 \end{bmatrix}.$$

Столь простой вид матрицы оператора можно получить, в основном, при раздельном выборе базисов для образов и прообразов. Если оператор действует в одном пространстве, то раздельный выбор базисов допускается довольно редко. Совпадение же базисов приводит к совпадению матриц P и Q в 7.8. В этом случае нахождение матрицы оператора простейшего вида становится сложной и трудоемкой задачей.

7.14. Две квадратные матрицы A и B одинаковых размеров называются *подобными*, если существует такая невырожденная матрица P, что $B = P^{-1}AP$.

7.15. Признак подобия матриц есть отношение эквивалентности.

7.16. Две матрицы подобны тогда и только тогда, когда они в разных базисах соответствуют одному и тому же линейному оператору, действующему в одном пространстве.

7.17. Подобные матрицы имеют одинаковые след и определитель.

7.18. Если хотя бы одна из двух квадратных матриц A, B одинакового размера невырождена, то матрицы AB и BA подобны.

Конечно, подобные матрицы всегда эквивалентны. Но если для установления эквивалентности матриц достаточно проверить равенство их рангов, то при исследовании множества подобных матриц приходится привлекать значительно более тонкий аппарат. Основным инструментом теперь являются собственные векторы и собственные значения матрицы.

7.19. Число λ называется *собственным значением* (*собственным числом*) матрицы A, если существует такой ненулевой вектор x, что

$$Ax = \lambda x.$$

Любой вектор $x \neq 0$, удовлетворяющий этому уравнению, называется *собственным вектором* матрицы A, соответствующим собственному значению λ. Совокупность всех собственных значений называется *спектром* матрицы.

7.20. Множество собственных векторов матрицы A, соответствующих одному и тому же собственному значению λ, становится подпространством, если к нему добавить нулевой вектор. Это подпространство называется *собственным подпространством* матрицы A, соответствующим собственному значению λ.

7.21. Для того чтобы ненулевой вектор x был собственным вектором матрицы A, соответствующим собственному значению λ, необходимо и достаточно, чтобы он являлся решением однородной системы линейных алгебраических уравнений $(-A + \lambda E)x = 0$ с матрицей $-A + \lambda E$.

7.22. Для того чтобы число λ было собственным значением матрицы A, необходимо и достаточно, чтобы выполнялось равенство $\det(-A + \lambda E) = 0$.

7.23. Функция $\det(\lambda E - A)$ относительно параметра λ есть многочлен, степень которого равна порядку матрицы A:

$$\det(\lambda E - A) = \lambda^m - a_1\lambda^{m-1} + a_2\lambda^{m-2} + \ldots + (-1)^m a_m.$$

Этот многочлен называется *характеристическим многочленом* матрицы A.

7.24. Коэффициент a_r, $1 \leqslant r \leqslant m$, характеристического многочлена матрицы A равен сумме всех главных миноров порядка r этой матрицы.

7.25. Коэффициент a_1 равен $\operatorname{tr} A$, коэффициент a_m равен $\det A$.

7.26. Корни характеристического многочлена, и только они, образуют спектр матрицы A.

Относительно последнего утверждения следует сделать небольшое замечание. Если матрица A комплексная, то в общем случае комплексными будут как коэффициенты характеристического многочлена, так и его корни. Если же матрица A вещественная, то характеристический многочлен может не иметь ни одного вещественного корня. В этом случае говорят, что матрица не имеет собственных значений, подразумевая под этим, что она не имеет вещественных собственных значений. Такое положение часто возникает тогда, когда матрица описывает оператор, действующий в вещественном пространстве. Характерным примером является матрица поворота на плоскости вокруг начала координат. Однако если рассматривать элементы матрицы как комплексные числа, то она всегда будет иметь собственные значения, вообще говоря, комплексные.

7.27. Для любых квадратных матриц A и B матрицы AB и BA имеют одинаковые характеристические многочлены.

7.28. Любая матрица порядка m имеет m собственных значений, в общем случае комплексных; при этом каждое собственное значение считается столько раз, какова его кратность как корня характеристического многочлена.

7.29. Кратность собственного значения как корня характеристического многочлена называется *алгебраической кратностью* собственного значения или просто его *кратностью*.

7.30. Максимальное число линейно независимых собственных векторов, относящихся к данному собственному значению, называется *геометрической кратностью* собственного значения.

7.31. Геометрическая кратность собственного значения λ матрицы A порядка m равна $m - \operatorname{rank}(\lambda E - A)$.

7.32. Геометрическая кратность любого собственного значения не превосходит его алгебраической кратности.

7.33. Для любой треугольной (нулевой, единичной, скалярной, диагональной) матрицы ее собственные значения совпадают с диагональными элементами.

7.34. Для нулевой, единичной и скалярной матриц любой ненулевой вектор является собственным.

7.35. Для диагональной матрицы любой единичный вектор является собственным.

7.36. Пусть $\lambda_1, \ldots, \lambda_m$ — собственные значения матрицы A. Имеют место соотношения

$$\operatorname{tr} A = \sum_{i=1}^{m} \lambda_i, \quad \det A = \prod_{i=1}^{m} \lambda_i.$$

7.37. Все подобные матрицы имеют одни и те же собственные значения.

7.38. Для того чтобы матрица была невырожденной, необходимо и достаточно, чтобы она не имела нулевого собственного значения.

7.39. Ядро любой матрицы есть собственное подпространство, соответствующее нулевому собственному значению.

7.40. Пусть $\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_s}$ — некоторая совокупность попарно различных собственных значений. Для каждого из собственных значений данной совокупности обязательно существует хотя бы один собственный вектор, и вся система этих собственных векторов линейно независима.

7.41. Матрица A порядка m называется *матрицей простой структуры*, если она имеет m линейно независимых собственных векторов. В противном случае матрица называется *дефектной*.

7.42. Матрица имеет простую структуру тогда и только тогда, когда она подобна диагональной матрице.

7.43. Матрица имеет простую структуру тогда и только тогда, когда для каждого ее собственного значения алгебраическая и геометрическая кратности совпадают.

7.44. Если все собственные значения матрицы попарно различны, то она имеет простую структуру.

7.45. Пусть матрица B имеет попарно различные собственные значения. Любая матрица A, перестановочная с матрицей B, имеет простую структуру.

7.46. Если матрица A имеет простую структуру, то простую структуру будут иметь матрицы A^{-1}, A', A^*, а также присоединенная матрица и все ассоциированные матрицы.

7.47. Для матрицы простой структуры образ есть линейная оболочка собственных векторов, соответствующих ненулевым собственным значениям.

7.48. Пересечение образа и ядра матрицы простой структуры состоит только из нулевого вектора.

7.49. Сумма собственных подпространств любой матрицы является прямой суммой.

7.50. Для матрицы простой структуры, и только для нее, все пространство может быть представлено как прямая сумма ее собственных подпространств.

7.51. Пусть матрица A невырожденная и ее собственные значения суть $\lambda_1, \ldots, \lambda_m$. Тогда собственные значения матрицы A^{-1} равны $\lambda_1^{-1}, \ldots, \lambda_m^{-1}$, а соответствующие собственные векторы совпадают.

7.52. Собственные значения (характеристические многочлены) матриц A и A' совпадают.

7.53. Собственные значения (коэффициенты характеристических многочленов) матриц A и A^* комплексно сопряжены.

7.54. Геометрические (алгебраические) кратности соответствующих собственных значений матриц A и A^* совпадают.

Формулировки некоторых из следующих утверждений связаны с ортогональностью векторов. Ортогональность в них понимается в смысле скалярного произведения вида 5.4 относительно естественного базиса.

7.55. Пусть x есть собственный вектор матрицы A, соответствующий собственному значению λ, y — собственный вектор матрицы A^*, соответствующий собственному значению μ. Если $\lambda \neq \overline{\mu}$, то векторы x и y ортогональны.

7.56. Пусть x есть собственный вектор матрицы A, соответствующий собственному значению λ. Существует такой собственный вектор y матрицы A^*, соответствующий собственному значению $\overline{\lambda}$, что $(x, y) = 1$.

7.57. Для любой системы $x_1, \ldots, x_s$ линейно независимых собственных векторов матрицы A существует биортогональная система $y_1, \ldots, y_s$ собственных векторов матрицы A^*.

7.58. Пусть A — произвольная матрица простой структуры порядка m. Всегда существуют биортогональные базисы $x_1, \ldots$ $\ldots, x_m$ и $y_1, \ldots, y_m$ в пространстве, состоящие из собственных векторов матриц A и A^*, относящихся соответственно к собственным значениям $\lambda_1, \ldots, \lambda_m$ и $\overline{\lambda}_1, \ldots, \overline{\lambda}_m$.

7.59. Если при переходе к некоторому базису матрица A преобразуется в подобную ей матрицу B, то при переходе к биортогональному базису матрица A^* преобразуется в подобную ей матрицу B^*.

7.60. В биортогональных базисах утверждения 7.58 матрицы A и A^* имеют диагональный вид Λ и $\overline{\Lambda}$ соответственно.

7.61. В обозначениях 7.58 матрица $G_i = x_i y_i^*$ ранга 1 называется матрицей, *соответствующей* собственному значению λ_i (*сопутствующей*).

7.62. Соответствующие матрицы обладают следующими свойствами:

$$\sum_{i=1}^{m} G_i = E; \qquad G_i G_k = \begin{cases} 0, & i \neq k, \\ G_i, & i = k. \end{cases}$$

7.63. Для любой матрицы A простой структуры имеет место разложение

$$A = \sum_{i=1}^{m} \lambda_i G_i.$$

7.64. Пусть λ — любое число, не являющееся собственным значением матрицы A простой структуры. Справедливо равенство

$$(\lambda E - A)^{-1} = \sum_{i=1}^{m} \frac{G_i}{\lambda - \lambda_i}.$$

7.65. В условиях и обозначениях 7.58 построим сопутствующие матрицы $G_1, \ldots, G_s$. Предположим, что собственные значения $\lambda_1, \ldots, \lambda_s, \lambda_{s+1}, \ldots, \lambda_m$ упорядочены таким образом, что λ_i, $1 \leqslant i \leqslant s$, соответствует x_i. Матрица

$$A - \sum_{i=1}^{s} \lambda_i G_i$$

имеет собственные значения $0, \ldots, 0, \lambda_{s+1}, \ldots, \lambda_m$.

7.66. Собственные векторы $x_1, \ldots, x_s$ матрицы A являются собственными векторами матрицы 7.65 и соответствуют нулевому собственному значению.

7.67. Любой собственный вектор матрицы A, ортогональный векторам $y_1, \ldots, y_s$, является собственным вектором матрицы 7.65, и наоборот. Каждый из этих собственных векторов соответствует одному из собственных значений $\lambda_{s+1}, \ldots, \lambda_m$.

7.68. Пусть матрица A имеет простую структуру. Это означает, что существуют невырожденная матрица P и диагональная матрица Λ такие, что $\Lambda = P^{-1}AP$. В этом случае столбцы матрицы $P(P^{-1*})$ образуют полную систему собственных векторов матрицы $A(A^*)$, диагональные элементы матрицы $\Lambda(\overline{\Lambda})$ совпадают с соответствующими собственными значениями.

Тот факт, что среди всех матриц выделяются матрицы простой структуры, объясняется очень просто. Эти, и только эти, матрицы в некотором базисе имеют диагональные матрицы. Такой базис может быть составлен лишь из собственных векторов матрицы. Действие любой матрицы простой структуры всегда сводится к «растяжению» координат вектора в данном базисе. Коэффициентами «растяжения» являются соответствующие собственные значения. Если бы все матрицы имели простую структуру, то вопрос о выборе базиса, в котором матрица приводится к простейшему виду, был бы полностью решен. Однако матрицами простой структуры не исчерпываются все матрицы.

§ 8. Инвариантные подпространства

8.1. Подпространство L линейного пространства X называется *инвариантным* относительно квадратной матрицы A, если для каждого вектора x из L его образ Ax также принадлежит L.

8.2. Нулевое подпространство и все пространство являются инвариантными. Они называются *тривиальными* инвариантными подпространствами.

8.3. Любое собственное подпространство является инвариантным.

8.4. Образ и ядро матрицы являются ее инвариантными подпространствами.

8.5. Если матрицы A и B перестановочны, то образ и ядро матрицы B инвариантны относительно матрицы A.

8.6. Если матрица A невырожденная, то A и A^{-1} имеют одни и те же инвариантные подпространства.

8.7. Если матрица A вырожденная, то любое подпространство, содержащее ее образ, является инвариантным.

8.8. Сумма и пересечение любого числа инвариантных подпространств являются инвариантными подпространствами.

8.9. В любом инвариантном подпространстве комплексной матрицы содержится хотя бы один ее собственный вектор, соответствующий, в общем случае, комплексному собственному значению.

Инвариатные подпространства широко используются при выборе базисов, в которых матрица приобретает в каком-либо смысле более простой вид. Вид этих матриц удобно описывать в терминах так называемых *блочных* или *клеточных* матриц. Пусть дана некоторая, вообще говоря, прямоугольная матрица A. Рассечем при помощи горизонтальных и вертикальных линий эту матрицу на прямоугольные блоки:

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1t} \\ A_{21} & A_{22} & \dots & A_{2t} \\ \dots & \dots & \dots & \dots \\ A_{s1} & A_{s2} & \dots & A_{st} \end{bmatrix},$$

каждый из которых представляет собой матрицу $A_{\alpha\beta}$ размера $m_\alpha \times m_\beta$. Про матрицу A будем говорить, что она разбита на *блоки* (или *клетки*). Важно подчеркнуть, что в блочном представлении матрицы все блоки в каждом столбце (строке) имеют одинаковое число столбцов (строк). Наиболее часто используется такое блочное разбиение, при котором диагональные блоки оказываются квадратными. Действия над блочными матрицами производятся по тем же формальным правилам, как и в случае численных матриц. Определенную осторожность необходимо лишь соблюдать из-за некоммутативности перемножения блоков.

Мы будем называть матрицу *блочно диагональной*, *блочно треугольной* и т. п., если она приобретает соответствующий вид при замене ее блоков числами.

8.10. Если матрица имеет блочно треугольный (блочно диагональный) вид, то характеристический многочлен матрицы равен произведению характеристических многочленов всех диагональных блоков.

8.11. Если матрица имеет блочно треугольный (блочно диагональный) вид, то собственные значения матрицы совпадают с множеством собственных значений всех диагональных блоков.

8.12. Представим пространство X размерности m в виде прямой суммы своих подпространств L и M размерностей k и $m-k$. Пусть базис $e_1, \dots, e_m$ в X выбран таким образом, что векторы

$e_1, \ldots, e_k$ принадлежат L, а векторы $e_{k+1}, \ldots, e_m$ принадлежат M. При переходе к этому базису матрица A преобразуется в подобную ей матрицу B, которую представим в блочном виде:

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

где B_{11}, B_{22} — квадратные блоки размеров k, $m-k$. Справедливы следующие утверждения:

— блок B_{21} — нулевой тогда и только тогда, когда подпространство L инвариантно относительно матрицы A;

— блок B_{12} — нулевой тогда и только тогда, когда подпространство M инвариантно относительно матрицы A.

8.13. Для того чтобы в базисе $e_1, \ldots, e_m$ матрица имела верхний блочно треугольный вид с блоками на диагонали размеров $k_1, \ldots, k_s$, необходимо и достаточно, чтобы линейные оболочки векторов $e_1, \ldots, e_p$ образовывали инвариантные подпространства L_p для всех $p = k_1 + \ldots + k_t$, $1 \leqslant t \leqslant s$.

8.14. Для того чтобы в базисе $e_1, \ldots, e_m$ матрица имела блочно диагональный вид с блоками на диагонали размеров $k_1, \ldots, k_s$, необходимо и достаточно, чтобы линейные оболочки векторов $e_r, e_{r+1}, \ldots, e_p$ образовывали инвариантные подпространства L_p для всех r, p таких, что $p = k_1 + \ldots + k_t$, $r = p - k_t + 1$, $1 \leqslant t \leqslant s$.

8.15. В условиях 8.13 инвариантные подпространства L_p удовлетворяют соотношениям вложения

$$L_1 \subset L_2 \subset \ldots \subset L_s = X.$$

8.16. В условиях 8.14 инвариантные подпространства L_p обеспечивают разложение пространства в прямую сумму:

$$L_1 \dotplus L_2 \dotplus \ldots \dotplus L_s = X.$$

8.17. Если хотя бы для одного нетривиального инвариантного подпространства L матрицы A не существует инвариантного подпространства M такого, что $X = L \dotplus M$, то матрица A не имеет простой структуры.

8.18. Для любого λ образ матрицы $\lambda E - A$ является инвариантным подпространством матрицы A.

8.19. Любая комплексная матрица порядка m имеет инвариантное подпространство размерности $m-1$.

8.20. Любая комплексная матрица порядка m имеет вложенные друг в друга инвариантные подпространства L_p размерности p для всех $0 \leqslant p \leqslant m$, т. е.

$$L_0 \subset L_1 \subset L_2 \subset \ldots \subset L_m = X.$$

8.21. Пусть задана произвольная комплексная матрица A порядка m. В соответствии с инвариантными подпространствами 8.20 выберем базис $e_1, \ldots, e_m$ так, чтобы $e_p \in L_p$ для всех p. Тогда в этом базисе матрица A будет иметь правую треугольную форму.

8.22. Любая комплексная матрица подобна треугольной матрице.

8.23. Если вещественная матрица A имеет комплексное собственное значение λ, то она имеет два комплексно сопряженных собственных вектора x, $\bar{x}$, соответствующих собственным значениям λ, $\bar{\lambda}$.

8.24. В условиях и обозначениях 8.23 множество векторов вида $\alpha x + \bar{\alpha}\bar{x}$, где α — произвольное комплексное число, есть вещественное инвариантное подпространство размерности 2 для матрицы A.

8.25. Для любой вещественной матрицы A и любого комплексного λ образ матрицы $(\lambda E - A)(\bar{\lambda} E - A)$ является вещественным инвариантным подпространством матрицы A.

8.26. Если вещественная матрица порядка m имеет вещественное собственное значение, то она имеет вещественные инвариантные подпространства размерностей 1 и $m-1$.

8.27. Если вещественная матрица порядка m имеет комплексное значение, то она имеет вещественные инвариантные подпространства размерностей 2 и $m-2$.

8.28. В любом вещественном инвариантном подпространстве вещественной матрицы содержится либо одномерное, либо двумерное вещественное инвариантное подпространство.

8.29. Любая вещественная матрица порядка m имеет вложенные друг в друга вещественные инвариантные подпространства L_{p_i} размерности p_i:

$$L_0 \subset L_{p_1} \subset L_{p_2} \subset \ldots \subset L_{p_s} = X;$$

при этом разность размерностей соседних подпространств равна 1 или 2.

8.30. Любая вещественная матрица вещественно подобна блочно треугольной матрице с диагональными блоками первого и второго порядков. Блоки второго порядка соответствуют парам комплексно сопряженных собственных значений.

8.31. Если вещественная матрица не имеет комплексных собственных значений, то она вещественно подобна треугольной матрице.

Дальнейшее упрощение вида, к которому можно привести произвольную матрицу подобным преобразованием, связано со специальным выбором инвариантных подпространств. Одним из важнейших способов построения инвариантных подпространств матрицы является использование матричных многочленов.

8.32. Пусть p — натуральное число. Тогда p-й *степенью* A^p квадратной матрицы A называется p-кратное ее произведение.

Если $-p$ есть натуральное число, то $A^p=(A^{-p})^{-1}$. По определению $A^0=E$.

8.33. Пусть A — произвольная квадратная матрица и $\varphi(\lambda)$ — произвольный многочлен

$$\varphi(\lambda)=a_0\lambda^p+\ldots+a_{p-1}\lambda+a_p$$

с комплексными или вещественными коэффициентами. Матрица вида

$$\varphi(A)=a_0A^p+\ldots+a_{p-1}A+a_pE$$

называется *матричным многочленом* или *многочленом от матрицы* A.

8.34. Над матричными многочленами от одной и той же матрицы можно выполнять алгебраические операции, аналогичные операциям над скалярными многочленами.

8.35. Любые два многочлена от одной матрицы перестановочны.

8.36. Множество всех многочленов от одной матрицы есть коммутативное кольцо.

8.37. Образ и ядро любого матричного многочлена $\varphi(A)$ являются инвариантными подпространствами матрицы A.

8.38. Пусть $\lambda_1, \ldots, \lambda_m$ — собственные значения матрицы A, выписанные подряд с учетом кратности. С учетом кратности собственными значениями матрицы $\varphi(A)$ при любом многочлене $\varphi(\lambda)$ являются числа $\varphi(\lambda_1), \ldots, \varphi(\lambda_m)$.

8.39. Любой собственный вектор матрицы A, соответствующий собственному значению λ_i, является собственным вектором матрицы $\varphi(A)$, соответствующим собственному значению $\varphi(\lambda_i)$.

8.40. Если собственное значение матрицы A является (не является) корнем многочлена $\varphi(\lambda)$, то все собственные векторы матрицы A, соответствующие этому собственному значению, принадлежат ядру (образу) матрицы $\varphi(A)$.

8.41. Предположим, что для некоторого многочлена $\varphi(\lambda)$ пространство X может быть разложено в прямую сумму инвариантных подпространств $\ker\varphi(A)$ и $\operatorname{im}\varphi(A)$. Пусть базис в X выбран как объединение базиса $\ker\varphi(A)$ и базиса $\operatorname{im}\varphi(A)$. Обозначим через P матрицу преобразования координат при переходе к этому базису. Матрица $B=P^{-1}AP$ будет блочно диагональной:

$$B=\begin{bmatrix} B_{11} & 0\\ 0 & B_{22}\end{bmatrix},$$

причем матрица $\varphi(B_{11})$ нулевая, а матрица $\varphi(B_{22})$ невырожденная.

8.42. Пусть характеристический многочлен $f(\lambda)$ матрицы A разложен в произведение многочленов $\varphi(\lambda)$ и $\psi(\lambda)$, не имеющих общих корней. Имеют место равенства

$$\ker\varphi(A)=\operatorname{im}\psi(A),\quad \ker\psi(A)=\operatorname{im}\varphi(A).$$

8.43. В обозначениях 8.42 для пространства X справедливо разложение

$$X = \ker \varphi(A) \dotplus \ker \psi(A).$$

8.44. Представим характеристический многочлен $f(\lambda)$ матрицы A порядка m в виде канонического разложения

$$f(\lambda) = (\lambda - \lambda_1)^{k_1} (\lambda - \lambda_2)^{k_2} \ \ldots \ (\lambda - \lambda_r)^{k_r},$$

где $\lambda_1, \ldots, \lambda_r$ — попарно различные собственные значения и $k_1 + \ldots + k_r = m$. Обозначим $\varphi_i(\lambda) = (\lambda - \lambda_i)^{k_i}$; тогда

$$X = \ker \varphi_1(A) \dotplus \ker \varphi_2(A) \dotplus \ldots \dotplus \ker \varphi_r(A).$$

8.45. Инвариантное подпространство $\ker \varphi_i(A)$ называется *корневым подпространством* матрицы A, соответствующим собственному значению. Векторы корневых подпространств называются *корневыми векторами.*

8.46. Размерность корневого подпространства, соответствующего некоторому собственному значению λ_i, равна алгебраической кратности этого собственного значения.

8.47. Корневые векторы, соответствующие попарно различным собственным значениям, линейно независимы.

8.48. Базис пространства X, составленный как последовательное объединение базисов всех корневых подпространств $\ker \varphi_i(A)$, называется *корневым базисом.*

8.49. Обозначим через P матрицу преобразования координат при переходе к какому-нибудь корневому базису. Матрица $B = P^{-1}AP$ будет блочно диагональной:

$$B = \begin{bmatrix} B_{11} & & 0 \\ & \ddots & \\ 0 & & B_{rr} \end{bmatrix}.$$

При этом выполняются следующие свойства:

— совокупность порядков диагональных блоков совпадает с совокупностью чисел $k_1, k_2, \ldots, k_r$;

— совокупность характеристических многочленов диагональных блоков совпадает с совокупностью многочленов $\varphi_i(\lambda)$;

— если характеристический многочлен диагонального блока B_{jj} равен $\varphi_i(\lambda)$, то матрица $\varphi_i(B_{jj})$ — нулевая.

8.50. Любая комплексная матрица подобна блочно диагональной матрице с размерами диагональных блоков, совпадающими с алгебраическими кратностями собственных значений.

8.51. Любая вещественная матрица вещественно подобна блочно диагональной матрице. Размер диагонального блока, соответствующего вещественному собственному значению λ, равен алгебраической кратности λ; если блок соответствует паре комплексно сопряженных собственных значений λ, $\bar{\lambda}$, то его размер равен удвоенной алгебраической кратности λ.

8.52. (*Теорема Кели — Гамильтона.*) Если $f(\lambda)$ есть характеристический многочлен матрицы A, то матрица $f(A)$ — нулевая.

8.53. Если матрица A невырожденная, то $A^{-1} = \varphi(A)$ для некоторого многочлена $\varphi(\lambda)$.

8.54. Матрица A называется *нильпотентной* индекса p, если существует такое целое число $p \geqslant 2$, что $A^p = 0$, но $A^{p-1} \neq 0$.

8.55. Нильпотентная матрица не имеет простую структуру.

8.56. Для того чтобы матрица не простой структуры была нильпотентной, необходимо и достаточно, чтобы все ее собственные значения были нулевыми.

8.57. Ненулевая треугольная матрица с нулевыми диагональными элементами является нильпотентной.

Получение простейшего вида матрицы при подобном преобразовании может осуществляться теперь лишь за счет специального построения базисов каждого из корневых подпространств. Конечно, корневые базисы можно выбрать таким образом, что каждый из диагональных блоков в матрице 8.49 будет треугольным. Однако этот вид матрицы не является самым простым.

8.58. *Высотой* корневого вектора x, соответствующего собственному значению λ матрицы A, называется наименьшее целое неотрицательное число m, для которого $(\lambda E - A)^m x = 0$.

8.59. Все корневые векторы, соответствующие одному собственному значению, имеют высоты, не превосходящие алгебраической кратности собственного значения.

8.60. Корневой вектор высоты 1 является собственным вектором.

8.61. Высоты всех корневых векторов матрицы простой структуры равны 1.

8.62. Если корневой вектор x, соответствующий собственному значению λ матрицы A, имеет высоту k, то вектор $(\lambda E - A)^s x$, $s \leqslant k$, будет корневым вектором высоты $k - s$, соответствующим тому же собственному значению λ.

8.63. Если корневой вектор x, соответствующий собственному значению λ матрицы A, имеет высоту k, то система векторов

$$x,\ (\lambda E - A)x,\ \ldots,\ (\lambda E - A)^{k-1}x$$

является линейно независимой.

8.64. Линейная оболочка корневых векторов 8.63 называется *циклическим подпространством*, порожденным вектором x.

8.65. Любое циклическое подпространство является инвариантным.

8.66. Корневое подпространство разложимо в прямую сумму циклических подпространств.

8.67. Корневой базис, составленный как последовательное объединение базисов 8.63 циклических подпространств, называется *корневым базисом Жордана*.

8.68. *Каноническим ящиком Жордана* называется матрица вида

$$\begin{bmatrix} \lambda & 1 & & & 0 \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ 0 & & & & \lambda \end{bmatrix}.$$

Диагональные элементы этой матрицы равны λ; элементы, расположенные рядом с диагональными справа, равны 1; все остальные элементы равны 0.

8.69. При переходе к корневому базису Жордана матрица приобретает так называемую *каноническую форму Жордана.* Это есть блочно диагональная матрица, составленная из ящиков Жордана.

8.70. С точностью до перестановки ящиков Жордана любая матрица может быть приведена подобным преобразованием к канонической форме Жордана единственного вида.

8.71. Существует подобное преобразование, приводящее матрицу к канонической форме Жордана с любым заданным порядком расположения ящиков Жордана на диагонали.

8.72. Матрицы A и A' подобны.

8.73. Если матрица A блочно диагональная, с блоками $A_{11}, \ldots, A_{rr}$ на диагонали, то говорят, что матрица A есть *прямая сумма* матриц $A_{11}, \ldots, A_{rr}$, и пишут $A = A_{11} \oplus \ldots \oplus A_{rr}$.

8.74. Процесс преобразования матрицы к блочно диагональному виду называется *разложением матрицы* в прямую сумму матриц меньшего размера.

Формулировки следующих утверждений связаны с ортогональностью векторов. Снова ортогональность в них понимается в смысле скалярного произведения вида 5.4 относительно естественного базиса.

8.75. Если некоторое подпространство инвариантно относительно матрицы A, то его ортогональное дополнение инвариантно относительно матрицы A^*.

8.76. (*Теорема Шура.*) Для любой матрицы существует ортонормированный базис, в общем случае комплексный, при переходе к которому матрица будет иметь правую треугольную форму.

8.77. Для любых перестановочных матриц A, B существует ортонормированный базис, в общем случае комплексный, при переходе к которому обе матрицы будут иметь одноименные треугольные формы.

8.78. Корневое подпространство матрицы A, соответствующее собственному значению λ, ортогонально корневому подпространству матрицы A^*, соответствующему собственному значению, не равному $\bar{\lambda}$.

8.79. Для любого многочлена $\varphi(\lambda)$ в обозначениях и условиях 7.63 имеет место разложение

$$\varphi(A)=\sum_{i=1}^{m}\varphi(\lambda_i)G_i.$$

§ 9. λ-матрицы

9.1. Квадратная матрица, элементами которой являются многочлены от λ, называется *λ-матрицей* и обозначается $A(\lambda)$.

9.2. *Степенью* λ-матрицы называется максимальная из степеней многочленов, образующих элементы матрицы.

9.3. λ-матрица $A(\lambda)$ степени l может быть записана в виде

$$A(\lambda)=\lambda^l A_0+\lambda^{l-1}A_1+\ldots+\lambda A_{l-1}+A_l,$$

где $A_0\neq 0$.

9.4. λ-матрица $A(\lambda)$ называется *регулярной*, если $\det A_0\neq 0$.

9.5. *Сумма* λ-матриц есть λ-матрица, степень которой не превосходит максимальной из степеней слагаемых.

9.6. *Произведение* λ-матриц есть λ-матрица, степень которой не превосходит суммы степеней сомножителей и равна этой сумме, если, например, один из сомножителей есть регулярная λ-матрица.

9.7. Произведение регулярных λ-матриц есть регулярная λ-матрица, степень которой равна сумме степеней сомножителей.

9.8. Множество λ-матриц одного порядка, вообще говоря, есть некоммутативное кольцо.

Рассматривая далее любые вопросы, касающиеся λ-матриц, мы будем предполагать, что все λ-матрицы не только квадратные, но и имеют один и тот же порядок. Тем не менее отметим, что многие утверждения остаются в силе и для прямоугольных матриц.

9.9. Пусть $A(\lambda)$ — произвольная λ-матрица, $B(\lambda)$ — регулярная λ-матрица. Тогда существуют единственные λ-матрицы $Q(\lambda)$ и $R(\lambda)$ такие, что

$$A(\lambda)=Q(\lambda)B(\lambda)+R(\lambda),$$

и при этом либо степень $R(\lambda)$ меньше степени $B(\lambda)$, либо $R(\lambda)\equiv 0$. Здесь $Q(\lambda)$ называется *правым частным* $A(\lambda)$ при делении на $B(\lambda)$, $R(\lambda)$ — *правым остатком* $A(\lambda)$ при делении на $B(\lambda)$.

9.10. Пусть $A(\lambda)$ — произвольная λ-матрица, $B(\lambda)$ — регулярная λ-матрица. Тогда существуют единственные λ-матрицы $\hat{Q}(\lambda)$ и $\hat{R}(\lambda)$ такие, что

$$A(\lambda)=B(\lambda)\hat{Q}(\lambda)+\hat{R}(\lambda),$$

и при этом либо степень $\hat{R}(\lambda)$ меньше степени $B(\lambda)$, либо $\hat{R}(\lambda)\equiv 0$. Здесь $\hat{Q}(\lambda)$ называется *левым частным* $A(\lambda)$ при делении на $B(\lambda)$, $\hat{R}(\lambda)$ — *левым остатком* $A(\lambda)$ при делении на $B(\lambda)$.

9.11. Если остаток $R(\lambda)$ $(\hat{R}(\lambda))$ равен нулю, то говорят, что λ-матрица $A(\lambda)$ *делится справа* (*слева*) на λ-матрицу $B(\lambda)$.

9.12. *Правым значением* $A(B)$ λ-матрицы $A(\lambda)$ на числовой матрице B называется матрица

$$A(B) = A_0B^l + A_1B^{l-1} + \ldots + A_l.$$

Левым значением $\hat{A}(B)$ λ-матрицы $A(\lambda)$ на B называется матрица

$$\hat{A}(B) = B^lA_0 + B^{l-1}A_1 + \ldots + A_l.$$

9.13. Для любой числовой матрицы B матрица $\lambda E - B$ есть регулярная матрица степени 1.

9.14. Правым и левым остатками λ-матрицы $A(\lambda)$ при делении на $\lambda E - B$ являются матрицы $A(B)$ и $\hat{A}(B)$ соответственно.

9.15. λ-матрица $A(\lambda)$ делится справа (слева) на $\lambda E - B$ тогда и только тогда, когда $A(B)$ $(\hat{A}(B))$ есть нулевая матрица.

9.16. Пусть A — квадратная матрица порядка m. Матрица $B(\lambda)$, присоединенная для λ-матрицы $\lambda E - A$, есть λ-матрица степени $m - 1$.

9.17. Если $f(\lambda)$ есть характеристический многочлен матрицы A, то имеют место равенства

$$(\lambda E - A)B(\lambda) = B(\lambda)(\lambda E - A) = f(\lambda)E.$$

9.18. λ-матрица $f(\lambda)E$ делится справа и слева на λ-матрицу $\lambda E - A$.

9.19. (*Теорема Кели — Гамильтона.*) Если $f(\lambda)$ есть характеристический многочлен матрицы A, то матрица $f(A)$ — нулевая.

9.20. Пусть заданы произвольная квадратная матрица A порядка n и произвольный многочлен $\varphi(\lambda)$. Либо существует многочлен $\psi(\lambda)$ степени, меньшей n, для которого $\varphi(A) = \psi(A)$, либо $\varphi(A) = 0$.

9.21. Скалярный многочлен $\psi(\lambda)$ называется *аннулирующим* многочленом для матрицы A, если $\psi(A) = 0$.

9.22. Характеристический многочлен является аннулирующим.

9.23. Многочлен называется *приведенным*, если коэффициент при старшей степени переменной равен 1.

9.24. Приведенный аннулирующий для матрицы A многочлен наименьшей степени называется *минимальным* многочленом матрицы A.

9.25. Любой аннулирующий многочлен матрицы A делится на ее минимальный многочлен.

9.26. Для любой матрицы ее минимальный многочлен единствен.

9.27. Подобные матрицы имеют один и тот же минимальный многочлен.

9.28. Если все собственные значения матрицы различны, то ее характеристический и минимальный многочлены совпадают.

9.29. Минимальный многочлен блочно диагональной матрицы равен наименьшему общему кратному минимальных многочленов ее диагональных блоков.

9.30. Пусть A — квадратная матрица порядка m. Обозначим через $d_{m-1}(\lambda)$ наибольший общий делитель всех миноров порядка $m-1$ λ-матрицы $\lambda E - A$. Многочлен $d_{m-1}(\lambda)$ является делителем характеристического многочлена.

9.31. Минимальный многочлен матрицы A равен частному от деления ее характеристического многочлена $f(\lambda)$ на многочлен $d_{m-1}(\lambda)$.

9.32. Пусть задана матрица A. Скалярный многочлен $\psi(\lambda)$ называется *аннулирующим вектор x* многочленом, если $\psi(A)x = 0$.

9.33. Приведенный аннулирующий вектор x многочлен наименьшей степени называется *минимальным аннулирующим вектор x* многочленом.

9.34. Степень минимального аннулирующего вектор x многочлена равна размерности линейной оболочки векторов x, Ax, A^2x, ...

9.35. Для любого вектора x минимальный многочлен матрицы A делится на минимальный аннулирующий вектор x многочлен.

9.36. Для любого делителя минимального многочлена матрицы A найдется вектор x, для которого этот делитель будет минимальным аннулирующим вектор x многочленом.

9.37. Пусть $\lambda_1, \ldots, \lambda_r$ — попарно различные собственные значения матрицы A, $s_1, \ldots, s_r$ — максимальные высоты соответствующих им корневых векторов. Тогда минимальный многочлен матрицы A имеет вид

$$(\lambda - \lambda_1)^{s_1} (\lambda - \lambda_2)^{s_2} \ldots (\lambda - \lambda_r)^{s_r},$$

причем $s_i \leqslant k_i$ для всех i (см. 8.44).

9.38. Разложим вектор x по векторам корневого базиса, и пусть $l_1, \ldots, l_r$ — высоты соответствующих корневых векторов, определяющих это разложение. Тогда минимальный аннулирующий вектор x многочлен имеет вид

$$(\lambda - \lambda_1)^{l_1} (\lambda - \lambda_2)^{l_2} \ldots (\lambda - \lambda_r)^{l_r},$$

причем $l_i \leqslant s_i$ для всех i.

Приведение матрицы A подобным преобразованием к тому или иному простейшему виду, вообще говоря, связано с разложением минимального многочлена матрицы A на множители. Однако эта же задача оказывается тесно связанной с приведением λ-матрицы $\lambda E - A$ к некоторому простому виду с помощью так называемых элементарных преобразований. При этом многие факты для произвольных λ-матриц устанавливаются так же легко, как и для матриц $\lambda E - A$.

9.39. *Элементарными операциями* над λ-матрицами называются следующие операции над ее строками и столбцами:

— умножение любой i-й строки (столбца) на ненулевое число d_i;

— прибавление к любой i-й строке (столбцу) любой другой j-й строки (столбца), умноженной на произвольный многочлен $\alpha_{ij}(\lambda)$;

— перестановка любых двух i-й и j-й строк (столбцов).

9.40. Реализация любой из описанных элементарных операций над строками (столбцами) λ-матрицы равносильна умножению λ-матрицы слева (справа) на некоторую матрицу S, также называемую *элементарной.* Все элементарные λ-матрицы лишь несколькими элементами s_{kl} отличаются от единичной матрицы. Для соответствующих операций 9.39 эти элементы таковы:

$$s_{ii}=\alpha_i \ (s_{ii}=\alpha_i); \qquad s_{ij}=\alpha_{ij}(\lambda) \quad (s_{ji}=\alpha_{ij}(\lambda));$$
$$s_{ii}=s_{jj}=0, \quad s_{ij}=s_{ji}=1 \ (s_{ii}=s_{jj}=0,\ s_{ij}=s_{ji}=1).$$

9.41. Определитель любой элементарной λ-матрицы не зависит от λ и не равен нулю.

9.42. λ-матрица, определитель которой не зависит от λ и не равен нулю, называется *унимодулярной.*

9.43. Унимодулярные λ-матрицы, и только они, могут быть представлены в виде произведения конечного числа элементарных λ-матриц.

9.44. Если λ-матрица $Q(\lambda)$ является унимодулярной, то матрица $Q^{-1}(\lambda)$ также будет унимодулярной λ-матрицей.

9.45. Две λ-матрицы называются *эквивалентными,* если одна из них может быть получена из другой с помощью последовательности элементарных операций.

9.46. Признак эквивалентности λ-матриц есть отношение эквивалентности.

9.47. λ-матрицы $A(\lambda)$ и $B(\lambda)$ эквивалентны тогда и только тогда, когда существуют такие унимодулярные λ-матрицы $P(\lambda)$ и и $Q(\lambda)$, что $B(\lambda)=Q^{-1}(\lambda)A(\lambda)P(\lambda)$.

9.48. Порядок наибольшего минора, не равного нулевому многочлену, называется *рангом* λ-матрицы.

9.49. Эквивалентные λ-матрицы имеют один и тот же ранг.

9.50. Пусть λ-матрица $A(\lambda)$ имеет ранг r. Обозначим через $d_j(\lambda)$, $j \leqslant r$, приведенный наибольший общий делитель всех миноров j-го порядка матрицы $A(\lambda)$. Будем считать, что $d_0(\lambda)\equiv 1$. Тогда многочлен $d_j(\lambda)$ делится на $d_{j-1}(\lambda)$ для всех $j \leqslant r$.

9.51. Эквивалентные λ-матрицы имеют одни и те же многочлены $d_j(\lambda)$ для всех $j \leqslant r$.

9.52. *Инвариантными многочленами* λ-матрицы $A(\lambda)$ ранга r называются приведенные многочлены

$$i_1(\lambda)=\frac{d_1(\lambda)}{d_0(\lambda)}, \quad i_2(\lambda)=\frac{d_2(\lambda)}{d_1(\lambda)}, \quad \ldots, \quad i_r(\lambda)=\frac{d_r(\lambda)}{d_{r-1}(\lambda)}.$$

9.53. Степени инвариантных многочленов не убывают с ростом их номеров.

9.54. Каждый из инвариантных многочленов делится на предыдущий инвариантный многочлен.

9.55. Для того чтобы λ-матрицы были эквивалентны, необходимо и достаточно, чтобы они имели одни и те же инвариантные многочлены.

9.56. Любая λ-матрица $A(\lambda)$ ранга r эквивалентна диагональной матрице с элементами $i_1(\lambda), \ldots, i_r(\lambda), 0, \ldots, 0$. Эта диагональная матрица называется *канонической формой Смита* для $A(\lambda)$.

9.57. Пусть λ-матрица $A(\lambda)$ эквивалентна диагональной матрице с элементами $e_1(\lambda), \ldots, e_r(\lambda), 0, \ldots, 0$, где многочлены e_j приведенные и каждый из многочленов e_j делится на многочлен e_{j-1}. Тогда $e_j(\lambda) = i_j(\lambda)$ для всех j.

9.58. Для унимодулярных λ-матриц, и только для них, каноническая форма Смита совпадает с единичной матрицей.

9.59. Для любой матрицы A порядка m ранг λ-матрицы $\lambda E - A$ равен m.

9.60. Для любой матрицы A порядка m сумма степеней инвариантных многочленов λ-матрицы $\lambda E - A$ равна m.

9.61. Характеристический многочлен матрицы A порядка m равен произведению всех инвариантных многочленов λ-матрицы $\lambda E - A$.

9.62. Минимальный многочлен матрицы A порядка m совпадает с инвариантным многочленом $i_m(\lambda)$ λ-матрицы $\lambda E - A$.

9.63. Матрицы A и B подобны тогда и только тогда, когда λ-матрицы $\lambda E - A$ и $\lambda E - B$ эквивалентны.

9.64. Пусть J — канонический ящик Жордана порядка m с диагональными элементами μ. Инвариантные многочлены λ-матрицы $\lambda E - J$ равны $1, \ldots, 1, (\lambda - \mu)^m$.

9.65. Пусть $N(\lambda)$ — диагональная λ-матрица с элементами $n_1(\lambda), \ldots, n_s(\lambda)$, представляющими собой попарно взаимно простые приведенные многочлены. Инвариантные многочлены этой матрицы равны $1, \ldots, 1, \prod_{i=1}^{s} n_i(\lambda)$.

9.66. Рассмотрим произвольный скалярный многочлен

$$f(\lambda) = \lambda^m + a_1\lambda^{m-1} + \ldots + a_{m-1}\lambda + a_m$$

и построим по его коэффициентам квадратную матрицу

$$F = \begin{bmatrix} -a_1 & -a_2 & & -a_{m-1} & -a_m \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & & 1 & 0 \end{bmatrix}.$$

Эта матрица называется *матрицей Фробениуса*.

9.67. Если F есть матрица Фробениуса для многочлена $f(\lambda)$, то она обладает следующими свойствами:

— характеристический многочлен матрицы F равен $f(\lambda)$;

— минимальный многочлен матрицы F равен $f(\lambda)$;

— инвариантные многочлены λ-матрицы $\lambda E - F$ равны $1, \ldots, 1, f(\lambda)$;

— если λ_i есть корень многочлена $f(\lambda)$, то независимо от кратности ему соответствует только один собственный вектор, и его координаты равны $1, \lambda_i, \lambda_i^2, \ldots, \lambda_i^{m-1}$.

9.68. Пусть λ-матрица $\lambda E - A$ имеет инвариантные многочлены $i_r(\lambda), \ldots, i_m(\lambda)$ ненулевой степени. Тогда матрица A подобна блочно диагональной матрице с блоками $F_r, \ldots, F_m$, где F_j — матрица Фробениуса для многочлена $i_j(\lambda)$ при $r \leqslant j \leqslant m$. Эта блочно диагональная матрица называется *канонической формой Фробениуса* для матрицы A.

9.69. Пусть $\lambda_1, \lambda_2, \ldots, \lambda_r$ — попарно различные собственные значения матрицы A. Представим инвариантные многочлены λ-матрицы $\lambda E - A$ в виде их канонических разложений:

$$i_1(\lambda) = (\lambda - \lambda_1)^{\alpha_{11}} (\lambda - \lambda_2)^{\alpha_{12}} \ldots (\lambda - \lambda_r)^{\alpha_{1r}},$$
$$i_2(\lambda) = (\lambda - \lambda_1)^{\alpha_{21}} (\lambda - \lambda_2)^{\alpha_{22}} \ldots (\lambda - \lambda_r)^{\alpha_{2r}},$$
$$\cdots \quad \cdots \quad \cdots \quad \cdots$$
$$i_m(\lambda) = (\lambda - \lambda_1)^{\alpha_{m1}} (\lambda - \lambda_2)^{\alpha_{m2}} \ldots (\lambda - \lambda_r)^{\alpha_{mr}}.$$

Имеют место соотношения

$$0 \leqslant \alpha_{1s} \leqslant \alpha_{2s} \leqslant \ldots \leqslant \alpha_{ms} \leqslant k_s, \quad \sum_{j=1}^{m} \alpha_{js} = k_s$$

для всех $1 \leqslant s \leqslant r$ (см. 8.44).

9.70. Любой множитель $(\lambda - \lambda_s)^{\alpha_{js}}$, входящий в разложения 9.69 при $\alpha_{js} > 0$, называется *элементарным делителем* матрицы A. Элементарный делитель, для которого $\alpha_{js} = 1$, называется *линейным;* в противном случае он называется *нелинейным.*

9.71. Пусть J — канонический ящик Жордана порядка m с диагональными элементами μ. Совокупность элементарных делителей этой матрицы состоит из одного многочлена $(\lambda - \mu)^m$, равного ее характеристическому многочлену.

9.72. Канонический ящик Жордана порядка m с диагональными элементами μ подобен матрице Фробениуса для многочлена $(\lambda - \mu)^m$.

9.73. Матрицы A и B подобны тогда и только тогда, когда они имеют одни и те же элементарные делители.

9.74. Пусть $e_1(\lambda), \ldots, e_p(\lambda)$ — элементарные делители матрицы A. Тогда матрица A подобна блочно диагональной матрице с диагональными блоками $F_1, \ldots, F_p$, где F_s — матрица Фробениуса для многочлена $e_s(\lambda)$ при $1 \leqslant s \leqslant p$.

9.75. Обозначим через J_s ящик Жордана с характеристическим многочленом $e_s(\lambda)$. Матрица A подобна блочно диагональной матрице с диагональными блоками $J_1, \ldots, J_p$. Эта блочно диагональная матрица называется *канонической формой Жордана* (см. 8.68—8.71).

9.76. Матрица A имеет простую структуру тогда и только тогда, когда все ее элементарные делители линейные.

9.77. Матрица A имеет простую структуру тогда и только тогда, когда ее минимальный многочлен имеет лишь простые корни.

9.78. Пусть $\varphi(\lambda)$ — минимальный многочлен матрицы A, $f(\lambda)$ — любой многочлен. Матрица $f(A)$ будет невырожденной в том и только в том случае, когда $\varphi(\lambda)$ и $f(\lambda)$ взаимно просты.

§ 10. Нормальные матрицы

10.1. Матрица A называется *нормальной*, если она перестановочна со своей сопряженной матрицей A^*, т. е. $AA^* = A^*A$.

10.2. Блочно треугольная нормальная матрица является блочно диагональной.

10.3. Диагональная матрица является нормальной.

10.4. Пусть A — произвольная матрица, α, β — равные по модулю комплексные числа. Тогда матрица $\alpha A + \beta A^*$ будет нормальной.

10.5. Пусть A — нормальная матрица. Для любого многочлена $\varphi(\lambda)$ матрица $\varphi(A)$ будет нормальной.

Мы уже неоднократно отмечали ортогональность различных векторов, связанных с матрицами A и A^*. Эта ортогональность особенно характерна для нормальных матриц. Опять будем понимать ее в смысле скалярного произведения вида 5.4 относительно естественного базиса, если не сделано какой-либо специальной оговорки. Начиная изучение нормальных матриц, полезно вспомнить утверждения 8.76 и 5.75.

10.6. Для того чтобы матрица была нормальной, необходимо и достаточно, чтобы она имела базисную систему ортонормированных собственных векторов.

10.7. Если матрица A нормальная, то матрицы A и A^* имеют одинаковые системы собственных векторов.

10.8. Если матрица A нормальная, то собственные значения матриц A и A^*, соответствующие общему собственному вектору, комплексно сопряжены.

10.9. Если матрица A нормальная, то в ортонормированном базисе из ее собственных векторов матрицы A и A^* одновременно принимают диагональный вид Λ и $\overline{\Lambda}$ соответственно.

10.10. Нормальная матрица всегда имеет простую структуру.

10.11. Пусть матрица A — нормальная и $x_1, \ldots, x_m$ — ее ортонормированные собственные векторы, соответствующие собственным значениям $\lambda_1, \ldots, \lambda_m$. Построим сопутствующие матрицы

$G_i = x_i x_i^*$ (см. 7.61) для всех i. Имеют место разложения

$$A = \sum_{i=1}^{m} \lambda_i G_i, \quad A^* = \sum_{i=1}^{m} \bar{\lambda}_i G_i.$$

10.12. Если матрица A нормальная, то в обозначениях 10.11 при $1 \leqslant s \leqslant m$ будет нормальной матрица (см. 7.65)

$$A - \sum_{i=1}^{s} \lambda_i G_i.$$

10.13. Матрица A является нормальной тогда и только тогда, когда для любого инвариантного подпространства L ортогональное дополнение $L^{\perp}$ также инвариантное.

10.14. Любое подпространство, инвариантное относительно нормальной матрицы A, инвариантно относительно A^*.

10.15. Для любой нормальной матрицы A в унитарном пространстве X имеют место разложения

$$X = \ker A \oplus \operatorname{im} A = \ker A \oplus (\ker A^*)^{\perp}.$$

10.16. Перестановочные нормальные матрицы имеют общую базисную систему ортонормированных собственных векторов.

10.17. Если A — нормальная матрица порядка m с собственными значениями $\lambda_1, \ldots, \lambda_m$, то существуют такие многочлены $p(\lambda)$, $q(\lambda)$, что

$$A^* = p(A), \quad A = q(A^*);$$

при этом $p(\lambda_i) = \overline{\lambda}_i$, $q(\overline{\lambda}_i) = \lambda_i$ для всех i.

10.18. Если существует хотя бы один из многочленов $p(\lambda)$, $q(\lambda)$, для которого выполняется соответствующее равенство 10.17, то матрица A — нормальная.

10.19. Комплексная матрица U называется *унитарной*, если сопряженная матрица U^* совпадает с обратной U^{-1}, т. е.

$$UU^* = U^*U = E.$$

10.20. Унитарная матрица является нормальной.

10.21. Среди нормальных матриц унитарная матрица выделяется тем и только тем, что все ее собственные значения по модулю равны 1.

10.22. Матрица U является унитарной в том и только в том случае, когда выполняется хотя бы одно (следовательно, все) из следующих условий:

— столбцы матрицы U, рассматриваемые как векторы унитарного арифметического пространства, образуют ортонормированную систему;

— строки матрицы U, рассматриваемые как векторы унитарного арифметического пространства, образуют ортонормированную систему;

— для любых двух векторов их скалярное произведение равно скалярному произведению их образов;

— для любых двух ортогональных векторов их образы ортогональны, и хотя бы для одного ненулевого вектора его длина равна длине его образа;

— для любого вектора его длина равна длине его образа;

— для любых двух векторов какого-нибудь базиса их скалярное произведение равно скалярному произведению их образов;

— образы векторов любого ортонормированного базиса образуют также ортонормированный базис;

— образы векторов хотя бы одного ортонормированного базиса образуют ортонормированный базис;

— матрица U является матрицей преобразования координат при переходе от ортонормированного базиса к ортонормированному;

— матрица U^* является матрицей преобразования координат при переходе от ортонормированного базиса к ортонормированному.

10.23. Любая матрица перестановок — унитарная.

10.24. Унитарные матрицы одного порядка образуют группу по умножению.

10.25. Определитель унитарной матрицы по модулю равен 1.

10.26. Любой элемент унитарной матрицы равен по модулю своему дополнительному минору.

10.27. Сумма квадратов модулей всех миноров порядка k, выбранных из произвольных k строк (столбцов) унитарной матрицы, равна 1.

10.28. Пусть ведущий минор порядка k унитарной матрицы порядка m по модулю равен 1. Тогда матрица является блочно диагональной с блоками порядков k и $m-k$.

10.29. Матрицы A и B называются *унитарно подобными*, если существует такая унитарная матрица U, что $B = U^*AU$.

10.30. Признак унитарного подобия на множестве квадратных матриц одного порядка есть отношение эквивалентности.

10.31. Любая комплексная матрица унитарно подобна треугольной матрице (см. 8.76).

10.32. Правая (левая) треугольная матрица унитарно подобна левой (правой) треугольной матрице.

10.33. При унитарно подобном преобразовании нормальная матрица переходит в нормальную.

10.34. При унитарно подобном преобразовании унитарная матрица переходит в унитарную.

10.35. Комплексные нормальные матрицы, и только они, унитарно подобны диагональным матрицам.

10.36. Вещественная унитарная матрица называется *ортогональной*.

Так как ортогональная матрица является унитарной, то почти все свойства унитарной матрицы переносятся на ортогональную матрицу. При этом меняется лишь терминология: «унитарное пространство» заменяется на «евклидово пространство», а слова «комплексный», «унитарный» и символ A^* — соответственно на слова «вещественный», «ортогональный» и символ A'. Некоторое отличие в формулировках утверждений может появиться только в тех случаях, когда по тем или иным причинам необходимо использовать лишь вещественные величины и преобразования.

10.37. Пусть $U = P + iQ$ — комплексная унитарная матрица порядка m с вещественными матрицами P, Q. Вещественная блочная матрица $D = \begin{bmatrix} P & -Q \\ Q & P \end{bmatrix}$ порядка $2m$ является ортогональной.

10.38. Пусть $A = B + iC$ — комплексная нормальная матрица порядка m с вещественными матрицами B, C. Вещественная блочная матрица $F = \begin{bmatrix} B & -C \\ C & B \end{bmatrix}$ порядка $2m$ является нормальной.

10.39. Любая вещественная матрица ортогонально подобна блочно треугольной матрице с диагональными блоками первого и второго порядков.

10.40. Вещественные нормальные матрицы ортогонально подобны блочно диагональным матрицам с диагональными блоками первого и второго порядков.

10.41. Комплексная матрица H называется *эрмитовой* или *самосопряженной*, если она совпадает со своей сопряженной матрицей H^*.

10.42. Эрмитова матрица является нормальной.

10.43. Среди нормальных матриц эрмитова матрица выделяется тем и только тем, что все ее собственные значения — вещественные.

10.44. Матрица H является эрмитовой в том и только в том случае, когда выполняется хотя бы одно (следовательно, все) из следующих условий:

— элементы h_{ij} матрицы H удовлетворяют равенствам $h_{ij} = \overline{h}_{ji}$ для всех i, j (диагональные элементы вещественные!);

— для любых комплексных векторов x, y справедливо равенство $(Hx, y) = (x, Hy)$;

— для любого комплексного вектора x скалярное произведение (Hx, x) есть вещественное число;

— матрица H унитарно подобна вещественной диагональной матрице.

10.45. Определитель эрмитовой матрицы есть вещественное число.

10.46. Произведение эрмитовых матриц есть эрмитова матрица тогда и только тогда, когда эти матрицы перестановочны.

10.47. Матрица, обратная к невырожденной эрмитовой матрице, есть эрмитова матрица.

10.48. При унитарно подобном преобразовании эрмитова матрица переходит в эрмитову.

10.49. Произвольную квадратную матрицу A всегда можно представить в виде суммы, называемой *эрмитовым разложением*:

$$A = H_1 + iH_2,$$

где матрицы H_1, H_2 эрмитовы. Эти матрицы называются *эрмитовыми компонентами* матрицы A. Они определяются однозначно, причем

$$H_1 = \frac{1}{2}(A + A^*), \quad H_2 = \frac{1}{2i}(A - A^*).$$

10.50. Если матрица A нормальная, то собственные значения матрицы H_1 (H_2) из разложения 10.49 являются вещественными (мнимыми) частями собственных значений матрицы A.

10.51. Матрица A является нормальной тогда и только тогда, когда ее эрмитовы компоненты перестановочны.

10.52. Комплексная матрица S называется *косоэрмитовой* или *кососамосопряженной*, если она совпадает с матрицей $-S^*$.

10.53. Косоэрмитова матрица является нормальной.

10.54. Среди нормальных матриц косоэрмитова матрица выделяется тем и только тем, что все ее собственные значения — чисто мнимые.

10.55. Матрица S является косоэрмитовой тогда и только тогда, когда матрица iS эрмитова.

10.56. Вещественная эрмитова (косоэрмитова) матрица называется *симметричной* (*кососимметричной*).

Снова есть очень много общего между этими вещественными матрицами и соответствующими комплексными матрицами. Конечно, здесь также могут появиться некоторые отличия в формулировках утверждений, если необходимо использовать лишь вещественные величины и преобразования.

10.57. Элементы симметричной (кососимметричной) матрицы A удовлетворяют соотношениям $a_{ij} = a_{ji}$ ($a_{ij} = -a_{ji}$) для всех i, j. В частности, все диагональные элементы кососимметричной матрицы равны нулю.

10.58. Симметричная матрица ортогонально подобна вещественной диагональной матрице.

10.59. Кососимметричная матрица ортогонально подобна вещественной блочно диагональной матрице с блоками первого и второго порядков. Все блоки первого порядка — нулевые, все блоки второго порядка — кососимметричные.

10.60. Определитель кососимметричной матрицы нечетного порядка равен нулю.

10.61. Кососимметричные матрицы H, и только они, обладают тем свойством, что для любого вещественного вектора x скалярное произведение (Hx, x) равно нулю.

10.62. Вещественную квадратную матрицу A всегда можно представить в виде суммы

$$A = B + C,$$

где матрица B — симметричная, C — кососимметричная. Это разложение единственно, причем

$$B = \frac{1}{2}(A + A'), \quad C = \frac{1}{2}(A - A').$$

Компонента B называется *симметричной* составляющей матрицы A, компонента C — *кососимметричной* составляющей.

10.63. Для любой матрицы A и разложения 10.49 справедливо равенство

$$(Ax, x) = (H_1x, x) + i(H_2x, x).$$

10.64. Для любой вещественной матрицы A, любого вещественного вектора x и разложения 10.62 справедливо равенство

$$(Ax, x) = (Bx, x).$$

10.65. Пусть A — произвольная прямоугольная комплексная матрица, Q, R — унитарные матрицы. Имеет место равенство

$$(QAR)^+ = R^*A^+Q^*.$$

§ 11. Мультипликативные представления матриц

11.1. Если прямоугольная матрица A представлена в клеточном виде:

$$A = \begin{bmatrix} B & Q \\ R & T \end{bmatrix},$$

где B — квадратная невырожденная матрица порядка r, то ранг матрицы A равен r в том и только в том случае, когда $T = RB^{-1}Q$.

11.2. Если A — прямоугольная матрица ранга r, то существуют такие матрицы перестановок P, H, что у матрицы PAH отличны от нуля ведущие миноры всех порядков от 1 до r.

11.3. Если матрица A имеет вид 11.1, то (см. 6.44)

$$A^+ = \begin{bmatrix} B^* \\ Q^* \end{bmatrix} [BB^* + QQ^*]^{-1} B [B^*B + R^*R]^{-1} [B^*, R^*].$$

11.4. Если ранг прямоугольной матрицы A равен числу строк (столбцов), то существует такая матрица перестановок P (H), что у матрицы PA (AH) отличны от нуля ведущие миноры всех возможных порядков.

11.5. Любую квадратную матрицу A порядка m, у которой отличны от нуля ведущие миноры всех порядков от 1 до $m-1$, можно представить в виде произведения левой треугольной матрицы L на правую треугольную матрицу U. Если

$$A = LU = \begin{bmatrix} l_{11} & & & \\ l_{21} & l_{22} & & 0 \\ \cdot & \cdot & \cdot & \\ l_{m1} & l_{m2} & \dots & l_{mm} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & & u_{1m} \\ & u_{22} & & u_{2m} \\ 0 & & & \cdot \\ & & & u_{mm} \end{bmatrix},$$

то (см. 4.18)

$$l_{11}u_{11} = A\begin{pmatrix}1\\1\end{pmatrix}, \quad l_{22}u_{22} = \frac{A\begin{pmatrix}1 & 2\\1 & 2\end{pmatrix}}{A\begin{pmatrix}1\\1\end{pmatrix}}, \quad \dots, \quad l_{mm}u_{mm} = \frac{A\begin{pmatrix}1 & 2 & \dots & m\\1 & 2 & \dots & m\end{pmatrix}}{A\begin{pmatrix}1 & 2 & \dots & m-1\\1 & 2 & \dots & m-1\end{pmatrix}},$$

$$l_{sk} = l_{kk}\frac{A\begin{pmatrix}1 & 2 & \dots & k-1 & s\\1 & 2 & \dots & k-1 & k\end{pmatrix}}{A\begin{pmatrix}1 & 2 & \dots & k\\1 & 2 & \dots & k\end{pmatrix}}; \quad u_{ks} = u_{kk}\frac{A\begin{pmatrix}1 & 2 & \dots & k-1 & k\\1 & 2 & \dots & k-1 & s\end{pmatrix}}{A\begin{pmatrix}1 & 2 & \dots & k\\1 & 2 & \dots & k\end{pmatrix}},$$

где $s = k+1, \dots, m$, $k = 1, 2, \dots, m-1$.

11.6. Разложение 11.5 единственно, если зафиксировать диагональные элементы матрицы L или матрицы U.

11.7. Разложение 11.5 с диагональными элементами матрицы L, равными 1, называется *LU-разложением* матрицы A.

11.8. Если имеет место разложение 11.5, то для всех k справедливо равенство

$$A\begin{pmatrix}1 & 2 & \dots & k\\1 & 2 & \dots & k\end{pmatrix} = \prod_{i=1}^{k} l_{ii}u_{ii}.$$

11.9. Если матрица A имеет положительные ведущие миноры, то ее LU-разложение существует и матрица U имеет положительные диагональные элементы.

11.10. Если матрица A эрмитова и имеет положительные ведущие миноры, то существует разложение $A = LL^*$, где L — левая треугольная матрица.

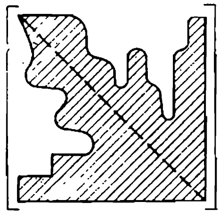
11.11. Разложение 11.10 единственно, если зафиксировать аргументы диагональных элементов матрицы L.

11.12. Если для некоторого j (i) элементы матрицы A удовлетворяют условиям

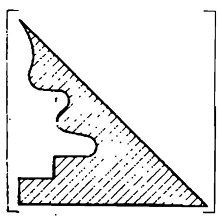
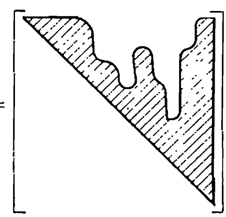
$$a_{ij} = 0, \qquad i = 1, 2, \dots, p < j, \quad j = 1, 2, \dots, r < i,$$

то будут равны нулю и элементы матрицы U (L) с соответствующими номерами.

Во многих прикладных задачах приходится иметь дело с *разреженными* матрицами, т. е. матрицами, имеющими много нулевых элементов. Утверждение 11.12 позволяет описать целый класс разреженных матриц, треугольные сомножители которых сохраняют специфику разреженности исходной матрицы. Пусть матрица A удовлетворяет условиям 11.5 и имеет вид

$A =$  ,

где все ее ненулевые элементы находятся в заштрихованной области. Граница этой области может быть произвольной. Требуется лишь, чтобы любая вертикальная (горизонтальная) прямая линия имела с правой (левой) частью границы односвязное множество общих точек. Как следует из 11.12, треугольные сомножители L и U будут иметь аналогичный вид, а именно:

$L =$  , $U =$ 

Все ненулевые элементы матриц L и U находятся в заштрихованных областях, границы которых такие же, как у матрицы A.

11.13. Если для матрицы B из 11.1 имеет место разложение $B = LU$, то справедливо разложение

$$\begin{bmatrix} B & Q \\ R & T \end{bmatrix} = \begin{bmatrix} L & 0 \\ RU^{-1} & E \end{bmatrix} \begin{bmatrix} U & L^{-1}Q \\ 0 & 0 \end{bmatrix}.$$

11.14. Пусть у матрицы A порядка m отличны от нуля ведущие миноры порядков $k_1 < k_2 < \ldots < k_s = m$. Тогда существует разложение $A = LU$, где L (U) — левая (правая) блочно треугольная матрица с невырожденными диагональными блоками размеров $k_1,\ k_2 - k_1,\ \ldots,\ k_s - k_{s-1}$. Это разложение называется *блочным LU-разложением.*

11.15. Блочное LU-разложение единственно, если зафиксировать диагональные блоки матрицы L или матрицы U.

11.16. Пусть у эрмитовой матрицы A порядка m отличны от нуля ведущие миноры порядков $k_1 < k_2 < \ldots < k_s = m$. Тогда существует разложение $A = LDL^*$, где L (D) — левая блочно треугольная (блочно диагональная) матрица с невырожденными диагональными блоками размеров $k_1,\ k_2 - k_1,\ \ldots,\ k_s - k_{s-1}$.

11.17. Все диагональные блоки матрицы D разложения 11.16 являются эрмитовыми матрицами.

11.18. Все ведущие миноры порядков $k_1,\ k_2,\ \ldots,\ k_s$ матриц A и D разложения 11.16 имеют одинаковые знаки.

11.19. Разложение 11.16 единственно, если зафиксировать D как эрмитову матрицу указанного в 11.16 строения со знаками ведущих миноров, удовлетворяющими условиям 11.18.

11.20. Если A — невырожденная эрмитова матрица, то существует такая матрица перестановок H, что у эрмитовой матрицы HAH' каждый нулевой ведущий минор имеет ненулевые соседние ведущие миноры.

11.21. Если для матрицы HAH' из 11.20 выполнить разложение 11.16, то диагональные блоки матрицы D будут первого и второго порядков.

11.22. Любую квадратную матрицу A можно представить в виде произведения $A = QR$, где Q — унитарная матрица, R — правая треугольная.

11.23. Разложение 11.22 называется *QR-разложением* матрицы A.

11.24. Для матрицы R из 11.22 имеет место равенство $A^*A = R^*R$.

11.25. Пусть матрица A — невырожденная порядка m. Тогда для элементов r_{ks} матрицы R справедливы формулы

$$r_{11} = \left(A^*A\begin{pmatrix}1\\1\end{pmatrix}\right)^{1/2} e^{i\varphi_1}, \quad \ldots, \quad r_{mm} = \left(\frac{A^*A\begin{pmatrix}1 & 2 & \cdots & m\\1 & 2 & \cdots & m\end{pmatrix}}{A^*A\begin{pmatrix}1 & 2 & \cdots & m-1\\1 & 2 & \cdots & m-1\end{pmatrix}}\right)^{1/2} e^{i\varphi_m},$$

$$r_{ks} = r_{kk}\frac{A^*A\begin{pmatrix}1 & 2 & \cdots & k-1 & k\\1 & 2 & \cdots & k-1 & s\end{pmatrix}}{A^*A\begin{pmatrix}1 & 2 & \cdots & k\\1 & 2 & \cdots & k\end{pmatrix}},$$

где $s = k+1, \ldots, m$, $k = 1, 2, \ldots, m-1$, $\varphi_1, \ldots, \varphi_m$ — произвольные вещественные числа.

11.26. QR-разложение невырожденной матрицы единственно, если зафиксировать аргументы диагональных элементов матрицы R.

11.27. Любую квадратную матрицу A можно представить в виде произведения $A = HU$, где H — эрмитова матрица с неотрицательными ведущими минорами, U — унитарная матрица.

11.28. Разложение 11.27 называется *полярным* разложением матрицы A.

11.29. В полярном разложении $A = HU$ матрицы A эрмитова матрица H определяется единственным образом.

11.30. Невырожденная матрица имеет единственное полярное разложение.

11.31. Каково бы ни было полярное разложение $A = HU$ матрицы A, унитарная матрица U переводит ортонормированный базис из собственных векторов матрицы A^*A в ортонормированный базис из собственных векторов матрицы AA^*.

11.32. Для любого полярного разложения $A = HU$ унитарная матрица U переводит $\operatorname{im} A^*$ в $\operatorname{im} A$, а $\ker A$ — в $\ker A^*$.

11.33. Матрица A является нормальной тогда и только тогда, когда в ее полярном разложении $A = HU$ матрицы H и U перестановочны.

11.34. Если матрица A нормальная, то собственные значения матрицы H (аргументы собственных значений матрицы U) полярного разложения $A = HU$ являются модулями собственных значений (аргументами ненулевых собственных значений) матрицы A.

11.35. (*Формулы Кели.*) Между произвольными эрмитовыми матрицами H и унитарными матрицами U, не имеющими собственных значений, равных -1, существует взаимно однозначное соответствие, определяемое формулами

$$U = (E + iH)(E - iH)^{-1}, \quad H = i(E - U)(E + U)^{-1}.$$

11.36. Для вещественной матрицы существует вещественное полярное разложение.

11.37. Какова бы ни была прямоугольная матрица A, матрицы A^*A и AA^* эрмитовы и имеют неотрицательные ведущие миноры.

11.38. Ненулевые собственные значения матриц A^*A и AA^* всегда совпадают.

11.39. Арифметические значения квадратных корней из общих собственных значений матриц A^*A и AA^* называются *сингулярными* (*главными*) *числами* матрицы A.

Всюду в дальнейшем будем обозначать ненулевые сингулярные числа матрицы A через $\rho_1, \ldots, \rho_t$ и предполагать, что они занумерованы в порядке убывания, т. е. $\rho_1 \geqslant \rho_2 \ldots \geqslant \rho_t > 0$. Сингулярные числа $\rho_{t+1}, \ldots$ будем считать нулевыми.

11.40. Сингулярные числа матрицы не меняются от ее умножения слева и справа на любые унитарные матрицы.

11.41. Квадратная матрица является невырожденной тогда и только тогда, когда все ее сингулярные числа отличны от нуля.

11.42. Модуль определителя матрицы равен произведению всех сингулярных чисел.

11.43. Квадратная матрица является нормальной тогда и только тогда, когда все ее собственные значения равны по модулю сингулярным числам.

11.44. Пусть A — прямоугольная матрица размера $m \times n$. Обозначим через $x_1, \ldots, x_n$ ортонормированные собственные векторы матрицы A^*A. Тогда:

— система векторов $Ax_1, \ldots, Ax_n$ является ортогональной;

— ненулевой вектор Ax_k является собственным вектором матрицы AA^* и соответствует собственному значению ρ_k^2;

— для всех k выполняется равенство $|Ax_k| = \rho_k$.

11.45. Пусть A — прямоугольная матрица размера $m\times n$. Всегда существуют ортонормированные системы векторов $x_1, \ldots$ $\ldots, x_n$ и $y_1, \ldots, y_m$ такие, что

$$Ax_k = \begin{cases} \rho_k y_k, & k \leqslant t, \\ 0 & k > t; \end{cases} \quad A^* y_k = \begin{cases} \rho_k x_k, & k \leqslant t, \\ 0, & k > t. \end{cases}$$

11.46. Ортонормированные системы $x_1, \ldots, x_n$ и $y_1, \ldots, y_m$ из 11.45 называются *сингулярными базисами* матрицы A.

11.47. Какова бы ни была прямоугольная матрица A размера $m\times n$, всегда существует разложение $A = U\Lambda V$, где U, V — унитарные матрицы, Λ — прямоугольная диагональная матрица размера $m\times n$ с невозрастающими неотрицательными элементами на диагонали.

11.48. Разложение 11.47 называется *сингулярным разложением* матрицы A.

11.49. Если задано сингулярное разложение $A = U\Lambda V$ матрицы A, то:

— диагональные элементы матрицы Λ являются сингулярными числами матрицы A;

— столбцы матрицы U образуют ортонормированный базис из собственных векторов матрицы AA^*;

— столбцы матрицы V^* образуют ортонормированный базис из собственных векторов матрицы A^*A;

— столбцы матриц V^*, U образуют в совокупности сингулярные базисы матрицы A.

11.50. Для вещественной матрицы существует вещественное сингулярное разложение.

11.51. Пусть A и B — прямоугольные матрицы размеров $m\times$ $\times n$ и $p\times q$ соответственно. *Кронекеровым* или *тензорным произведением* $A\times B$ матриц A и B называется матрица C размера $mp\times nq$ следующего блочного строения:

$$C = \begin{bmatrix} a_{11}B & a_{12}B & & a_{1n}B \\ a_{21}B & a_{22}B & & a_{2n}B \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ a_{m1}B & a_{m2}B & \ldots & a_{mn}B \end{bmatrix}.$$

11.52. Следующие соотношения справедливы при любом числе α и любых матрицах, для которых соответствующие операции имеют смысл:

$$(\alpha A)\times B = A\times(\alpha B) = \alpha\,(A\times B),$$
$$(A+B)\times C = A\times C + B\times C,$$
$$A\times(B+C) = A\times B + A\times C,$$
$$A\times(B\times C) = (A\times B)\times C,$$
$$(AB)\times(CD) = (A\times C)(B\times D),$$

$$(A\times B)^* = A^*\times B^*,$$
$$(A\times B)^+ = A^+\times B^+.$$

11.53. Для квадратных матриц A, B порядков m, n справедливы соотношения

$$\operatorname{tr}(A\times B) = (\operatorname{tr} A)\cdot(\operatorname{tr} B),$$
$$\det(A\times B) = (\det A)^n (\det B)^m,$$
$$(A\times B)^{-1} = A^{-1}\times B^{-1}.$$

11.54. Матрицу $A\times B$ перестановками строк и столбцов можно привести к матрице $B\times A$. Если матрицы A, B квадратные, то строки и столбцы переставляются одинаково.

11.55. Кронекерово произведение квадратных матриц A, B любых порядков является диагональной (треугольной того же наименования, нормальной, эрмитовой, унитарной) матрицей, если матрицы A, B диагональные (треугольные одного наименования, нормальные, эрмитовы, унитарные).

11.56. Если матрицы A, B порядков m, n подобны соответственно матрицам C, D, то:

— матрица $A\times B$ подобна матрице $C\times D$;

— матрица $A\times E_n+E_m\times B$ подобна матрице $C\times E_n+E_m\times D$.

11.57. Если матрицы A, B порядков m, n имеют простую структуру, то имеют простую структуру матрицы $A\times B$ и $A\times E_n+E_m\times B$.

11.58. Пусть λ и x — собственное значение и собственный вектор матрицы A порядка m; μ и y — собственное значение и собственный вектор матрицы B порядка n. Тогда:

— кронекерово произведение $x\times y$ является собственным вектором матрицы $A\times B$ и соответствует собственному значению $\lambda\mu$;

— кронекерово произведение $x\times y$ является собственным вектором матрицы $A\times E_n+E_m\times B$ и соответствует собственному значению $\lambda+\mu$.

11.59. Все mn собственных значений и соответствующие им собственные векторы матриц $A\times B$ и $A\times E_n+E_m\times B$ имеют вид, указанный в 11.58.

11.60. Пусть для матриц A, B определены: LU-разложения, QR-разложения, полярные разложения, сингулярные разложения, скелетные разложения, т. е.

$$A=L_1U_1,\quad B=L_2U_2,\quad A=U_1\Lambda_1V_1,\quad B=U_2\Lambda_2V_2,$$
$$A=Q_1R_1,\quad B=Q_2R_2,\quad A=C\cdot D,\quad B=F\cdot G.$$
$$A=H_1U_1,\quad B=H_2U_2.$$

Тогда соответствующие разложения для матрицы $A \times B$ имеют вид

$$(A \times B) = (L_1 \times L_2)(U_1 \times U_2),$$
$$(A \times B) = (Q_1 \times Q_2)(R_1 \times R_2),$$
$$(A \times B) = (H_1 \times H_2)(U_1 \times U_2),$$
$$(A \times B) = (U_1 \times U_2)(\Lambda_1 \times \Lambda_2)(V_1 \times V_2),$$
$$(A \times B) = (C \times F)(D \times G).$$

§ 12. Билинейные формы

12.1. Пусть задано комплексное (вещественное) линейное пространство. Числовая функция $\varphi(x, y)$ называется *билинейной формой* в этом пространстве, если для любых векторов x, y, z и любого комплексного (вещественного) числа α она принимает комплексное (вещественное) значение, и при этом выполняются соотношения

$$\varphi(x + z, y) = \varphi(x, y) + \varphi(z, y), \quad \varphi(\alpha x, y) = \alpha\varphi(x, y),$$
$$\varphi(x, y + z) = \varphi(x, y) + \varphi(x, z), \quad \varphi(x, \alpha y) = \alpha\varphi(x, y).$$

Мы уже встречались ранее с функцией такого вида. Сравнивая 5.1 и 12.1, легко заметить, что скалярное произведение в евклидовом пространстве является билинейной формой. Однако не каждая билинейная форма даже в вещественном пространстве будет скалярным произведением, так как могут оказаться невыполненными важные свойства симметричности скалярного произведения и его положительной определенности. Вспоминая, какую важную роль играло скалярное произведение при изучении евклидовых пространств и действующих в них линейных операторов, нетрудно понять и важность изучения билинейных форм, которые можно рассматривать как скалярные произведения с расширенными свойствами.

Многие свойства билинейных форм в комплексных и вещественных пространствах совпадают, хотя в ряде случаев имеются, и очень существенные, различия. В тех случаях, когда свойства билинейных форм совпадают, мы не будем указывать тип пространств.

12.2. С билинейной формой можно выполнять формальные алгебраические преобразования, т. е.

$$\varphi\left(\sum_{i=1}^{r} \alpha_i x_i, \ \sum_{j=1}^{s} \beta_j y_j\right) = \sum_{i=1}^{r} \sum_{j=1}^{s} \alpha_i \beta_j \varphi(x_i, y_j).$$

12.3. Для любой билинейной формы выполняются соотношения $\varphi(0, y) = \varphi(x, 0) = 0$.

12.4. Билинейная форма называется *нулевой*, если она принимает нулевое значение для всех пар векторов.

12.5. Множество всех билинейных форм, заданных над одним и тем же линейным пространством, есть линейное пространство.

12.6. Билинейная форма $\varphi(x, y)$ называется *симметричной*, если для любых векторов x, y выполняется равенство $\varphi(x, y) = \varphi(y, x)$.

12.7. Билинейная форма $\varphi(x, y)$ называется *кососимметричной*, если для любых векторов x, y выполняется равенство $\varphi(x, y) = -\varphi(y, x)$.

12.8. Множества симметричных и кососимметричных билинейных форм образуют подпространства в линейном пространстве всех билинейных форм.

12.9. Любая билинейная форма однозначно разложима в сумму симметричной и кососимметричной билинейных форм, а именно:

$$\varphi(x, y) = \frac{1}{2}\{\varphi(x, y) + \varphi(y, x)\} + \frac{1}{2}\{\varphi(x, y) - \varphi(y, x)\}.$$

Первые два слагаемых в правой части дают симметричную билинейную форму, последние два — кососимметричную.

12.10. Для любой билинейной формы имеет место тождество

$$\varphi(x, y) + \varphi(y, x) = \varphi(x+y, x+y) - \varphi(x, x) - \varphi(y, y).$$

12.11. Кососимметричные билинейные формы, и только они, принимают нулевые значения при всех совпадающих аргументах.

12.12. Симметричная билинейная форма однозначно определяется своими значениями при совпадающих аргументах.

12.13. Для несимметричной билинейной формы ее симметричная часть однозначно определяется значениями формы при совпадающих аргументах.

12.14. *Квадратичной формой* в линейном пространстве K называется числовая функция $\varphi(x, x)$ от одного векторного аргумента $x \in K$, которая получается из билинейной формы $\varphi(x, y)$ заменой вектора y на вектор x.

Вообще говоря, нельзя однозначно восстановить по квадратичной форме породившую ее билинейную форму. Но, как вытекает из 12.10, существует, и притом только одна, симметричная билинейная форма, из которой может быть получена исходная квадратичная форма. Эта билинейная форма называется *полярной* по отношению к заданной квадратичной форме. Множество всех билинейных форм, порождающих одну и ту же квадратичную форму, может быть получено путем сложения полярной билинейной формы и произвольной кососимметричной формы. Поэтому при использовании билинейных форм для изучения свойств квадратичных форм достаточно ограничиться рассмотрением симметричных билинейных форм. Невозможность восстановления билинейной формы по квадратичной объясняется тем, что согласно 12.7 квадратичная форма не дает никакой информации о кососимметричной части любой билинейной формы.

12.15. Вектор z называется *изотропным* для квадратичной формы $\varphi(x, x)$, если $\varphi(z, z) = 0$.

Как вытекает из свойств линейности билинейных форм по каждому аргументу, $\varphi(0, 0) = 0$ для любой билинейной формы. Однако в общем случае могут существовать и ненулевые векторы, на которых квадратичная форма принимает нулевое значение. Понятие изотропности связано только с квадратичной формой. Поэтому векторы, изотропные для одной квадратичной формы, могут быть не изотропными для другой квадратичной формы, и наоборот. Множество изотропных векторов, вообще говоря, образует некоторую поверхность второго порядка в линейном пространстве, но в ряде важных случаев оказывается линейным подпространством.

12.16. Для квадратичной формы, порожденной кососимметричной билинейной формой, все векторы пространства являются изотропными.

12.17. Любая квадратичная форма в комплексном пространстве размерности больше 1 всегда имеет ненулевые изотропные векторы.

12.18. Для квадратичной формы в комплексном пространстве множество изотропных векторов является подпространством в том и только в том случае, когда эта форма порождена кососимметричной билинейной формой.

12.19. Квадратичная форма называется *вещественной*, если для любых векторов пространства она принимает вещественные значения.

12.20. Квадратичная форма в комплексном пространстве является вещественной тогда и только тогда, когда эта форма порождена кососимметричной билинейной формой.

12.21. Вещественная квадратичная форма называется *положительно* (*отрицательно*, *неположительно*, *неотрицательно*) *определенной*, если $\varphi(x, x) > 0$ (<0, $\leqslant 0$, $\geqslant 0$) для всех $x \neq 0$.

Как правило, только положительно и отрицательно определенные квадратичные формы называются *знакопостоянными*. Но иногда знакопостоянными называются также неотрицательные и неположительные квадратичные формы. Во избежание появления недоразумений в нужных случаях положительно и отрицательно определенные квадратичные формы мы будем называть *строго* знакопостоянными, а неположительно и неотрицательно определенные — *нестрого* знакопостоянными.

12.22. Вещественная квадратичная форма является строго знакопостоянной тогда и только тогда, когда она не имеет ненулевых изотропных векторов.

12.23. Вещественная квадратичная форма является нестрого знакопостоянной тогда и только тогда, когда множество ее изотропных векторов есть нетривиальное линейное пространство.

Сравнение свойств скалярного произведения 5.2 и соотношений 12.1 показывает, что в комплексном пространстве скалярное произведение, строго говоря, не является билинейной формой. В комплексном пространстве со скалярным произведением тесно связаны эрмитовы билинейные формы.

12.24. Пусть задано комплексное (вещественное) линейное пространство. Числовая функция $\varphi(x, y)$ называется *эрмитовой билинейной формой* в этом пространстве, если для любых векторов x, y, z и комплексного (вещественного) числа α она принимает комплексное (вещественное) значение, и при этом выполняются соотношения

$$\varphi(x+z,\ y)=\varphi(x,\ y)+\varphi(z,\ y),\quad \varphi(\alpha x,\ y)=\alpha\varphi(x,\ y),$$
$$\varphi(x,\ y+z)=\varphi(x,\ y)+\varphi(x,\ z),\quad \varphi(x,\ \alpha y)=\bar{\alpha}\varphi(x,\ y).$$

Здесь черта означает комплексное сопряжение.

12.25. Эрмитова билинейная форма в вещественном пространстве является обыкновенной билинейной формой.

12.26. Если эрмитова билинейная форма в комплексном пространстве является обыкновенной билинейной формой, то эта форма — нулевая.

12.27. С эрмитовой билинейной формой можно выполнять формальные алгебраические преобразования, т. е.

$$\varphi\left(\sum_{i=1}^{r}\alpha_i x_i,\ \sum_{j=1}^{s}\beta_j y_j\right)=\sum_{i=1}^{r}\sum_{j=1}^{s}\alpha_i\bar{\beta}_j\varphi(x_i,\ y_j).$$

12.28. Эрмитова билинейная форма называется *эрмитовой симметричной* (*эрмитовой кососимметричной*), если для любых векторов x, y

$$\varphi(x,\ y)=\overline{\varphi(y,\ x)}\qquad (\varphi(x,\ y)=-\varphi(y,\ x)).$$

12.29. Если $\varphi(x,\ y)$ — эрмитова симметричная (кососимметричная) билинейная форма, то $i\varphi(x,\ y)$ будет эрмитовой кососимметричной (симметричной) билинейной формой.

12.30. Любая эрмитова билинейная форма однозначно разложима в сумму эрмитовой симметричной и эрмитовой кососимметричной билинейных форм. Именно:

$$\varphi(x,y)=\frac{1}{2}\{\varphi(x,y)+\overline{\varphi(y,x)}\}+\frac{1}{2}\{\varphi(x,y)-\overline{\varphi(y,x)}\}.$$

Первые два слагаемых в правой части дают эрмитову симметричную билинейную форму, последние два — эрмитову кососимметричную.

12.31. Для любой эрмитовой билинейной формы имеет место тождество

$$\varphi(x,y)=\frac{1}{4}\{\varphi(x+y,x+y)-\varphi(x-y,x-y)+$$
$$+i\varphi(x+iy,x+iy)-i\varphi(x-iy,x-iy)\}.$$

12.32. Любая эрмитова билинейная форма однозначно определяется своими значениями при совпадающих аргументах.

12.33. Среди эрмитовых билинейных форм нулевая форма, и только она, принимает нулевые значения при всех совпадающих аргументах.

12.34. *Эрмитовой квадратичной формой* в комплексном пространстве называется числовая функция $\varphi(x,\ x)$ от одного векторного аргумента x, которая получается из эрмитовой билинейной формы $\varphi(x, y)$ заменой вектора y на вектор x.

В отличие от квадратичных форм, по эрмитовой квадратичной форме однозначно восстанавливается порождающая ее эрмитова билинейная форма. Это восстановление осуществляется согласно 12.31, и соответствующая билинейная форма также называется *полярной* по отношению к исходной квадратичной форме. Возможность однозначного восстановления объясняется тесной связью между эрмитовыми симметричными и эрмитовыми кососимметричными билинейными формами.

Несмотря на имеющиеся различия в свойствах билинейных и эрмитовых билинейных форм, для обоих типов форм многие понятия вводятся одинаково. В частности, аналогично вводятся вещественные и знакоопределенные квадратичные формы, изотропные векторы и т. п.

12.35. Среди эрмитовых билинейных форм симметричные (кососимметричные) формы, и только они, порождают вещественные (чисто мнимые) эрмитовы квадратичные формы.

12.36. Никакая эрмитова несимметричная билинейная форма не может порождать вещественную эрмитову квадратичную форму.

12.37. Вещественная эрмитова квадратичная форма является строго знакопостоянной тогда и только тогда, когда она не имеет ненулевых изотропных векторов.

12.38. Вещественная эрмитова квадратичная форма является нестрого знакопостоянной тогда и только тогда, когда множество ее изотропных векторов есть нетривиальное линейное пространство.

12.39. Для того чтобы эрмитова квадратичная форма не имела ненулевых изотропных векторов, достаточно, чтобы ее вещественная или мнимая часть была строго знакопостоянной.

12.40. Симметричные эрмитовы билинейные формы, порождающие положительно определенные квадратичные формы, и только они, задают скалярное произведение как в вещественном, так и в комплексном пространстве.

Если в линейном пространстве задан какой-либо базис, то билинейная и эрмитова билинейная формы являются функциями координат векторов. По существу, вид этих функций полностью определяется формулами 12.2, 12.27. Особенно просто билинейные формы записываются в том случае, когда при заданном базисе в линейном пространстве согласно 5.4 введено скалярное произведение.

12.41. Пусть $e_1, \ldots, e_n$ и $q_1, \ldots, q_n$ — два базиса пространства и для векторов x, y заданы разложения

$$x = \sum_{i=1}^{n} \xi_i e_i, \quad y = \sum_{j=1}^{n} \eta_j q_j.$$

Тогда имеет место представление

$$\varphi(x, y) = \sum_{i=1}^{n} \sum_{j=1}^{n} \varphi(e_i, q_j)\xi_i \eta_j$$

для билинейной формы и представление

$$\varphi(x, y) = \sum_{i=1}^{n} \sum_{j=1}^{n} \varphi(e_i, q_j)\, \xi_i \bar{\eta}_j$$

для эрмитовой билинейной формы.

12.42. Матрица Φ_{eq} с элементами $\varphi_{ij} = \varphi(e_i, q_j)$ **называется** *матрицей билинейной (эрмитовой билинейной) формы* $\varphi(x, y)$ в выбранной паре базисов.

12.43. Если x_e, y_q — вектор-столбцы арифметического пространства, составленные из координат векторов x, y в разложениях 12.41, то $\varphi(x, y) = x_e'\Phi_{eq}y_q$ для билинейной формы и $\varphi(x, y) = x_e'\Phi_{eq}\bar{y}_q$ для эрмитовой билинейной формы.

12.44. В евклидовом и унитарном пространствах любая билинейная (эрмитова билинейная) форма имеет вид

$$\varphi(x, y) = (\Phi x, \bar{y}) \qquad (\varphi(x, y) = (\Phi x, y)).$$

Здесь Φ — матрица формы в паре совпадающих естественных базисов.

Заметим, что матрица билинейной формы однозначно определяется упорядоченной парой базисов. Поэтому при перемене базисов местами матрица формы меняется.

12.45. Между линейными пространствами билинейных и эрмитовых билинейных форм в комплексном (вещественном) n-мерном пространстве и линейным пространством квадратных матриц порядка n с комплексными (вещественными) элементами существуют изоморфные соответствия.

12.46. В изоморфизме 12.45 симметричной и кососимметричной (эрмитовой симметричной и эрмитовой кососимметричной) билинейной форме соответствуют симметричная и кососимметричная (эрмитова и косоэрмитова) матрица.

Итак, если в пространстве введено скалярное произведение, то изучение билинейных и эрмитовых билинейных форм означает изучение функций вида $(Ax, \bar{y})$, (Bx, y), и ничего другого. Поэтому все рассмотренные свойства форм сразу же переносятся на их матрицы, и наоборот.

12.47. Матрица A называется *положительно определенной* и обозначается символом $A > 0$, если $(Ax, x) > 0$ для всех векторов $x \neq 0$. Аналогично вводятся *отрицательно*, *неотрицательно* и *неположительно* определенные матрицы. Неотрицательно определенная матрица называется также *положительно полуопределенной.*

12.48. Если матрица положительно определенная, то:

— все главные миноры положительные;

— все коэффициенты характеристического многочлена отличны от нуля и имеют чередующиеся знаки;

— все собственные значения имеют положительные вещественные части.

12.49. Для того чтобы комплексная матрица была определенной в каком-либо смысле 12.47, необходимо и достаточно, чтобы она была эрмитовой и определенной в том же смысле.

12.50. Для того чтобы вещественная матрица была определенной в каком-либо смысле 12.47, необходимо и достаточно, чтобы ее симметричная составляющая в разложении 10.62 была определенной в том же смысле.

12.51. (*Критерий Сильвестра.*) Для того чтобы эрмитова матрица была положительно определенной, необходимо и достаточно, чтобы все ведущие миноры этой матрицы были положительными.

12.52. Для того чтобы эрмитова матрица была отрицательно определенной, необходимо и достаточно, чтобы все ведущие миноры нечетного порядка были отрицательными, четного — положительными.

12.53. (*Критерий Якоби.*) Для того чтобы эрмитова матрица была положительно определенной, необходимо и достаточно, чтобы все коэффициенты характеристического многочлена были отличны от нуля и имели чередующиеся знаки.

12.54. Для того чтобы эрмитова матрица была отрицательно определенной, необходимо и достаточно, чтобы все коэффициенты ее характеристического многочлена были положительны.

12.55. Для того чтобы эрмитова матрица A удовлетворяла условию $A > 0$ ($\geqslant$, $<$, $\leqslant$), необходимо и достаточно, чтобы ее собственные значения λ_i удовлетворяли условию $\lambda_i > 0$ ($\geqslant$, $<$, $\leqslant$) при всех i.

Заметим, что критерии знакоопределенности 12.51—12.55 в целом справедливы только для эрмитовых матриц. Для вещественных несимметричных матриц достаточность уже не обязательно будет иметь место. Поэтому, например, вещественная несимметричная матрица с положительными собственными значениями может не быть положительно определенной.

12.56. Неравенство $A > B$ или $A - B > 0$ ($\geqslant$, $<$, $\leqslant$) для матриц A и B означает, что $(Ax, x) > (Bx, x)$ ($\geqslant$, $<$, $\leqslant$) для всех $x \neq 0$.

12.57. Для любой матрицы A существуют числа α, β, зависящие только от A и такие, что

$$\alpha E \leqslant A \leqslant \beta E.$$

12.58. Если $A > 0$ ($\geqslant 0$), то число α в неравенствах 12.57 может быть выбрано положительным (неотрицательным).

12.59. Если $A > 0$ (< 0), то матрица A — невырожденная и $A^{-1} > 0$ (< 0).

12.60. Пусть A, B — положительно определенные матрицы. Существуют положительные числа δ, γ, зависящие только от A, B и такие, что

$$\delta B \leqslant A \leqslant \gamma B.$$

12.61. Если $A > 0$ $(\geqslant, <, \leqslant)$, то для любой невырожденной матрицы B имеем $B^*AB > 0$ $(\geqslant, <, \leqslant)$.

12.62. Если $A > 0$ $(\geqslant, <, \leqslant)$, то для любой матрицы B имеем $B^*AB \geqslant 0$ $(\geqslant, \leqslant, \leqslant)$.

12.63. Если A, $B > 0$ $(\geqslant, <, \leqslant)$, то для любых неотрицательных чисел α, β, не равных нулю одновременно, имеем $\alpha A + \beta B > 0$ $(\geqslant, <, \leqslant)$.

12.64. Если $A > 0$ $(\geqslant, <, \leqslant)$, то такому же условию удовлетворяет матрица любого главного минора матрицы A.

12.65. Если $A, B > 0$, то $A \times B > 0$.

12.66. Для любой эрмитовой положительно (неотрицательно) определенной матрицы A существует, и притом единственная, эрмитова положительно (неотрицательно) определенная матрица S такая, что $S^2 = A$. Матрица S называется *(арифметическим) квадратным корнем* из матрицы A и обозначается $A^{1/2}$.

12.67. Собственные векторы матриц A и $A^{1/2}$ совпадают, а собственные значения матрицы A равны квадратам собственных значений матрицы $A^{1/2}$.

12.68. Пусть A, B — эрмитовы положительно определенные матрицы. Тогда для любых вещественных чисел α, β неравенства $\alpha A \geqslant \beta B$ и $\alpha B^{-1} \geqslant \beta A^{-1}$ эквивалентны.

12.69. Пусть A, B — эрмитовы матрицы, причем B — положительно определенная. Тогда матрица AB имеет простую структуру и все ее собственные значения вещественные.

12.70. В условиях 12.69 собственные значения матрицы AB положительны (неотрицательны) тогда и только тогда, когда матрица A положительно (неотрицательно) определенная.

12.71. Пусть $A = S + iK$ — эрмитова положительно определенная матрица порядка m с вещественными матрицами S, K. Вещественная блочная матрица

$$R = \begin{bmatrix} S & -K \\ K & S \end{bmatrix}$$

порядка $2m$ является симметричной и положительно определенной.

При переходе от одной пары базисов к другой матрица билинейной формы меняется. Характер этого изменения полностью определяется видом формы 12.44 и соотношением 7.2, описывающим связь координат векторов в разных базисах.

12.72. Пусть P — матрица преобразования координат при переходе от базиса $e_1, \ldots, e_n$ к базису $f_1, \ldots, f_n$, а Q — матрица преобразования координат при переходе от $q_1, \ldots, q_n$ к $t_1, \ldots, t_n$. Тогда для матриц билинейной (эрмитовой билинейной) формы $\varphi(x, y)$ выполняются соотношения

$$\Phi_{ft} = P'\Phi_{eq}Q \qquad (\Phi_{ft} = P'\Phi_{eq}\bar{Q}).$$

12.73. Множество матриц одной и той же формы $\varphi(x, y)$ в различных парах базисов есть множество эквивалентных матриц.

12.74. Для любой билинейной (эрмитовой билинейной) формы всегда существует такая пара базисов, в которых матрица формы является диагональной, с элементами 0 и 1 на диагонали.

12.75. Ранг матрицы билинейной формы не зависит от выбранных базисов. Он называется *рангом билинейной формы.*

12.76. Разность между размерностью пространства и рангом билинейной формы называется *дефектом билинейной формы.*

12.77. Билинейная форма называется *невырожденной*, если ее дефект равен 0.

Раздельный выбор базисов для каждой переменной билинейной формы применяется довольно редко. Значительно чаще используется общий базис.

12.78. Пусть P — матрица преобразования координат при переходе от базиса $e_1, \ldots, e_n$ к базису $f_1, \ldots, f_n$. Тогда для матриц билинейной (эрмитовой билинейной) формы $\varphi(x, y)$ выполняются соотношения

$$\Phi_{ff} = P'\Phi_{ee}P \qquad (\Phi_{ff} = P'\Phi_{ee}\bar{P}).$$

12.79. Матрицы, связанные соотношением 12.78, называются *конгруэнтными* (*эрмитово конгруэнтными*).

12.80. Отношение конгруэнтности есть отношение эквивалентности.

12.81. Правая (левая) треугольная матрица ранга r с первыми r ненулевыми диагональными элементами называется *правой* (*левой*) *трапециевидной матрицей.*

12.82. Трапециевидная матрица с диагональными элементами, равными 1 или 0, называется *канонической.*

12.83. В правой (левой) трапециевидной матрице ранга r все строки (столбцы), начиная с $r+1$-й ($r+1$-го), являются нулевыми.

12.84. Любая некососимметричная матрица конгруэнтна правой трапециевидной матрице.

12.85. Любая кососимметричная матрица конгруэнтна блочно диагональной матрице с блоками 2-го и 1-го порядков. При этом все блоки 2-го порядка являются невырожденными кососимметричными матрицами, блоки 1-го порядка — нулевые.

12.86. Любая матрица эрмитово конгруэнтна правой трапециевидной матрице.

12.87. Симметричная матрица конгруэнтна диагональной матрице.

12.88. Эрмитова матрица эрмитово конгруэнтна вещественной диагональной матрице.

В общем случае не всегда можно сказать заранее, какой вид будет иметь матрица конгруэнтного преобразования при переходе к трапециевидной матрице. Однако при некоторых дополнительных ограничениях, накладываемых на исходную матрицу, на этот вопрос можно дать вполне определенный ответ. При этом широко используются различные мультипликативные разложения матрицы.

12.89. Для того чтобы некососимметричная (произвольная) матрица A могла быть приведена к правой трапециевидной с помощью конгруэнтного (эрмитово конгруэнтного) преобразования с правой треугольной матрицей, необходимо и достаточно, чтобы число первых ненулевых ведущих миноров матрицы A равнялось ее рангу.

12.90. Для того чтобы кососимметричная матрица A ранга r могла быть приведена к правой трапециевидной с помощью конгруэнтного преобразования с правой треугольной матрицей, необходимо и достаточно, чтобы число первых ненулевых ведущих миноров четного порядка матрицы A равнялось $r/2$.

12.91. Любая нормальная матрица приводится к диагональной эрмитово конгруэнтным преобразованием с унитарной матрицей.

12.92. (*Закон инерции квадратичных форм.*) Если вещественная симметричная (эрмитова) матрица приводится вещественным конгруэнтным (эрмитово конгруэнтным) преобразованием к диагональному виду, то число положительных, отрицательных и нулевых элементов на диагонали не зависит от способа приведения.

12.93. Число положительных (отрицательных) элементов на диагонали в утверждении 12.92 называется *положительным* (*отрицательным*) *индексом инерции*. Разность между положительным и отрицательным индексами называется *сигнатурой*.

12.94. Если матрица A имеет вид 11.1, то ее индексы инерции совпадают с индексами инерции матрицы B из 11.1.

12.95. Пусть для эрмитовой матрицы A существует LU-разложение. Тогда число нулевых, положительных и отрицательных диагональных элементов матрицы U совпадает с числом нулевых, положительных и отрицательных собственных значений матрицы A.

§ 13. Билинейно метрические пространства

Изучение евклидовых и унитарных пространств сводилось к исследованию дополнительных свойств как самих пространств, так и матриц по отношению к билинейным формам, определяющим скалярные произведения. Как уже отмечалось, далеко не всегда возникает необходимость введения именно скалярного произведения. Для решения многих задач достаточно задать в пространстве билинейную или эрмитову билинейную форму, причем не обязательно симметричную и положительно определенную.

13.1. Линейное пространство называется *билинейно метрическим* (*эрмитовым билинейно метрическим*), если в нем задана билинейная (эрмитова билинейная) форма.

Многие определения и факты будут одинаковыми как для билинейных, так и для эрмитовых билинейных пространств. Поэтому всюду, где это не вызывает недоразумений, слово «эрмитово» в данном параграфе мы будем опускать и будем проводить соответствующие выкладки только для билинейных пространств, подразумевая, что для эрмитовых пространств они проводятся аналогично. Желая подчеркнуть общность между билинейными формами и скалярными произведениями, мы будем оба вида форм называть также скалярным произведением и обозначать соответствующим символом.

13.2. Если для заданной системы векторов $x_1, \ldots, x_m$ и вектора x имеет место разложение

$$x = \alpha_1 x_1 + \ldots + \alpha_m x_m,$$

то коэффициенты этого разложения удовлетворяют системе линейных алгебраических уравнений

$$\begin{gathered}
\alpha_1(x_1, x_1) + \alpha_2(x_2, x_1) + \ldots + \alpha_m(x_m, x_1) = (x, x_1), \\
\alpha_1(x_1, x_2) + \alpha_2(x_2, x_2) + \ldots + \alpha_m(x_m, x_2) = (x, x_2), \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
\alpha_1(x_1, x_m) + \alpha_2(x_2, x_m) + \ldots + \alpha_m(x_m, x_m) = (x, x_m).
\end{gathered}$$

13.3. Матрица G, являющаяся транспонированной матрицей системы 13.2, имеет вид

$$G = \begin{bmatrix} (x_1, x_1) & (x_1, x_2) & \ldots & (x_1, x_m) \\ (x_2, x_1) & (x_2, x_2) & \ldots & (x_2, x_m) \\ \cdot & \cdot & \cdot & \cdot \\ (x_m, x_1) & (x_m, x_2) & \ldots & (x_m, x_m) \end{bmatrix}$$

и называется *матрицей Грама* системы векторов $x_1, x_2, \ldots, x_m$. Ее определитель $G(x_1, \ldots, x_m)$ называется *определителем Грама*.

13.4. Если векторы $x_1, \ldots, x_m$ образуют базис пространства, то матрица Грама для них является матрицей основной билинейной формы (x, y) в этом базисе.

13.5. Матрицы Грама для различных базисов конгруэнтны.

13.6. Ранг матриц Грама является инвариантом билинейно метрического пространства и называется его *рангом*.

13.7. Разность между размерностью и рангом пространства называется *дефектом* пространства.

13.8. Билинейно метрическое пространство называется *вырожденным* (*невырожденным*), если его дефект отличен от нуля (равен нулю).

13.9. Для невырожденного пространства система 13.2, где $x_1, \ldots, x_m$ — базис, всегда имеет, и притом единственное, решение.

13.10. Если для двух векторов x, y билинейно метрического пространства выполняется равенство $(x, y) = 0$, то вектор y называется *ортогональным справа* к вектору x, а вектор x — *ортогональным слева* к вектору y.

13.11. Если $(x, y) = (y, x) = 0$, то векторы x, y называются *ортогональными*.

13.12. Для того чтобы вектор пространства был ортогонален в каком-либо смысле ко всем векторам линейного подпространства, необходимо и достаточно, чтобы этот вектор был ортогонален в том же смысле к векторам какого-нибудь его базиса.

13.13. Множество векторов F билинейно метрического пространства ортогонально справа, слева или просто ортогонально множеству векторов G того же пространства, если аналогичное отношение ортогональности выполняется для каждой пары векторов x, y, где $x \in F$, $y \in G$.

13.14. Множество векторов пространства, ортогональных справа (слева) каждому из векторов множества F, называется *ортогональным дополнением F справа (слева)* и обозначается $F^{\perp}({}^{\perp}F)$.

13.15. Ортогональное дополнение есть подпространство.

13.16. Для любого множества F имеют место включения $F \subseteq {}^{\perp}(F^{\perp})$, $F \subseteq ({}^{\perp}F)^{\perp}$.

13.17. Если скалярное произведение задано симметричной или кососимметричной билинейной формой, то для любого множества F выполняется равенство $F^{\perp} = {}^{\perp}F$.

13.18. Если матрица Грама системы векторов $x_1, \ldots, x_m$ вырожденная, то существуют такие векторы u, v, являющиеся нетривиальными линейными комбинациями векторов $x_1, \ldots, x_m$, что u ортогонален справа, а v — слева ко всем векторам линейной оболочки векторов $x_1, \ldots, x_m$.

13.19. Если матрица Грама для линейно независимой системы векторов вырожденная, то квадратичная форма (x, x) имеет ненулевой изотропный вектор, принадлежащий линейной оболочке заданной системы и ортогональный справа (слева) ко всем векторам этой оболочки.

13.20. Для любой линейно зависимой системы векторов определитель Грама равен нулю.

13.21. Если квадратичная форма (x, x) не имеет ненулевых изотропных векторов, то определитель Грама не равен нулю тогда и только тогда, когда его система векторов линейно независима.

13.22. Если квадратичная форма (x, x) строго знакопостоянная, то определитель Грама равен нулю тогда и только тогда, когда система векторов линейно зависима.

13.23. Если билинейная форма (x, y) симметричная, а квадратичная форма (x, x) строго знакопостоянная, то для любых двух векторов x, y выполняется неравенство Коши — Буняковского $|(x, y)|^2 \leqslant (x, x)(y, y)$, причем равенство достигается тогда и только тогда, когда векторы x, y линейно зависимы.

13.24. Определитель Грама не изменяется при перемене местами любых двух векторов в системе $x_1, \ldots, x_m$.

13.25. Определитель Грама не изменяется от прибавления к любому вектору системы $x_1, \ldots, x_m$ любой линейной комбинации остальных векторов.

13.26. Если какой-либо вектор системы $x_1, \ldots, x_m$ умножить на число α, то определитель Грама умножается на α^2, если билинейная форма (x, y) обыкновенная, и на $|\alpha|^2$, если форма (x, y) эрмитова.

13.27. Если каждый из векторов $x_1, \ldots, x_m$ ортогонален слева (справа) ко всем предшествующим векторам, то для определителя Грама справедливо равенство

$$G(x_1, \ldots, x_m) = \prod_{i=1}^{m} (x_i, x_i).$$

13.28. В евклидовом и унитарном пространствах определитель Грама для любой линейно независимой системы векторов является положительным.

13.29. В евклидовом и унитарном пространствах определитель Грама системы векторов равен квадрату модуля определителя матрицы, столбцы которой составлены из координат этих векторов в любом ортонормированном базисе.

13.30. Для любой системы векторов $x_1, \ldots, x_m$ евклидова или унитарного пространства справедливы неравенства

$$0 \leqslant G(x_1, \ldots, x_m) \leqslant \prod_{i=1}^{m} (x_i, x_i),$$

причем равенство слева достигается тогда и только тогда, когда система векторов линейно зависима, а равенство справа — тогда и только тогда, когда система векторов либо ортогональна, либо содержит нулевой вектор.

13.31. Для любой системы векторов $x_1, \ldots, x_m$ евклидова или унитарного пространства справедливо неравенство

$$G(x_1, \ldots, x_l, x_{l+1}, \ldots, x_m) \leqslant G(x_1, \ldots, x_l) G(x_{l+1}, \ldots, x_m),$$

причем равенство достигается тогда и только тогда, когда либо множества векторов $x_1, \ldots, x_l$ и $x_{l+1}, \ldots, x_m$ ортогональны, либо одно из этих множеств представляет собой линейно зависимую систему.

13.32. Для любой матрицы A в евклидовом или унитарном пространстве отношение

$$k(A) = \frac{G(Ax_1, \ldots, Ax_m)}{G(x_1, \ldots, x_m)}$$

не зависит от векторов $x_1, \ldots, x_m$, если они образуют базис, и равно произведению квадратов модулей собственных значений матрицы A.

13.33. Для любой линейно независимой системы векторов $x_1, \ldots, x_m$ евклидова или унитарного пространства и любого вектора z выполняется неравенство

$$\frac{G(x_1, \ldots, x_m, z)}{G(x_1, \ldots, x_m)} \leqslant \frac{G(x_1, \ldots, x_{m-1}, z)}{G(x_1, \ldots, x_{m-1})}.$$

Любое линейное подпространство можно рассматривать как билинейно метрическое пространство относительно того же скалярного произведения.

При этом на подпространства переносится вся введенная ранее терминология.

13.34. Для того чтобы в пространстве были невырожденными все его подпространства, необходимо и достаточно, чтобы квадратичная форма (x, x) не имела ненулевых изотропных векторов.

13.35. Подпространства ${}^{\perp}K$ и $K^{\perp}$ билинейно метрического пространства K называются соответственно *левым* и *правым нулевыми подпространствами* в K.

13.36. Имеют место соотношения ${}^{\perp}(K^{\perp}) = ({}^{\perp}K)^{\perp} = K$.

13.37. Для любого множества векторов F всегда справедливы включения $K^{\perp} \subseteq F^{\perp}$, ${}^{\perp}K \subseteq {}^{\perp}F$.

13.38. Для любых систем векторов из ${}^{\perp}K$ или $K^{\perp}$ матрицы Грама являются нулевыми.

13.39. Размерности левого и правого нулевых подпространств совпадают и равны дефекту билинейной формы (x, y).

13.40. Для того чтобы пространство было невырожденным, необходимо и достаточно, чтобы правое и левое нулевые подпространства состояли только из нулевого вектора.

13.41. Пусть L — подпространство в K. Для того чтобы существовали разложения

$$K = L \dotplus L^{\perp} = L \dotplus {}^{\perp}L,$$

необходимо и достаточно, чтобы подпространство L было невырожденным.

13.42. Если невырожденное подпространство L имеет размерность m, то размерность подпространств $L^{\perp}$ и ${}^{\perp}L$ равна $n-m$, где n — размерность всего пространства.

13.43. Если невырожденное подпространство L имеет максимально возможную размерность, то $L^{\perp} = K^{\perp}$, ${}^{\perp}L = {}^{\perp}K$.

13.44. Пусть L — невырожденное подпространство максимальной размерности. Разложения 13.41 будут ортогональными тогда и только тогда, когда левое и правое нулевые подпространства совпадают.

13.45. Если L — невырожденное подпространство максимальной размерности, то подпространства ${}^{\perp}L$ и $L^{\perp}$ состоят только из изотропных векторов.

13.46. Пусть векторы $x_1, \ldots, x_m$ образуют базис невырожденного подпространства L и x — произвольный вектор. Тогда вектор

$$z = \alpha_1 x_1 + \ldots + \alpha_m x_m,$$

где коэффициенты $\alpha_1, \ldots, \alpha_m$ удовлетворяют системе линейных алгебраических уравнений

$$\begin{gathered}
\alpha_1(x_1, x_1) + \alpha_2(x_2, x_1) + \ldots + \alpha_m(x_m, x_1) = (x, x_1), \\
\alpha_1(x_1, x_2) + \alpha_2(x_2, x_2) + \ldots + \alpha_m(x_m, x_2) = (x, x_2), \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
\alpha_1(x_1, x_m) + \alpha_2(x_2, x_m) + \ldots + \alpha_m(x_m, x_m) = (x, x_m),
\end{gathered}$$

есть проекция вектора x на подпространство L параллельно подпространству ${}^{\perp}L$.

13.47. Если коэффициенты $\alpha_1, \ldots, \alpha_m$ удовлетворяют системе с матрицей, транспонированной по отношению к матрице системы 13.46, и с правой частью 13.46, то вектор z из 13.46 есть проекция вектора x на подпространство L параллельно подпространству $L^{\perp}$.

13.48. Для эрмитова билинейно метрического пространства утверждение 13.47 остается в силе, если в системе заменить коэффициенты $\alpha_1, \ldots, \alpha_m$ на $\overline{\alpha}_1, \ldots, \overline{\alpha}_m$ и $(x, x_1), \ldots, (x, x_m)$ на $(x_1, x), \ldots, (x_m, x)$.

В билинейно метрических пространствах базисы неравноправны. Среди них имеются такие, для которых системы 13.46—13.48 решаются и исследуются особенно просто. Так будет, например, в случае, когда значительная часть матрицы Грама состоит из нулевых элементов. Соответственно тому, какой вид имеют матрицы Грама, мы будем рассматривать различные классы базисов в билинейно метрических пространствах. Класс базисов определяется, конечно, простейшим видом матриц, задающих основную форму (x, y). Напомним, что все эти матрицы конгруэнтны.

13.49. Базис называется *ортогональным*, если его матрица Грама диагональная.

13.50. Среди билинейно метрических пространств имеют ортогональные базисы те и только те пространства, в которых билинейная форма (x, y) симметричная.

13.51. Среди эрмитовых билинейно метрических пространств имеют ортогональные базисы пространства с эрмитовой симметричной и эрмитовой кососимметричной формой (x, y), а также с формой (x, y), имеющей знакопостоянную вещественную или мнимую часть квадратичной формы (x, x).

Отметим сразу одно принципиальное различие между обыкновенными и эрмитовыми билинейно метрическими пространствами с ортогональными базисами. В обыкновенном билинейно метрическом пространстве наличие ортогонального базиса влечет за собой симметрию скалярного произведения (x, y), а это, в свою очередь, обеспечивает существование ортогонального базиса в любом подпространстве. В эрмитовом билинейно метрическом пространстве в общем случае из существования ортогонального базиса в самом пространстве не вытекает автоматически существование ортогонального базиса в любом его подпространстве. Однако если скалярное произведение задано эрмитовой симметричной или эрмитовой кососимметричной билинейной формой, то существование ортогонального базиса в любом подпространстве снова имеет место.

13.52. Если в пространстве существует ортогональный базис, то правое и левое нулевые подпространства совпадают.

13.53. В любом ортогональном базисе изотропные векторы, и только они, образуют базис общего нулевого подпространства.

Ортогональные базисы существуют не во всяком билинейно метрическом и эрмитовом билинейно метрическом пространстве. Это обстоятельство заставляет искать другие классы базисов, более удобные с точки зрения заданного в пространстве скалярного произведения. Решение подсказывается каноническим видом матрицы билинейной формы.

13.54. Базис называется *псевдоортогональным*, если его матрица Грама правая трапециевидная.

13.55. Псевдоортогональный базис существует в любом эрмитовом билинейно метрическом пространстве, а также в любом обыкновенном билинейно метрическом пространстве, кроме пространств с кососимметричной билинейной формой (x, y).

13.56. Для того чтобы базис был псевдоортогональным, необходимо и достаточно, чтобы каждый из его неизотропных векторов был ортогонален слева ко всем предшествующим векторам базиса, а каждый из его изотропных векторов был ортогонален слева ко всем векторам базиса.

13.57. Изотропные векторы псевдоортогонального базиса образуют базис левого нулевого подпространства.

13.58. Существуют пространства, в которых имеются как ортогональный, так и псевдоортогональный базис, не являющийся ортогональным.

13.59. Система векторов, образующих псевдоортогональный базис в своей линейной оболочке, называется *псевдоортогональной*.

Псевдоортогональный базис является достаточно общим типом базиса, так как существует почти во всех пространствах. Как мы уже отмечали, он не существует только в обыкновенных билинейно метрических пространствах с кососимметричной формой (x, y). Вообще говоря, можно ввести тип базиса, покрывающий все рассмотренные типы базисов и существующий во всяком пространстве со скалярным произведением. Однако его введение дает мало новых фактов, и мы не будем на нем останавливаться.

13.60. Пусть в паре базисов матрица билинейной формы является диагональной с элементами 1 или 0 на диагонали. Первый (второй) из этих базисов называется *левым* (*правым*) *двойственным* для второго (первого) базиса.

13.61. Пусть в паре базисов матрица билинейной формы (x, y) является канонической правой (левой) трапециевидной с элементами 1 или 0 на диагонали. Первый (второй) из этих базисов называется *левым* (*правым*) *псевдодвойственным* для второго (первого) базиса.

13.62. В любом невырожденном пространстве каждый базис имеет правый и левый двойственные базисы, и притом единственные.

13.63. В любом невырожденном пространстве каждый базис имеет левый и правый псевдодвойственные базисы.

13.64. В невырожденном пространстве матрица преобразования координат при переходе от одного базиса, псевдодвойственного к заданному, к любому другому псевдодвойственному базису того же наименования является левой треугольной.

13.65. Пусть A — некоторая квадратная матрица в евклидовом или унитарном пространстве. Системы векторов, ортогональные (псевдоортогональные, двойственные, псевдодвойственные) относительно формы (Ax, y), называются *A-ортогональными* (*A-псевдоортогональными*, *A-двойственными*, *A-псевдодвойственными*).

13.66. A-ортогональные (A-псевдоортогональные) системы векторов называются также системами, *сопряженными* (псевдосопряженными) относительно матрицы A.

§ 14. Векторные и матричные нормы

Одним из основных понятий математического анализа является понятие предела. Основано оно на том, что для точек числовой оси определено понятие «близости» или, точнее, расстояния между точками.

Сравнение на «близость» можно ввести и в множествах совсем иной природы. Мы уже определили в § 5 расстояние между векторами линейных пространств со скалярным произведением. При этом было обнаружено, что оно обладает теми же свойствами 5.49, что и расстояние между точками числовой оси. Мы рассмотрим теперь другой, более общий способ введения расстояния между векторами и распространим его также на матрицы. Однако вначале целесообразно привести некоторые общие определения и факты, касающиеся измерений в различных множествах.

14.1. Множество называется *метрическим пространством*, если каждой паре его элементов поставлено в соответствие неотрицательное вещественное число, называемое *расстоянием*, причем выполнены следующие аксиомы:

$$\rho(x, y)=\rho(y, x),$$
$$\rho(x, y)>0, \text{ если } x\neq y,\ \rho(x, y)=0, \text{ если } x=y,$$
$$\rho(x, y)\leqslant\rho(x, z)+\rho(z, y)$$

для любых элементов x, y, z. Эти аксиомы называются *аксиомами метрики*, причем первая из них называется аксиомой *симметрии*, третья — аксиомой *треугольника* (*неравенством треугольника*).

14.2. Элемент x_0 метрического пространства X называется *пределом* последовательности $\{x_i\}$ элементов $x_1, \ldots, x_i, \ldots$ из X, если последовательность расстояний $\rho(x_0, x_1), \ldots, \rho(x_0, x_i), \ldots$ сходится к нулю. Последовательность $\{x_i\}$ называется *сходящейся в X* или просто *сходящейся*. Для указания сходимости последовательности используется символика

$$x_i\to x_0 \quad \text{или} \quad \lim_{i\to\infty} x_i = x_0.$$

14.3. В зависимости от введенной метрики одна и та же последовательность может быть и сходящейся, и несходящейся.

14.4. Если последовательность сходится, то сходится (и имеет тот же предел) любая ее подпоследовательность.

14.5. Последовательность не может иметь более одного предела.

14.6. *Шаром* $S(a, r)$ в метрическом пространстве X называется множество элементов $x \in X$, удовлетворяющих условию $\rho(a, x) < r$. Элемент a называется *центром* шара, число r — *радиусом* шара.

14.7. Любой шар с центром в a называется *окрестностью* элемента a.

14.8. Множество элементов называется *ограниченным*, если оно целиком принадлежит некоторому шару.

14.9. Элемент x_0 является пределом последовательности $\{x_n\}$ тогда и только тогда, когда любая окрестность элемента x_0 содержит все элементы рассматриваемой последовательности, начиная с некоторого номера.

14.10. Элемент x называется *предельной точкой* множества M, если любая окрестность элемента x содержит хотя бы один элемент множества M, не совпадающий с x.

14.11. Множество, полученное присоединением к M всех его предельных точек, называется *замыканием* множества M и обозначается $\overline{M}$.

14.12. Множество M называется *замкнутым*, если $M = \overline{M}$.

14.13. *Замкнутым шаром* $\overline{S}(a, r)$ называется множество элементов x, удовлетворяющих условию $\rho(a, x) \leqslant r$.

14.14. Последовательность $\{x_n\}$ элементов метрического пространства называется *фундаментальной* или *сходящейся в себе*, если для любого числа $\varepsilon > 0$ найдется такое N, что $\rho(x_n, x_m) < \varepsilon$ при $n, m > N$.

14.15. Если последовательность — сходящаяся, то она — фундаментальная.

14.16. Любая фундаментальная последовательность ограничена.

14.17. Метрическое пространство называется *полным*, если любая фундаментальная последовательность в нем является сходящейся.

14.18. Пусть дана некоторая последовательность шаров. Эти шары называются *вложенными* друг в друга, если каждый последующий шар содержится внутри предыдущего.

14.19. Пусть в полном метрическом простанстве X задана последовательность $\{\overline{S}(a_n, r_n)\}$ замкнутых шаров, вложенных друг в друга. Если последовательность радиусов стремится к нулю, то существует единственный элемент из X, принадлежащий всем этим шарам.

При исследовании метрического пространства основное внимание сосредоточивается лишь на одном свойстве множества — наличии в нем расстояния. При исследовании линейного пространства изучаются лишь операции в множестве. Теперь мы рассмотрим линейные пространства с метрикой.

Очевидно, что если понятие расстояния никак не связано с операциями над элементами, то нельзя построить содержательной теории, факты которой соединяли бы вместе алгебраические и метрические понятия. Поэтому мы будем накладывать на метрику, введенную в линейном пространстве, дополнительные условия.

В действительности мы уже встречались с метрическими линейными пространствами. Таковы, например, евклидово и унитарное пространства с метрикой 5.48. Однако необходимость в такой метрике возникает далеко не всегда. Введение скалярного произведения означает, по существу, введение не только расстояния между элементами, но и углов между ними. Чаще же всего в линейном пространстве требуется дать приемлемое определение лишь расстояния. Важнейшими линейными пространствами такого рода являются так называемые нормированные пространства.

Мы снова не акцентируем внимание на арифметических пространствах, так как соответствующие иллюстрации почти очевидны. Единственное, что существенно,— это предположение о конечномерности линейного пространства.

14.20. Вещественное или комплексное линейное пространство K называется *нормированным* пространством, если каждому вектору $x \in K$ поставлено в соответствие вещественное число $\|x\|$, называемое *нормой* вектора x, причем выполнены следующие аксиомы:

$$\|x\| > 0, \text{ если } x \neq 0, \quad \|0\| = 0,$$
$$\|\lambda x\| = |\lambda| \|x\|,$$
$$\|x + y\| \leqslant \|x\| + \|y\|$$

для любых векторов x, y и любого числа λ. Вторая аксиома называется аксиомой *абсолютной однородности* нормы, третья аксиома — *аксиомой треугольника* (*неравенством треугольника*).

14.21. Нормированное пространство становится метрическим, если положить $\rho(x, y) = \|x - y\|$. При этом метрика будет обладать двумя дополнительными к 7.1 свойствами:

$$\rho(x + z, y + z) = \rho(x, y), \quad \rho(\lambda x, \lambda y) = |\lambda| \rho(x, y)$$

для любых векторов x, y и любого числа λ.

14.22. Если в метрическом линейном пространстве K какая-либо метрика обладает дополнительными свойствами 14.21, то K можно рассматривать как нормированное пространство, если определить норму равенством $\|x\| = \rho(x, 0)$ для всех $x \in K$.

14.23. Любое линейное пространство со скалярным произведением становится нормированным, если под нормой вектора понимать его длину.

14.24. Если в пространстве со скалярным произведением задан положительно определенный оператор L, то функция

$$\|x\|_L = (Lx, x)^{1/2}$$

является нормой. Норма этого типа называется *энергетической* нормой, порожденной оператором L, если оператор L самосопряженный.

14.25. Пусть в линейном пространстве векторы заданы своими координатами относительно некоторого базиса. Если $x = (\alpha_1, \ldots$

..., α_n), то функция

$$\|x\|_p=\left(\sum_{k=1}^{n}|\alpha_k|^p\right)^{1/p}$$

при любом $p\geqslant 1$ является нормой. Норма этого типа называется *нормой Гельдера* с показателем p.

14.26. Наиболее распространенными среди гельдеровых норм являются следующие:

$$\|x\|_1=\sum_{k=1}^{n}|\alpha_k|,\quad \|x\|_2=\left(\sum_{k=1}^{n}|\alpha_k|^2\right)^{1/2},\ \|x\|_\infty=\max_{1\leqslant k\leqslant n}|\alpha_k|.$$

Вторая из этих норм часто называется *евклидовой* нормой и обозначается $\|x\|_E$.

14.27. Вектор, норма которого равна единице, называется *нормированным.*

14.28. Любой ненулевой вектор x можно нормировать, умножив его на число $\lambda=\|x\|^{-1}$.

14.29. Для любых векторов x, y, z, u выполняются неравенства

$$|\|x\|-\|y\||\leqslant\|x-y\|,\quad |\|x-y\|-\|z-u\||\leqslant\|x-z\|+\|y-u\|.$$

14.30. Для любых чисел λ_i и векторов x_i выполняется неравенство

$$\left\|\sum_{i=1}^{n}\lambda_i x_i\right\|\leqslant\sum_{i=1}^{n}|\lambda_i|\,|x_i|.$$

14.31. Если в метрике 14.21 для последовательностей векторов $\{x_i\}$, $\{y_i\}$ и чисел $\{\lambda_i\}$ справедливы предельные соотношения

$$x_i\to x,\quad y_i\to y,\quad \lambda_i\to\lambda,$$

то имеют место предельные соотношения

$$\|x_i\|\to\|x\|,\quad x_i+y_i\to x+y,\quad \lambda_i x_i\to\lambda x.$$

14.32. Сходимость последовательности векторов в метрике 14.21 называется *сходимостью по норме*, ограниченность множества векторов — *ограниченностью по норме*, и т. д.

14.33. Пусть последовательность векторов $x_i=\left(\alpha_1^{(i)},\ldots,\alpha_n^{(i)}\right)$ и вектор $x=(\alpha_1,\ldots,\alpha_n)$ заданы своими координатами в некотором базисе. Если для всех j выполняются предельные соотношения $\alpha_j^{(i)}\to\alpha_j$, то говорят, что последовательность векторов x_i сходится *покоординатно* к вектору x.

14.34. Если $x_i\to x$ покоординатно в каком-нибудь одном базисе, то сходимость имеет место и в любом другом базисе.

14.35. Если в нормированном пространстве последовательность векторов ограничена по норме, то ограничены и числовые последовательности всех координат в разложении векторов по любому базису.

14.36. В нормированном пространстве из сходимости по норме вытекает координатная сходимость, и наоборот.

Координатная сходимость эффективно используется в теоретических исследованиях; в практических же приложениях удобнее пользоваться сходимостью по норме. Это объясняется, в основном, тем, что при исследовании линейных пространств большой размерности трудно иметь дело с большим числом координатных последовательностей. К тому же не всегда бывает известен хотя бы один базис. Но даже если базис известен, его использование чаще всего приводит к неоправданно громоздким вычислениям.

14.37. Из всякой ограниченной по норме последовательности векторов нормированного пространства можно выбрать подпоследовательность, сходящуюся по норме в этом пространстве.

14.38. Любое нормированное пространство является полным.

14.39. Любое подпространство нормированного пространства является замкнутым множеством.

14.40. Пусть K — нормированное пространство и L — его подпространство, не совпадающее с K. Существует нормированный вектор $x \notin L$ такой, что $\|x-y\| \geqslant 1$ для любого вектора $y \in L$.

14.41. Если в произвольном нормированном пространстве из всякой ограниченной по норме последовательности векторов можно выбрать сходящуюся последовательность, то пространство конечномерно.

14.42. В нормированном пространстве любые две нормы *эквивалентны*. Это означает, что для любых норм $\|\cdot\|_{\mathrm{I}}$, $\|\cdot\|_{\mathrm{II}}$ существуют такие положительные числа α, β, не зависящие от векторов пространства, что

$$\alpha\|x\|_{\mathrm{I}} \leqslant \|x\|_{\mathrm{II}} \leqslant \beta\|x\|_{\mathrm{I}}$$

для всех векторов x.

14.43. В замкнутом, ограниченном по какой-либо норме множестве векторов существуют векторы, на которых достигаются как нижняя, так и верхняя грани значений любой нормы.

Множество матриц одинаковых размеров является линейным пространством. Его можно сделать нормированным, если ввести норму любым из описанных способов. При этом, конечно, нормированное пространство матриц будет полным, и имеют место все вытекающие отсюда следствия. Однако по отношению к матричным операциям эти метрические факты оказываются несколько обедненными, так как не отражают мультипликативных свойств матриц. Вводя матричные нормы и проводя метрические исследования, обычно в той или иной форме принимают во внимание операции умножения матриц, умножения матрицы на вектор и т. п.

14.44. Пусть каждой матрице A поставлено в соответствие вещественное число $\|A\|$. Это число называется *нормой матрицы*, если выполнены следующие аксиомы:

$$\|A\| > 0, \text{ если } A \neq 0, \quad \|0\| = 0,$$

$$\|\lambda A\| = |\lambda|\,\|A\|,$$

$$\|A+B\| \leqslant \|A\| + \|B\|,$$
$$\|AC\| \leqslant \|A\| \|C\|$$

для любого числа λ и любых матриц A, B, C, для которых соответствующие операции имеют смысл.

14.45. Норма матрицы называется *мультипликативной*, если она удовлетворяет всем четырем аксиомам 14.44.

14.46. Норма матрицы называется *аддитивной* или *обобщенной матричной нормой*, если она удовлетворяет первым трем аксиомам 14.44.

14.47. Умножением на достаточно большую положительную константу всякую аддитивную матричную норму можно превратить в мультипликативную. Наименьшей из таких констант является

$$q = \max_{\|A\|=\|C\|=1} \|AC\|.$$

Последнее утверждение означает, что при изучении матричных норм можно ограничиться рассмотрением лишь одного вида норм, например мультипликативных. Поэтому в дальнейшем, если не сделано специальной оговорки, под нормой матрицы будет пониматься, как правило, мультипликативная норма.

14.48. Пусть матрица A с элементами a_{ij} имеет размер $n \times m$. Наиболее употребительными являются следующие нормы:

$$\|A\|_1 = \max_{1 \leqslant j \leqslant m} \sum_{i=1}^{n} |a_{ij}|;$$

$$\|A\|_\infty = \max_{1 \leqslant i \leqslant n} \sum_{j=1}^{m} |a_{ij}|;$$

$$M(A) = (nm)^{1/2} \max_{i,j} |a_{ij}|;$$

$\|A\|_2 =$ максимальному сингулярному числу матрицы;

$$\|A\|_E = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} |a_{ij}|^2 \right)^{1/2}.$$

Первые три нормы называются, соответственно 1-, ∞- и M-нормой, четвертая норма называется *спектральной*, последняя — *евклидовой*. Однако иногда всем этим нормам даются и другие названия. Например, евклидова норма называется также *сферической*, спектральная — *нижней гранью матрицы*, 1-норма — *второй*, ∞-норма — *первой* и т. п.

14.49. Имеют место следующие соотношения эквивалентности:

$$(nm)^{-1/2} M(A) \leqslant \|A\|_1 \leqslant \left(\frac{n}{m}\right)^{1/2} M(A),$$

$$(nm)^{-1/2} M(A) \leqslant \|A\|_\infty \leqslant \left(\frac{m}{n}\right)^{1/2} M(A),$$

$$(nm)^{-1/2} M(A) \leqslant \|A\|_2 \leqslant M(A),$$
$$(nm)^{-1/2} M(A) \leqslant \|A\|_E \leqslant M(A),$$
$$(\min\{m, n\})^{-1/2} \|A\|_E \leqslant \|A\|_2 \leqslant \|A\|_E,$$
$$m^{-1/2} \|A\|_E \leqslant \|A\|_1 \leqslant n^{1/2} \|A\|_E,$$
$$n^{-1/2} \|A\|_E \leqslant \|A\|_\infty \leqslant m^{1/2} \|A\|_E,$$
$$m^{-1/2} \|A\|_2 \leqslant \|A\|_1 \leqslant n^{1/2} \|A\|_2,$$
$$n^{-1/2} \|A\|_2 \leqslant \|A\|_\infty \leqslant m^{1/2} \|A\|_2,$$
$$m^{-1} \|A\|_\infty \leqslant \|A\|_1 \leqslant n \|A\|_\infty.$$

14.50. Для евклидовой нормы матрицы справедливы представления

$$\|A\|_E^2 = \operatorname{tr} AA^* = \operatorname{tr} A^*A = \left(\sum_{i=1}^{t} \rho_i^2\right)^{1/2},$$

где $\rho_1, \ldots, \rho_t$ — ненулевые сингулярные числа матрицы A.

14.51. Евклидова и спектральная нормы не меняются при умножении матрицы справа и слева на любые унитарные матрицы.

14.52. Для евклидовой нормы произведения матриц имеют место более точные оценки:

$$\|AC\|_E \leqslant \|A\|_2 \|C\|_E, \quad \|AC\|_E \leqslant \|A\|_E \|C\|_2.$$

14.53. Евклидова и спектральная нормы матрицы совпадают тогда и только тогда, когда ранг матрицы равен 1.

14.54. Аддитивная или мультипликативная норма матрицы называется *согласованной* с векторными нормами, если $\|Ax\| \leqslant \leqslant \|A\| \|x\|$ для всех векторов x.

14.55. Всякая норма матриц согласована с какими-нибудь нормами векторов.

14.56. Пусть заданы любые векторные нормы. Числовая функция

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|=1} \|Ax\|$$

является, по крайней мере, аддитивной матричной нормой и называется нормой матрицы, *подчиненной* заданным векторным нормам.

14.57. Среди всех норм, согласованных с заданными векторными нормами, подчиненная норма является минимальной.

14.58. 1-, ∞- и спектральная нормы из 14.48 являются подчиненными по отношению, соответственно, к 1-, ∞- и 2-нормам векторов.

14.59. M-норма из 14.48 является согласованной по отношению к 1-, ∞- и 2-нормам векторов.

14.60. Евклидова норма из 14.48 согласована только с 2-нормами векторов.

14.61. Для квадратных матриц и одинаковых норм в пространствах векторов образов и прообразов подчиненная норма 14.56 всегда будет мультипликативной.

14.62. Для любой мультипликативной нормы $\|E\| \geqslant 1$.

14.63. В условиях 14.56 всегда $\|E\| = 1$.

§ 15. Функционалы в евклидовом пространстве

15.1. *Функционалом* в линейном пространстве называется числовая функция, аргументы которой суть векторы пространства.

Функционалы очень широко используются в линейной алгебре. Так, функционалами являются норма, скалярное произведение, объем системы векторов, определитель Грама, билинейная и квадратичная формы и т. п. Мы рассмотрим сейчас более подробно некоторые специальные функционалы от одного векторного аргумента. При этом ограничимся изучением функционалов, в основном, в евклидовом пространстве. Наличие скалярного произведения дает возможность задавать многие функционалы в простой форме, а вещественность функционалов и пространств позволяет упростить исследования. Комплексные функционалы применяются значительно реже вещественных, вещественные же функционалы в унитарном пространстве довольно часто сводятся к вещественным функционалам в евклидовом пространстве.

15.2. Функционал $f(x)$ называется *линейным*, если для любых векторов x, y и любого числа α выполняются соотношения

$$f(x+y) = f(x) + f(y), \quad f(\alpha x) = \alpha f(x).$$

Функционал, не являющийся линейным, называется *нелинейным*.

15.3. В пространстве со скалярным произведением любой линейный функционал $f(x)$ может быть представлен в виде скалярного произведения (x, f) для некоторого вектора f, однозначно определяемого функционалом.

15.4. Среди нелинейных функционалов наиболее употребительными являются следующие:

$$(Ax, x) - 2(b, x) + c \quad (\textit{функционал ошибки});$$

$$\frac{(Ax, x)}{(x, x)} \quad (\textit{отношение Релея});$$

$$\frac{(Ax, x)}{(Bx, x)} \quad (\textit{обобщенное отношение Релея});$$

$$(Ax - b, Ax - b) \quad (\textit{функционал невязки}).$$

Здесь A, B — квадратные матрицы, b — вектор, c — число.

15.5. Первые три функционала из 15.4 не изменятся, если к матрицам A и B прибавить любые кососимметричные матрицы.

Последнее утверждение означает, что без ограничения общности можно считать, что матрицы A и B в первых трех функционалах из 15.4 симметричные. Поэтому всюду в дальнейшем мы будем предполагать выполненным это условие, если не сделано какой-либо оговорки.

15.6. *Производной* функционала $F(x)$ в точке x по направлению y называется выражение

$$\frac{\partial F(x)}{\partial y} = \lim_{t \to 0} \frac{F(x+ty) - F(x)}{t},$$

если предел существует.

15.7. Для функционалов $F(x)$ из 15.4 в каждой точке x, где эти функционалы определены, существует производная по любому направлению y. Как функция от y, она является линейным функционалом, т. е.

$$\frac{\partial F}{\partial y} = (y, z).$$

Соответствующий вектор z называется *градиентом* функционала $F(x)$ в точке x и обозначается $\operatorname{grad} F(x)$.

15.8. Для того чтобы производная функционала в некоторой точке по направлению y равнялась нулю, необходимо и достаточно, чтобы градиент в этой точке был ортогонален вектору y.

15.9. Для того чтобы градиент функционала в некоторой точке равнялся нулю, необходимо и достаточно, чтобы он был ортогонален ко всем векторам какого-нибудь базиса.

15.10. В окрестности точки x функционал $F(x)$ растет (убывает) быстрее всего в направлении $\operatorname{grad} F(x)$ $(-\operatorname{grad} F(x))$.

15.11. Имеют место следующие формулы:

$$\operatorname{grad}(x, f) = f, \quad \operatorname{grad}(Ax, x) = 2Ax,$$
$$\operatorname{grad}((Ax, x) - 2(b, x) + c) = 2(Ax - b),$$
$$\operatorname{grad}\frac{(Ax, x)}{(x, x)} = \frac{2}{(x, x)}\left(Ax - \frac{(Ax, x)}{(x, x)}x\right),$$
$$\operatorname{grad}\frac{(Ax, x)}{(Bx, x)} = \frac{2}{(Bx, x)}\left(Ax - \frac{(Ax, x)}{(Bx, x)}Bx\right),$$
$$\operatorname{grad}(Ax - b, Ax - b) = 2A^*(Ax - b).$$

Один из эффективных способов изучения нелинейных функционалов состоит в исследовании их поведения на прямых линиях в пространстве.

15.12. Если $(Al, l) > 0$ (<0) для некоторого вектора l, то на любой прямой с направляющим вектором l функционал ошибки достигает минимума (максимума).

15.13. Если $(Al, l) \neq 0$, то множество экстремальных точек x функционала ошибки на прямых линиях с направляющим вектором l есть гиперплоскость $(Ax - b, l) = 0$.

15.14. Для гиперплоскости $(Ax - b, l) = 0$ вектор Al является нормальным.

15.15. Прямая линия с направляющим вектором l пересекает гиперплоскость $(Ax-b,\ l)=0$ в одной точке тогда и только тогда, когда $(Al,\ l)\neq 0$.

15.16. Если $(Al,\ l)=0$, то прямая с направляющим вектором l принадлежит гиперплоскости $(Ax-b,\ l)=0$ в том и только в том случае, когда функционал ошибки остается постоянным на этой прямой.

15.17. Если $(Al,\ l)=0$, то прямая с направляющим вектором l параллельна гиперплоскости $(Ax-b,\ l)=0$ в том и только в том случае, когда функционал ошибки изменяется линейно на этой прямой.

15.18. Множество точек, в которых градиент функционала ошибок 15.4 равен нулю, совпадает с множеством решений системы $Ax=b$.

15.19. Если матрица A является неотрицательно (неположительно) определенной, то множество точек x минимума (максимума) функционала ошибок 15.4 совпадает с множеством решений системы $Ax=b$.

15.20. Функционал ошибок 15.4 имеет единственную точку минимума (максимума) в том и только в том случае, когда матрица A положительно (отрицательно) определена.

15.21. Пусть $l_1, \ldots, l_n$ — произвольный A-ортогональный базис пространства. Число его векторов, на которых квадратичная форма $(Al,\ l)$ принимает положительное (отрицательное, нулевое) значение, не зависит от выбора базиса и совпадает с числом положительных (отрицательных, нулевых) собственных значений матрицы A.

15.22. Предположим, что задан A-ортогональный базис $l_1, \ldots$ $\ldots, l_n$, причем на векторах $l_1, \ldots, l_r$ квадратичная форма $(Al,\ l)$ не равна нулю, на остальных — равна нулю. Рассмотрим гиперплоскости

$$(Ax-b,\ l_1)=0,\quad (Ax-b,\ l_2)=0, \ldots, (Ax-b,\ l_r)=0.$$

Пересечение этих гиперплоскостей есть плоскость, размерность которой равна размерности ядра матрицы A.

15.23. Пересечение гиперплоскостей 15.22 совпадает с множеством псевдорешений системы $Ax=b$ тогда и только тогда, когда векторы $l_1, \ldots, l_r$ принадлежат образу матрицы A.

15.24. Пусть в условиях и обозначениях 15.22 вектор x_s принадлежит первым s гиперплоскостям, где $0\leqslant s<r$. Проведем через x_s прямую с направляющим вектором l_{s+1} и на этой прямой найдем экстремальную точку x_{s+1} функционала ошибки 15.4. Вектор x_{s+1} принадлежит первым $s+1$ гиперплоскостям 15.22.

15.25. При любом векторе x_0 процесс 15.24 после r шагов для $s=0, 1, \ldots, r-1$ приводит к какому-нибудь вектору, принадлежащему пересечению гиперплоскостей 15.22. Число «спусков» и «подъемов» по функционалу ошибки не зависит от вектора x_0 и совпадает соответственно с числом положительных и отрицательных собственных значений матрицы A.

15.26. Перпендикуляр, опущенный из вектора, полученного согласно 15.25, на линейную оболочку векторов $l_{r+1}, \ldots, l_n$ из 15.22, есть нормальное псевдорешение системы $Ax = b$, если выполнено условие 15.23.

15.27. Для любого вектора x^* справедливо тождество

$$(Ax, x) - 2(b, x) + c \equiv (A(x - x^*), x - x^*) + \\ + 2(Ax^* - b, x - x^*) + (Ax^*, x^*) - 2(b, x^*) + c.$$

15.28. Вектор x^* называется *центром симметрии* функционала $F(x)$, если $F(x^* + y) = F(x^* - y)$ для любого вектора y.

15.29. Решения системы $Ax = b$, и только они, являются центрами симметрии функционала ошибки 15.4.

15.30. Если $(Al, l) \neq 0$ для некоторого вектора l, то на любой прямой с направляющим вектором l функционал ошибки есть симметричная функция. Множество центров симметрии на таких прямых линиях принадлежит гиперплоскости $(Ax - b, l) = 0$. Эта гиперплоскость называется *диаметральной* гиперплоскостью, сопряженной вектору l относительно функционала ошибки 15.4.

15.31. Если матрица A невырожденная, то A-ортогональный базис называется также *системой сопряженных диаметров* функционала ошибки 15.4.

Если матрица A положительно определенная, то согласно 15.20, 15.27 минимизация отклонения вектора x от решения системы $Ax = b$ равносильна минимизации функционала ошибки. Собственно говоря, только данным обстоятельством и объясняется название рассмотренного функционала как функционала ошибки. Этот функционал в ряде случаев называется также *функционалом энергии.*

15.32. Каковы бы ни были матрица A и вектор x, минимальное значение $\|Ax - \lambda x\|_2$ достигается на числе λ, равном отношению Релея, т. е.

$$\left\| Ax - \frac{(Ax, x)}{(x, x)} x \right\|_2 \leqslant \| Ax - \lambda x \|_2$$

для любого числа λ.

Если вектор x и число λ являются собственным вектором и собственным значением матрицы A, то выполняется равенство $Ax - \lambda x = 0$. Если вектор x и число λ или одно из них не является таковым, то норма невязки $Ax - \lambda x$ есть мера того, насколько вектор x и число λ близки совместно к собственному вектору и собственному значению матрицы A. Утверждение 15.32 говорит о том, что если вектор x считается «приближением» к собственному вектору, то отношение Релея оказывается лучшим «приближением» к соответствующему собственному значению в смысле малости евклидовой нормы невязки $\|Ax - \lambda x\|_2$, чем любые другие числа.

Многие свойства отношения Релея в значительной мере связаны с собственными векторами и собственными значениями матрицы. Далее в этом параграфе мы будем считать, что собственные значения $\lambda_1, \ldots, \lambda_n$ симметричной матрицы A упорядочены по алгебраическому возрастанию. Именно:

$$\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n.$$

15.33. Градиент ξ отношения Релея в точке x обладает следующими свойствами:

$$(\xi, x) = 0, \quad (\xi, \xi) = (\xi, Ax).$$

15.34. Как функция векторов x отношение Релея есть однородная функция нулевой степени.

15.35. Градиент отношения Релея 15.4 обращается в нуль на собственных векторах матрицы A. Значение отношения Релея в этих точках равно соответствующему собственному значению матрицы A.

15.36. На любой прямой линии, направляющий вектор которой не является собственным вектором матрицы A, отношение Релея имеет два экстремума.

15.37. Для любого ненулевого вектора x справедливы неравенства

$$\lambda_1 \leqslant \frac{(Ax, x)}{(x, x)} \leqslant \lambda_n.$$

15.38. Для минимального и максимального собственных значений имеют место следующие представления:

$$\lambda_1 = \min_{x \neq 0} \frac{(Ax, x)}{(x, x)}, \quad \lambda_n = \max_{x \neq 0} \frac{(Ax, x)}{(x, x)}.$$

15.39. Для линейного подпространства L, натянутого на собственнные векторы $x_{i_1}, \ldots, x_{i_k}$ $(i_1 < \ldots < i_k)$ матрицы A, выполняются соотношения

$$\lambda_{i_1} = \min_{\substack{x \neq 0 \\ x \in L}} \frac{(Ax, x)}{(x, x)}, \quad \lambda_{i_k} = \max_{\substack{x \neq 0 \\ x \in L}} \frac{(Ax, x)}{(x, x)}.$$

15.40. (*Теорема Куранта — Фишера.*) Для собственного значения λ_k симметричной матрицы A имеют место представления

$$\lambda_k = \max_{L_{n-k+1}} \min_{\substack{x \neq 0 \\ x \in L_{n-k+1}}} \frac{(Ax, x)}{(x, x)}, \quad \lambda_k = \min_{L_k} \max_{\substack{x \neq 0 \\ x \in L_k}} \frac{(Ax, x)}{(x, x)}.$$

Здесь L_k, L_{n-k+1} — произвольные подпространства размерностей соответственно k и $n-k+1$.

15.41. Если матрица B положительно определенная, то обобщенное отношение Релея сводится к обычному отношению Релея, так как

$$\frac{(Ax, x)}{(Bx, x)} = \frac{(B^{-1/2}AB^{-1/2}y, y)}{(y, y)},$$

где $y = B^{1/2}x$.

15.42. Для любой матрицы A функционал невязки сводится к функционалу ошибки, так как

$$(Ax - b, Ax - b) = (A^*Ax, x) - 2(A^*b, x) + (b, b).$$

15.43. По отношению к функционалу ошибки общего вида функционал невязки обладает следующими свойствами:

— система линейных алгебраических уравнений $A^*Ax = A^*b$ всегда имеет решение;

— матрица A^*A неотрицательно определенная;

— если векторы $l_1, \ldots, l_n$ образуют A^*A-ортогональную систему, то векторы $Al_1, \ldots, Al_n$ образуют ортогональную систему.

Все изложенные здесь факты без существенного изменения переносятся на функционалы с комплексными эрмитовыми матрицами. Несколько меняется лишь вид функционала ошибки. Теперь он выглядит так: $(Ax, x) - 2\operatorname{Re}(b, x) + c$. Естественно, что в функционале невязки матрица A может быть произвольной.

§ 16. Возмущения и локализация

Различные объекты линейной алгебры, такие, как решение системы линейных алгебраических уравнений, собственные и сингулярные числа и векторы, определитель, функционалы и т. п., можно рассматривать как некоторые функции от определяющих их матриц и векторов. При изменении матриц и векторов изменяются и сами объекты. Теория возмущений акцентирует внимание на исследовании взаимосвязи этих изменений, или, как их называют иначе, *возмущений*. Как правило, изучение больших возмущений связывается с локализацией объектов, малых возмущений — с асимптотическими разложениями. Всюду в данном параграфе мы рассматриваем только согласованные нормы матриц.

16.1. Если для квадратной матрицы H имеем $\|H\| < 1$ для какой-либо нормы, то матрица $E + H$ — невырожденная, и при этом

$$(E + H)^{-1} = E + \sum_{k=1}^{\infty} (-H)^k.$$

16.2. В условиях 16.1 выполняются следующие неравенства:

$$\|(E + H)^{-1}\| \leqslant \frac{1}{1 - \|H\|}, \quad \|(E + H)^{-1} - E\| \leqslant \frac{\|H\|}{1 - \|H\|}.$$

16.3. Пусть матрица A — невырожденная и $\|\mathscr{E}\| < \|A^{-1}\|^{-1}$ для возмущения $\mathscr{E}$. Тогда матрица $A + \mathscr{E}$ — невырожденная.

16.4. Для невырожденной матрицы A выражение $\|A\|\|A^{-1}\|$ называется *числом обусловленности* матрицы A и обозначается $\operatorname{cond} A$.

16.5. Всегда $\operatorname{cond} A \geqslant 1$.

16.6. Для любой матрицы A и любого числа α выполняется равенство $\operatorname{cond} A = \operatorname{cond}(\alpha A)$.

16.7. Для спектральной нормы равенство $\operatorname{cond} A = 1$ имеет место тогда и только тогда, когда $A = \alpha U$, где U — унитарная матрица, α — ненулевое число.

16.8. Для спектральной и евклидовой норм число обусловленности не меняется от умножения матрицы слева и справа на любые унитарные матрицы.

16.9. Для 1-, 2-, ∞-норм и евклидовой нормы число обусловленности не меняется при перестановке строк и столбцов.

16.10. Для спектральной нормы число обусловленности любой нормальной матрицы равно отношению максимального модуля собственных значений к их минимальному модулю.

16.11. Для спектральной нормы число обусловленности любой матрицы равно отношению максимального сингулярного числа к минимальному.

16.12. Имеют место неравенства

$$\max\left\{\frac{\operatorname{cond} A}{\operatorname{cond} B}, \frac{\operatorname{cond} B}{\operatorname{cond} A}\right\} \leqslant \operatorname{cond}(AB) \leqslant \operatorname{cond} A \operatorname{cond} B.$$

16.13. Пусть A — положительно определенная матрица, B — любая ее главная подматрица. Тогда для спектральной нормы $\operatorname{cond} B \leqslant \operatorname{cond} A$.

16.14. Если матрица A невырожденная, а матрица $A+B$ вырожденная, то $\operatorname{cond} A \geqslant \|A\|\|B\|^{-1}$.

16.15. Для любых невырожденных матриц A, B справедливо неравенство

$$\frac{\|B^{-1}-A^{-1}\|}{\|B^{-1}\|} \leqslant \operatorname{cond} A \frac{\|B-A\|}{\|A\|}.$$

16.16. Пусть дана последовательность матриц A_k фиксированного порядка, причем $\|A_k\| = 1$. Тогда условия $\operatorname{cond} A_k \to \infty$ и $\det A_k \to 0$ при $k \to \infty$ эквивалентны.

16.17. Существуют последовательности матриц A_k возрастающего порядка, для которых $\|A_k\| = 1$ и $\operatorname{cond} A_k \to \infty$ при $k \to \infty$, но $\det A_k = 1$ для всех k.

16.18. Пусть в условиях и обозначениях 16.3

$$\delta A = \frac{\|\mathscr{E}\|}{\|A\|}, \quad \delta A^{-1} = \frac{\|(A+\mathscr{E})^{-1}-A^{-1}\|}{\|A^{-1}\|}.$$

Имеет место неравенство

$$\delta A^{-1} \leqslant \frac{\operatorname{cond} A\cdot\delta A}{1-\operatorname{cond} A\cdot\delta A}.$$

16.19. Предположим, что решаются системы линейных алгебраических уравнений $Ax = b$ и $(A+\mathscr{E})\hat{x} = b + \varepsilon$. Обозначим

$$\delta x = \frac{\|\hat{x}-x\|}{\|x\|}, \quad \delta b = \frac{\|\varepsilon\|}{\|b\|}.$$

В условиях и обозначениях 16.3, 16.18 имеет место неравенство

$$\delta x \leqslant \frac{\operatorname{cond} A}{1-\operatorname{cond} A\cdot\delta A}(\delta A + \delta b).$$

Неравенства 16.18, 16.19 дают количественные оценки возмущения обратной матрицы и решения системы линейных алгебраических уравнений при возмущении матрицы и правой части. Из них вытекает, что в окрестности любой невырожденной матрицы обратная матрица и решение системы являются непрерывными функциями входных данных. Непрерывность решения по правой части имеет место всюду. При этом величина возмущения как обратной матрицы, так и решения системы существенно зависит от числа обусловленности матрицы.

16.20. *Спектральным радиусом* $\rho(A)$ квадратной матрицы A называется максимальный из модулей ее собственных значений.

16.21. Для нормальной матрицы A порядка n с элементами a_{ij}

$$\frac{1}{n}\left|\sum_{i,j=1}^{n} a_{ij}\right| \leqslant \rho(A) = \max_{x\neq 0}\frac{|Ax|}{|x|}.$$

16.22. Для любой согласованной нормы и любой матрицы A

$$|\det A|^{1/n} \leqslant \rho(A) \leqslant \min\{\|A\|,\ \|A^*\|\}.$$

16.23. По крайней мере одно собственное значение матрицы A находится в круге $|\lambda| \leqslant |\det A|^{1/n}$.

16.24. Для любой матрицы A порядка n справедливы неравенства

$$\sum_{i=1}^{n}|\lambda_i|^2 \leqslant \|A\|_E^2, \quad \sum_{i=1}^{n}|\mathrm{Re}\,\lambda_i|^2 \leqslant \|B\|_E^2, \quad \sum_{i=1}^{n}|\mathrm{Im}\,\lambda_i|^2 \leqslant \|C\|_E^2,$$

где $B = 0{,}5(A + A^*)$, $C = 0{,}5(A - A^*)$, причем равенства достигаются тогда и только тогда, когда матрица A нормальная.

16.25. Рассмотрим нормальную матрицу A, нормированный вектор v, $\|v\|_2 = 1$, и пусть $Av - \mu v = \eta$, $\|\eta\|_2 = \varepsilon$, для некоторого числа μ. Предположим, что одно из собственных значений, например λ_1, изолировано и близко к μ, т. е. существует такая константа a, что $|\lambda_i - \mu| \geqslant a > 0$ для $i = 2, \ldots, n$. Если x_1 — нормированный собственный вектор, соответствующий λ_1, то найдется такое вещественное число θ, что

$$\left\|e^{i\theta}v - x_1\right\|_2 \leqslant \frac{\varepsilon^2}{a^2}\left(1 + \frac{\varepsilon^2}{a^2}\right).$$

16.26. Если в условиях и обозначениях 16.25 число μ есть отношение Релея, вычисленное для вектора v, то

$$|\mu - \lambda_1| \leqslant \frac{\varepsilon^2}{a}\left(1 - \frac{\varepsilon^2}{a^2}\right)^{-1}.$$

Общий принцип построения областей, локализующих собственные значения, основан на следующей идее. Пусть A — произвольная матрица и $B(A)$ — некоторое арифметическое условие, выполнение которого достаточно для невырожденности матрицы A. Если λ является собственным значением, то матрица $A - \lambda E$ — вырожденная. Поэтому для того, чтобы λ было собственным значением матрицы A, необходимо невыполнение условия $B(A - \lambda E)$. Это и определяет некоторую область, в которой должны находиться все собственные значения.

16.27. Для того чтобы матрица A была невырожденной, достаточно выполнения неравенств

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|$$

для всех i. Матрица, элементы которой удовлетворяют этим неравенствам, называется матрицей с *доминирующей диагональю*.

16.28. Эрмитова матрица с положительной доминирующей диагональю является положительно определенной.

16.29. Любое собственное значение матрицы A лежит по крайней мере в одном из кругов

$$|\lambda - a_{ii}| \leqslant \sum_{j \neq i} |a_{ij}|$$

для всех i. Эти круги называются *кругами Гершгорина*.

16.30. Если s кругов Гершгорина образуют область G, изолированную от остальных кругов, то в G находится ровно s собственных значений матрицы A.

16.31. Если какой-либо круг Гершгорина изолирован, то он содержит точно одно собственное значение.

16.32. Если при некотором i для всех $k \neq i$ выполняются неравенства

$$|a_{kk} - a_{ii}| > \sum_{j \neq k} |a_{kj}| + \sum_{j \neq i} |a_{ij}|,$$

то круг Гершгорина

$$|\lambda - a_{ii}| \leqslant \sum_{j \neq i} |a_{ij}|$$

содержит точно одно собственное значение.

16.33. Пусть у матрицы A элементы главной диагонали и коэффициенты характеристического многочлена — вещественные. Если круги Гершгорина попарно не пересекаются, то все собственные значения матрицы A — вещественные.

16.34. Если λ — собственное значение матрицы A и дефект матрицы $A - \lambda E$ равен m, то λ лежит по крайней мере в m кругах Гершгорина.

16.35. Пусть $A = B + C$, где B — диагональная матрица с элементами $b_1, \ldots, b_n$, C — матрица с элементами c_{kl}. Если порядок матриц равен n и для некоторого j имеем

$$\min_{k \neq j} |b_j - b_k| = \beta_j > 0, \quad \max_{k,l} |c_{kl}| = \varepsilon < \frac{\beta_j}{2n},$$

то существует собственное значение λ матрицы A в круге

$$|\lambda - b_j - c_{jj}| \leqslant 2n(n-1)\varepsilon^2/\beta_j.$$

16.36. Пусть для элементов матрицы A известны оценки $|a_{ij}| < b_{ij}$ и μ есть максимальное собственное значение матрицы

B с элементами b_{ij}. Тогда любое собственное значение матрицы A лежит по крайней мере в одном из кругов

$$|\lambda - a_{ii}| \leqslant \mu - b_{ii}$$

для всех i.

16.37. Любое собственное значение матрицы A лежит по крайней мере в одном из кругов

$$|\lambda - a_{ii}| \leqslant \Big(\sum_{j \neq i} |a_{ij}|\Big)^{\alpha} \Big(\sum_{j \neq i} |a_{ji}|\Big)^{1-\alpha}$$

для всех i при $0 \leqslant \alpha \leqslant 1$.

16.38. Матрица, которая перестановкой одноименных строк и столбцов не может быть приведена к блочно треугольному виду, называется *неразложимой*.

16.39. Пусть для неразложимой матрицы A выполняются неравенства

$$|a_{ii}| \geqslant \sum_{j \neq i} |a_{ij}|$$

для всех i, причем хотя бы для одного i имеет место строгое неравенство. Тогда матрица A — невырожденная.

16.40. Все собственные значения неразложимой матрицы лежат внутри объединения кругов Гершгорина, за исключением того случая, когда собственное значение является общей граничной точкой всех кругов,

16.41. Для того чтобы матрица A была невырожденной, достаточно выполнения неравенств

$$|a_{ii}||a_{jj}| > \sum_{k \neq i} |a_{ik}| \sum_{k \neq j} |a_{jk}|$$

для всех $i \neq j$.

16.42. Любое собственное значение матрицы A лежит по крайней мере в одной из областей

$$|\lambda - a_{ii}||\lambda - a_{jj}| \leqslant \sum_{k \neq i} |a_{ik}| \sum_{k \neq j} |a_{jk}|$$

для всех $i \neq j$. Эти области называются *овалами Кассини*.

16.43. Любое собственное значение матрицы A лежит по крайней мере в одном из овалов

$$|\lambda - a_{ii}||\lambda - a_{jj}| \leqslant \Big(\sum_{k \neq i} |a_{ik}| \sum_{k \neq j} |a_{jk}|\Big)^{\alpha} \Big(\sum_{k \neq i} |a_{ki}| \sum_{k \neq j} |a_{kj}|\Big)^{1-\alpha}$$

для всех $i \neq j$ при $0 \leqslant \alpha \leqslant 1$.

Мы рассмотрели различные способы локализации собственных значений матрицы A. Естественно, что аналогично можно локализовать и собственные значения транспонированной матрицы A'. Во многих случаях совместное рассмотрение обеих матриц позволяет получить лучшие оценки.

Локализация, вообще говоря, позволяет одинаковым образом исследовать как большие, так и малые возмущения. Однако, как показывают утверждения 16.26, 16.35, в случае малых возмущений можно ожидать получения более точных оценок. Поэтому мы приведем сейчас некоторые результаты, касающиеся асимптотически малых возмущений. По существу, они связаны с разложениями в ряды по малым возмущениям и выделением главных членов этих разложений. Асимптотически верные соотношения будем отмечать символами $\cong$, $\lessapprox$ и т. п., возмущенные объекты — символом $\wedge$ вверху.

16.44. Если λ_i — простое собственное значение матрицы A, x_i, y_i — соответствующие собственные векторы матриц A и A^*, то величина

$$c_i = \frac{|x_i|\,|y_i|}{|(x_i, y_i)|}$$

называется *коэффициентом перекоса* матрицы A, соответствующим собственному значению λ_i.

16.45. В вещественном случае коэффициент перекоса c_i равен обратной величине косинуса угла между x_i и y_i.

16.46. Всегда $c_i \geqslant 1$, причем равенство достигается тогда и только тогда, когда вектор x_i ортогонален корневому базису, соответствующему собственным значениям, не равным λ_i.

16.47. Пусть матрица A имеет только простые собственные значения. Все коэффициенты перекоса равны единице тогда и только тогда, когда матрица A нормальная.

16.48. Обозначим через X матрицу, составленную из собственных векторов матрицы A. Число обусловленности матрицы X, выраженное в спектральной норме, не меньше любого из коэффициентов перекоса.

16.49. Матрицу X из 16.48 за счет умножения собственных векторов на подходящие множители всегда можно выбрать такой, что ее число обусловленности, выраженное в евклидовой норме, равно сумме всех коэффициентов перекоса.

16.50. Пусть матрица A имеет только простые собственные значения и матрица $\hat{A} = A + \Delta$ близка к ней. Предположим, что собственные векторы матриц A и A^* нормированы по 2-норме. С точностью до членов второго порядка имеют место следующие соотношения:

$$\hat{\lambda}_i - \lambda_i \cong \frac{(\Delta x_i, y_i)}{(x_i, y_i)}, \quad |\hat{\lambda}_i - \lambda_i| \lessapprox c_i \|\Delta\|_2,$$

$$\hat{x}_i - x_i \cong \sum_{j \neq i} \frac{(\Delta x_i, y_j)}{(x_j, y_j)(\lambda_i - \lambda_j)} x_j,$$

$$\|\hat{x}_i - x_i\|_2 \lessapprox \|\Delta\|_2 \sum_{j \neq i} \frac{c_j}{|\lambda_i - \lambda_j|}.$$

При исследовании влияния малого возмущения удобно предварительно привести задачу к некоторому простейшему виду с помощью специальных преобразований. Такими простейшими задачами почти всегда оказываются задачи с диагональными матрицами и лишь иногда с двухдиагональными. В большинстве случаев подобные преобразования можно считать унитарными. Тогда спектральная и евклидова нормы возмущений исходной задачи и приведенной будут одинаковыми. Мы не будем специально останавливаться на способах сведения задач к простейшему виду в силу их очевидности, но рассмотрим асимптотическое влияние малых возмущений в некоторых задачах с простыми матрицами.

16.51. Пусть даны матрицы A, B, C размеров соответственно $n\times n$, $m\times m$, $n\times m$. Матричное уравнение $AZ-ZB=C$ относительно матрицы Z размера $n\times m$ имеет единственное решение тогда и только тогда, когда матрицы A и B не имеют общих собственных значений.

16.52. Пусть Λ — блочно диагональная матрица с блоками $\Lambda_1, \ldots, \Lambda_r$ и $\Lambda+\Omega$ — возмущенная матрица. Обозначим

$$\Lambda=\begin{bmatrix}\Lambda_1 & & & 0\\ & \Lambda_2 & & \\ & & \ddots & \\ 0 & & & \Lambda_r\end{bmatrix},\quad \Omega=\begin{bmatrix}\Omega_{11} & \Omega_{12} & \ldots & \Omega_{1r}\\ \Omega_{21} & \Omega_{22} & \ldots & \Omega_{2r}\\ \cdot & \cdot & \cdot & \cdot\\ \Omega_{r1} & \Omega_{r2} & \ldots & \Omega_{rr}\end{bmatrix},$$

где диагональные блоки матриц Λ и Ω имеют одинаковые размеры. Предположим, что попарно блоки матрицы Λ не имеют общих собственных значений. Тогда матрица $\Lambda+\Omega$ асимптотически подобна блочно диагональной матрице

$$\widehat{\Lambda}=\begin{bmatrix}\widehat{\Lambda}_1 & & & 0\\ & \widehat{\Lambda}_2 & & \\ & & \ddots & \\ 0 & & & \widehat{\Lambda}_r\end{bmatrix},$$

где блоки $\widehat{\Lambda}$ и Λ имеют одинаковые размеры и с точностью до членов второго порядка $\widehat{\Lambda}_k\cong\Lambda_k+\Omega_{kk}$ для всех k.

16.53. Пусть в условиях и обозначениях 16.52 выполняется соотношение

$$\widehat{\Lambda}\cong(E+H)^{-1}(\Lambda+\Omega)(E+H).$$

Представим матрицу H в блочном виде, аналогично Ω. Тогда для ее блоков H_{lk} с точностью до членов второго порядка выполняются соотношения

$$H_{kk}=0,\quad H_{lk}\Lambda_k-\Lambda_l H_{lk}=\Omega_{lk}$$

для всех k и $l\neq k$.

16.54. В условиях и обозначениях 16.52 с точностью до членов третьего порядка

$$\widehat{\Lambda}_k \cong \Lambda_k + \Omega_{kk} + \sum_{\substack{i=1\\ i\neq k}}^{r} \Omega_{ki} H_{ik}$$

для всех k.

16.55. Пусть матрица Λ — диагональная, с элементами $\lambda_1, \ldots, \lambda_n$. Обозначим через $\widehat{\lambda}_i$ собственные значения матрицы $\Lambda+\Omega$, через ω_{ij}, η_{ij} — элементы матриц Ω, H. Тогда с точностью до членов второго порядка

$$\eta_{ij} \cong \begin{cases} 0, & \lambda_i = \lambda_j, \\ \omega_{ij}/(\lambda_j - \lambda_i), & \lambda_i \neq \lambda_j. \end{cases}$$

16.56. В условиях и обозначениях 16.52, 16.55

$$\sum_{i=1}^{n} |\lambda_i - \widehat{\lambda}_i|^2 \lessapprox \sum_{k=1}^{r} \|\Omega_{kk}\|_E^2.$$

Для нормальной матрицы асимптотическое неравенство переходит в асимптотическое равенство.

16.57. Если λ_p — простое собственное значение матрицы Λ, то с точностью до членов третьего порядка

$$\widehat{\lambda}_p \cong \lambda_p + \omega_{pp} + \sum_{i\neq p} \frac{\omega_{pi}\omega_{ip}}{(\lambda_p - \lambda_i)}.$$

При этом формулы 16.55 при $j=p$, $i\neq p$ дают асимптотические выражения для координат нормированного собственного вектора матрицы $\Lambda+\Omega$, соответствующего $\widehat{\lambda}_p$.

16.58. Если собственное значение λ_p соответствует жордановой клетке порядка s матрицы Λ, а все элементы матрицы Ω — величины порядка ω, то асимптотически может достигаться равенство $\widehat{\lambda}_p = \lambda_p + O(\omega^{1/s})$.

16.59. Пусть даны матрицы A, B, C, D размеров соответственно $n\times n$, $m\times m$, $n\times m$, $m\times n$. Система матричных уравнений

$$AU - VB = C, \quad UB^* - A^*V = D$$

относительно матриц U, V размеров $n\times m$ имеет единственное решение тогда и только тогда, когда матрицы A и B не имеют общих сингулярных чисел.

16.60. Пусть P — блочно диагональная матрица с блоками $P_1, \ldots, P_r$ и $P+\Omega$ — возмущенная матрица. Обозначим P и Ω аналогично 16.52. Предположим, что попарно блоки матрицы P не имеют общих сингулярных чисел. Тогда существуют унитарные матрицы $E+H$ и $E+T$ такие, что матрица

$$\widehat{P} = (E+T)^*(P+\Omega)(E+H)$$

асимптотически является блочно диагональной с блоками $\widehat{P}_1, \ldots, \widehat{P}_r$, размеры которых совпадают с размерами блоков $P_1, \ldots, P_r$, и с точностью до членов второго порядка $\widehat{P}_k \cong P_k + \Omega_{kk}$ для всех k.

16.61. Если матрицы H и T разбить на блоки H_{ij} и T_{ij} аналогично Ω, то с точностью до членов второго порядка

$$H_{kk} = T_{kk} = 0, \quad P_k H_{kl} - T_{kl} P_l = -\Omega_{kl}, \quad H_{kl} P_l^* - P_k^* T_{kl} = \Omega_{lk}$$

для всех k и $l \neq k$.

16.62. Пусть матрица P — диагональная, с неотрицательными элементами $\rho_1, \ldots, \rho_n$. Обозначим через $\widehat{\rho}_i$ сингулярные числа матрицы $P + \Omega$, через ω_{ij}, η_{ij}, τ_{ij} — элементы матриц Ω, H, T. Тогда с точностью до членов второго порядка

$$\eta_{ij} \cong \begin{cases} 0, & \rho_i = \rho_j, \\ \dfrac{\omega_{ij}\rho_i + \bar{\omega}_{ji}\rho_j}{\rho_j^2 - \rho_i^2}, & \rho_i \neq \rho_j; \end{cases}$$

$$\tau_{ij} \cong \begin{cases} 0, & \rho_i = \rho_j, \\ \dfrac{\omega_{ij}\rho_j + \bar{\omega}_{ji}\rho_i}{\rho_j^2 - \rho_i^2}, & \rho_i \neq \rho_j. \end{cases}$$

16.63. В условиях и обозначениях 16.60, 16.62

$$\sum_{i=1}^{n} |\rho_i - \widehat{\rho}_i|^2 \cong \sum_{k=1}^{r} \|\Omega_{kk}\|_E^2.$$

§ 17. Матрицы типа теплицевых

Во многих приложениях нередко приходится решать задачи линейной алгебры с матрицами, все элементы которых хотя и отличны, в основном, от нуля, но определяются небольшим числом независимых параметров. Такие матрицы имеют существенные особенности в своем строении, использование которых дает возможность строить эффективные численные методы.

В этом параграфе будут рассмотрены так называемые теплицевы матрицы и некоторые похожие на них матрицы. Будем считать, что элементы матриц принадлежат некоторому кольцу Ω с единицей I.

От кольца будем требовать выполнение следующего условия: из единственности решений уравнений $Ax = y$ и $z'A = y'$ для матрицы A и вектора y с элементами из Ω должно вытекать существование решений системы при любой правой части и, следовательно, обратимость матрицы A. Дальнейшие свойства кольца конкретизируются по мере необходимости.

17.1. Матрица A_n порядка $n+1$, элементы a_{ij} которой зависят только от разности индексов $i - j$, называется *теплицевой*. Такая матрица и ее элементы обозначаются следующим образом:

$$A_n = \begin{bmatrix} a_0 & a_{-1} \ldots a_{-n} \\ a_1 & a_0 \ldots a_{-n+1} \\ \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot \\ a_n & a_{n-1} \ldots a_0 \end{bmatrix}.$$

17.2. Множество теплицевых матриц одного порядка с элементами из одного кольца Ω замкнуто относительно операций сложения матриц, а также левого и правого умножения на скалярные множители из Ω.

Произведение теплицевых матриц уже не является теплицевой матрицей. Однако матрица, обратная к теплицевой, выражается через сумму произведений теплицевых матриц специального вида. Это представление обратной матрицы не однозначно и зависит от различных дополнительных условий, накладываемых на теплицеву матрицу.

17.3. Пусть A_n — теплицева матрица. Рассмотрим величины α_i, β_i, γ_i, $\delta_i \in \Omega$, удовлетворяющие уравнениям

$$\begin{aligned} A_n(\alpha_0, \ldots, \alpha_n)' &= (\varphi_0, \ldots, \varphi_n)', \\ A_n(\beta_0, \ldots, \beta_n)' &= (I, \ldots, I)', \\ (\gamma_0, \ldots, \gamma_n) A_n &= (I, \ldots, I), \\ (\delta_0, \ldots, \delta_n) A_n &= (\psi_0, \ldots, \psi_n), \end{aligned}$$

где $\varphi_0 = \psi_0 = 0$, $\varphi_i = a_1 + \ldots + a_i$, $\psi_i = a_{-1} + \ldots + a_{-i}$, и теплицевы матрицы

$$Q_l = \begin{bmatrix} I & & 0 \\ -I & \cdot & \\ 0 & \cdot & \\ & -I & I \end{bmatrix}, \quad Q_u = \begin{bmatrix} I & -I & 0 \\ & & \\ & \cdot & -I \\ 0 & & I \end{bmatrix},$$

$$\alpha = \begin{bmatrix} \alpha_n & \cdots & \alpha_0 \\ & \ddots & \vdots \\ 0 & & \alpha_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_n & \cdots & \beta_0 \\ & \ddots & \vdots \\ 0 & & \beta_n \end{bmatrix},$$

$$\gamma = \begin{bmatrix} \gamma_n & & 0 \\ \vdots & \ddots & \\ \gamma_0 & \cdots & \gamma_n \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_n & & 0 \\ \vdots & \ddots & \\ \delta_0 & \cdots & \delta_n \end{bmatrix},$$

Если уравнения, определяющие α_i, β_i, γ_i, δ_i, разрешимы, то матрица A_n — невырожденная и A_n^{-1} имеет вид

$$A_n^{-1} = Q_u [\beta a_0 \gamma + \alpha\gamma + \beta\delta] Q_l.$$

17.4. Если A_n — невырожденная теплицева матрица с элементами из кольца Ω, то матрица A_n' — также невырожденная теплицева.

17.5. Пусть A_n — теплицева матрица и величины $\bar{\alpha}_i$, $\bar{\delta}_i$ удовлетворяют уравнениям

$$A_n(\bar{\alpha}_0, \ldots, \bar{\alpha}_n)' = (\psi_n, \ldots, \psi_0)',$$
$$(\bar{\delta}_0, \ldots, \bar{\delta}_n) A_n = (\varphi_n, \ldots, \varphi_0).$$

Если уравнения, определяющие $\bar{\alpha}_i$, $\bar{\delta}_i$, β_i, γ_i, разрешимы, то матрица A_n — невырожденная и A_n^{-1} имеет вид

$$A_n^{-1} = Q_l\left[\bar{\beta} a_0 \bar{\gamma} + \bar{\alpha}\bar{\gamma} + \bar{\beta}\bar{\delta}\right] Q_u,$$

где

$$\bar{\alpha} = \begin{bmatrix} \bar{\alpha}_0 & & 0 \\ \vdots & \ddots & \\ \bar{\alpha}_n & \ldots & \bar{\alpha}_0 \end{bmatrix}, \quad \bar{\beta} = \begin{bmatrix} \beta_0 & & 0 \\ \vdots & \ddots & \\ \beta_n & \ldots & \beta_0 \end{bmatrix},$$

$$\bar{\gamma} = \begin{bmatrix} \gamma_0 & \ldots & \gamma_n \\ & \ddots & \vdots \\ 0 & & \gamma_0 \end{bmatrix}, \quad \bar{\delta} = \begin{bmatrix} \bar{\delta}_0 & \ldots & \bar{\delta}_n \\ & \ddots & \vdots \\ 0 & & \bar{\delta}_0 \end{bmatrix},$$

Представления 17.3, 17.5 для A_n^{-1} легко преобразуются к виду, содержащему суммы двух произведений треугольных теплицевых матриц, причем в 17.3 левые и правые сомножители являются соответственно правыми и левыми треугольными матрицами, а в 17.5 — левыми и правыми. Эти представления матрицы, обратной к теплицевой, являются самыми общими. Они полностью определяются через решения некоторых систем с матрицами A и A' при специальном выборе правых частей, при одном лишь предположении о разрешимости этих систем. Если же на матрицу A наложить дополнительные ограничения, то для представления A_n^{-1} в качестве решений таких систем можно взять некоторые строки и столбцы A_n^{-1}.

17.6. Пусть A_n — теплицева матрица и величины x_i, y_i, z_i, w_i удовлетворяют уравнениям

$$A_n(x_0, \ldots, x_n)' = (I, 0, \ldots, 0)',$$
$$A_n(y_0, \ldots, y_n)' = (0, \ldots, 0, I)',$$
$$(z_0, \ldots, z_n) A_n = (I, 0, \ldots, 0)',$$
$$(w_0, \ldots, w_n) A_n = (0, \ldots, 0, I).$$

Если эти уравнения разрешимы и существуют элементы x_0^{-1}, y_n^{-1}, то матрица A_n — невырожденная и A_n^{-1} имеет вид

$$A_n^{-1} = \begin{bmatrix} x_0 & & 0 \\ \vdots & \ddots & \\ x_n & \ldots & x_0 \end{bmatrix} x_0^{-1} \begin{bmatrix} z_0 & \ldots & z_n \\ & \ddots & \vdots \\ 0 & & z_0 \end{bmatrix} -$$

$$-\begin{bmatrix} 0 & & & 0 \\ y_0 & & & \\ \vdots & \ddots & & \\ y_{n-1} & \cdots & y_0 & 0 \end{bmatrix} y_n^{-1} \begin{bmatrix} 0 & w_0 & \cdots & w_{n-1} \\ & \ddots & \ddots & \vdots \\ & & \ddots & w_0 \\ 0 & & & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} y_n & \cdots & y_0 \\ & \ddots & \vdots \\ 0 & & y_n \end{bmatrix} y_n^{-1} \begin{bmatrix} w_n & & 0 \\ \vdots & \ddots & \\ w_0 & \cdots & w_n \end{bmatrix} - \begin{bmatrix} 0 & x_n & \cdots & x_1 \\ & \ddots & \ddots & \vdots \\ & & \ddots & x_n \\ 0 & & & 0 \end{bmatrix} x_0^{-1} \begin{bmatrix} 0 & & & 0 \\ z_n & \ddots & & \\ \vdots & \ddots & \ddots & \\ z_1 & \cdots & z_n & 0 \end{bmatrix}.$$

Заметим, что x_i, y_i, z_i, w_i — элементы соответственно первого и последнего столбцов, а также первой и последней строк матрицы A_n^{-1}. Аналогичные представления можно получить с привлечением элементов второго и предпоследнего столбцов, а также второй и предпоследней строк матрицы A_n^{-1}. Необходимость исследования других представлений связана только с тем, что не всегда для невырожденной матрицы A_n существуют элементы x_0^{-1}, y_n^{-1}. Однако мы не будем останавливаться на этих представлениях обратной матрицы. Все они в конечном счете аналогичны представлениям 17.3, 17.5, 17.6.

17.7. Элементы x_0, y_n и ведущий минор порядка n обратимой матрицы A_n обратимы или не обратимы все одновременно.

17.8. Если невырожденная теплицева матрица является треугольной, то обратная к ней матрица теплицева.

17.9. Теплицева матрица C_n называется *циркулянтной* или *циркулянтом*, если при всех $i \neq 0$ для ее элементов выполняются соотношения $c_{-i} = c_{n-i+1}$.

Если циркулянтная матрица — невырожденная, то обратная к ней матрица — циркулянтная.

17.10. Матрица A называется *ленточной*, если ее элементы a_{ij} удовлетворяют соотношениям

$$a_{ij} = 0, \qquad l < i - j,\ j - i > m,$$

для некоторых неотрицательных чисел l, m. В случае $m = 0$ матрица называется *левой* ленточной, в случае $l = 0$ — *правой* ленточной; величина $l + m + 1$ называется *шириной* ленты. Если $l = m = 1$, матрица называется *трехдиагональной.*

17.11. Пусть A_n — невырожденная ленточная теплицева матрица. При условии $n > l + m$ матрица A_n^{-1} будет теплицевой тогда и только тогда, когда A_n — треугольная.

17.12. Матрицы A_n и A_n^{-1}, отличающиеся от треугольных, являются одновременно теплицевыми тогда и только тогда, когда для некоторого $\varphi \neq 0$ при всех $i > 0$ для элементов a_{i-j} и $a_{i-j}^{(-1)}$ этих матриц выполняются соотношения

$$a_{-i} = \varphi a_{n-i+1}, \qquad a_i^{-1} = \varphi a_{n-i+1}^{(-1)}.$$

Для практических вычислений особое значение имеют случаи, когда Ω — поле комплексных чисел или кольцо числовых матриц. В первом случае формулы упрощаются за счет того, что между параметрами соответствующих описаний существует простая зависимость, порожденная свойством персимметричности рассматриваемых матриц.

17.13. Матрица A называется *персимметричной*, если она симметрична относительно второй диагонали.

17.14. Матрица A персимметрична тогда и только тогда, когда $A' = PAP$, где P есть матрица перестановок вида

$$P = \begin{bmatrix} 0 & \dots & 0 & I \\ 0 & \dots & I & 0 \\ \cdot & \cdot & \cdot & \cdot \\ I & \dots & 0 & 0 \end{bmatrix}.$$

17.15. Теплицева матрица является персимметричной для любого кольца Ω.

17.16. Если элементы невырожденной теплицевой матрицы принадлежат полю комплексных чисел, то обратная матрица является персимметричной. В этом случае для элементов x_i, y_i, z_i, w_i из 17.6 для всех i выполняются соотношения $x_i = w_{n-i}$, $y_i = z_{n-i}$.

В случае произвольного кольца Ω свойство персимметричности не обязательно переносится на обратную матрицу. К сокращению параметров, описывающих обратную матрицу, приводят и другие дополнительные свойства теплицевой матрицы, например симметричность.

17.17. Если Ω есть поле комплексных чисел, то в условиях и обозначениях 17.6 матрица A_n^{-1} имеет следующий вид:

$$A_n^{-1} = x_0^{-1}\left(\begin{bmatrix} x_0 & & 0 \\ \vdots & \ddots & \\ x_n & \dots & x_0 \end{bmatrix}\begin{bmatrix} y_n & \dots & y_0 \\ & \ddots & \vdots \\ 0 & & y_n \end{bmatrix} - \begin{bmatrix} 0 & & & 0 \\ y_0 & & & \\ \vdots & & & \\ y_{n-1} & \dots & y_0 & 0 \end{bmatrix}\begin{bmatrix} 0 & x_n & \dots & x_1 \\ & & \ddots & \vdots \\ & & & x_n \\ 0 & & & 0 \end{bmatrix}\right) =$$

$$= x_0^{-1}\left(\begin{bmatrix} y_n & \dots & y_0 \\ & \ddots & \vdots \\ 0 & & y_n \end{bmatrix}\begin{bmatrix} x_0 & & 0 \\ \vdots & \ddots & \\ x_n & \dots & x_0 \end{bmatrix} - \begin{bmatrix} 0 & x_n & \dots & x_1 \\ & & \ddots & \vdots \\ & & & x_n \\ 0 & & & 0 \end{bmatrix}\begin{bmatrix} 0 & & & 0 \\ y_0 & & & \\ \vdots & & & \\ y_{n-1} & \dots & y_0 & 0 \end{bmatrix}\right).$$

Таким образом, если элементы невырожденной теплицевой матрицы являются комплексными числами, то элементы обратной матрицы в случае $x_0 \neq 0$ полностью определяются ее первым и последним столбцом.

17.18. Если Ω есть поле комплескных чисел, то:

— множество треугольных теплицевых матриц одного наименования есть коммутативное кольцо;

— множество невырожденных треугольных теплицевых матриц есть коммутативная группа;

— множество циркулянтных матриц есть коммутативное кольцо;

— множество невырожденых циркулянтных матриц есть коммутативная группа по умножению;

— между параметрами α_i, β_i, γ_i, δ_i утверждения 17.3 для всех i существует следующая зависимость:

$$\gamma_i = \beta_{n-i}, \quad \delta_i = 1 - \alpha_{n-i} - \beta_{n-i}a_0.$$

Дополнительные свойства, которые приобретают различные виды теплицевых матриц в поле комплексных чисел, очень важны. Поэтому всюду в дальнейшем, если не сделано особой оговорки, мы будем считать выполненным это условие.

17.19. Пусть $\varepsilon_n = \exp\{2\pi i/n\}$ (i — мнимая единица). Матрица F_n порядка n с элементами

$$f_{kl} = \varepsilon_n^{(k-1)(l-1)}, \quad 1 \leqslant k, \quad l \leqslant n,$$

называется матрицей *дискретного преобразования Фурье.*

17.20. Матрица дискретного преобразования Фурье, умноженная на скалярный множитель $n^{-1/2}$, является комплексной симметричной унитарной матрицей.

17.21. Вектор $y = F_n x \left(x = F_n^* y\right)$ называется *прямым* (*обратным*) *дискретным преобразованием Фурье вектора* x (y).

17.22. Имеет место соотношение $F_n^* y = \overline{F_n \bar{y}}$, где черта означает комплексное сопряжение.

17.23. Пусть $y(k)$, $x(l)$ — периодические функции целочисленных переменных k, l. Функция $y(k)$ называется *дискретным преобразованием Фурье функции* $x(l)$, если эти функции имеют общий период n и $(y(0), \ldots, y(n-1)) = (x(0), \ldots, x(n-1))F_n$. В этом случае будем писать $y(k) = [F_n x](k)$ или $y(k) = [F_n x(l)](k)$.

17.24. Функция $y(k)$:

— является четной;

— является нечетной;

— принимает вещественные значения;

— принимает чисто мнимые значения

тогда и только тогда, когда функция $x(l)$:

— является четной;

— является нечетной;

— $x(l) = \overline{x(-l)} = \overline{x(n-l)}$;

— $x(l) = -\overline{x(-l)} = -\overline{x(n-l)}$.

17.25. Функция $y(k)$ принимает:

— вещественные (чисто мнимые) значения и является четной (нечетной)

— чисто мнимые (вещественные) значения и является четной (нечетной)

тогда и только тогда, когда функция $x(l)$ принимает:

— вещественные значения и является четной (нечетной);

— чисто мнимые значения и является четной (нечетной).

17.26. Имеют место соотношения

$$[F_n x(l+m)](k) = \varepsilon_n^{-km} y(k),$$
$$\left[F_n\left(\varepsilon_n^{lm} x(l)\right)\right](k) = y(k+m).$$

17.27. Пусть задана последовательность $z(j)$, $0 \leqslant j \leqslant n$. Продолжим ее периодически и предположим, что $x^{(j)}(l)$ — такие функции с периодом n, что $x^{(j)}(l) = z(j+l)$, $0 \leqslant l \leqslant n$. Тогда

$$\left[F_n x^{(j+1)}(l)\right](k) = \varepsilon_n^{-k}\left(\left[F_n x^{(j)}(l)\right](k) + z(j+n) - z(j)\right).$$

17.28. Обозначим через $x_m(l)$ и $y_m(k)$ такие функции с периодом nm, что

$$x_m(l) = x(l/m), \quad y_m(l) = y(l/m),$$

если l делится на m, и

$$x_m(l) = y_m(l) = 0,.$$

если l не делится на m. Тогда

$$[F_{nm} x_m(l)](k) = y(k), \quad [F_{nm} x(l)](k) = m y_m(k).$$

17.29. Если $n = \gamma\delta$, где n, γ, δ — целые, то имеют место соотношения

$$[F_\gamma x(l\delta)](k) = \delta^{-1} \sum_{j=0}^{\delta-1} y(k + j\gamma),$$

$$\left[F_\gamma\left(\sum_{j=0}^{\delta-1} x(l + j\gamma)\right)\right](k) = y(k\delta).$$

Если $y(k) = 0$ при $\gamma \leqslant k \leqslant n-1$, то

$$x(l) = \sum_{j=0}^{\gamma-1} x(j\delta)\, G(l - j\delta),$$

где $G(j) = 1$, если j кратно n, и $G(j) = \delta^{-1}\left(1 - \varepsilon_\gamma^j\right)\left(1 - \varepsilon_n^j\right)$ в других случаях.

Прямое умножение матрицы F_n преобразования Фурье на вектор требует выполнения n^2 операций комплексного умножения и n^2 операций комплексного сложения. Однако громадная практическая значимость преобразования Фурье и специальный вид его матрицы стимулировали многочисленные исследования, направленные на создание более быстрых алгоритмов. В конечном счете различные варианты этих алгоритмов связаны с различными представлениями матрицы преобразования Фурье. Мы рассмотрим здесь два вида представления. Одна из основных идей связана со свойством 17.29.

17.30. Пусть введены следующие обозначения:

F_γ — матрица дискретного преобразования Фурье порядка γ;

E_γ — единичная матрица порядка γ;

W_δ^γ — диагональная матрица $W_\delta^\gamma = \operatorname{diag}\left(\varepsilon_{\gamma\delta}^{0\cdot 0}, \ldots, \varepsilon_{\gamma\delta}^{(\gamma-1)0}, \ldots\right.$
$\left.\ldots, \varepsilon_{\gamma\delta}^{0\cdot(\delta-1)}, \ldots, \varepsilon_{\gamma\delta}^{(\gamma-1)(\delta-1)}\right)$;

P_δ^γ — матрица перестановок порядка $\gamma\delta$, в которой единицы находятся в тех позициях (k, l), где выполняются соотношения

$$l = r_1\gamma + r_0 + 1, \quad k = r_0\delta + r_1 + 1, \quad 0 \leqslant r_1 < \delta, \quad 0 \leqslant r_0 < \gamma.$$

Тогда, если $n = \gamma\delta$, то

$$F_{\gamma\delta} = P_\delta^\gamma (E_\delta \times F_\gamma) W_\delta^\gamma (F_\delta \times E_\gamma).$$

17.31. Пусть $n = \gamma\delta$ и требуется выполнить дискретное преобразование Фурье $y = F_n x$ порядка n. Запишем координаты вектора x по столбцам прямоугольной матрицы размера $\gamma \times \delta$ подряд сверху вниз и слева направо; координаты вектора y запишем аналогично по столбцам матрицы размера $\delta \times \gamma$. Тогда преобразование вектора x в вектор y сводится к следующим операциям:

— вычисление преобразований Фурье порядка δ вектор-строк матрицы, содержащей x;

— умножение элементов полученной матрицы с индексами i, j на так называемые *поворачивающие множители* $\varepsilon_n^{(i-1)(j-1)}$;

— вычисление преобразований Фурье порядка γ вектор-столбцов новой матрицы;

— транспонирование матрицы.

17.32. Пусть $n = n_1 \ldots n_p$. Обозначим

$$\alpha_j = \prod_{l=1}^{j-1} n_l, \quad \beta_j = \prod_{l=j+1}^{p} n_l.$$

Если верхний индекс меньше нижнего, то будем считать произведения равными 1. Тогда в обозначениях 17.30

$$F_n = P\left(E_{\alpha_p} \times W_{n_p}^{\beta_p}\right)\left(E_{\alpha_p} \times F_{n_p} \times E_{\beta_p}\right) \ldots \left(E_{\alpha_1} \times W_{n_1}^{\beta_1}\right)\left(E_{\alpha_1} \times F_{n_1} \times E_{\beta_1}\right) =$$
$$= \left(E_{\alpha_1} \times F_{n_1} \times E_{\beta_1}\right)\left(E_{\alpha_1} \times W_{n_1}^{\beta_1}\right) \ldots \left(E_{\alpha_p} \times F_{n_p} \times E_{\beta_p}\right)\left(E_{\alpha_p} \times W_{n_p}^{\beta_p}\right) P',$$

где

$$P = \left(E_{\alpha_1} \times P_{n_1}^{\beta_1}\right) \ldots \left(E_{\alpha_p} \times P_{n_p}^{\beta_p}\right).$$

17.33. Если $n = 2^s$, то умножение матрицы F_n преобразования Фурье на вектор может быть выполнено за $(1/2) n \log_2 n$ операций комплексного умножения и $n \log_2 n$ операций комплексного сложения.

17.34. Если $n=4^s$, то умножение матрицы F_n преобразования Фурье на вектор может быть выполнено за $(3/8)n\log_2 n$ операций комплексного умножения и $n\log_2 n$ операций комплексного сложения.

Утверждения 17.29 — 17.34 связаны со случаем составного n и зависят от вида сомножителей. Однако существует способ сведения преобразования Фурье любого порядка к преобразованию Фурье, порядок которого равен, например, степени двойки.

17.35. Пусть $c=(c_0, \ldots, c_{n-1})'$ — первый столбец циркулянтной матрицы C_{n-1}. Имеет место разложение

$$C_{n-1}=n^{-1}F_n^*\operatorname{diag}(F_nc)\,F_n,$$

где $\operatorname{diag}(F_nc)$ означает диагональную матрицу, элементы которой совпадают с элементами вектора F_nc.

17.36. Умножение вектора на циркулянтную матрицу сводится к трехкратному выполнению преобразования Фурье и двукратному умножению на диагональную матрицу.

17.37. Для матрицы преобразования Фурье любого порядка имеет место разложение $F_n=G_nN_nG_n$, где N_n — теплицева матрица вида

$$N_n=\begin{bmatrix} \varepsilon_{2n}^{-0^2} & \varepsilon_{2n}^{-1^2} & \ldots & \varepsilon_{2n}^{-(n-1)^2} \\ \varepsilon_{2n}^{-1^2} & \varepsilon_{2n}^{-0^2} & \ldots & \varepsilon_{2n}^{-(n-2)^2} \\ \cdots & \cdots & \cdots & \cdots \\ \varepsilon_{2n}^{-(n-1)^2} & \varepsilon_{2n}^{-(n-2)^2} & \ldots & \varepsilon_{2n}^{-0^2} \end{bmatrix},$$

G_n — диагональная матрица: $G_n=\operatorname{diag}\left(\varepsilon_{2n}^{0^2}, \ldots, \varepsilon_{2n}^{(n-1)^2}\right)$.

17.38. Любая теплицева матрица A_{n-1} вида 17.1 может быть представлена в виде произведения

$$A_{n-1}=[E_n \quad 0],C_m\begin{bmatrix} E_n \\ 0 \end{bmatrix},$$

где C_m — циркулянтная матрица порядка $m\geqslant 2n-1$, первый столбец которой имеет элементы

$$a_0,\ a_1,\ \ldots,\ a_{n-1},\ 0,\ \ldots,\ 0,\ a_{-n+1},\ \ldots,\ a_{-1}.$$

17.39. В обозначениях 17.38 матрица A_{n-1} есть матрица ведущего минора порядка n матрицы C_m.

17.40. Число m в 17.38 всегда можно выбрать равным степени двойки и удовлетворяющим неравенствам $2n\leqslant m\leqslant 4n$.

17.41. Умножение вектора на теплицеву матрицу любого порядка n на основе представлений 17.35, 17.38 может быть выполнено за $6n\log_2 n$ операций комплексного умножения и $12n\log_2 n$ операций комплексного сложения.

17.42. Преобразование Фурье любого порядка n на основе представлений 17.37, 17.38, 17.35 может быть выполнено за

$6n\log_2 n$ операций комплексного умножения и $12n\log_2 n$ операций комплексного сложения.

В различных задачах приходится иметь дело с блочными матрицами, устроенными по типу теплицевых. Блоки этих матриц, в свою очередь, могут быть разбиты на более мелкие блоки и также устроены по типу теплицевых, и т. д. При некоторых условиях на такие *многоуровневые разбиения* задачи с подобными матрицами могут решаться весьма эффективно. Снова будем считать, что элементы матриц принадлежат произвольному кольцу Ω с единицей.

17.43. Пусть квадратная матрица A разбита на $n_1\times n_1$ квадратных блоков $A_{i_1j_1}$. Предположим, что уже определено разбиение матрицы A на блоки $A_{i_1j_1;\ldots;i_{k-1},j_{k-1}}$, $k>1$, и пусть снова каждый из этих блоков разбит на $n_k\times n_k$ квадратных блоков $A_{i_1j_1;\ldots;i_kj_k}$. Тогда k-м *уровнем разбиения* матрицы A называется множество блоков $A_{i_1j_1;\ldots;i_kj_k}$, а число n_k называется *порядком* k-го уровня.

17.44. Будем говорить, что матрица A имеет p уровней, если $A_{i_1j_1;\ldots;i_{p+1}j_{p+1}}$ представляют собой элементы из Ω и $n_{p+1}=1$.

17.45. Пусть зафиксированы элементы $\alpha_{ij}^{(q)},\ \beta_{ij}^{(q)},\ \gamma^{(q)}\in\Omega$, $q=1,\ldots,Q,\quad i,j=1,\ldots,n$, и рассматриваются равенства

$$\sum_{i,j}^{n}\alpha_{ij}^{(q)}z_{ij}\beta_{ij}^{(q)}=\gamma^{(q)}.$$

Говорят, что блочная матрица $A=\{A_{ij}\}_{ij=1}^{n}$ имеет *блочный тип* Γ, если равенства выполняются при $z_{ij}=A_{ij}$.

Рассмотрим важные для нас примеры. Пусть $A=\{A_{ij}\}$, $A_{ij}\in\Omega$,— теплицева матрица порядка n. Тогда равенства 17.45 имеют вид $A_{i+1,\,j+1}-A_{ij}=0$, $i,j=1,\ldots,n-1$. Тип этой матрицы обозначим T_n. Если матрица A является циркулянтной, то к выписанным равенствам добавляются и такие: $A_{i1}-A_{i-1,\,n}=0$, $i=2,3,\ldots,n$. Тип этой матрицы обозначим C_n. Можно было бы ввести обозначения для типов матриц, являющихся симметричными, треугольными, трехдиагональными, разреженными, с какой-либо определенной структурой нулевых элементов и т. д. Мы будем использовать в основном лишь обозначения D_n, T_n, C_n, G_n соответственно для типов диагональной, теплицевой, циркулянтной и общего вида матрицы блочного порядка n. Кроме того, полезным оказывается введение блочной матрицы типа Φ_m, у которой все блоки являются скалярными матрицами с $\varepsilon_m^{(k-1)(l-1)}$ в блоке Φ_{kl}, где $\varepsilon_m=\exp\{2\pi i/m\{$ (i — мнимая единица).

17.46. Матрица A имеет *композиционный тип* $\alpha_1\ldots\alpha_p$, если она является p-уровневой и блоки k-го уровня удовлетворяют при $1\leqslant k\leqslant p$ соотношениям 17.45, определяющим тип α_k.

17.47. Операция композиции типов ассоциативна.

17.48. Пусть заданы типы α_i, $i=1,2,\ldots,p$, и их композиции

$$\Gamma_1=\alpha_1\ldots\alpha_k\alpha_{k+1}\ldots\alpha_p,\quad \Gamma_2=\alpha_1\ldots\alpha_{k+1}\alpha_k\ldots\alpha_p.$$

Существует такая матрица перестановок P, определяемая лишь порядками типов α_i, что для любой матрицы A типа Γ_1 матрица

$B=PAP'$ имеет тип Γ_2 и для любой матрицы B типа Γ_2 матрица $A=P'BP$ имеет тип Γ_1.

17.49. Всякая многоуровневая матрица, содержащая разбиения типов D, T, C и G в любом числе и любом порядке, путем перестановки строк и столбцов может быть преобразована в подобную ей матрицу типа $D\ldots DT\ldots TC\ldots CG\ldots G$, где все разбиения одинакового типа стоят подряд.

17.50. Всегда выполняются равенства

$$D_{n_1}\ldots D_{n_p}=D_{n_1\ldots n_p},\quad G_{n_1}\ldots G_{n_p}=G_{n_1\ldots n_p},$$

т. е. всегда все уровни типа D или G можно объединить в один уровень, причем блочный порядок объединенного уровня равен произведению блочных порядков отдельных уровней.

Заметим, что если матрица A имеет тип $D\Gamma$, то это означает, что она является блочно диагональной с блоками типа Γ на диагонали. В этом случае решение всех основных задач линейной алгебры с матрицей типа $D\Gamma$ сводится к решению нескольких аналогичных задач с матрицей типа Γ. Число таких задач совпадает с блочным порядком матрицы D. Поэтому в композиции типов, вообще говоря, можно опускать все типы вида D. Объединение уровней типа G в конце композиции типов можно трактовать как переход к другому кольцу Ω.

17.51. Пусть кольцо Ω содержит не менее n различных элементов. Путем перестановок соответствующих строк и столбцов уровни типов $C_{n_1},\ldots,C_{n_p}$ объединяются в один уровень типа C_n, где $n=n_1\ldots n_p$, тогда и только тогда, когда числа $n_1,\ldots,n_p$ попарно взаимно простые.

Далее будем опять считать, что Ω есть поле комплексных чисел или кольцо числовых матриц. Рассмотрим матрицу типа $C_{n_1}\ldots C_{n_p}\Gamma$. Основой решения задач с такими матрицами служит следующее утверждение.

17.52. Пусть A — матрица типа $C_m\Gamma_n$, где Γ_n означает тип матрицы порядка n, заданный соотношениями 17.45 при $\gamma^{(q)}=0$. Тогда матрица $B=\Phi_m^*A\Phi_m$ имеет тип $D_m\Gamma_n$.

17.53. Для всякой матрицы A типа $C_{n_1}\ldots C_{n_p}\Gamma$ матрица $B=U^*AU$ имеет тип $D_{n_1\ldots n_p}\Gamma$, где U является произведением унитарных матриц типов Φ_{n_1}, $D_{n_1}\Phi_{n_2}$, $\ldots$, $D_{n_1\ldots n_{p-1}}\Phi_{n_p}$.

В заключение укажем еще один тип матрицы, тесно связанный с теплицевыми матрицами.

17.54. Матрица H_n, элементы h_{ij} которой зависят только от суммы индексов $i+j$, называется *ганкелевой*.

17.55. Ганкелева матрица является симметричной.

17.56. Если H_n — ганкелева матрица, то $A_n=P_nH_n=H_nP_n$ — теплицева матрица, если P_n — матрица перестановок 17.14.

Иногда приходится рассматривать матрицы, представленные суммой парных произведений теплицевых матриц. Отметим некоторые свойства таких представлений, при этом матрицы будем считать вещественными или комплексными.

17.57. Наименьшее число слагаемых во всевозможных представлениях матрицы суммой парных произведений теплицевых левых и правых (правых и левых) треугольных матриц называется ее *левым* (*правым*) *теплицевым рангом.*

17.58. Левый (правый) теплицев ранг квадратной матрицы порядка n с элементами a_{ij} совпадает с обычным рангом матрицы такого же порядка, которая строится по исходной матрице следующим образом: первые (последние) строка и столбец берутся без изменений, а при всех $i, j > 1$ $(i, j < n)$ соответствующий элемент заменяется разностью $a_{ij} - a_{i-1,\, j-1}$ $(a_{ij} - a_{i+1,\, j+1})$.

17.59. Матрица выражается суммой парных произведений теплицевых левых и правых (правых и левых) треугольных матриц S_iT_i тогда и только тогда, когда матрица, построенная в 17.58 при рассмотрении левого (правого) теплицева ранга, разлагается в сумму матриц s_it_i, где s_i — первый (последний) столбец матрицы S_i, а t_i — первая (последняя) строка матрицы T_i.

17.60. Для любой матрицы левый и правый теплицевы ранги не превосходят ее порядка и отличаются между собой не более чем на 2.

17.61. Левый и правый теплицевы ранги, вычисленные для произвольной теплицевой матрицы, всегда равны и не превосходят 2.

17.62. Для всякой невырожденной матрицы левый (правый) теплицев ранг равен правому (левому) теплицеву рангу обратной матрицы.

17.63. В невырожденной матрице левые теплицевы ранги матриц любых ненулевых ведущих миноров не превосходят левого теплицевого ранга исходной матрицы.

§ 18. Матрицы с неотрицательными элементами

18.1. Вещественную квадратную матрицу A будем называть *неотрицательной* (*положительной*), если все ее элементы неотрицательны (положительны). Для неотрицательных матриц используется обозначение $A \gg 0$, которое не нужно путать с обозначением $A \geqslant 0$ для положительно полуопределенных матриц.

При изучении свойств неотрицательных матриц большую роль играют понятия разложимости и неразложимости, вместо которых часто используются соответственно термины *приводимая* и *неприводимая* матрицы. Интересная и удобная на практике характеризация неразложимых матриц дается с помощью ее направленного графа.

18.2. Пусть $P_1, \ldots, P_n$ — различные точки двумерной плоскости и A — матрица порядка n. Для каждого ненулевого элемента a_{ij} матрицы A соединим точку P_i с точкой P_j направленной линией $\overrightarrow{P_iP_j}$. Получающаяся в результате фигура называется *направленным* (*ориентированным*) *графом* матрицы A.

18.3. Направленный граф матрицы A *сильно связен*, если для любых двух точек P_i и P_j существует связывающий их *ориентированный путь* $\overrightarrow{P_iP_{i_1}}, \ldots, \overrightarrow{P_{i_l}P_j}$. Целая величина $l+1$ называется *длиной* этого пути.

18.4. Матрица A неразложима тогда и только тогда, когда ее направленный граф сильно связен.

18.5. Матрица A, все элементы которой отличны от нуля, неразложима.

18.6. Если $A \geqslant 0$ — неразложимая квадратная матрица порядка n, то $(E+A)^{n-1}$ — положительная матрица.

18.7. Вещественный вектор x называется *неотрицательным* (*положительным*), если все его координаты неотрицательны (положительны). Аналогично матрицам, для неотрицательного вектора x используется обозначение $x \geqslant 0$.

18.8. (*Теорема Перрона — Фробениуса.*) Пусть $A \geqslant 0$ — неразложимая матрица. Тогда:

— A имеет положительное собственное число, равное ее спектральному радиусу $\rho(A)$;

— $\rho(A)$ соответствует положительный собственный вектор;

— $\rho(A)$ возрастает при возрастании любого из элементов A;

— $\rho(A)$ имеет кратность 1.

18.9. Если квадратная матрица $A=(a_{ij}) \geqslant 0$ порядка n неразложима, то либо

$$\rho(A) = \sum_{j=1}^{n} a_{ij}$$

для всех $i=1, \ldots, n$, либо

$$\min_{1 \leqslant i \leqslant n} \sum_{j=1}^{n} a_{ij} < \rho(A) < \max_{1 \leqslant i \leqslant n} \sum_{j=1}^{n} a_{ij}.$$

18.10. Пусть $A \geqslant 0$ — неразложимая квадратная матрица порядка n и K — множество векторов размерности n с положительными координатами. Тогда для любого вектора $d=(d_1, \ldots, d_n) \in K$ выполняются неравенства

$$\min_{1 \leqslant i \leqslant n} \frac{1}{d_i} \sum_{j=1}^{n} a_{ij}d_j \leqslant \rho(A) \leqslant \max_{1 \leqslant i \leqslant n} \frac{1}{d_i} \sum_{j=1}^{n} a_{ij}d_j;$$

более того,

$$\sup_{d \in K} \min_{1 \leqslant i \leqslant n} \frac{1}{d_i} \sum_{j=1}^{n} a_{ij}d_j = \rho(A) = \inf_{d \in K} \max_{1 \leqslant i \leqslant n} \frac{1}{d_i} \sum_{j=1}^{n} a_{ij}d_j.$$

Равенства достигаются, когда d — собственный вектор A, соответствующий $\rho(A)$.

18.11. Если $A \geqslant 0$ — неразложимая квадратная матрица, то

$$\rho(A+B) > \rho(A)$$

для любой ненулевой матрицы $B \geqslant 0$ того же порядка.

18.12. Если $A \geqslant 0$ — неразложимая матрица, то для любой ненулевой матрицы $B \geqslant 0$ справедливо соотношение

$$\rho(A+tB) \underset{t\to+\infty}{\longrightarrow} +\infty.$$

18.13. Пусть неразложимая квадратная матрица $A \geqslant 0$ имеет k собственных чисел, равных по модулю $\rho(A)$. Тогда, если $k > 1$, то матрица A называется *циклической* индекса k. В противном случае ($k=1$) матрица A называется *примитивной*. Циклическая индекса k матрица называется также *импримитивной* матрицей, а соответствующее число k — *ее индексом импримитивности.*

18.14. Если $A \geqslant 0$ — неразложимая циклическая индекса k матрица, то существует такая матрица перестановок P, что

$$PAP' = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1k} \\ A_{21} & A_{22} & \dots & A_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ A_{k1} & A_{k2} & \dots & A_{kk} \end{bmatrix},$$

где A_{ij} — блоки размеров $n_i \times n_j$ $(n_1+n_2+\ldots+n_k=n)$, причем имеется только k ненулевых блоков: $A_{12}, A_{23}, \ldots, A_{k-1,k}, A_{k1}$, т. е.

$$PAP' = \begin{bmatrix} 0 & A_{12} & & 0 \\ & & \ddots & \\ 0 & & & A_{k,k-1} \\ A_{k1} & & 0 & \end{bmatrix}.$$

18.15. Если $A \geqslant 0$ — неразложимая циклическая индекса k матрица, то множество ее собственных чисел инвариантно относительно вращения комплексной плоскости на углы $2\pi l/k$, $l = 1, 2, \ldots, k$, т. е. для любого собственного числа λ матрицы A величины $\lambda e^{i2\pi l/k}$, $l=1, \ldots, k-1$, также являются собственными числами A; в частности, ее собственными числами являются величины $\rho(A)e^{i2\pi l/k}$, $l=0, \ldots, k-1$.

18.16. Если $A \geqslant 0$ — неразложимая циклическая индекса k матрица, то существуют такое целое положительное число $r \geqslant 1$ и такие величины $\sigma_1 = 1, \sigma_2, \ldots, \sigma_r$, удовлетворяющие неравенствам $|\sigma_r| \leqslant |\sigma_{r-1}| \leqslant \ldots \leqslant \sigma_1$, что

$$\det(tE-A) = t^{n-kr}\prod_{i=1}^{r}\left(t^k - \sigma_i\rho^k(A)\right).$$

18.17. (*Критерий цикличности.*) Пусть G — направленный граф неразложимой матрицы $A \geqslant 0$, S — множество всех замкнутых путей $\overrightarrow{P_iP_{i_1}}, \ldots, \overrightarrow{P_{i_l}P_i}$ этого графа и k — наибольший общий делитель соответствующих величин $l+1$, являющихся длинами замкнутых путей. Тогда, если $k>1$, то матрица A является циклической индекса k; если же $k=1$, то A — примитивная матрица.

18.18. Если все элементы матрицы $A \geqslant 0$ положительны, то A — примитивная матрица.

18.19. Если матрица $A \geqslant 0$ неразложима и хотя бы один ее диагональный элемент положителен, то A — примитивная матрица.

18.20. Если $A \geqslant 0$ — неразложимая примитивная матрица, то матрица A^s неразложима и примитивна для любого $s \geqslant 1$, причем, начиная с некоторого значения s, все элементы матрицы A^s положительны.

18.21. Квадратная матрица $A \geqslant 0$ примитивна тогда и только тогда, когда некоторая ее степень является положительной матрицей.

18.22. Если $A \geqslant 0$ — неразложимая циклическая индекса k матрица, приведенная к виду 18.14, то матрица $B = A^k$ является блочно диагональной матрицей вида

$$B = B_1 \oplus B_2 \oplus \ldots \oplus B_k$$

с квадратными матрицами

$$B_i \doteq A_{i,\,i+1} \ldots A_{k-1,\,k} A_{k1} \ldots A_{i-1,\,i}.$$

Эти матрицы неразложимы, примитивны, и множества их ненулевых собственных чисел (вплоть до кратностей) совпадают; в частности, их спектральные радиусы равны $\rho^k(A)$.

18.23. Для любой разложимой матрицы A существует такая матрица перестановок P, что

$$PAP' = \begin{bmatrix} A_{11} & A_{12} & \ldots & A_{1s} \\ 0 & A_{22} & \ldots & A_{2s} \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & A_{ss} \end{bmatrix},$$

где A_{ij} — матрицы размеров $n_i \times n_j$, $n_1 + \ldots + n_s = n$, а A_{ii} — либо неразложимые матрицы, либо нулевые матрицы порядка единица. Построенное представление называется *нормальной формой* разложимой матрицы A.

18.24. Пусть $A \geqslant 0$ — некоторая квадратная матрица. Тогда:

— A имеет неотрицательное собственное число, равное ее спектральному радиусу $\rho(A)$, причем $\rho(A) = 0$ только в том случае, если ее нормальная форма является строго верхней треугольной матрицей;

— $\rho(A)$ соответствует неотрицательный собственный вектор;

— $\rho(A)$ не убывает при возрастании любого из элементов матрицы A.

18.25. Пусть A и B — квадратные матрицы с неотрицательными элементами, причем матрица $A+B$ неразложима. Тогда

$$\rho(A) < \rho(A+B).$$

18.26. Пусть $A=(a_{ij})$ — произвольная квадратная матрица порядка n. Тогда матрица B порядка n с элементами $b_{ij}=|a_{ij}|$ называется *модулем матрицы* A и обозначается $B=|A|$.

18.27. Для любой квадратной матрицы A выполняется неравенство

$$\rho(A) \leqslant \rho(|A|).$$

18.28. Матрица $B \geqslant 0$ называется *стохастической*, если сумма элементов каждой ее строки равна единице.

18.29. Матрица переходных вероятностей однородной цепи Маркова с конечным числом состояний является стохастической матрицей.

18.30. Любая матрица перестановок, в том числе единичная, является стохастической матрицей.

18.31. Матрица $B \geqslant 0$ тогда и только тогда является стохастической, когда она имеет единицу своим собственным числом и ему соответствует собственный вектор, все координаты которого равны единице. Это собственное число является спектральным радиусом стохастической матрицы.

18.32. Матрица $A \geqslant 0$, имеющая положительное собственное число, которому соответствует положительный собственный вектор, всегда подобна некоторой стохастической матрице, умноженной на это собственное число. Точнее говоря, если $Az=\lambda z$, где $\lambda=\operatorname{Re}\lambda>0$, $z=(z_1, \ldots, z_n)$ — положительный вектор и $Z=\operatorname{diag}\{z_1, \ldots, z_n\}$, то матрица $\frac{1}{\lambda}Z^{-1}AZ$ является стохастической.

18.33. Собственному числу 1 стохастической матрицы соответствуют элементарные делители только первой степени.

18.34. Матрица $B \geqslant 0$ называется *двоякостохастической*, если сумма элементов каждой ее строки и сумма элементов каждого ее столбца равны единице.

18.35. Любая матрица перестановок является двоякостохастической матрицей.

18.36. Симметричная стохастическая матрица является двоякостохастической.

18.37. Множество всех двоякостохастических матриц порядка n представляет собой в n-мерном вещественном пространстве выпуклый многогранник с матрицами перестановок в качестве его вершин.

18.38. Если двоякостохастическая матрица B порядка n неразложима, то индекс ее цикличности является делителем n.

18.39. Если $B \geqslant 0$ — вещественная симметричная положительно полуопределенная матрица, то существует вещественная диагональная матрица D такая, что DBD является двоякостохастической матрицей.

18.40. Квадратная матрица A называется *осцилляционной*, если $A \geqslant 0$ и существует такое целое $s > 0$, что A^s является вполне положительной матрицей (см. 4.63).

18.41. Трехдиагональная матрица

$$A = \begin{bmatrix} b_1 & c_1 & & 0 \\ a_2 & \cdot & & \\ & & & c_{n-1} \\ 0 & & a_n & b_n \end{bmatrix}$$

является осцилляционной в том и только в том случае, если все числа a_i и c_i и все ее главные миноры — положительные.

18.42. Осцилляционная матрица A порядка n всегда имеет n различных положительных собственных чисел λ_i, т. е.

$$0 < \lambda_1 < \lambda_2 < \ldots < \lambda_n.$$

18.43. Наибольшему характеристическому собственному числу $\lambda_n = \rho(A)$ осцилляционной матрицы A соответствует собственный вектор, все координаты которого — только положительные (или только отрицательные) числа, а у каждого собственного вектора, соответствующего собственному числу λ_s $(s > 1)$ в ряду координат, упорядоченных по возрастанию индексов, имеется точно $s - 1$ перемен знака.

§ 19. Матрицы специального вида

19.1. Трехдиагональная вещественная матрица

$$A = \begin{bmatrix} b_1 & c_1 & & & 0 \\ a_2 & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & c_{n-1} \\ 0 & & & a_n & b_n \end{bmatrix}$$

порядка n называется *якобиевой*, если $a_i c_{i-1} > 0$ для $i = 2, \ldots, n$.

19.2. Любая якобиева матрица является неразложимой.

19.3. Для любой якобиевой матрицы A и диагональной невырожденной матрицы D матрица DAD^{-1} является якобиевой.

19.4. Для любой якобиевой матрицы A существует такая диагональная невырожденная матрица $D = \operatorname{diag}(d_1, \ldots, d_n)$, что матрица DAD^{-1} — симметричная. В частности, для якобиевой матрицы A из 19.1 необходимо и достаточно, чтобы элементы d_i

матрицы D были связаны равенствами

$$d_i^2 = d_{i-1}^2 \frac{c_{i-1}}{a_i}, \qquad i = 2, \ldots, n,$$

при произвольном $d_1 \neq 0$.

19.5. Все собственные числа якобиевой матрицы вещественны и она обладает полной D^2-ортонормированной системой собственных векторов, где D — диагональная матрица из 19.4.

19.6. Для любой якобиевой матрицы A существуют такая диагональная невырожденная матрица D и такое вещественное число a, что матрица $DAD^{-1} + aE$ является осцилляционной матрицей. В обозначениях 19.1 диагональные элементы d_i такой матрицы D вычисляются по формулам

$$d_i = \operatorname{sign}(a_i)\, d_{i-1} \sqrt{\frac{c_{i-1}}{a_i}}, \qquad i = 2, \ldots, n,$$

при произвольном $d_1 \neq 0$, а в качестве величины a может быть взято любое число, удовлетворяющее неравенству

$$a > \max_{1 \leqslant i \leqslant n} (|a_i| + |b_i| + |c_i|),$$

где $a_1 = c_n = 0$.

19.7. Все собственные числа якобиевой матрицы A различны, т. е. $\lambda_1 < \lambda_2 < \ldots < \lambda_n$, и если все элементы a_i положительны (отрицательны), то для любого $k \geqslant 1$ число перемен знака в ряду координат собственного вектора u^k матрицы A, соответствующего собственному числу λ_k (нулевые значения координат собственного вектора не учитываются), в точности равно $n - k$ (равно $k - 1$).

19.8. Произведение якобиевой матрицы и диагональной матрицы с положительными диагональными элементами является якобиевой матрицей.

19.9. Пусть A и B — некоторые квадратные матрицы. Тогда задача нахождения чисел λ и ненулевых векторов u (вообще говоря, комплексных), являющихся решениями уравнения

$$Au = \lambda Bu,$$

называется *обобщенной проблемой собственных значений*. Искомые числа λ называются *собственными числами* обобщенной проблемы, а ненулевые векторы u — *собственными векторами*, соответствующими этим собственным числам λ.

19.10. Если матрица B невырожденная, то обобщенная проблема собственных значений 19.9 эквивалентна стандартной проблеме собственных значений

$$\tilde{A}u = \lambda u,$$

где $\tilde{A} = B^{-1}A$.

19.11. Если матрица B симметричная и положительно определенная, то обобщенная проблема собственных значений 19.9 эквивалентна стандартной проблеме собственных значений

$$\tilde{A}v = \lambda v,$$

где $\tilde{A} = B^{-1/2}AB^{-1/2}$, $v = B^{1/2}u$.

19.12. Если B — симметричная и положительно определенная матрица, A — симметричная матрица, то все собственные числа обобщенной проблемы собственных значений 19.9 вещественны и она обладает полной B-ортонормированной системой собственных векторов.

19.13. Если матрица B удовлетворяет условиям 19.11, а A — симметричная и положительно определенная (полуопределенная) матрица, то все собственные числа обобщенной проблемы собственных значений 19.9 положительны (неотрицательны).

19.14. Пусть A — симметричная якобиева матрица, B — диагональная матрица с положительными диагональными элементами. Тогда все собственные числа обобщенной проблемы собственных значений $Au = \lambda Bu$ вещественны и различны, и им соответствует полная B-ортонормированная система собственных векторов.

Дальнейший материал настоящего параграфа будет связан с симметричными положительно определенными и полуопределенными матрицами специального вида, с которыми очень часто приходится иметь дело на практике. Выпишем эти матрицы:

$$A_1 = \frac{1}{h_1}\begin{bmatrix} 2 & -1 & & 0 \\ -1 & \ddots & & \\ & & \ddots & -1 \\ 0 & & -1 & 2 \end{bmatrix}, \qquad A_2 = \frac{1}{h_2}\begin{bmatrix} 1 & -1 & & & 0 \\ -1 & 2 & & & \\ & & \ddots & & \\ & & & 2 & -1 \\ 0 & & & -1 & 1 \end{bmatrix},$$

$$A_3 = \frac{1}{h_3}\begin{bmatrix} 1 & -1 & & 0 \\ -1 & 2 & & \\ & & \ddots & -1 \\ 0 & & -1 & 2 \end{bmatrix}, \qquad A_4 = \frac{1}{h_4}\begin{bmatrix} 2 & -1 & & -1 \\ -1 & & 0 & \\ 0 & & \ddots & -1 \\ -1 & & -1 & 2 \end{bmatrix},$$

а также введем в рассмотрение следующие матрицы:

$$B_1 = h_1E, \quad B_2 = h_2 \operatorname{diag}(1/2, 1, \ldots, 1, 1/2),$$
$$B_3 = h_3 \operatorname{diag}(1/2, 1, \ldots, 1), \quad B_4 = h_4E,$$

$$\hat{B}_1 = \frac{h_1}{6}\begin{bmatrix} 4 & 1 & & 0 \\ 1 & \ddots & & \\ & & \ddots & 1 \\ 0 & & 1 & 4 \end{bmatrix}, \qquad \hat{B}_2 = \frac{h_2}{6}\begin{bmatrix} 2 & 1 & & & 0 \\ 1 & 4 & & & \\ & & \ddots & & \\ & & & 4 & 1 \\ 0 & & & 1 & 2 \end{bmatrix},$$

$$\widehat{B}_3 = \frac{h_3}{6}\begin{bmatrix} 2 & 1 & & & 0 \\ 1 & 4 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & \cdot & 1 \\ 0 & & & 1 & 4 \end{bmatrix}, \quad \widehat{B}_4 = \frac{h_4}{6}\begin{bmatrix} 4 & 1 & & & 1 \\ 1 & \cdot & & & 0 \\ & & \cdot & & \\ & & & \cdot & \\ & 0 & & 4 & 1 \\ 1 & & & 1 & 2 \end{bmatrix},$$

где $h_1 = X/(n+1)$, $h_2 = X/(n-1)$, $h_3 = h_4 = X/n$ (n — порядок всех выписанных матриц).

Перечисленные матрицы возникают, например, в связи с решением разностным методом (матрицы A_α и B_α, $\alpha = 1, 2, 3, 4$) или методом конечных элементов со стандартными кусочно линейными базисными функциями (матрицы A_α и $\widehat{B}_\alpha$, $\alpha = 1, 2, 3, 4$) одномерной задачи Штурма — Лиувилля

$$-d^2v/dx^2 = \mu v, \quad 0 < x < X,$$

при граничных условиях следующего типа:

$$v(0) = v(X) = 0 \qquad \text{(условия Дирихле, } \alpha = 1\text{);}$$

$$\left.\frac{dv}{dx}\right|_{x=0} = \left.\frac{dv}{dx}\right|_{x=X} = 0 \qquad \text{(условия Неймана, } \alpha = 2\text{);}$$

$$\left.\frac{dv}{dx}\right|_{x=0} = v(X) = 0 \qquad \text{(смешанные условия, } \alpha = 3\text{);}$$

$$v(0) = v(X), \quad \left.\frac{dv}{dx}\right|_{x=0} = \left.\frac{dv}{dx}\right|_{x=X} \qquad \text{(периодические условия, } \alpha = 4\text{).}$$

При этом величина h_α, $\alpha = 1, 2, 3, 4$ является шагом равномерной сетки, а величина n — размерностью алгебраической задачи для конкретных граничных условий.

Для удобства интерпретации некоторых излагаемых далее фактов нужно заметить, что собственные числа μ_k ($\mu_1 \leqslant \mu_2 \leqslant \ldots$, $\mu_k \to \infty$ при $k \to \infty$) и соответствующие им ортонормированные (в смысле нормы пространства $L_2(0, X)$ собственные функции сформулированной задачи Штурма — Лиувилля при перечисленных граничных условиях, $\alpha = 1, 2, 3, 4$, легко выписываются в явном виде:

1) $\mu_k = \left(\frac{k\pi}{X}\right)^2, \quad v_k = \sqrt{\frac{2}{X}}\sin\frac{k\pi x}{X};$

2) $\mu_k = \left(\frac{(k-1)\pi}{X}\right)^2, \quad v_1 = \frac{1}{\sqrt{X}}, \quad v_k = \sqrt{\frac{2}{X}}\cos\frac{(k-1)\pi x}{X}$ для $k > 1$;

3) $\mu_k = \left(\frac{(k-1/2)\pi}{X}\right)^2, \quad v_k = \sqrt{\frac{2}{X}}\cos\frac{(k-1/2)\pi x}{X};$

4) $$\mu_k = \begin{cases} \left(\frac{(k-1)\pi}{X}\right)^2, & k \text{ — нечетное;} \\ \left(\frac{k\pi}{X}\right)^2, & k \text{ — четное;} \end{cases}$$

$$v_k = \begin{cases} 1/\sqrt{X}, & k = 1; \\ \sqrt{\frac{2}{X}}\sin\frac{k\pi x}{X}, & k \text{ — четное;} \\ \sqrt{\frac{2}{X}}\cos\frac{(k-1)\pi x}{X}, & k \text{ — нечетное, } k > 1. \end{cases}$$

19.15. Матрицы A_1, A_3, B_α, $\widehat{B}_\alpha$, $\alpha=1, 2, 3, 4$,— симметричные и положительно определенные.

19.16. Матрицы A_2 и A_4 симметричные и положительно полуопределенные.

19.17. Все собственные числа обобщенных проблем собственных значений

$$A_\alpha u=\lambda B_\alpha u, \quad A_\alpha \widehat{u}=\widehat{\lambda}\widehat{B}_\alpha \widehat{u}$$

для $\alpha=1, 3$ положительны.

19.18. Все собственные числа обобщенных проблем собственных значений

$$A_\alpha u=\lambda B_\alpha u, \quad A_\alpha \widehat{u}=\widehat{\lambda}\widehat{B}_\alpha \widehat{u}$$

для $\alpha=2, 4$ неотрицательны.

19.19. Все собственные числа обобщенных проблем собственных значений $A_\alpha u=\lambda B_\alpha u$, $\alpha=1, 2, 3$, различны.

19.20. Обобщенные проблемы собственных значений 19.17 (19.18) обладают полными B_α-ортонормированными (соответственно $\widehat{B}_\alpha$-ортонормированными) системами собственных векторов.

19.21. Собственные числа $\lambda_1<\lambda_2<\ldots<\lambda_n$ обобщенных проблем собственных значений

$$A_\alpha u=\lambda B_\alpha u$$

и компоненты соответствующих им B_α-ортонормированных собственных векторов для $\alpha=1, 2, 3$ вычисляются по формулам ($k=1, \ldots, n$):

$$\alpha=1:\ \lambda_k=\frac{4}{h_1^2}\sin^2\frac{k\pi}{2(n+1)}=\frac{4}{h_1^2}\sin^2\frac{k\pi h_1}{2X},$$

$$u_{k,j}=\sqrt{\frac{2}{X}}\sin\frac{k\pi j}{n+1}, \quad j=1, \ldots, n;$$

$$\alpha=2:\ \lambda_k=\frac{4}{h_2^2}\sin^2\frac{(k-1)\pi}{2(n-1)},$$

$$u_{k,j}=\begin{cases}1/\sqrt{X}, & k=1;\\ \sqrt{\dfrac{2}{X}}\cos\dfrac{(k-1)\pi(j-1)}{n-1}, & k>1,\ j=1,\ldots,n;\end{cases}$$

$$\alpha=3:\ \lambda_k=\frac{4}{h_3^2}\sin^2\frac{(2k-1)\pi}{4n},$$

$$u_{k,j}=\sqrt{\frac{2}{X}}\cos\frac{(k-1/2)\pi(j-1)}{n}, \quad j=1, \ldots, n.$$

19.22. Собственные числа $\lambda_1<\lambda_2\leqslant\lambda_3<\lambda_4\leqslant\ldots\leqslant\lambda_n$ обобщенной проблемы собственных значений

$$A_4 u=\lambda B_4 u$$

и компоненты соответствующих им B_4-ортонормированных собственных векторов вычисляются по формулам:

$$\lambda_k = \begin{cases} \dfrac{4}{h_4^2}\sin^2\dfrac{(k-1)\pi}{2n}, & k \text{ — нечетное},\\[2mm] \dfrac{4}{h_4^2}\sin^2\dfrac{k\pi}{2n}, & k \text{ — четное};\end{cases}$$

$$u_{k,j} = \begin{cases} 1/\sqrt{X}, & k=1;\\ \sqrt{\dfrac{2}{X}}\sin\dfrac{k\pi j}{n}, & k \text{ — четное}, \quad k<n;\\ \sqrt{\dfrac{2}{X}}\cos\dfrac{(k-1)\pi j}{n}, & k \text{ — нечетное}, \quad k>1;\\ (-1)^j/\sqrt{X}, & k \text{ — четное}, \quad k=n;\end{cases}$$

$$j = 1, \ldots, n.$$

19.23. Трехдиагональная матрица

$$C = \begin{bmatrix} 0 & c_1 & & 0\\ a_2 & \cdot & & \\ & & \cdot & c_{n-1}\\ 0 & & a_n & 0 \end{bmatrix}$$

с положительными элементами a_i, c_{i-1}, $i = 2, \ldots, n$, является неотрицательной якобиевой циклической индекса 2 матрицей и, следовательно, если ν — собственное число C, то величина $-\nu$ также является собственным числом C.

19.24. Собственные числа ν_k матриц $C_\alpha = 2E - h_\alpha^2 B_\alpha^{-1} A_\alpha$ вычисляются по формулам ($k = 1, 2, \ldots, n$):

$$\nu_k = 2\cos\frac{k\pi}{n+1}, \qquad \alpha = 1;$$

$$\nu_k = 2\cos\frac{(k-1)\pi}{n-1}, \qquad \alpha = 2;$$

$$\nu_k = 2\cos\frac{(k-1/2)\pi}{n}, \qquad \alpha = 3.$$

19.25. Матрица $C_4 = 2E - h_4^2 B_4^{-1} A_4$ является неотрицательной неразложимой циклической индекса 2 матрицей, собственные числа ν_k которой вычисляются по формулам:

$$\nu_k = \begin{cases} 2\cos\dfrac{(k-1)\pi}{n}, & k \text{ — нечетное};\\[2mm] 2\cos\dfrac{k\pi}{n}, & k \text{ — четное};\end{cases}$$

т. е., если ν является собственным числом матрицы C_4, то величина $-\nu$ также является ее собственным числом.

19.26. Справедливы равенства

$$\widehat{B}_\alpha = B_\alpha\left(E - \frac{h_\alpha^2}{6} B_\alpha^{-1} A_\alpha\right), \qquad \alpha = 1, 2, 3, 4.$$

19.27. Собственные числа обобщенных проблем собственных значений

$$A_\alpha \widehat{u} = \widehat{\lambda}\widehat{B}_\alpha \widehat{u}$$

вычисляются по формулам

$$\widehat{\lambda}_k = \lambda_k \Big/ \left(1 - \frac{h_\alpha^2}{6}\lambda_k\right), \quad k = 1, \ldots, n,$$

а соответствующие им $\widehat{B}_\alpha$-ортонормированные собственные векторы вычисляются по формулам:

$$\widehat{u}_k = \frac{1}{\left(1 - \frac{h_\alpha^2}{6}\lambda_k\right)^{1/2}} u_k,$$

где λ_k и u_k — соответствующие собственные числа и B_α-ортонормированные собственные векторы обобщенных проблем собственных значений $A_\alpha u = \lambda B_\alpha u$ из 19.20, 19.21, $\alpha = 1, 2, 3, 4$.

19.28. Пусть $\mu_k^{(\alpha)}$ — собственные числа, $v_k^{(\alpha)} = v_k^{(\alpha)}(x)$ — соответствующие им собственные функции одномерных задач Штурма — Лиувилля, $\alpha = 1, 2, 3, 4$. Тогда для всех значений $\alpha = 1, 2, 3, 4$ и $k \geqslant 1$, для которых $\mu_k^{(\alpha)} \neq 0$, выполняются неравенства

$$\lambda_k^{(\alpha)} < \mu_k < \widehat{\lambda}_k^{(\alpha)},$$

причем для любого фиксированного k выполняются соотношения

$$\lim_{n\to\infty} \lambda_k^{(\alpha)} = \mu_k^{(\alpha)} - 0, \quad \lim_{n\to\infty} \widehat{\lambda}_k^{(\alpha)} = \mu_k^{(\alpha)} + 0,$$

где $\lambda_k^{(\alpha)}$ и $\widehat{\lambda}_k^{(\alpha)}$ — собственные числа обобщенных проблем собственных значений

$$A_\alpha u^{(\alpha)} = \lambda^{(\alpha)} B_\alpha u^{(\alpha)}, \quad A_\alpha \widehat{u}^{(\alpha)} = \widehat{\lambda}^{(\alpha)} \widehat{B}_\alpha \widehat{u}^{(\alpha)}.$$

Кроме того, для всех $k \geqslant 1$ компоненты B_α-ортонормированных собственных векторов $u_k^{(\alpha)}$ вычисляются по формулам

$$u_{k,j}^{(1)} = v_k^{(1)}(h_1 j), \qquad u_{k,j}^{(2)} = v_k^{(2)}(h_2(j-1)),$$

$$u_{k,j}^{(3)} = v_k^{(3)}(h_3(j-1)), \quad u_{k,j}^{(4)} = v_k^{(4)}(h_4 j), \qquad j = 1, \ldots, n,$$

19.29. Для любой симметричной трехдиагональной матрицы ($b_1, \ldots, b_n$ — некоторые вещественные числа)

$$A = \begin{bmatrix} b_1 & -1 & & 0 \\ -1 & \cdot & & \\ & & \cdot & \\ & & \cdot & -1 \\ 0 & & -1 & b_n \end{bmatrix},$$

главные миноры которой отличны от нуля, справедливо разложение

$$A = LDL^T,$$

где

$$D = \operatorname{diag}\{1/\beta_1, \ldots, 1/\beta_n\}, \quad L = \begin{bmatrix} 1 & & & 0 \\ -\beta_1 & \cdot & & \\ & & \cdot & \\ & & \cdot & \\ 0 & & -\beta_{n-1} & 1 \end{bmatrix},$$

и величины β_i вычисляются по формулам

$$\beta_i = \begin{cases} 1/b_1, & i = 1, \\ 1/(b_i - \beta_{i-1}), & i = 2, \ldots, n. \end{cases}$$

19.30. Для матрицы A_1 имеем $\beta_i = ih_1^{-1}/(i+1)$, $i = 1, \ldots, n$.

19.31. Для матрицы A_3 имеем $\beta_i = h_3^{-1}$, $i = 1, \ldots, n$.

19.32. Элементы $\tilde{a}_{ij}^{(\alpha)}$ матриц A_α^{-1}, $\alpha = 1, 3$, вычисляются по формулам:

$$\tilde{a}_{ij}^{(1)} = \begin{cases} \dfrac{(n-i+1)}{(n+1)} jh_1, & j < i, \\ \dfrac{(n-j+1)}{n+1} ih_1, & j \geqslant i, \end{cases} \qquad \tilde{a}_{ij}^{(3)} = \begin{cases} (n-i+1)h_3, & j < i, \\ (n-j+1)h_3, & j \geqslant i, \end{cases}$$

$$i, j = 1, \ldots, n.$$

§ 20. Неравенства и оценки

20.1. Положительные числа p, q называются *сопряженными*, если $p^{-1} + q^{-1} = 1$.

20.2. (*Неравенство Гельдера.*) Для любых неотрицательных чисел $x_1, \ldots, x_k$ и $y_1, \ldots, y_k$

$$\sum_{i=1}^{n} x_i y_i \leqslant \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} \left(\sum_{i=1}^{n} y_i^q\right)^{1/q},$$

причем равенство имеет место тогда и только тогда, когда векторы с координатами $x_1, \ldots, x_n$ и $y_1, \ldots, y_n$ коллинеарны.

20.3. (*Неравенство Коши — Буняковского.*) При $p = q = 2$ неравенство Гельдера называется неравенством Коши — Буняковского.

20.4. (*Неравенство Минковского.*) Для любой матрицы A размера $m\times n$ с неотрицательными элементами a_{ij} и любого числа $p\geqslant 1$

$$\left(\sum_{j=1}^{n}\left(\sum_{i=1}^{m}a_{ij}\right)^{p}\right)^{1/p}\leqslant\sum_{i=1}^{m}\left(\sum_{j=1}^{n}a_{ij}^{p}\right)^{1/p},$$

причем равенство имеет место тогда и только тогда, когда ранг матрицы A не превышает 1 или $p=1$.

20.5. Для любых неотрицательных чисел $x_1, \ldots, x_n$ и $y_1, \ldots, y_n$

$$\left(\prod_{i=1}^{n}(x_i+y_i)\right)^{1/n}\geqslant\left(\prod_{i=1}^{n}x_i\right)^{1/n}+\left(\prod_{i=1}^{n}y_i\right)^{1/n},$$

причем равенство имеет место тогда и только тогда, когда векторы с координатами $x_1, \ldots, x_n$ и $y_1, \ldots, y_n$ коллинеарны.

20.6. (*Тождество Лагранжа.*) Для любых чисел $a_1, \ldots, a_n$ и $b_1, \ldots, b_n$

$$\left(\sum_{i=1}^{n}a_i^2\right)\left(\sum_{i=1}^{n}b_i^2\right)-\left(\sum_{i=1}^{n}a_ib_i\right)^2=\sum_{j=1}^{n}\sum_{i<j}(a_ib_j-a_jb_i)^2.$$

20.7. Пусть $\lambda_1\leqslant\lambda_2\leqslant\ldots\leqslant\lambda_n$ — собственные значения эрмитовой матрицы A. Тогда для любой ортонормированной системы векторов $x_1, \ldots, x_k$

$$\sum_{i=1}^{k}\lambda_i\leqslant\sum_{i=1}^{k}(Ax_i, x_i)\leqslant\sum_{i=1}^{k}\lambda_{n-i+1}.$$

20.8. Пусть $\lambda_1\leqslant\lambda_2\leqslant\ldots\leqslant\lambda_n$ — собственные значения неотрицательной эрмитовой матрицы A. Тогда для любой ортонормированной системы векторов $x_1, \ldots, x_k$

$$\prod_{i=1}^{k}\lambda_i\leqslant\prod_{i=1}^{k}(Ax_i, x_i)\leqslant\left(\frac{1}{k}\sum_{i=1}^{k}\lambda_{n-i+1}\right)^{k}.$$

20.9. (*Теорема Виландта — Гофмана.*) Пусть A, B, C — нормальные матрицы порядка n, причем $C=A+B$. Обозначим через λ_i, β_i, γ_i их собственные значения, занумерованные в порядке неубывания модулей. Тогда

$$\sum_{i=1}^{n}|\gamma_i-\lambda_i|^2\leqslant\sum_{i=1}^{n}|\beta_i|^2.$$

20.10. Пусть A, B, C — эрмитовы матрицы порядка n, причем $C=A+B$. Обозначим через λ_i, β_i, γ_i их собственные значения, занумерованные в порядке неубывания. Тогда для любого набора m натуральных чисел $1 \leqslant i_1 < \ldots < i_m \leqslant n$

$$\sum_{k=1}^{m} \gamma_{i_k} \leqslant \sum_{k=1}^{m} \lambda_{i_k} + \sum_{k=1}^{m} \beta_{n-k+1}.$$

При $m = n$ достигается равенство.

20.11. В обозначениях 20.10 для любого m, $1 \leqslant m \leqslant n$,

$$\gamma_m \leqslant \lambda_n + \beta_m, \quad \gamma_m \leqslant \lambda_m + \beta_n,$$
$$\gamma_m \geqslant \lambda_1 + \beta_m, \quad \gamma_m \geqslant \lambda_m + \beta_1.$$

20.12. Если матрицы A, B, C эрмитовы, $C = A + B$ и $B > 0$, то при нумерации в порядке неубывания каждое собственное значение матрицы C больше соответствующего собственного значения матрицы A.

20.13. Пусть A, B, C — квадратные матрицы порядка n, причем $C = A + B$. Обозначим через ρ_i, μ_i, ν_i их сингулярные числа, занумерованные в порядке неубывания. Тогда для любого набора m натуральных чисел $1 \leqslant i_1 < \ldots < i_m \leqslant n$

$$\sum_{k=1}^{m} \nu_{i_k} \leqslant \sum_{k=1}^{m} \rho_{i_k} + \sum_{k=1}^{m} \mu_{n-k+1}.$$

20.14. В обозначениях 20.13 для любого m, $1 \leqslant m \leqslant n$,

$$\nu_m \leqslant \rho_n + \mu_m, \quad \nu_m \leqslant \rho_m + \mu_n,$$
$$\nu_m \geqslant -\rho_n + \mu_m, \quad \nu_m \geqslant \rho_m - \mu_n.$$

20.15. В условиях и обозначениях 20.13

$$\sum_{k=1}^{m} |\nu_{i_k} - \rho_{i_k}| \leqslant \sum_{k=1}^{m} \mu_{n-k+1}.$$

20.16. Пусть A, B, C — квадратные матрицы порядка n, причем $C = AB$. Обозначим через ρ_i, μ_i, τ_i их сингулярные числа, занумерованные в порядке неубывания. Тогда для любого набора m натуральных чисел $1 \leqslant i_1 < \ldots < i_m \leqslant n$

$$\prod_{k=1}^{m} \tau_{i_k} \leqslant \prod_{k=1}^{m} \rho_{i_k} \prod_{k=1}^{m} \mu_{n-k+1}, \quad \prod_{k=1}^{m} \tau_{i_k} \leqslant \prod_{k=1}^{m} \rho_{n-k+1} \prod_{k=1}^{m} \mu_{i_k}.$$

20.17. В обозначениях 20.16 при любом m, $1 \leqslant m \leqslant n$,

$$\tau_m \leqslant \rho_n \mu_m, \quad \tau_m \leqslant \rho_m \mu_n,$$
$$\tau_m \geqslant \rho_1 \mu_m, \quad \tau_m \geqslant \rho_m \mu_1.$$

20.18. Пусть A, B — эрмитовы положительно определенные матрицы порядка n и $C = AB$. Обозначим через λ_i, β_i, δ_i их собственные значения, занумерованные в порядке неубывания. Тогда для любого набора m натуральных чисел $1 \leqslant i_1 < \ldots < i_m \leqslant n$

$$\prod_{k=1}^{m} \delta_{i_k} \leqslant \prod_{k=1}^{m} \lambda_{i_k} \prod_{k=1}^{m} \beta_{n-k+1}, \quad \prod_{k=1}^{m} \delta_{i_k} \leqslant \prod_{k=1}^{m} \beta_{i_k} \prod_{k=1}^{m} \lambda_{n-k+1}.$$

20.19. В обозначениях 20.18 при любом m, $1 \leqslant m \leqslant n$,

$$\delta_m \leqslant \lambda_n \beta_m, \quad \delta_m \leqslant \lambda_m \beta_n,$$
$$\delta_m \geqslant \lambda_1 \beta_m, \quad \delta_m \geqslant \lambda_m \beta_1.$$

20.20. (*Неравенства Г. Вейля.*) Пусть собственные значения матрицы A порядка n занумерованы так, что $|\lambda_1| \leqslant |\lambda_2| \leqslant \ldots \leqslant |\lambda_n|$, и $\rho_1 \leqslant \rho_2 \leqslant \ldots \leqslant \rho_n$ — ее сингулярные числа. Тогда при любом $m \leqslant n$ и при $\sigma > 0$

$$\prod_{i=1}^{m} |\lambda_{n-i+1}| \leqslant \prod_{i=1}^{m} \rho_{n-i+1},$$
$$\prod_{i=1}^{m} |\lambda_i| \geqslant \prod_{i=1}^{m} \rho_i,$$
$$\sum_{i=1}^{m} |\lambda_{n-i+1}|^{\sigma} \leqslant \sum_{i=1}^{m} \rho_{n-i+1}^{\sigma}.$$

20.21. Рассмотрим две неубывающие последовательности по n неотрицательных чисел:

$$\alpha_1 \leqslant \alpha_2 \leqslant \ldots \leqslant \alpha_n, \quad \beta_1 \leqslant \beta_2 \leqslant \ldots \leqslant \beta_n.$$

Говорят, что последовательность $\beta_1, \ldots, \beta_n$ *мажорируется* последовательностью $\alpha_1, \ldots, \alpha_n$, если

$$\sum_{k=1}^{m} \beta_{n-k+1} \leqslant \sum_{k=1}^{m} \alpha_{n-k+1}$$

при $1 \leqslant m < n$, причем при $m = n$ выполняется равенство. Если через α и β обозначить векторы с соответствующими координатами, то в этом случае пишут $\beta \leqslant \alpha$.

20.22. Последовательность β мажорируется последовательностью α тогда и только тогда, когда существует двоякостохастическая матрица A такая, что $\beta = A\alpha$.

20.23. Пусть $\varphi(t)$ — непрерывная выпуклая монотонно возрастающая функция. Если для последовательностей 20.21 нестрогие неравенства имеют место при всех m, включая $m = n$, то

$$\sum_{k=1}^{n} \varphi(\beta_k) \leqslant \sum_{k=1}^{n} \varphi(\alpha_k).$$

20.24. Если последовательность β мажорируется последовательностью α, то требование монотонности $\varphi(t)$ в 20.23 является излишним.

20.25. Пусть A — эрмитова матрица порядка n, представленная в блочном виде:

$$A = \begin{bmatrix} H & B^* \\ B & U \end{bmatrix},$$

где H — эрмитова матрица порядка m. Обозначим через $\lambda_1 \leqslant \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n$ собственные значения A, через $\theta_1 \leqslant \theta_2 \leqslant \ldots \leqslant \theta_m$ — собственные значения матрицы H и будем для удобства считать, что $\theta_i = -\infty$ для $i \leqslant 0$ и $\theta_i = +\infty$ для $i > m$. Для всех j имеют место неравенства

$$\lambda_j \leqslant \theta_j \leqslant \lambda_{n-m+j}, \quad \theta_{m-n+j} \leqslant \lambda_j \leqslant \theta_j.$$

20.26. Пусть A — эрмитова матрица, представленная в блочном виде:

$$A = \begin{bmatrix} H & C^* & 0^* \\ C & V & Z^* \\ 0 & Z & W \end{bmatrix}; \quad \begin{matrix} H\colon m \times m, \\ V\colon k \times k, \\ A\colon n \times n. \end{matrix}$$

Наряду с обозначениями 20.25, обозначим через $\mu_i = \mu_i(X)$ собственные значения эрмитовой матрицы

$$M(X) = \begin{bmatrix} H & C^* \\ C & X \end{bmatrix}.$$

Если для некоторой матрицы X, такой, что матрица $V - X$ невырожденная, известны собственные значения, то:

— каждый интервал $[\mu_j, \mu_{j+k}]$, $j = 1, \ldots, m$, содержит собственное значение λ_J матрицы A;

— между интервалами и числами λ_J можно установить взаимно однозначное соответствие;

— во внешности каждого открытого интервала (μ_i, μ_{i+m}), $i = 1, \ldots, k$, имеется собственное значение λ_I;

— между интервалами и числами λ_I можно установить взаимно однозначное соответствие.

20.27. Пусть ζ — произвольное число, удовлетворяющее неравенствам $\theta_j < \zeta < \theta_{j+1}$. Рассмотрим матрицу

$$X_\zeta = \zeta E + C(H - \zeta E)^{-1} C^*.$$

Каждый из интервалов $[\mu_j(X_\zeta), \zeta]$ и $[\zeta, \mu_{j+k+1}(X_\zeta)]$ содержит по крайней мере одно собственное значение матрицы A. Кроме того, существует матрица A, собственными значениями которой являются только концевые точки этих интервалов, помимо собственных значений, лежащих вне обоих интервалов.

20.28. Пусть матрица A вида 20.26 такова, что отрезок $[\alpha', \alpha'']$ содержит то же число собственных значений λ матрицы A, что и собственных значений θ матрицы H, т. е. для индексов π, k, l

$$\lambda_\pi < \alpha' \leqslant \lambda_{\pi+1} \leqslant \ldots \leqslant \lambda_{\pi+l} \leqslant \alpha'' < \lambda_{\pi+l+1},$$

$$\theta_\nu < \alpha' \leqslant \theta_{\nu+1} \leqslant \ldots \leqslant \theta_{\nu+l} < \alpha'' < \theta_{\nu+l+1};$$

тогда $\nu \leqslant \pi$ и для $1 \leqslant j \leqslant l$

$$\mu_{\nu+j}(X_{\alpha''}) \leqslant \lambda_{\pi+j} \leqslant \mu_{\nu+k+j}(X_{\alpha'}).$$

20.29. Пусть $\|C\|_E$ настолько мала, что в обозначениях 20.28 выполняется неравенство $\alpha' + \|C\|_E \leqslant \theta_{q+1}$. Тогда для $1 \leqslant j \leqslant l$

$$\theta_{q+j} \leqslant \mu_{q+k+j}(X_{\alpha'}) \leqslant \theta_{q+j} + \|C\|^2/(\theta_{q+j} - \alpha').$$

ГЛАВА 2

ЧИСЛЕННЫЕ МЕТОДЫ

§ 21. Математические особенности машинной арифметики

Общий эффект от решения задачи и даже возможность ее решения во многом определяются тем, как в действительности выполняются операции над числами. Это, в свою очередь, зависит от принятой системы записи чисел, или, как говорят, *системы счисления*.

21.1. Пусть заданы целое число $p > 1$ и целые числа $\alpha_0, \alpha_1, \ldots$ $\ldots, \alpha_{p-1}$. Если любое число x может быть представлено в виде ряда

$$x = \pm(b_n p^n + b_{n-1}p^{n-1} + \ldots + b_0 + b_{-1}p^{-1} + b_{-2}p^{-2} + \ldots),$$

где каждый из коэффициентов b_i может принимать одно из значений $\alpha_0, \alpha_1, \ldots, \alpha_{p-1}$, то запись вида

$$x = \pm b_n b_{n-1} \ldots b_0, \ b_{-1} b_{-2} \ldots$$

называется *позиционной системой* счисления.

Название этих систем связано с тем, что роль, которую играет каждый коэффициент в записи 21.1, зависит от занимаемой им позиции. Отсчет позиции определяется положением запятой или, что то же самое, положением коэффициента b_0. Число p в 21.1 называется *основанием* системы счисления, числа $\alpha_0, \alpha_1, \ldots, \alpha_{p-1}$ — *базисными*, правая часть записи числа x — *p-ичной дробью*, коэффициенты b_i — *разрядами*. Дробь называется *бесконечной*, если в ее записи содержится бесконечно много ненулевых разрядов, и *конечной* — в противном случае. Если для разрядов в записи 21.1 используется единая нумерация, то они нумеруются справа налево и им приписываются номера, соответствующие степеням p ряда 21.1.

Обычно в записи 21.1 опускаются все первые и последние нулевые разряды. Опускается и запятая, если разряды после нее являются нулевыми.

21.2. Если базисные числа образуют совокупность $\alpha_k = k$, то любое вещественное число может быть представлено в виде p-ичной дроби.

21.3. Если p — нечетное и базисные числа образуют совокупность $\alpha_k = (1 + 2k - p)/2$, то любое вещественное число может быть представлено в виде p-ичной дроби.

21.4. Позиционные системы 21.3 называются *сокращенными*.

21.5. В сокращенной позиционной системе базисные числа симметричны относительно нуля.

21.6. В сокращенной позиционной системе знак числа совпадает со знаком старшего разряда.

Арифметические операции над числами, заданными в любой позиционной системе счисления, производятся по таким же правилам, что и в десятичной системе. Это объясняется тем, что все операции основаны на правилах выполнения действий над соответствующими полиномами. Позиционные системы счисления широко применяются для представления чисел в современной вычислительной технике. Наиболее часто используется простейшая из них — двоичная система счисления. Среди сокращенных систем более употребительна троичная система.

21.7. *Округлением* числа до s разрядов в заданной системе счисления называется операция замены этого числа таким числом, все разряды которого в той же системе счисления, начиная с $s-1$-го, являются нулевыми. Разность между округленным и округляемым числами называется *ошибкой округления.*

Заметим, что в данном определении ничего не говорится ни о способе выполнения операции округления, ни о том, насколько округленное число близко к округляемому. Это не случайно. В практике конструирования ЭВМ операции округления реализуются различными способами. Единственное, что их объединяет,— это малость ошибки округления по крайней мере для большинства чисел.

21.8. Каков бы ни был способ округления чисел x до s-разрядных чисел x_s, всегда найдутся такие числа x, что $|x_s-x| \geqslant 0{,}5p^s$ в любой системе с базисными числами $|\alpha_k| \leqslant p$.

21.9. *Усечением* числа x до s разрядов называется такое его округление до s разрядов, при котором сохраняются все разряды слева до s-го включительно.

21.10. Если x_s есть усечение числа x до s разрядов, то $|x_s-x| \leqslant p^s$ в любой системе с базисными числами $\alpha_k=k$, и $|x_s-x| \leqslant 0{,}5p^s$ в любой сокращенной системе.

Хотя усечение чисел не является в общем случае лучшим способом округления, тем не менее именно с ним тесно связаны все другие способы. В самом деле, как бы ни выполнялась операция округления, ее результатом будет число, все разряды которого, начиная с $s-1$-го, являются нулевыми. Следовательно, операцию округления всегда можно практиковать как отбрасывание всех разрядов, начиная с $s-1$-го, и последующее добавление или вычитание некоторого числа, кратного p^s. Для того чтобы ошибка округления была малой, необходимо, чтобы было малым и это число.

21.11. Округление до s разрядов называется *правильным*, если для любого числа x и округленного числа x_s имеем $|x_s-x| \leqslant 0{,}5p^s$.

21.12. В любой сокращенной системе счисления усечение чисел обеспечивает правильное округление.

21.13. В любой p-ичной системе счисления с базисными числами $\alpha_k=k$ возможна реализация правильного округления.

21.14. В условиях 21.13 правильное округление обеспечивает минимальную величину модуля ошибки округления среди всех способов округления.

Правильное округление есть не что иное, как обычное школьное правило округления чисел, перенесенное на другие позиционные системы счисления. Конечно, оно несколько сложнее, чем усечение чисел, так как приходится анализировать отбрасываемую часть числа и дополнительно прибавлять или вычитать единицу последнего оставляемого разряда. Однако усечение чисел обладает серьезными недостатками. Они связаны не столько с большей величиной ошибки округления, сколько с тем фактом, что ошибка округления всегда имеет один и тот же знак, противоположный знаку округляемого числа. Это явление нежелательно, так как приводит во многих случаях к быстрому накоплению ошибок в вычислительных процессах. Правильное округление таких недостатков не имеет. Всюду в дальнейшем в этой книге, где придется иметь дело с ошибками округления, будем предполагать, без дополнительного пояснения, что округление является правильным.

21.15. При любом целом $p > 1$ любое ненулевое число x можно представить в виде $x = ap^b$, где b — целое число и $1/p \leqslant |a| < 1$. Число a называется *мантиссой* числа x, число b — его *порядком*.

21.16. Представление числа x в виде записи 21.1 называется представлением с *фиксированной запятой*, если фиксирована позиция нулевого разряда.

21.17. Представление числа x в виде 21.15, где мантисса представлена с фиксированной запятой, называется представлением *с плавающей запятой*.

В любой вычислительной машине на представление каждого числа отводится одно и то же количество базисных элементов, моделирующих числовой разряд p-ичной системы счисления. Пусть число таких элементов равно τ. Этими элементами можно распорядиться по-разному. Можно, например, воспользоваться представлением с фиксированной запятой и отвести r элементов на разряды слева от запятой и $\tau - r$ элементов на разряды справа от запятой. Но можно воспользоваться представлением с плавающей запятой и отвести r элементов на порядок и $\tau - r$ элементов на мантиссу. Естественно спросить, какой из этих вариантов лучше? Ответ на этот вопрос не однозначен.

Если числа, с которыми приходится иметь дело, имеют приблизительно одинаковые порядки, то нет необходимости использовать представление с плавающей запятой с большим диапазоном изменения порядка. В этом случае гораздо выгоднее использовать представление с фиксированной запятой, так как можно обеспечить бо́льшую точность. Если же числа могут иметь очень большой разброс своих величин, то представление с фиксированной запятой невыгодно, так как малые числа будут представлены с очень малой точностью. В целом для представления с фиксированной запятой характерна высокая абсолютная точность, с плавающей запятой — высокая относительная точность.

Для фиксированной и плавающей запятой разряды нумеруются несколько иначе, чем указано ранее. Именно, для фиксированной запятой разряды слева от запятой нумеруются подряд справа налево, начиная с нулевого, справа от запятой — подряд слева направо, начиная с первого. Для плавающей запятой нумеруются только разряды мантиссы. Если говорят, что число представлено с t разрядами, то, как правило, имеют в виду либо t разрядов справа от запятой, либо t разрядов мантиссы.

21.18. Через $\mathrm{fi}(x)$ обозначается конечная дробь, которая получается после округления числа x с фиксированной запятой до t-го разряда справа от запятой.

21.19. Через $\mathrm{fl}(x)$ обозначается конечная дробь, которая получается после округления мантиссы числа x с плавающей запятой до t-го разряда.

Для фиксированной запятой округление до t-го разряда обеспечивает равномерную малость абсолютной ошибки для всех чисел. Для плавающей запятой положение несколько сложнее. Если числа могут быть представлены в ЭВМ, то для них округление мантиссы до t-го разряда обеспечивает равномерную малость относительной ошибки. Если же числа настолько малы, что их порядок не может быть изображен с помощью отведенного числа разрядов, то такие числа представляются в ЭВМ нулем. В этом случае относительная ошибка будет по модулю равна единице. Минимальное положительное число ω, которое может быть представлено с плавающей запятой, иногда называется *машинным нулем.*

Для современных ЭВМ величина p^{-t} обычно находится в пределах $10^{-10}-10^{-20}$, величина ω — в пределах $10^{-18}-10^{-40}$. Вообще говоря, p^{-t} и ω не связаны между собой. Однако на всех ЭВМ, за исключением ЭВМ с переменной длиной мантиссы, выполняется соотношение $\omega < p^{-1,5t}$. Несмотря на «малость» ω, вычисления со столь малыми числами приходится проводить значительно чаще, чем может показаться на первый взгляд.

На ЭВМ, работающей с фиксированной запятой, обычно допускаются числа, не превосходящие по модулю единицы, на ЭВМ с плавающей запятой,— не превосходящие ω^{-1}. Если в процессе вычислений появляются числа, выходящие за эти границы, то в большинстве случаев вычислительный процесс останавливается. Это явление принято называть *переполнением*. Конечно, его надо учитывать при реализации алгоритмов на ЭВМ.

Отмеченные особенности представления чисел не могут быть устранены какими-либо техническими средствами. Можно сконструировать ЭВМ со сколь угодно малым числом ω. Но все равно оно будет отлично от нуля. Можно построить ЭВМ, у которой операция округления будет реализована самым лучшим способом. Но все равно ошибки округления останутся. Нельзя, по существу, избежать и переполнения.

Чтобы не заниматься излишними деталями, связанными с особенностями представления чисел на конкретных ЭВМ, мы предположим выполнение двух основных гипотез, касающихся операции округления. Отклонения от этих гипотез в большинстве случаев приводят лишь к небольшому изменению числовых коэффициентов в итоговых оценках.

21.20. (Гипотеза.) Если числа округляются до t разрядов, то выполняются следующие соотношения:

$$\mathrm{fi}(x) \equiv x + \nu, \qquad |\nu| \leqslant \frac{1}{2}p^{-t},$$

$$\mathrm{fl}(x) \equiv x(1+\varepsilon), \quad |\varepsilon| \leqslant \frac{1}{2}p^{-t+1}, \text{ если } |x| \geqslant \omega, \ \varepsilon = -1,$$

$$\text{если } |x| < \omega, \ \omega < p^{-1,5t}.$$

21.21. (Гипотеза.) Выполнение любой отдельной арифметической операции ($+$, $-$, $\times$, $:$, $\sqrt{\ }$) эквивалентно точному выполнению этой операции и последующему округлению согласно 21.20. Операции с нулевым аргументом выполняются точно.

Символы fi и fl используются не только для указания способов представления и округления чисел, но и для указания режимов вычисления сложных выражений.

21.22. Если под символом fi или fl стоит сложное выражение, то это означает следующее: выражение каким-либо способом представляется в виде последовательности алгебраических операций, и каждая операция выполняется согласно 21.21.

При реализации некоторых алгоритмов возникает необходимость выполнять отдельные промежуточные вычисления с существенно большей точностью, чем допускается принятой системой представления чисел. На большинстве современных ЭВМ такие вычисления организуются следующим образом. Во-первых, имеется техническая возможность получать результаты выполнения основных арифметических операций над t-разрядными числами не с t разрядами, а с $2t$ разрядами. При этом ошибка округления обычно искажает лишь последние из этих разрядов. Во-вторых, имеется возможность программного доступа как к старшим t разрядам результата, так и к младшим его t разрядам. Используя эту возможность, можно программным путем осуществлять любые вычисления со сколь угодно большой точностью.

21.23. Пусть требуется вычислить некоторое сложное выражение, аргументами которого являются t-разрядные числа. Режим вычислений, при котором все промежуточные вычисления осуществляются с $2t$ разрядами и лишь результат округляется до t разрядов, называется режимом *накопления*. Он обозначается соответственно символом fi_2 или fl_2.

Какими бы малыми ни были ошибки округления, возникающие при выполнении арифметических операций, их появление существенно меняет математические свойства самих операций. Точные операции умножения и сложения, например, являются коммутативными, ассоциативными и связаны между собой законом дистрибутивности. Эти операции на ЭВМ, вообще говоря, не являются таковыми.

21.24. При любом способе округления на ЭВМ с плавающей запятой операции сложения и умножения не являются ассоциативными и не связаны между собой законом дистрибутивности.

21.25. Коммутативность операций сложения и умножения может быть достигнута в том случае, если ошибка округления однозначно определяется результатом точного выполнения операций.

21.26. Операция сложения на ЭВМ с фиксированной запятой коммутативна, но не связана с операцией умножения законом дистрибутивности.

С точки зрения точного выполнения операций расстановка скобок обычно не бывает однозначной. Но при реализации в условиях ошибок округления различные расстановки скобок в одном и том же арифметическом выражении будут приводить к различным результатам. Поэтому всякая задача при постановке на ЭВМ определяет в действительности целую совокупность вычислительных алгоритмов, отличающихся друг от друга порядком выполнения операций. Несмотря на математическую эквивалентность всех этих модификаций в точном смысле, различие в вычислительном эффекте может быть огромным, в особенности с точки зрения численной устойчивости.

Какой бы ни была эта совокупность «приближенно эквивалентных» алгоритмов, среди них находится алгоритм, обеспечивающий наибольшую точность. Найти его или близкий к нему — трудная задача, особенно в сложных вычислениях. Тем не менее о принципиальной возможности оптимизации алгоритмов в отношении точности по порядку выполнения операций забывать не следует.

За исключением редких случаев, ошибка округления появляется в каждой арифметической операции. При реализации на ЭВМ сложного вычислительного алгоритма на его окончательный результат будет оказывать влияние очень большое число ошибок округления результатов промежуточных вычислений. Общий эффект влияния ошибок обычно описывается одним из следующих двух способов.

Обозначим через A входные данные задачи, через $B = \varphi(A)$ — результат их обработки по некоторому точному алгоритму φ. При реализации на ЭВМ алгоритм φ будет заменен другим, «близким» алгоритмом φ_t, в силу неизбежных отличий машинной арифметики от точной. Следовательно, вместо B будет получен результат $B_t = \varphi_t(A)$.

21.27. Исследование отклонения приближенно вычисленного решения B_t от точного решения B называется *прямым анализом ошибок*.

Во многих задачах реально вычисленное решение B_t можно рассматривать как результат обработки некоторых возмущенных входных данных A_t по точному алгоритму φ, т. е. $B_t = \varphi(A_t)$. В этом случае ошибку вычисленного решения характеризует также отклонение A_t от A.

21.28. Если точное решение B есть результат реализации некоторого алгоритма над входными данными A, а приближенно вычисленное решение B_t можно рассматривать как результат реализации того же точного алгоритма над входными данными A_t, то отклонение A_t от A называется *эквивалентным возмущением*.

21.29. Исследование эквивалентного возмущения называется *обратным анализом ошибок*.

Заметим, что эквивалентное возмущение, вообще говоря, определяется не единственным образом. Поэтому в обратном анализе ошибок исключительно важным и самым трудным моментом является доказательство существования эквивалентного возмущения, соответствующего той или иной оценке. Конечно, при этом необходимо стремиться к тому, чтобы установить по возможности существование наименьшего эквивалентного возмущения.

Оценивая в дальнейшем эффект влияния ошибок округления, мы будем указывать лишь главные члены оценок по малости величины p^{-t}. Для главных членов соотношений снова будем использовать символы $\cong$, $\lessapprox$ и т. д. Всюду предполагаем, что при осуществлении вычислений выполняются гипотезы 21.20, 21.21.

§ 22. Элементарные матрицы и преобразования

22.1. Вещественные матрицы, отличающиеся от единичной матрицы четырьмя элементами, расположенными на пересечении строк и столбцов с номерами i, j, и имеющие вид

$$T_{ij} = \begin{array}{cc} \begin{array}{ccccccc} & & & i & & j & \end{array} & \\ \begin{bmatrix} 1 & & & \vdots & & \vdots & 0 \\ & \ddots & & \vdots & & \vdots & \\ & & & c & \dots & -s & \\ & & & \vdots & \ddots & & \\ & & & \pm s & \dots & \pm c & \\ & & & & & & \ddots \\ 0 & & & & & & 1 \end{bmatrix} & \begin{array}{c} \\ \\ i \\ \\ j \\ \\ \\ \end{array} \end{array},$$

где $c^2 + s^2 = 1$, называются *матрицами вращения* (*простого поворота, Гивенса*).

22.2. Матрицы вращения с элементами c, c (c, $-c$) на диагонали называются *правыми* (*левыми*) матрицами вращения.

22.3. При любых c, s, где $c^2 + s^2 = 1$, матрицы вращения являются ортогональными.

22.4. При умножении вектора на матрицу вращения T_{ij} меняются только i-я и j-я координаты вектора.

Как правые, так и левые матрицы вращения с одинаковым успехом могут быть использованы в самых различных вычислительных процессах. Различие между ними проявляется лишь при компактном хранении информации о выполненном преобразовании. Как правило, если не сделано специальной оговорки, в дальнейшем под матрицами вращения будут пониматься правые матрицы вращения. Прежде чем приступать к исследованию последовательностей матриц вращения общего вида, изучим подробнее вещественное преобразование второго порядка.

22.5. При любых c и s, не равных нулю одновременно, вещественная матрица

$$T = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$$

становится ортогональной после деления на множитель $\tau = (c^2 + s^2)^{1/2}$.

22.6. Каков бы ни был вектор b с координатами x, y, существует ортогональная матрица T, для которой вторая координата вектора Tb равна нулю, например:

$$c = \begin{cases} 1, & a = 0, \\ \dfrac{x}{(x^2 + y^2)^{1/2}}, & a \neq 0, \end{cases} \qquad s = \begin{cases} 0, & a = 0, \\ -\dfrac{y}{(x^2 + y^2)^{1/2}}, & a \neq 0. \end{cases}$$

22.7. В условиях 22.6 первая координата вектора Tb равна

$$z = \begin{cases} 0, & a = 0, \\ (x^2 + y^2)^{1/2}, & a \neq 0. \end{cases}$$

Из условия $c^2+s^2=1$ следует, что величины c и s можно рассматривать как $\cos\alpha$ и $\sin\alpha$ для некоторого угла α. В этом случае матрицу T можно трактовать как матрицу преобразования, заключающегося в повороте плоскости вокруг начала координат на угол α. Название матрицы объясняется именно этим фактом. Как вытекает из 22.6, всегда можно подобрать такой поворот, при котором любой заданный вектор перейдет в вектор, коллинеарный координатному.

22.8. Пусть $\tilde{T}$ — реально заданная или реально вычисленная по некоторому вектору b согласно 22.6 матрица вращения. Возьмем любой вещественный вектор a. Если $\tilde{\tau}$ и $\|a\|_E$ много больше машинного нуля, то при умножении матрицы $\tilde{T}$ на вектор a выполняются соотношения

$$\mathrm{fl}(\tilde{T}a) \equiv \tilde{T}a + f, \qquad \|f\|_E \lessapprox \sqrt{2}\tilde{\tau}p^{-t+1}\|a\|_E,$$
$$\mathrm{fl}(\tilde{T}a) \equiv \tilde{T}(a+\varepsilon), \qquad \|\varepsilon\|_E \lessapprox \sqrt{2}p^{-t+1}\|a\|_E.$$

22.9. Для матрицы $\tilde{T}$, реально вычисленной согласно 22.6, матрица $\tilde{T}\tilde{T}'$ является скалярной, и при этом выполняются следующие оценки:

$$\|\tilde{T}\tilde{T}' - E\|_2 \lessapprox \frac{5}{2}p^{-t+1}, \qquad \|\tilde{T} - T\|_2 \lessapprox \frac{5}{4}p^{-t+1},$$
$$\tilde{\tau} = (\tilde{c}^2 + \tilde{s}^2)^{1/2} = 1 + \nu, \qquad |\nu| \lessapprox \frac{5}{4}p^{-t+1}.$$

22.10. Пусть по вектору b реально вычисляется матрица $\tilde{T}$ (согласно 22.6) и согласно 22.7 вычисляется единственная ненулевая координата вектора Tb; тогда

$$\mathrm{fl}(Tb) \equiv \tilde{T}(b+\varepsilon), \qquad \|\varepsilon\|_E \lessapprox \frac{\sqrt{5}}{2}p^{-t+1}\|b\|_E.$$

Приведенные оценки говорят о том, что реально вычисленная матрица вращения $\tilde{T}$ с высокой степенью точности не только вообще близка к какой-то ортогональной матрице, но даже близка к ортогональной матрице, получаемой при точных вычислениях. При этом оказываются малыми и эквивалентные возмущения преобразуемых векторов.

Если $\|b\|_E$ по своей величине соизмерима с машинным нулем, то матрица $\tilde{T}$ уже может не быть близкой к ортогональной матрице T, получающейся при точных вычислениях. Однако вычисления всегда можно организовать так, что $\tilde{T}$ будет близка к некоторой ортогональной матрице. Собственно говоря, только это мы и будем использовать в дальнейшем.

В связи со сказанным мы хотим еще раз подчеркнуть одно исключительно важное обстоятельство. Приведенные оценки имеют место только в том случае, когда при выполнении каждой операции обеспечивается относительная погрешность результата не более $0{,}5\ p^{-t+1}$ в соответствии с 21.24, 21.25. Для достижения этих оценок формулы 22.6, 22.7 приходится преобразовывать. Необходимость подобных преобразований связана, например, с возможностью появления очень малых чисел x, y. В этом случае нужно быть особенно осторожным при вычислении величины $(x^2+y^2)^{1/2}$. Если x, y находятся вблизи машинного нуля, то эта величина может оказаться равной нулю или иметь очень малую относительную точность. Тогда реально вычисленная матрица $\tilde{T}$ может оказаться далекой от любой ортогональной матрицы. Такие ситуации часто встречаются в практических вычислениях.

Конечно, было бы наиболее правильно точно описывать ту последовательность операций, для которой имеют место приводимые оценки. Но это, по существу, означает, что нужно приводить программы алгоритмов на одном из алгоритмических машинно-независимых языков. Такая задача не ставится в данной книге. Приводя эти и все последующие оценки, мы хотим в первую очередь показать, какого эффекта в отношении точности можно добиться при правильной реализации алгоритмов.

22.11. Пусть n-мерный вектор z умножается на последовательность из N матриц вращения $T_{i_1j_1}, \ldots, T_{i_Nj_N}$. Предположим, что $\tilde{T}_{i_1j_1}, \ldots, \tilde{T}_{i_Nj_N}$ — реально заданные или реально вычисленные матрицы вращения с соответствующими матрицами второго порядка, удовлетворяющими 22.9. Тогда для любой последовательности пар индексов $i_1j_1, \ldots, i_Nj_N$ имеют место соотношения

$$\mathrm{fl}\,(\tilde{T}_{i_Nj_N} \ldots \tilde{T}_{i_1j_1}z) \equiv (\tilde{T}_{i_Nj_N} \ldots \tilde{T}_{i_1j_1})(z + \mathscr{E}),$$

$$\|\mathscr{E}\|_E \lessapprox \sqrt{2}Np^{-t+1}\|z\|_E.$$

22.12. Последовательность матриц вращения называется *сильно связанной*, если любые две соседние матрицы имеют хотя бы один общий индекс.

22.13. Оценка эквивалентного возмущения 22.11 для произвольных векторов z достигается на множестве сильно связанных последовательностей матриц вращения.

22.14. Последовательность матриц вращения называется *несвязанной*, если все индексы матриц различны.

22.15. Для того чтобы последовательность $T_{i_1j_1}, \ldots, T_{i_Nj_N}$ была несвязанной, необходимо, чтобы N не превосходило $n/2$.

22.16. Результат выполнения несвязанной последовательности преобразований с матрицами вращения, включая всю совокупность ошибок округления, не зависит от порядка выполнения самих преобразований.

22.17. В условиях и обозначениях 22.11 при любой несвязанной последовательности матриц вращения для эквивалентного возмущения $\mathscr{E}$ справедлива оценка

$$\|\mathscr{E}\|_E \lessapprox \sqrt{2}p^{-t+1}\|z\|_E.$$

22.18. Оценка эквивалентного возмущения 22.17 для произвольных векторов z достигается на множестве несвязанных последовательностей матриц вращения.

22.19. Если последовательность матриц вращения можно разбить на k групп так, что внутри каждой группы матрицы вращения не имеют одинаковых индексов, то в условиях и обозначениях 22.11 справедлива оценка

$$\|\mathscr{E}\|_E \lessapprox \sqrt{2}kp^{-t+1}\|z\|_E.$$

Выполнение любой последовательности из N преобразований вращения можно трактовать как выполнение N несвязанных последовательностей, каждая из которых содержит лишь одно преобразование. При этом для суммарного эквивалентного возмущения $\mathcal{E}$ справедливы как неравенство 22.11, так и неравенство 22.19. В данном случае обе оценки, конечно, совпадают.

В общем случае выполнение преобразований вращения можно сводить к несвязанным последовательностям не единственным способом, что видно на примере самой несвязанной последовательности. Чтобы на основе формулы 22.19 получить наилучшую оценку эквивалентного возмущения, необходимо определить минимальное число несвязанных последовательностей, к которым сводится исходная последовательность.

22.20. Две последовательности матриц вращения называются *эквивалентными*, если одну последовательность можно получить из другой с помощью перестановок соседних матриц, не имеющих общих индексов.

22.21. Признак эквивалентности последовательностей матриц вращения есть отношение эквивалентности.

22.22. Результат выполнения эквивалентных преобразований вращения будет одним и тем же, включая всю совокупность ошибок округления.

22.23. Минимальное число несвязанных групп, на которые распадается последовательность матриц вращения при эквивалентном преобразовании, называется *индексом* эквивалентности.

22.24. Рассмотрим две последовательности матриц вращения, индексы которых описываются одной и той же совокупностью пар:

$$\begin{array}{l}
1,\ 2;\\
1,\ 3;\ 2,\ 3;\\
1,\ 4;\ 2,\ 4;\ 3,\ 4;\\
1,\ 5;\ 2,\ 5;\ 3,\ 5;\ 4,\ 5;\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
1,\ m+1;\ 2,\ m+1;\ 3,\ m+1;\ \ldots;\ m,\ m+1;\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
1,\ n-1;\ 2,\ n-1;\ 3,\ n-1;\ \ldots;\ m,\ n-1;\\
1,\ n;\ 2,\ n;\ 3,\ n;\ \ldots;\ m,\ n,
\end{array}$$

где $m<n$. Для первой последовательности совокупность пар индексов упорядочивается по строкам, для второй — по столбцам, причем сами строки и столбцы упорядочиваются сверху вниз и слева направо. Обе последовательности называются *циклическими*.

22.25. Циклические последовательности состоят из $0{,}5\cdot m(2n-m-1)$ матриц вращения.

22.26. Каждая из циклических последовательностей эквивалентна последовательности, индексы которой задаются строками таблицы

$$\begin{array}{llllll}
1,\ 2; & & & & & \\
1,\ 3; & & & & & \\
1,\ 4; & 2,\ 3; & & & & \\
1,\ 5; & 2,\ 4; & & & & \\
1,\ 6; & 2,\ 5; & 3,\ 4; & & & \\
\ldots & \ldots & \ldots & \ldots & \ldots & \\
1,\ n; & 2,\ n-1; & 3,\ n-2; & \ldots; & m,\ n-m+1; \\
 & 2,\ n; & 3,\ n-1; & \ldots; & m,\ n-m+2; \\
 & & 3,\ n; & \ldots; & m,\ n-m+3; \\
 & & & & \ldots \\
 & & & & m,\ n.
\end{array}$$

22.27. Циклические последовательности эквивалентны между собой, и их индекс эквивалентности равен $m+n-2$.

22.28. В условиях и обозначениях 22.11 при обеих циклических последовательностях для эквивалентного возмущения $\mathscr{E}$ справедлива оценка $\|\mathscr{E}\|_E \lesssim \sqrt{2}\,(m+n-2)\,p^{-t+1}\|z\|_E$.

22.29. Пусть $\tilde{R}$ есть точное произведение реально вычисленных матриц вращения, соответствующих любой из циклических последовательностей. Существует такая ортогональная матрица R, что

$$\|R-\tilde{R}\|_E \lesssim \frac{5}{4}\,(m+n-2)\,\sqrt{n}\,p^{-t+1}.$$

22.30. В условиях и обозначениях 22.11, 22.28 при $m=n-1$ имеет место оценка $\|\mathscr{E}\|_E \lesssim \sqrt{2}\,(2n-3)\,p^{-t+1}\|z\|_E$.

Полученные оценки говорят о том, что общий эффект влияния ошибок округления зависит не только от числа выполненных преобразований вращения, но и от того, в какой последовательности осуществляются преобразования. В некоторых задачах мы сможем в известной мере выбирать эту последовательность и, следовательно, строить лучшие по точности методы.

22.31. Для любого n-мерного вектора z существуют такие последовательности матриц вращения $T_{i_1j_1}, \ldots, T_{i_Nj_N}$, $N \leqslant n-1$, что их произведение U переводит вектор z в вектор, коллинеарный вектору $e=(1, 0, \ldots, 0)'$, т. е. $Uz=\alpha e$ для некоторого числа α.

22.32. Пусть $j_1, j_2, \ldots, j_{n-1}$ — любая перестановка из чисел $2, 3, \ldots, n$. Возьмем любые целые положительные числа $i_1, i_2, \ldots, i_{n-1}$, не превосходящие n и такие, что для всех $1 \leqslant k \leqslant n-1$ число j_k не входит в совокупность чисел $i_1, i_2, \ldots, i_k$. Рассмотрим последовательность векторов $z_k = T_{i_kj_k}z_{k-1}$, $z_0 = z$, и подберем согласно 22.6, 22.7 параметры матриц $T_{i_kj_k}$ так, чтобы при всех k в векторе z_k обращалась в нуль j_k-я координата. Тогда матрицы $T_{i_kj_k}$ удовлетворяют условию 22.31.

22.33. Если $j_k = k+1$, $i_k = 1$, то последовательность будет сильно связанной и для эквивалентного возмущения $\mathscr{E}$ из 22.11 справедлива оценка

$$\|\mathscr{E}\|_E \lesssim \sqrt{2}\,(n-1)\,p^{-t+1}\,\|z\|_E.$$

22.34. Построим последовательность матриц вращения из 22.32 с наименьшим индексом эквивалентности. Для этого возьмем индексы j_k, равные 2, 4, 6, ..., затем 3, 7, 11, ..., далее 5, 13, 21, ... и т. д.; индексы i_k — соответственно равные 1, 3, 5, ..., затем 1, 5, 9, ..., далее 1, 9, 17, ... и т. д. Число несвязанных групп в этой последовательности будет не более $\log_2(2n)$, поэтому

$$\|\mathscr{E}\|_E \lesssim \sqrt{2}\,\log_2(2n)\,p^{-t+1}\,\|z\|_E.$$

22.35. Предположим, что при каждом умножении на матрицу вращения исключается наименьшая по модулю координата вектора. Пусть вторая преобразуемая координата является наименьшей по модулю из оставшихся ненулевых координат. В этом случае

$$\|\mathscr{E}\|_E \lesssim 2\sqrt{2(n-1)}\,p^{-t+1}\,\|z\|_E.$$

22.36. Пусть при каждом умножении на матрицу вращения исключается наибольшая по модулю координата. Тогда снова имеет место оценка 22.33.

22.37. Пусть w — произвольный вещественный вектор единичной длины, т. е. $\|w\|_E = 1$. Вещественная матрица $U = E - 2ww'$ называется *матрицей отражения* (*матрицей Хаусхолдера*).

Название этой матрицы объясняется следующим обстоятельством. Если в трехмерном пространстве рассмотреть зеркальную плоскость с нормальным вектором w, проходящую через начало координат, то преобразование зеркального отражения от такой плоскости задается именно матрицей отражения. Аналогичное свойство сохраняется и в n-мерном пространстве.

Некоторые свойства матрицы отражения и преобразования с такой матрицей удобно описывать через скалярное произведение, заданное как сумма попарных произведений координат. Всюду, где встречается скалярное произведение при использовании матриц отражения, имеется в виду именно это скалярное произведение.

22.38. Матрица отражения является симметричной и ортогональной.

22.39. Одно собственное значение матрицы отражения равно -1, остальные равны $+1$.

22.40. Для любых матриц отражения U и вектора z выполняются соотношения

$$Uz = (E - 2ww')z = z - 2(z, w)w.$$

22.41. Вектор z, коллинеарный вектору w, переводится матрицей отражения в вектор $-z$.

22.42. Вектор z, ортогональный вектору w, оставляется матрицей отражения без изменения.

22.43. Пусть заданы любые ненулевые векторы q, s. Существует единственный, с точностью до множителя ± 1, вектор w единичной длины такой, что определяемая им матрица отражения переводит вектор q в вектор αs для некоторого числа α.

22.44. В качестве вектора w из 22.43 можно взять любой из векторов, определяемых соотношениями

$$w = \frac{1}{\rho}(q - \alpha s), \quad \rho^2 = 2(q, q - \alpha s), \quad \alpha = \pm \|q\|_E / \|s\|_E.$$

22.45. Бо́льшая точность и бо́льшее значение ρ в 22.44 обеспечиваются в том случае, когда знак α выбирается противоположным знаку скалярного произведения (q, s).

22.46. Пусть $s = (1, 0, \ldots, 0)'$. Для повышения устойчивости реальные вычисления, соответствующие 22.44, выполняются по следующей схеме. Предположим, что $\|q\|_E \neq 0$. Обозначим через $u_1, \ldots, u_n$ координаты вектора $\|q\|_E^{-1} q$ и определим координаты $v_1, \ldots, v_n$ вектора v:

$$v_1 = \frac{u_1}{|u_1|}(1 + |u_1|), \quad v_i = u_i, \quad i \geqslant 2.$$

Если $u_1 = 0$, то будем считать, что $u_1/|u_1|$ есть любое из чисел ± 1. Теперь матрицу отражения можно представить в виде

$$U = E - \frac{1}{\gamma} vv',$$

где $\gamma = 1 + |u_1|$. Если $\|q\|_E = 0$, то берем $v = e$, $\gamma = 0{,}5$.

22.47. Для любых матрицы отражения U вида 22.46 и вектора z выполняются соотношения

$$Uz = \left(E - \frac{1}{\gamma} vv'\right) z = z - \frac{1}{\gamma}(z, v) v.$$

22.48. Пусть $\tilde{\gamma}$, $\tilde{v}$ — реально вычисленные параметры матрицы отражения из 22.46, причем $\|q\|_E$ вычисляется в режиме накопления. Тогда реально вычисленная матрица $\tilde{U}$ удовлетворяет соотношению

$$\|\tilde{U}\tilde{U}' - E\|_2 \lessapprox 4p^{-t+1}.$$

22.49. Пусть $\tilde{W}$ есть точное произведение N реально вычисленных матриц отражения. Существует такая ортогональная матрица W, что

$$\|\tilde{W} - W\|_E \lessapprox 2Np^{-t+1}.$$

Обращаем внимание на то, что в 22.48, 22.49 ничего не говорится о близости реально вычисленных матриц к матрицам, полученным при точных вычислениях. Такой близости может не быть. Однако подчеркнем еще раз, что в дальнейшем используется лишь близость этих матриц к некоторым ортогональным матрицам и малость эквивалентных возмущений при преобразованиях с реально вычисленными матрицами.

22.50. Пусть выполняется преобразование 22.47 с матрицей $\widetilde{U}$, определяемой реально вычисленными параметрами $\widetilde{\gamma}$, $\widetilde{v}$, причем скалярное произведение $(z, \widetilde{v})$ вычисляется в режиме накопления. Имеют место соотношения

$$\mathrm{fl}\,(\widetilde{U}z) \equiv \widetilde{U}\,(z + \rho), \quad \|\rho\|_E \lesssim 2{,}5 \cdot p^{-t+1} \|z\|.$$

22.51. Пусть параметры матрицы $\widetilde{U}$ вычисляются согласно 22.46 и образ αs вектора q задается реально вычисленным вектором $\widetilde{\alpha s}$. Если $\|q\|_E$ вычисляется в режиме накопления, то имеют место соотношения

$$\widetilde{\alpha s} \equiv \widetilde{U}\,(q + \hat{\rho}), \quad \|\hat{\rho}\|_E \lesssim 2{,}5 \cdot \sqrt{2}\, p^{-t+1} \|q\|_E.$$

22.52. Пусть с помощью преобразования отражения меняются только последние r координат вектора z. Такое преобразование может быть задано матрицей вида

$$U = \left[\begin{array}{c:c} E & 0 \\ \hdashline 0 & E - \dfrac{1}{\gamma}\,\hat{v}\hat{v}' \end{array}\right],$$

где клетка в нижнем правом углу имеет порядок r. Матрицу U можно представить в виде матрицы отражения $U = E - \frac{1}{\gamma} vv'$, где последние r координат вектора v совпадают с координатами вектора $\hat{v}$, а остальные его координаты нулевые.

Мы рассмотрели вещественные матрицы вращения и отражения. Без особого изменения они переносятся на комплексный случай, при этом в приведенных оценках ошибок округления в 2—3 раза увеличиваются числовые множители. Свойства устойчивости вычислений с этими матрицами объясняются довольно просто. Евклидова норма и 2-норма инвариантны k унитарным преобразованиям, поэтому не может происходить существенного увеличения в целом элементов преобразуемых векторов и матриц. Это очень важно, так как на каждом шаге ошибки округления в основном пропорциональны величинам элементов.

Неунитарные преобразования не обладают естественной устойчивостью, однако иногда вычисления можно организовать так, что в некоторой ограниченной форме устойчивость все же будет иметь место. Вопросами глобальной устойчивости этих преобразований мы будем заниматься только при изучении конкретных методов. Сейчас же ограничимся описанием некоторых свойств двух основных типов элементарных неунитарных матриц. Снова будем изучать лишь вещественные матрицы.

22.53. *Элементарными неунитарными матрицами* называются матрицы вида $V = E + ab'$, где a, b — векторы размерности n.

22.54. Вещественная элементарная матрица $E + ab'$ является унитарной тогда и только тогда, когда либо один из векторов

a, b является нулевым, либо эти векторы коллинеарны и $ab' = -2ww'$ для некоторого вектора w единичной длины.

22.55. Для того чтобы элементарная неунитарная матрица была невырожденной, необходимо и достаточно, чтобы $(a, b) \neq -1$, при этом

$$(E + ab')^{-1} = E - \frac{1}{1+(a, b)}\, ab'.$$

22.56. Для любого вектора s имеем $Vs = s + (s, b)a$.

Среди матриц 22.53 наиболее часто используются те, в которых либо вектор a, либо вектор b является координатным. Второй из векторов определяется условиями задачи.

22.57. Матрицы 22.53, в которых в качестве вектора b взят координатный вектор e_r, а первые r координат вектора a нулевые, называются *матрицами типа* N_r.

22.58. Матрицы типа N_r отличаются от единичных лишь поддиагональными элементами в r-м столбце, т. е.

$$N_r = \begin{bmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & 1 & & \\ & & n_{r+1,r} & \ddots & \\ & & \vdots & & \\ & & n_{n,r} & & 1 \end{bmatrix}.$$

22.59. Для матриц типа N_r матрица N_r^{-1} имеет вид

$$N_r^{-1} = \begin{bmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & 1 & & \\ & & -n_{r+1,r} & \ddots & \\ & & \vdots & & \\ & & -n_{n,r} & & 1 \end{bmatrix}.$$

22.60. Произведение матриц $N_1N_2\ldots N_r$ имеет вид

$$N_1N_2\ldots N_r = \begin{bmatrix} 1 & & & & & & 0 \\ n_{21} & 1 & & & & & \\ n_{31} & n_{32} & 1 & & & & \\ \vdots & \vdots & & \ddots & & & \\ & & & & 1 & & \\ & & & & \vdots & \ddots & \\ n_{n1} & n_{n2} & n_{n3} & \ldots & n_{nr} & \ldots & 1 \end{bmatrix},$$

т. е. ненулевые поддиагональные элементы расположены лишь в первых r столбцах и они совпадают с поддиагональными элементами матриц $N_1, N_2, \ldots, N_r$.

22.61. Для матриц типа N_r при $k<n$ выполняются соотношения

$$N_1N_2 \ldots N_k = N_1 + N_2 + \ldots + N_k - (k-1)E,$$
$$N_1^{-1}N_2^{-1} \ldots N_k^{-1} = (k+1)E - N_1 - N_2 - \ldots - N_k.$$

22.62. Пусть задан вектор z размерности n и находится последовательность векторов

$$z_k = N_k z_{k-1}, \quad z_0 = z, \qquad k = 1, 2, \ldots, r < n.$$

Предположим, что в реальных вычислениях получаются векторы

$$\tilde{z}_k = \mathrm{fl}\,(\tilde{N}_k \tilde{z}_{k-1}) \equiv \tilde{N}_k \tilde{z}_{k-1} + \mu_{k-1},$$

где $\tilde{N}_k$ означает реально заданную или реально вычисленную матрицу, μ_{k-1} — вектор ошибок, возникающий из-за неточного вычисления произведения $\tilde{N}_k \tilde{z}_{k-1}$. Не ограничивая общности, можно считать, что первые k координат векторов $\tilde{z}_{k-1}$, $\tilde{z}_k$ совпадают и, следовательно, первые k координат вектора ошибок μ_{k-1} являются нулевыми. Тогда

$$\tilde{z}_r = \tilde{N}_r \tilde{N}_{r-1} \ldots \tilde{N}_1 (z + \mu), \quad \mu = \sum_{k=1}^{r} \mu_{k-1}.$$

Эквивалентное возмущение процесса 22.62 не зависит явно от матриц N_k. Поэтому опасность неустойчивости может возникнуть лишь в том случае, если велики сами векторы ошибок μ_{k-1}. Мы уже отмечали, что ошибки, сделанные на отдельных шагах, в целом пропорциональны величинам координат векторов. Поэтому и важно, чтобы эти координаты были по возможности меньше. Такую задачу часто выполняет подходящий выбор матриц перестановок.

22.63. Пусть в процессе 22.62 вместо матриц $\tilde{N}_k$ стоят произведения $\tilde{N}\, P_{kk'}$, где $P_{kk'}$, $k' \geqslant k$, есть матрица перестановок столбцов с номерами k и k'. Тогда

$$\mu = \sum_{k=1}^{r} P_{11'} \ldots P_{kk'} \mu_{k-1}.$$

22.64. Матрицы 22.53, в которых в качестве вектора a взят координатный вектор e_r, а первые r координат вектора b нулевые, называются *матрицами типа* M_r.

22.65. Матрицы типа M_r отличаются от единичных лишь элементами в r-й строке, находящимися левее диагонали, т. е.

$$M_r = \begin{bmatrix} 1 & & & & & 0 \\ 0 & & & & & \\ & & & \ddots & & \\ m_{r1} & \cdots & m_{r,r-1} & 1 & & \\ & & & & & \\ 0 & & & & & 1 \end{bmatrix}.$$

22.66. Для матриц типа M_r матрица M_r^{-1} имеет вид

$$M_r^{-1} = \begin{bmatrix} 1 & & & & & 0 \\ 0 & & & & & \\ -m_{r1} & \cdots & -m_{r,r-1} & 1 & & \\ & & & & & \\ 0 & & & & & 1 \end{bmatrix}.$$

22.67. Произведение матриц $M_2M_3 \ldots M_r$ имеет вид

$$M_2M_3 \ldots M_r = \begin{bmatrix} 1 & & & & & 0 \\ m_{21} & 1 & & & & \\ & & & & & \\ m_{r1} & m_{r2} & \cdots & m_{r,r-1} & 1 & \\ & & & & & \\ & 0 & & & & 1 \end{bmatrix},$$

т. е. ненулевые внедиагональные элементы расположены лишь в первых r строках и они совпадают с внедиагональными элементами матриц $M_2, M_3, \ldots, M_r$.

22.68. Пусть задан вектор z размерности n, и пусть находится последовательность векторов

$$z_k = M_k z_{k-1}, \quad z_1 = z, \quad k = 2, 3, \ldots, r \leqslant n.$$

Предположим, что в реальных вычислениях получаются векторы

$$\tilde{z}_k = \mathrm{fl}\,(\tilde{M}_k \tilde{z}_{k-1}) \equiv \tilde{M}_k \tilde{z}_{k-1} + \nu_{k-1},$$

где $\tilde{M}_k$ означает реально заданную или реально вычисленную матрицу, ν_{k-1} — вектор ошибок, возникающих из-за неточного вычисления произведения $\tilde{M}_k \tilde{z}_{k-1}$. Не ограничивая общности, можно считать, что векторы $\tilde{z}_{k-1}$ и $\tilde{z}_k$ различаются только k-ми координатами и, следовательно, только k-я координата вектора ошибок ν_{k-1} является ненулевой. Тогда

$$\tilde{z}_r = \widetilde{M}_r \widetilde{M}_{r-1} \ldots \widetilde{M}_2 (z + \nu), \quad \nu = \sum_{k=2}^{r} \nu_{k-1}.$$

Несмотря на то, что процессы 22.62 и 22.68 внешне очень похожи, процесс 22.68 представляет большой интерес с точки зрения устойчивости, так как строение векторов ν_{k-1} значительно проще, чем μ_{k-1}.

Это дает основание надеяться, что в вычислительных алгоритмах, использующих преобразования с матрицами типа M_r, можно достичь высокой точности.

Заметим, что простой вид суммарного эквивалентного возмущения в 22.62, 22.68 связан только с указанным выше порядком выполнения элементарных преобразований. Если хотя бы одно преобразование с матрицей, имеющей больший номер, выполняется раньше преобразования, имеющего меньший номер, то суммарное эквивалентное возмущение будет уже зависеть от матриц преобразования.

Элементарные неунитарные матрицы можно использовать в задачах преобразования векторов так же, как и матрицы вращения и отражения.

22.69. Пусть задан вектор s с координатами $s_1, \ldots, s_n$ и $s_r \neq 0$ для некоторого r. Существует единственная матрица N_r такая, что у вектора $N_r s$ последние $r-1$ координат нулевые, а остальные координаты совпадают с соответствующими координатами вектора s. При этом поддиагональные элементы $n_{i,r}$ матрицы N_r равны $-s_i/s_r$ для всех $i > r$.

Матрицы типа M_r также можно использовать для исключения каких-либо элементов векторов, но значительно чаще они применяются для других целей.

§ 23. Ортогонализация

Одним из важнейших понятий, связанных с линейным пространством, является понятие ортогональности. Мы уже неоднократно убеждались в том, какую важную роль играют ортогональные и псевдоортогональные системы векторов и особенно базисные системы такого типа. До сих пор большинство наших рассуждений было связано с доказательством существования подобных систем, но не с процессами их построения. Ввиду важности ортогональных, псевдоортогональных и других аналогичных систем для конструирования самых различных вычислительных алгоритмов рассмотрим сейчас общий процесс построения этих систем, называемый процессом *ортогонализации*.

23.1. Пусть в линейном пространстве задано скалярное произведение (x, y) с помощью невырожденной эрмитовой билинейной формы.

Пусть для базиса $e_1, \ldots, e_n$ существует базис $f_1, \ldots, f_n$, обладающий следующими свойствами:

— для любого $k \geq 1$ линейные оболочки L_k векторов $e_1, \ldots, e_k$ и $f_1, \ldots, f_k$ совпадают;

— базис $f_1, \ldots, f_n$ псевдоортогональный.

Тогда этот базис — единственный, с точностью до нормировки векторов.

23.2. Предположим, что $(e_1, e_1) \neq 0$, и положим $f_1 = e_1$. Пусть уже построена система псевдоортогональных векторов $f_1, \ldots, f_k$, причем линейные оболочки этих векторов и векторов $e_1, \ldots, e_k$ совпадают. Будем искать вектор f_{k+1} в виде

$$f_{k+1} = e_{k+1} + \sum_{i=1}^{k} \alpha_{i,k+1} f_i,$$

где $\alpha_{1,k+1}, \ldots, \alpha_{k,k+1}$ — неизвестные коэффициенты. Условия ортогональности вектора f_{k+1} слева к векторам $f_1, \ldots, f_k$ дают для определения $\alpha_{1,k+1}, \ldots, \alpha_{k,k+1}$ следующую систему уравнений:

$$\begin{aligned}
&\alpha_{1,k+1}(f_1, f_1) && = -(e_{k+1}, f_1),\\
&\alpha_{1,k+1}(f_1, f_2) + \alpha_{2,k+1}(f_2, f_2) && = -(e_{k+1}, f_2),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\alpha_{1,k+1}(f_1, f_k) + \alpha_{2,k+1}(f_2, f_k) + \ldots + \alpha_{k,k+1}(f_k, f_k) && = -(e_{k+1}, f_k).
\end{aligned}$$

Матрица этой системы — треугольная. Если $(f_i, f_i) \neq 0$ для $1 \leqslant i \leqslant n$, то после n шагов будет построена система векторов $f_1, \ldots, f_n$, удовлетворяющая обоим свойствам 23.1.

23.3. Процесс ортогонализации 23.2 осуществим на всех шагах тогда и только тогда, когда матрица билинейной формы (x, y) в базисе $e_1, \ldots, e_n$ или, что то же самое, матрица Грама этой системы имеет ненулевые ведущие миноры.

23.4. Если процесс ортогонализации 23.2 применить к линейно зависимой системе $e_1, \ldots, e_n$, то $f_k = 0$ для некоторого $k \leqslant n$.

23.5. Процесс ортогонализации 23.2 позволяет получить разложения векторов e_{k+1} по векторам $f_1, \ldots, f_{k+1}$. Если согласно 23.2

$$e_{k+1} = \left(-\sum_{i=1}^{k} \alpha_{i,k+1} f_i \right) + f_{k+1},$$

то вектор в скобках принадлежит линейной оболочке L_k векторов $e_1, \ldots, e_k$, вектор f_{k+1} принадлежит ${}^{\perp}L_k$. Поэтому решение систем 23.2 дает разложение каждого вектора e_{k+1} на проекцию и левый перпендикуляр по отношению к подпространству L_k.

23.6. Если скалярное произведение (x, y) задано симметричной формой, то система 23.2 становится системой с диагональной матрицей и

$$\alpha_{i,k+1} = -(e_{k+1}, f_i)/(f_i, f_i)$$

для всех i. Построенный базис $f_1, \ldots, f_n$ будет не только псевдоортогональным, но и ортогональным.

23.7. В евклидовом и унитарном пространствах процесс ортогонализации 23.2, 23.6 называется *ортогонализацией Грама — Шмидта*.

23.8. При реализации процесса ортогонализации Грама — Шмидта для всех k выполняется неравенство $\|f_k\|_E \leqslant \|e_k\|_E$, причем равенство достигается тогда и только тогда, когда вектор e_k ортогонален векторам $e_1, \ldots, e_{k-1}$.

Во многих задачах нет необходимости сохранять связь нового базиса $f_1, \ldots, f_n$ с исходным базисом $e_1, \ldots, e_n$, так как требуется построить лишь какой-нибудь псевдоортогональный базис в пространстве. В этом случае при каждом появлении равенства $(f_i, f_i) = 0$ нужно заменить вектор e_i другим и снова вычислить вектор f_i, повторяя эту процедуру до тех пор, пока не выполнится условие $(f_i, f_i) \neq 0$. При этом векторы $f_1, \ldots, f_{i-1}$ не изменяются. Необходимый для замены вектор e_i обязательно найдется.

Процессы, аналогичные описанным, позволяют строить не только ортогональные и псевдоортогональные базисы, но и базисы, псевдодвойственные для заданного. Независимо от их конкретного содержания все они называются процессами ортогонализации.

23.9. Пусть $e_1, \ldots, e_n$ — заданный базис, и пусть нужно построить какой-нибудь псевдодвойственный для него базис, например левый. Возьмем еще один базис $q_1, \ldots, q_n$. Предположим, что $(q_1, e_1) \neq 0$, и положим $t_1 = q_1$. Допустим, что уже построена система векторов $t_1, \ldots, t_k$ так, что их линейная оболочка совпадает с линейной оболочкой векторов $q_1, \ldots, q_k$ и выполняются условия $(t_i, e_i) \neq 0$ для $1 \leqslant i \leqslant k$ и $(t_i, e_j) = 0$ для $j < i$. Будем искать вектор t_{k+1} в виде

$$t_{k+1} = q_{k+1} + \sum_{i=1}^{k} \beta_{i,k+1} t_i,$$

где $\beta_{1,k+1}, \ldots, \beta_{k,k+1}$ — неизвестные коэффициенты. Условия ортогональности вектора t_{k+1} слева к векторам $e_1, \ldots, e_k$ дают для определения $\beta_{1,k+1}, \ldots, \beta_{k,k+1}$ следующую систему уравнений с треугольной матрицей:

$$\begin{aligned}
&\beta_{1,k+1}(t_1, e_1) && = -(q_{k+1}, e_1),\\
&\beta_{1,k+1}(t_1, e_2) + \beta_{2,k+1}(t_2, e_2) && = -(q_{k+1}, e_2),\\
&\ldots\ldots\ldots\ldots\ldots\ldots && \ldots\ldots\\
&\beta_{1,k+1}(t_1, e_k) + \beta_{2,k+1}(t_2, e_k) + \ldots + \beta_{k,k+1}(t_k, e_k) && = -(q_{k+1}, e_k).
\end{aligned}$$

Если при всех k будет выполнено неравенство $(t_k, e_k) \neq 0$, то базис $t_1, \ldots, t_n$ с точностью до нормировки будет левым псевдодвойственным для базиса $e_1, \ldots, e_n$.

23.10. Для того чтобы процесс 23.9 был осуществим при всех k, необходимо и достаточно, чтобы матрица билинейной формы (x, y) в базисах $q_1, \ldots, q_n$ и $e_1, \ldots, e_n$ имела ненулевые ведущие миноры.

23.11. В процессе 23.9 система векторов $t_1, \ldots, t_n$ не изменится, если к каждому вектору e_s (q_s) прибавить линейную комбинацию векторов $e_1, \ldots, e_{s-1}$ $(q_1, \ldots, q_{s-1})$.

23.12. Предположим, что после выполнения процесса 23.9 аналогичный процесс проводится над векторами $e_1, e_2, \ldots, e_n$ с целью получить из них базис $p_1, \ldots, p_n$, правый псевдодвойственный для базиса $t_1, \ldots, t_n$. С точностью до нормировки базис $p_1, \ldots, p_n$ $(t_1, \ldots, t_n)$ будет правым (левым) двойственным для базиса $t_1, \ldots, t_n$ $(p_1, \ldots, p_n)$.

Процессы 23.9, 23.12 можно выполнять не только последовательно, но и параллельно, строя новые последовательности векторов p_i, t_i одновременно. Такое объединение процессов приводит к эффективному методу биортогонализации построения двойственных базисов.

23.13. Пусть заданы базисы $e_1, \ldots, e_n$ и $q_1, \ldots, q_n$. Предположим, что из этих базисов строится пара двойственных базисов $p_1, \ldots, p_n$ и $t_1, \ldots, t_n$. Будем искать векторы p_{k+1}, t_{k+1} в следующем виде:

$$p_{k+1} = e_{k+1} + \sum_{i=1}^{k} \gamma_{i,k+1} p_i, \quad t_{k+1} = q_{k+1} + \sum_{i=1}^{k} \delta_{i,k+1} t_i.$$

Условия $(t_i, p_j) = 0$ для $i \neq j$ позволяют определить неизвестные коэффициенты:

$$\gamma_{i,k+1} = -(t_i, e_{k+1})/(t_i, p_i), \quad \delta_{i,k+1} = -(q_{k+1}, p_i)/(t_i, p_i).$$

С точностью до нормировки базис $p_1, \ldots, p_n$ $(t_1, \ldots, t_n)$ будет правым (левым) двойственным базисом для базиса $t_1, \ldots, t_n$ $(p_1, \ldots, p_n)$.

23.14. Во всех описанных процессах ортогонализации матрица преобразования координат при переходе от старого базиса к новому является треугольной.

23.15. Если в евклидовом или унитарном пространствах эрмитова билинейная форма задается с помощью скалярного произведения вида (Ax, y), то с помощью описанных процессов можно строить A-ортогональные, A-псевдоортогональные, A-двойственные и A-псевдодвойственные системы векторов.

Согласно формулам, определяющим изменение векторов в процессе ортогонализации, в общем случае нельзя надеяться на более простой вид матрицы преобразования координат, чем треугольный. Однако если исходный базис выбрать согласованным с билинейной формой, порождающей скалярное произведение, то можно получить более простые представления для этой матрицы.

23.16. Пусть A — квадратная матрица, x — вектор. Последовательность векторов $x, Ax, \ldots, A^k x, \ldots$ называется *степенной последовательностью*, порожденной вектором x.

23.17. Приведенный многочлен $\varphi(\lambda)$ минимальной степени, для которого выполняется равенство $\varphi(A)x = 0$, называется *минимальным аннулирующим* вектор x многочленом.

23.18. Минимальный аннулирующий вектор x многочлен единствен.

23.19. Минимальный аннулирующий вектор x многочлен является делителем характеристического многочлена.

23.20. Степень минимального аннулирующего вектор x многочлена равна максимальному числу первых векторов степенной последовательности, образующих линейно независимую систему.

23.21. Степень минимального аннулирующего вектор x многочлена равна сумме максимальных высот корневых векторов матрицы A, присутствующих в разложении вектора x по корневому базису и соответствующих попарно различным собственным значениям.

23.22. Пусть в унитарном пространстве скалярное произведение введено согласно 5.4 через естественный базис и задана невырожденная билинейная форма (Cx, y). Предположим, что матрица A удовлетворяет условию

$$(CAC^{-1})^* = \alpha E + \beta A$$

для некоторых чисел α, β. Пусть векторы $e_i = A^{i-1}x$ линейно независимы для $1 \leqslant i \leqslant k$ и векторы $f_1, \ldots, f_k$ получены из векторов $e_1, \ldots, e_k$ с помощью процесса псевдоортогонализации по отношению к форме (Cx, y). Тогда имеют место следующие соотношения:

$$\begin{aligned} f_1 &= x, \\ f_2 &= Af_1 - \alpha_1 f_1, \\ &\ldots\ldots\ldots\ldots \\ f_{i+1} &= Af_i - \alpha_i f_i - \beta_{i-1} f_{i-1}, \qquad i > 1, \end{aligned}$$

где

$$\alpha_1 = \frac{(CAf_1, f_1)}{(Cf_1, f_1)}, \quad \beta_{i-1} = \frac{(CAf_i, f_{i-1})}{(Cf_{i-1}, f_{i-1})}, \quad i > 1,$$

$$\alpha_i = \frac{(Cf_{i-1}, f_{i-1})(CAf_i, f_i) - (CAf_i, f_{i-1})(Cf_{i-1}, f_i)}{(Cf_{i-1}, f_{i-1})(Cf_i, f_i)}, \qquad i > 1.$$

23.23. Если $A = A^*$ и $C = E$, то для коэффициентов из 23.22 имеем

$$\alpha_i = \frac{(Af_i, f_i)}{(f_i, f_i)}, \quad \beta_{i-1} = \frac{(Af_i, f_{i-1})}{(f_{i-1}, f_{i-1})} = \frac{(f_i, Af_{i-1})}{(f_{i-1}, f_{i-1})} = \frac{(f_i, f_i)}{(f_{i-1}, f_{i-1})} > 0.$$

23.24. В условиях и обозначениях 23.22 в базисе $f_1, \ldots, f_n$ матрица A имеет трехдиагональный вид:

$$A_f = \begin{bmatrix} \alpha_1 & \beta_1 & & & & 0 \\ 1 & \alpha_2 & \beta_2 & & & \\ & 1 & \alpha_3 & \beta_3 & & \\ & & \cdot & \cdot & \cdot & \\ & & & 1 & \alpha_{n-1} & \beta_{n-1} \\ 0 & & & & 1 & \alpha_n \end{bmatrix}.$$

23.25. В условиях 23.23 существует такая диагональная матрица D, что для матрицы A_f из 23.24 матрица $D^{-1}A_fD$ будет вещественной симметричной трехдиагональной.

23.26. Пусть A — произвольная матрица и заданы два вектора x, y в унитарном пространстве. Предположим, что системы векторов $e_i = A^{i-1}x$, $q_i = (A^*)^{i-1}y$ линейно независимы и двойственные системы векторов $p_1, \ldots, p_k$, $t_1, \ldots, t_k$ получены из них с помощью процесса биортогонализации 23.13. Тогда имеют место соотношения

$$
\begin{array}{ll}
p_1 = x, & t_1 = y, \\
p_2 = Ap_1 - \alpha_1 p_1, & t_2 = A^* t_1 - \bar{\alpha}_1 t_1, \\
\ldots\ldots\ldots & \ldots\ldots\ldots \\
p_{i+1} = Ap_i - \alpha_i p_i - \beta_{i-1} p_{i-1}, & t_{i+1} = A^* t_i - \bar{\alpha}_i t_i - \bar{\beta}_{i-1} t_{i-1}, \quad i > 1,
\end{array}
$$

где

$$\alpha_i = \frac{(Ap_i, t_i)}{(p_i, t_i)}, \quad \beta_{i-1} = \frac{(Ap_i, t_{i-1})}{(p_{i-1}, t_{i-1})} = \frac{(Ap_{i-1}, t_i)}{(p_{i-1}, t_{i-1})} = \frac{(p_i, t_i)}{(p_{i-1}, t_{i-1})}.$$

23.27. В условиях и обозначениях 23.26 в базисе $p_1, \ldots, p_n$ ($t_1, \ldots, t_n$) матрица A (A^*) имеет трехдиагональный вид.

Заметим, что в общем случае в процессах ортогонализации для построения каждого последующего вектора необходимо привлекать все ранее построенные векторы. В процессах 23.22, 23.26 ситуация существенно проще, так как нужно привлекать лишь два последних вектора из строящихся систем. Однако при этом приходится на каждом шаге процесса выполнять одну или две операции умножения матрицы на вектор. К процессам этого типа мы будем обращаться неоднократно.

Процессы ортогонализации очень чувствительны к ошибкам округления. Мы исследуем подробно их влияние в процессе 23.2 для случая 23.6.

23.28. Пусть $\tilde{f}_i$, $\tilde{\alpha}_{i,k+1}$ — реально вычисленные величины процесса ортогонализации 23.2, 23.6. Если для $k+1$-го шага выполняется неравенство

$$\|e_{k+1}\|_E \leqslant \left\| \sum_{i=1}^{k} \tilde{\alpha}_{i,k+1} \tilde{f}_i \right\|_E,$$

то вектор $\hat{f}_{k+1} = e_{k+1} + \sum\limits_{i=1}^{k} \tilde{\alpha}_{i,k+1} \tilde{f}_i$ имеет относительную ошибку больше единицы.

23.29. Пусть на всех шагах процесса выполняются неравенства

$$\left\| \sum_{i=1}^{k} \tilde{\alpha}_{i,k+1} \tilde{f}_i \right\|_E \leqslant \|e_{k+1}\|_E$$

и правые части равенств, определяющих векторы $\tilde{f}_{k+1}$, вычисляются в режиме накопления. Тогда, независимо от точности вычисления коэффициентов $\alpha_{i,k+1}$, линейные оболочки реально вычисленных векторов $\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_k$ для всех k совпадают с линейными оболочками векторов e_1, $e_2+\varepsilon_2$, ..., $e_k+\varepsilon_k$, где $\|\varepsilon_k\|_E \leqslant \leqslant p^{-t+1}\|e_k\|_E$.

В любом процессе ортогонализации для построенных систем векторов должны иметь место как совпадение линейных оболочек соответствующих подсистем, так и ортогональность векторов. Утверждение 23.29 означает, что независимо от того, насколько реально вычисленные векторы ортогональны, линейные оболочки этих векторов при небольших ограничениях на вычислительный процесс совпадают с линейными оболочками слабо возмущенных исходных векторов. Эти возмущения настолько малы, что их оценки только в два раза превосходят оценки ошибок округления векторов при их вводе в память ЭВМ. Поэтому в отношении выполнения первого условия 23.1 процессы ортогонализации ведут себя исключительно устойчиво. Заметим, что в данном случае оценки эквивалентных возмущений ε_k не зависят ни от размерности векторов, ни от числа выполняемых над ними арифметических операций.

23.30. Пусть выполнено k шагов процесса ортогонализации 23.2, 23.6 и реально вычисленные векторы $\tilde f_1$, $\tilde f_2$, ..., $\tilde f_k$ незначительно отличаются от точно вычисленных векторов f_1, f_2, ..., f_k. Предположим, что

$$\|\tilde f_i\|_E = (1+\tau_{ii})\|f_i\|_E, \quad (\tilde f_i, \tilde f_j) = \tau_{ij}\|f_i\|_E\|f_j\|_E$$

для $i, j \leqslant k$ и все числа τ_{ii}, τ_{ij} являются малыми порядка τ. Пусть, далее, точно вычисляются коэффициенты $\alpha_{i,k+1}$ согласно 23.6 по реально вычисленным векторам $\tilde f_i$ и по этим коэффициентам и векторам $\tilde f_i$ точно вычисляется вектор $\tilde f_{k+1}$. Если аналогично τ_{ij} определить $\tau_{k+1,j}$, $j \leqslant k$, то в этом случае

$$\tau_{k+1,j} = -\sum_{i=1}^{k} \frac{(e_{k+1}, f_i)}{\|f_i\|_E}\tau_{ij}\left[\left(\|e_{k+1}\|_E^2 - \sum_{i=1}^{k}\frac{(e_{k+1}, f_i)^2}{\|f_i\|_E^2}\right)^{1/2}\right]^{-1} + O(\tau^2).$$

23.31. Пусть линейное пространство — вещественное. Обозначим через $\{e_{k+1}, L_k\}$ угол между вектором e_{k+1} и линейной оболочкой L_k векторов e_1, ..., e_k или, что то же самое, векторов f_1, ..., f_k. Справедливо равенство

$$\operatorname{ctg}^2\{e_{k+1}, L_k\} = \sum_{i=1}^{k}\frac{(e_{k+1}, f_i)^2}{\|f_i\|_E^2}\left(\|e_{k+1}\|_E^2 - \sum_{i=1}^{k}\frac{(e_{k+1}, f_i)^2}{\|f_i\|_E^2}\right)^{-1}$$

Формула для $\tau_{k+1,j}$ из 23.30 показывает, как неортогональность векторов $\tilde f_i$, реально вычисленных на первых k шагах процесса ортогонализации, влияет на неортогональность к ним вектора $\tilde f_{k+1}$ в том случае, когда все вычисления на $k+1$-м шаге осуществляются точно. Ясно, что в реальных условиях $\tau_{k+1,j}$ будет, как правило, не меньше. Если $\operatorname{ctg}^2\{e_{k+1}, L_k\}$ является большим, то будут большими по модулю и некоторые коэффициенты при τ_{ij} в формуле для $\tau_{k+1,j}$. Следовательно, даже при точном выполнении операций на $k+1$-м шаге процесса величины ошибок $\tau_{k+1,j}$ могут стать значительными по сравнению с ошибками τ_{ij}, полученными на предыдущих шагах.

Если мы проследим распространение ошибок на несколько последующих шагов, то положение окажется еще более серьезным. Для того чтобы нарушилась ортогональность системы векторов $\tilde f_1$, $\tilde f_2$, ..., $\tilde f_n$, совсем не обязательно, чтобы для какого-нибудь k был большим $\operatorname{ctg}^2\{e_{k+1}, L_k\}$. Достаточно, чтобы большим было произведение $\operatorname{ctg}^2\{e_{k+1}, L_k\}$ для разных k. В реальных задачах, как правило, имеет место именно эта ситуация. Поэтому в отношении второго условия 23.1 можно утверждать, что формальное выполнение процессов ортогонализации согласно описанным схемам в большинстве случаев должно приводить к значительному нарушению ортогональности строящихся систем векторов. Этот вывод полностью подтверждается на практике.

Для устранения указанной неортогональности используются различные приемы, основанные на следующей идее. Пусть на каком-то шаге вектор $\tilde{f}_{k+1}$ не ортогонален векторам $f_1, \tilde{f}_2, \ldots, \tilde{f}_k$, но все же более ортогонален, чем исходный вектор e_{k+1}. Тогда берем вектор $\tilde{f}_{k+1}$ в качестве e_{k+1} и снова проводим процесс его ортогонализации к векторам, построенным ранее. После нескольких таких итераций почти всегда можно достичь нужной степени ортогональности и сохранения линейной оболочки соответственно условиям 23.1.

23.32. Пусть $f_1, \ldots, f_k$ — линейно независимая система векторов. Рассмотрим последовательность векторов

$$f_{k+1}^{(s)} = f_{k+1}^{(s-1)} + \sum_{i=1}^{k} \alpha_{i,k+1}^{(s-1)} \tilde{f}_i, \quad f_{k+1}^{(0)} = e_{k+1},$$

где коэффициенты $\alpha_{i,k+1}^{(s-1)}$ определяются в соответствии с 23.6, исходя из $f_{k+1}^{(s-1)}$ вместо e_{k+1}. Обозначим через F_k матрицу Грама для системы, полученной после нормировки векторов $f_1, \ldots, f_k$. Если $\|E - F_k\| \leqslant q < 1$, то последовательность векторов $f_{k+1}^{(s)}$ сходится при $s \to \infty$ к некоторому вектору f_{k+1}, причём:

— скорость сходимости не меньше, чем скорость сходимости геометрической прогрессии со знаменателем q;

— вектор f_{k+1} ортогонален векторам $f_1, \ldots, f_k$;

— линейные оболочки векторов $f_1, \ldots, f_k, e_{k+1}$ и $f_1, \ldots, f_k, f_{k+1}$ совпадают.

23.33. Итерационные процессы типа 23.32 называются процессами *переортогонализации*.

На первых шагах процесса ортогонализации реально вычисленные векторы $f_1, \tilde{f}_2, \ldots, \tilde{f}_k$ близки к ортогональным, и мы находимся в условиях применимости процесса переортогонализации. Если переортогонализацию делать на каждом шаге, то для достижения максимально возможной ортогональности векторов приходится делать 1—2 итерации. При этом оценки эквивалентных возмущений из 23.29 увеличиваются не более чем в два раза.

При реализации процесса переортогонализации необходимо иметь доступ ко всем построенным ранее векторам ортогональной системы. Если переортогонализацию применять к процессам 23.22, 23,26, то эти процессы потеряют главное свое достоинство — необходимость доступа лишь к двум последним векторам.

§ 24. Основные разложения матрицы на множители

Элементарные унитарные и неунитарные преобразования широко используются для получения самых различных разложений матрицы на множители. Среди этих разложений наиболее важными являются разложения на два треугольных множителя и на треугольный и унитарный множители.

Снова для простоты будем рассматривать вещественные матрицы и преобразования. Теоретические основы этих разложений даны в § 11.

24.1. Пусть A — произвольная квадратная или прямоугольная матрица. При ее умножении слева (справа) на любую последовательность матриц типа N_r (транспонированных матриц типа N_r) все ведущие миноры не изменяются.

24.2. Пусть матрица A ранга r имеет ненулевые ведущие миноры до r-го порядка включительно. Рассмотрим последовательность матриц $A_k = N_k A_{k-1}$, $A_0 = A$, для $1 \leqslant k \leqslant r$. Предположим, что каждая из матриц N_k строится согласно 22.69 по k-му столбцу матрицы A_{k-1}. Все шаги этого процесса могут быть реализованы, и матрица A_r будет правой трапециевидной матрицей. Описанный процесс называется *методом Гаусса* разложения матрицы на множители.

Вообще говоря, название «метод Гаусса» является собирательным для большой группы алгоритмов, связанных с решением систем линейных алгебраических уравнений, вычислением определителей, разложением матрицы на множители и т. д. Процесс 24.2 — это то общее, что в той или иной мере присутствует в каждом из таких алгоритмов. Аналогичная ситуация имеет место и в отношении других методов.

24.3. Процесс 24.2 дает следующее разложение матрицы A на множители:

$$A = \left(N_1^{-1} \dots N_r^{-1}\right) A_r.$$

Здесь матрица, стоящая в скобках, есть левая треугольная матрица с единичными диагональными элементами, структура которой полностью определяется утверждениями 22.59, 22.60; матрица A_r — правая трапециевидная ранга r.

24.4. Для любой квадратной матрицы A порядка n, у которой отличны от нуля ведущие миноры всех порядков от 1 до $n-1$, процесс 24.2 позволяет получить ее LU-разложение (см. 11.5). При этом в соответствии с 24.3

$$L = N_1^{-1} \dots N_{n-1}^{-1}, \quad U = A_{n-1}.$$

24.5. Пусть $\widetilde{N}_1, \dots, \widetilde{N}_{n-1}$ и $\widetilde{A}_1, \dots, \widetilde{A}_{n-1}$ — реально вычисленные матрицы в процессе 24.2 при условиях 24.4. Предположим, что

$$\left(\widetilde{N}_1^{-1} \dots \widetilde{N}_{n-1}^{-1}\right) \widetilde{A}_{n-1} = A + M.$$

Обозначим через μ_{ij} элементы матрицы эквивалентного возмущения M, через $\widetilde{a}_{ij}^{(k)}$ — элементы матрицы $\widetilde{A}_k$. Если

$$\alpha = \max_{0 \leqslant k \leqslant n-1} \alpha_k, \quad \alpha_k = \max_{\substack{i > k \\ j \geqslant k}} \left| \widetilde{a}_{ij}^{(k)} \right|,$$

то имеют место оценки

$$|\mu_{ij}| \lessapprox \begin{cases} 0, & i=1, \\ 1{,}5\cdot(i-1)\,p^{-t+1}\alpha, & j\geqslant i, \\ 1{,}5\cdot(j-1)\,p^{-t+1}\alpha, & j<i. \end{cases}$$

24.6. В условиях и обозначениях 24.5 выполняется неравенство

$$\|M\|_E \lessapprox \frac{\sqrt{6}}{4} n^2 p^{-t+1}\alpha.$$

Оценки 24.5, 24.6 получены без каких-либо предположений относительно величины ведущих миноров матрицы A, кроме, конечно, предположения об осуществимости процесса. Они подтвердили высказанное ранее мнение о том, что существенным источником неустойчивости в процессах с неунитарными преобразованиями может быть лишь значительный рост элементов промежуточных матриц A_k.

Если не менять принципиально общую схему вычислений, то единственной возможностью в какой-то мере регулировать рост элементов является использование перестановок при реализации процесса 24.2.

24.7. Пусть A — произвольная прямоугольная матрица ранга r. Рассмотрим последовательность матриц $\widehat{A}_k = \widehat{N}_k\left(P_{ki_k}\widehat{A}_{k-1}R_{kj_k}\right)$, $A_0=A$, для $1\leqslant k\leqslant r$, где P_{ki_k}, R_{kj_k} — матрицы перестановок, причем i_k, $j_k\geqslant k$. Предположим, что при всех k матрицы перестановок таковы, что в позициях (k, k) матриц $P_{ki_k}\widehat{A}_{k-1}R_{kj_k}$ находятся ненулевые элементы и каждая из матриц $\widehat{N}_k$ строится согласно 22.69 по k-му столбцу матрицы $P_{ki_k}\widehat{A}_{k-1}R_{kj_k}$. После r шагов этого процесса матрица $\widehat{A}_r$ будет правой трапециевидной. Описанный процесс называется *методом Гаусса с перестановками* для разложения матрицы на множители или *методом Гаусса с выбором ведущего элемента*.

Заметим, что в позиции (k, k) матрицы $P_{ki_k}\widehat{A}_{k-1}R_{kj_k}$ стоит тот же элемент, который находится в позиции (i_k, j_k) матрицы $\widehat{A}_{k-1}$. Поэтому для осуществимости процесса 24.7 необходимо брать такие перестановки P_{ki_k}, R_{kj_k}, индексы (i_k, j_k) которых соответствуют позиции ненулевого элемента матрицы $\widehat{A}_{k-1}$.

24.8. Процесс 24.7 дает следующее разложение матрицы A на множители:

$$A = \left(P_{1i_1}\widehat{N}_1^{-1}\ldots P_{ri_r}\widehat{N}_r^{-1}\right)\left(\widehat{A}_r R_{rj_r}\ldots R_{1j_1}\right).$$

Матрицы, стоящие в скобках 24.8, уже не являются треугольными. Поэтому может показаться, что анализ ошибок для процесса 24.2 не переносится на процесс 24.7. Однако в действительности между обоими процессами имеется очень тесная связь.

24.9. Матрицы

$$\breve{N}_{k-1}^{-1} = P_{ri_r} \ldots P_{ki_k} \widehat{N}_{k-1}^{-1} P_{ki_k} \ldots P_{ri_r}$$

являются матрицами типа N_r и отличаются от матриц $\widehat{N}_{k-1}^{-1}$ лишь перестановкой поддиагональных элементов в $k-1$-м столбце.

24.10. Равенство 24.8 эквивалентно равенству

$$\breve{A} = (\breve{N}_1^{-1} \ldots \breve{N}_r^{-1}) \widehat{A}_r,$$

где

$$\breve{A} = (P_{ri_r} \ldots P_{1i_1}) A (R_{1j_1} \ldots R_{rj_r}).$$

Сравнивая 24.3, 24.10, можно заключить, что процесс 24.7 определяет разложение на треугольные множители матрицы $\breve{A}$, которая получается из матрицы A путем перестановок ее строк и столбцов. Так как перестановки не вносят никаких дополнительных ошибок, то оценки 24.5, 24.6 для процесса 24.2 без изменения переносятся на матрицу $\breve{A}$ и процесс 24.7. Теперь рост элементов матриц $\breve{A}_k$ полностью определяется стратегией выбора перестановок или ведущих элементов.

24.11. Элементы матриц $\widehat{A}_k$ в позициях (i_k, j_k) процесса 24.7 называются *ведущими элементами* метода Гаусса.

24.12. Существуют три наиболее распространенные *стратегии выбора* ведущих элементов:

— в качестве ведущего элемента k-го шага выбирается максимальный по модулю элемент $\widehat{a}_{ij}^{(k-1)}$ матрицы $\widehat{A}_{k-1}$ при условиях $i \geqslant k$, $j = k$; если имеется несколько максимальных по модулю элементов, то ведущим берется любой из них; эта стратегия называется выбором ведущего элемента *по столбцу*;

— в качестве ведущего элемента k-го шага выбирается максимальный по модулю элемент $\widehat{a}_{ij}^{(k-1)}$ матрицы $\widehat{A}_k$ при условиях $i = k$, $j \geqslant k$; если имеется несколько максимальных по модулю элементов, то ведущим берется любой из них; эта стратегия называется выбором ведущего элемента *по строке*;

— в качестве ведущего элемента k-го шага выбирается максимальный по модулю элемент $\widehat{a}_{ij}^{(k-1)}$ матрицы $\widehat{A}_{k-1}$ при условиях $i \geqslant k$, $j \geqslant k$; если имеется несколько максимальных по модулю элементов, то ведущим берется любой из них; эта стратегия называется выбором ведущего элемента *по всей матрице*.

24.13. Применение стратегий выбора ведущих элементов по столбцу и по всей матрице обеспечивает выполнение неравенств $|\widehat{n}_{ij}^{(k)}| \leqslant 1$ для элементов матриц $\widehat{N}_k$.

24.14. Существуют матрицы, для которых применение стратегии выбора ведущего элемента по столбцу в обозначениях 24.5 приводит к выполнению соотношений $\alpha_k = 2^k \alpha_0$ при всех k.

24.15. Какова бы ни была матрица A, применение стратегии выбора ведущего элемента по всей матрице в обозначениях 24.5

приводит к выполнению соотношений $\alpha_k \leqslant f(k)\alpha_0$ при всех k, где

$$f(k) \leqslant k^{1/2}(2^1 3^{1/2} 4^{1/3} \dots k^{1/(k-1)})^{1/2}.$$

Рост элементов, указанный в 24.14, достигается на матрицах очень специального вида. В практических вычислениях он оказывается, как правило, не очень большим. Оценка роста 24.15, по-видимому, сильно завышена, так как до сих пор не найдено ни одной матрицы, для которой $f(k) > k$. В ряде случаев можно гарантировать отсутствие роста элементов.

24.16. Если матрица имеет доминирующую диагональ и не осуществляется выбор ведущих элементов, то при получении треугольного разложения не происходит рост элементов.

Мы уже неоднократно подчеркивали, что применение режима накопления при реализации операций позволяет снизить общий уровень ошибок. В рассмотренных вариантах метода Гаусса, по существу, нет возможности для применения такого режима, так как нет больших групп явно выписанных операций. Однако это связано лишь с выбором вычислительной схемы для получения разложения матрицы на треугольные множители.

24.17. Предположим, что для матрицы A порядка n с элементами a_{ij} существует LU-разложение. Обозначим через l_{ij}, u_{ij} элементы матриц L, U. Для всех i, j имеют место равенства

$$a_{ij} = \sum_{p=1}^{\min(i,j)} l_{ip}u_{pj}.$$

24.18. Если ведущие миноры матрицы A отличны от нуля, то уравнения 24.17 рекуррентно разрешимы относительно элементов матриц L, U, при этом

$$u_{11} = a_{11},$$

$$u_{1j} = a_{1j}, \quad l_{j1} = \frac{a_{j1}}{u_{11}}, \qquad j = 2, 3, \dots, n,$$

$$u_{ii} = a_{ii} - \sum_{p=1}^{i-1} l_{ip}u_{pi}, \qquad i = 2, 3, \dots, n,$$

$$u_{ij} = a_{ij} - \sum_{p=1}^{i-1} l_{ip}u_{pj}, \quad l_{ji} = \frac{a_{ji} - \sum_{p=1}^{j-1} l_{jp}u_{pi}}{u_{ii}},$$

$$i = 2, 3, \dots, n, \quad j = i+1, i+2, \dots, n.$$

Приведенные формулы для определения треугольных сомножителей L, U называются *компактной схемой* метода Гаусса для разложения матрицы на множители.

24.19. Пусть все элементы LU-разложения в компактной схеме вычисляются в режиме накопления. Обозначим через $\tilde{L}$, $\tilde{U}$ реально вычисленные матрицы в процессе 24.18 и предположим, что $\tilde{L}\tilde{U} = A + M$. Тогда для элементов μ_{ij} эквивалентного возмущения M имеют место оценки

$$|\mu_{ij}| \lessapprox \begin{cases} 0, & i = 1, \\ \frac{1}{2} p^{-t+1} |\tilde{u}_{ij}|, & j \geqslant i, \\ \frac{1}{2} p^{-t+1} |\tilde{u}_{jj}| |\tilde{l}_{ij}|, & j < i. \end{cases}$$

24.20. Если матрица A имеет разреженный вид 11.12, то такой же вид имеет и матрица эквивалентного возмущения M в 24.19.

24.21. Если матрица A является трехдиагональной, то формулы 24.18 приобретают вид

$$\begin{aligned} &u_{11} = a_{11}, && u_{12} = a_{12}, && l_{21} = a_{21}/u_{11}, \\ &u_{ii} = a_{ii} - l_{i,i-1}u_{i-1,i}; && u_{i,i+1} = a_{i,i+1}; && l_{i+1,i} = a_{i+1,i}/u_{ii}, \\ &&&&& i = 2, 3, \ldots, n. \end{aligned}$$

Компактная схема более приспособлена к применению режима накопления, особенно скалярного произведения. Поэтому ее использование дает возможность снизить уровень влияния ошибок округления. Однако в ней несколько сложнее выполнять перестановки. Наиболее эффективно компактная схема применяется для получения треугольного разложения симметричных положительно определенных матриц.

24.22. Для любой симметричной положительно определенной матрицы A существует разложение $A = LL'$, где L — левая треугольная матрица с положительными диагональными элементами.

24.23. В условиях и обозначениях 24.22 справедливы соотношения

$$\|L\|_2 = \|A\|_2^{1/2}, \quad \|A\|_E^{1/2} \leqslant \|L\|_E \leqslant n^{1/4} \|A\|_E^{1/2}.$$

24.24. В условиях и обозначениях 24.21 компактная схема для определения элементов l_{ij} матрицы L имеет вид

$$\begin{aligned} &l_{11} = a_{11}^{1/2}, \quad l_{j1} = a_{j1}/l_{11}, && j > 1, \\ &l_{ii} = \left(a_{ii} - \sum_{p=1}^{i-1} l_{ip}^2\right)^{1/2}, && i > 1, \\ &l_{ji} = \left(a_{ji} - \sum_{p=1}^{i-1} l_{ip} l_{jp}\right) \Big/ l_{ii}, && j > i. \end{aligned}$$

Вычисления по этой схеме называются *методом квадратного корня* (*метод Холецкого*) для разложения матрицы на множители.

24.25. Пусть все элементы матрицы L разложения $A = LL'$ вычисляются в режиме накопления согласно 24.24. Обозначим через $\tilde{L}$ реально вычисленную матрицу и предположим, что $\tilde{L}\tilde{L}' = A + M$. Тогда для элементов μ_{ij} эквивалентного возмущения M имеют место оценки

$$|\mu_{ij}| \lessapprox \begin{cases} \frac{1}{2} p^{-t+1} |\tilde{c}_{ij}| |\tilde{c}_{ii}|, & j > i, \\ \frac{1}{2} p^{-t+1} |\tilde{c}_{ji}| |\tilde{c}_{jj}|, & j < i, \\ p^{-t+1} |\tilde{c}_{ii}|^2, & j = i. \end{cases}$$

24.26. В условиях и обозначениях 24.24 выполняется неравенство $\|M\|_E \lessapprox p^{-t+1} \|A\|_E$.

Обратим внимание на исключительную малость эквивалентного возмущения в методе квадратного корня. Согласно 24.26 его оценка лишь вдвое больше оценки возмущения, вносимого при вводе матрицы в память ЭВМ. Подобная малость эквивалентного возмущения объясняется отсутствием роста элементов, что является характерной чертой для симметричных положительно определенных матриц. Если матрица симметричная, но не положительно определенная, то для нее существует аналог метода квадратного корня, основанный на разложении 11.16. Однако в этом случае уже нельзя гарантировать существование оценок типа 24.26 для эквивалентного возмущения.

24.27. Пусть A — произвольная прямоугольная матрица размера $m \times n$. Умножим ее слева на циклическую по столбцам (см. 22.24) последовательность матриц вращения. Пусть в соответствии с 22.31, 22.33 каждая из матриц вращения выбирается из условия обращения в нуль после умножения на нее поддиагонального элемента, индексы которого совпадают с индексами матрицы вращения. Все шаги этого процесса могут быть реализованы, и после их выполнения исходная матрица A будет преобразована в правую треугольную матрицу A_N. Описанный процесс называется *методом вращений* (*методом Гивенса*) для разложения матрицы на множители.

24.28. Процесс 24.27 определяет соотношение $A_N = R_N A$, где $R_N = (\ldots T_{23} T_{1m} \ldots T_{13} T_{12})$ есть произведение всех участвующих в процессе матриц вращения.

24.29. Пусть $\tilde{T}_{ij}$, $\tilde{A}_N$ — реально вычисленные матрицы, $\tilde{R}_N$ — точное произведение реально вычисленных матриц $\tilde{T}_{ij}$. Если $\tilde{A}_N = \tilde{R}_N(A + M)$, то для эквивалентного возмущения M выполняются неравенства

$$\|M\|_E \lessapprox \begin{cases} \sqrt{2}\,(2m-3)\, p^{-t+1} \|A\|_E, & m \leqslant n, \\ \sqrt{2}\,(m+n-2)\, p^{-t+1} \|A\|_E, & m > n. \end{cases}$$

Несмотря на то, что процесс 24.27 связывается с разложением матрицы A на множители, утверждения 24.28, 24.29 представлены в несколько иной форме. С точки зрения точных вычислений равенство $A_N = R_N A$ эквивалентно равенству $A = R'_N A_N$, которое и означает разложение матрицы A на ортогональный множитель R'_N и треугольный множитель A_N. Однако если мы от равенства $\tilde{A}_N = \tilde{R}_N(A+M)$ перейдем к равенству $A + N = \tilde{R}'_N \tilde{A}_N$, то с учетом 22.29 оценка для N будет больше, чем оценка для

M. Как правило, в таком переходе нет особой необходимости, так как для большинства практических целей необходимо не разложение на множители, а именно представления вида 24.28. В методе Гаусса мы делали этот переход только по той причине, что он осуществляется точно и без изменения эквивалентных возмущений.

24.30. Пусть A — произвольная прямоугольная матрица размера $m \times n$. Умножим ее слева на последовательность матриц отражения $U_1, \ldots, U_{n-1}$ вида 22.52. Пусть каждая из матриц отражения U_k выбирается из условия обращения в нуль поддиагональных элементов k-го столбца. Все шаги этого процесса могут быть реализованы, и после их выполнения исходная матрица A будет преобразована в правую треугольную матрицу $\widehat{A}_N$. Описанный процесс называется *методом отражений* (*методом Хаусхолдера*) для разложения матрицы на множители.

24.31. Процесс 24.30 определяет соотношение $\widehat{A}_N = Q_N A$, где $Q_N = (U_{n-1} \ldots U_1)$ есть произведение всех участвующих в процессе матриц отражения.

24.32. Пусть $\widetilde{U}_k$, $\widetilde{\widetilde{A}}_N$ — реально вычисленные матрицы, $\widetilde{Q}_N$ — точное произведение реально вычисленных матриц $\widetilde{U}_k$, причем вычисление скалярных произведений всюду осуществляется в режиме накопления. Если $\widetilde{\widetilde{A}}_N = \widetilde{Q}_N(A + M)$, то для эквивалентного возмущения M выполняется неравенство

$$\|M\|_E \leqslant \frac{2{,}5 \cdot \sqrt{2} + 5}{3} np^{-t+1} \|A\|_E.$$

24.33. Пусть A — произвольная прямоугольная матрица размера $m \times n$. Переставим ее строки таким образом, чтобы на месте первой строки находилась строка наибольшей длины. Умножим далее справа эту матрицу на последовательность матриц вращения $T_{21}, T_{31}, \ldots, T_{n1}$, исключая при этом внедиагональные элементы первой строки. Переставим теперь строки матрицы, кроме первой, так, чтобы на месте второй строки находилась строка, имеющая наибольшую длину, без учета элементов первого столбца. Умножим справа матрицу на последовательность матриц $T_{32}, T_{42}, \ldots, T_{n2}$, исключая внедиагональные элементы второй строки, и т. д. Все шаги этого процесса могут быть реализованы, и после их выполнения исходная матрица A будет преобразована в левую трапециевидную матрицу $\widehat{A}_N$. Описанный процесс называется *нормализованным* методом вращения для разложения матрицы на множители.

24.34. Процесс 24.33 определяет соотношение $\widehat{A}_N = P_m A T_N$, где P_m есть матрица перестановок, а $T_N = (T_{21} T_{31} \ldots T_{n1} T_{32} \ldots)$ есть произведение всех участвующих в процессе матриц вращения.

24.35. Пусть $\widetilde{T}_{ij}$, $\widetilde{\widetilde{A}}_N$ — реально вычисленные матрицы, $\widetilde{T}_N$ — точное произведение реально вычисленных матриц $\widetilde{T}_{ij}$. Если $\widetilde{\widetilde{A}}_N = P_m(A+M)\widetilde{T}_N$, то для эквивалентного возмущения M выполняются неравенства

$$\|M\|_E \lessapprox \begin{cases} \sqrt{2}\,(2n-3)\,p^{-t+1}\|A\|_E, & n \leqslant m, \\ \sqrt{2}\,(m+n-2)\,p^{-t+1}\|A\|_E, & n > m. \end{cases}$$

24.36. Обозначим через $a_{ij}^{(N)}$ элементы матрицы A_N процесса 24.33. Для всех $1 \leqslant k < n$ справедливы соотношения

$$\begin{aligned} &|\,a_{kk}^{(N)}\,|^2 \geqslant \\ &\geqslant |\,a_{k+1,k}^{(N)}\,|^2 + |\,a_{k+1,k+1}^{(N)}\,|^2 \geqslant \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &\geqslant |\,a_{nk}^{(N)}\,|^2 + |\,a_{n,k+1}^{(N)}\,|^2 + \ldots + |\,a_{nn}^{(N)}\,|^2 \geqslant \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &\geqslant |\,a_{mk}^{(N)}\,|^2 + |\,a_{m,k+1}^{(N)}\,|^2 + |\,a_{m,k+2}^{(N)}\,|^2 + \ldots + |\,a_{mn}^{(N)}\,|^2. \end{aligned}$$

Матрица, удовлетворяющая этим соотношениям, называется *левой нормализованной трапециевидной матрицей.*

24.37. Процесс, аналогичный 24.33, но с использованием матриц отражения, называется *нормализованным методом отражений* для разложения матрицы на множители.

24.38. Пусть A — произвольная квадратная или прямоугольная матрица размера $m \times n$. Будем рассматривать ее столбцы слева направо как векторы и подвергнем их процессу ортогонализации Грама — Шмидта. Это означает, что строится последовательность матриц $S_r = S_{r-1}R_r$, $S_0 = A$, где матрицы R_r являются транспонированными матрицами типа M_r, внедиагональные элементы которых составлены из коэффициентов линейных комбинаций ортогонализуемых векторов. Если ранг матрицы A совпадает с числом ее столбцов, то все шаги этого процесса могут быть реализованы, и матрица S_n будет иметь ортогональные вектор-столбцы. Описанный процесс называется *методом ортогонализации* для разложения матрицы на множители.

24.39. Процесс 24.38 дает следующее разложение матрицы A на множители:

$$A = S_n\left(R_n^{-1} \ldots R_2^{-1}\right).$$

Здесь матрица, стоящая в скобках, есть левая треугольная с единичными диагональными элементами, структура которой с точностью до транспонирования полностью определяется утверждениями 22.66, 22.67, матрица S_n имеет ортогональные вектор-столбцы.

24.40. Пусть $\tilde{R}_2, \ldots, \tilde{R}_n$ и $\tilde{S}_1, \ldots, \tilde{S}_n$ — реально вычисленные матрицы в процессе 24.38. Предположим, что

$$\tilde{S}_n(\tilde{R}_n^{-1} \ldots \tilde{R}_2^{-1}) = A + M.$$

Тогда для эквивалентного возмущения M выполняется неравенство $\|M\|_E \lessgtr p^{-t+1}\|A\|_E$.

24.41. Методы вращений, отражений и ортогонализации позволяют получить QR-разложение любой матрицы, при этом эквивалентное возмущение увеличивается не более чем в 2—3 раза по сравнению с оценками, приведенными выше для этих методов.

С формальной точки зрения последнее утверждение почти очевидно, и требуется сделать совсем немного дополнительных вычислений, чтобы получить именно QR-разложение и дать для него оценку эквивалентного возмущения.

Разложение матрицы на множители является основой построения большинства численных методов линейной алгебры. Чем эффективнее осуществляется разложение, тем лучшими характеристиками обычно обладает и соответствующий метод. Нельзя дать однозначный ответ на вопрос «Какое из разложений лучше?», так как разные задачи предъявляют к разложениям различные требования. Тем не менее по некоторым характеристикам не только можно, но и нужно проводить сравнение. Для описанных разложений такие характеристики приведены в табл. 24.1. Предполагается, что все

Таблица 24.1

Сравнительная характеристика разложений

Способ получения сомножителей	Режим вычислений	Число операций	Точность	Дополнительная память
Метод Гаусса	fl	$(2/3)n^3$	αn	0
Компактная схема	fl_2	$(2/3)n^3$	β	0
Метод квадратного корня	fl_2	$(1/3)n^3$	1,0	0
Метод вращений	fl	$2n^3$	$2{,}9 \cdot n$	0
Метод отражений	fl_2	$(4/3)n^3$	$2{,}9 \cdot n$	$2n$
Нормализованный метод отражений	fl_2	$(4/3)n^3$	$2{,}9 \cdot n$	$3n$
Метод ортогонализации	fl_2	$2n^3$	1,0	$0{,}5 \cdot n^2$
Нормализованный метод вращений	fl	$2n^3$	$2{,}9 \cdot n$	n

матрицы квадратные и имеют один и тот же порядок, равный n. Данные для процесса ортогонализации указаны при отсутствии дополнительной переортогонализации.

Общее время, затрачиваемое на получение разложения, по существу, определяется числом арифметических операций, которые необходимо при этом выполнить. В графе «Число операций» табл. 24.1 приведены главные члены числа арифметических операций для всех разложений. В случае использования преобразований вращения одну треть от общего числа операций составляют операции сложения, две трети — операции умножения. Для остальных разложений число операций сложения и умножения примерно одинаково. Операции деления и извлечения квадратного корня главный член не определяют. Самое медленное из рассмотренных разложений выполняется в 6 раз дольше, чем самое быстрое.

Алгебраические задачи, особенно с матрицами большого порядка, требуют для своего решения значительных ресурсов памяти ЭВМ. Один из путей экономии памяти — размещение информации о сомножителях на месте исходной матрицы, в основном на местах получаемых нулевых элементов. Не всегда бывает достаточно этого места, нередко требуется дополнительная память. В графе «Дополнительная память» табл. 24.1 приведены главные члены числа полных слов памяти ЭВМ, которые необходимо добавить для размещения сомножителей. При этом предполагается, что вся память, отведенная для исходной матрицы, также используется для хранения сомножителей.

Со всех точек зрения невыгодно хранить ортогональный множитель разложения как одну квадратную матрицу. Поэтому его всегда запоминают в факторизованном виде — через параметры определяющих его матриц вращения или отражения. Хранение параметров матриц отражения не вызывает особых вопросов. Относительно же параметров матриц вращения стоит сказать несколько слов. Можно хранить оба параметра c, s. Однако это приводит к большому объему дополнительно требуемой памяти. Принимая во внимание соотношение $c^2 + s^2 = 1$, любую матрицу вращения можно задавать несколькими способами и одним параметром. В целях обеспечения высокой точности используют, например, представления c и s через тангенс половинного угла с указанием того, является ли соответствующая матрица вращения правой или левой. Первоначальный вид матрицы вращения восстанавливается при каждом случае ее использования. Обычно такое восстановление не приводит к изменению главного члена общего числа выполняемых операций. Именно эта ситуация отражена в табл. 24.1.

Точность является одной из важнейших, а чаще всего решающих характеристик любого численного метода, в том числе и разложения матрицы на множители. Для любого из рассмотренных разложений матрицы A эквивалентное возмущение M удовлетворяет неравенству $\|M\|_E \leqslant \leqslant f(n)p^{-t+1}\|A\|_E$, где функция $f(n)$ зависит только от n и от способа получения разложения. В графе «Точность» табл. 24.1 приведены главные члены функции $f(n)$. Если имеется рост элементов, то он определяется параметрами α, β.

Оценки точности, приведенные в табл. 24.1, гарантируются лишь в том случае, когда разложения во всех деталях осуществляются по рассмотренным выше вычислительным схемам. Любое изменение вычислительной схемы должно обосновываться соответствующим анализом ошибок, так как иначе возможна катастрофическая потеря точности. Символ fl в графе «Режим вычислений» табл. 24.1 означает, что для достижения соответствующей точности можно ограничиться вычислениями с одинарной точностью. Символ fl_2 означает, что необходимо использование операций накопления.

Выводы относительно точности касаются только величин эквивалентных возмущений, а не выполненных условий, определяющих вид множителей. Напомним, что QR-разложение матрицы в основном единственно. Поэтому в теоретическом плане ортогональный множитель в методе ортогонализации с точностью до нормировок должен совпадать с ортогональными множителями в методах вращений и отражений. На практике между ними нет ничего общего. В методе ортогонализации меньше эквивалентное возмущение, но плохо выполняются условия ортогональности. В методах вращений и отражений эквивалентное возмущение больше, но условия ортогональности выполняются с высокой точностью.

§ 25. Решение систем с невырожденными матрицами

Решение систем линейных алгебраических уравнений $Ax = b$ общего вида, основанное на разложении матрицы на множители, сводится к последовательному решению одной или нескольких систем с матрицами специального вида. Поэтому, прежде чем переходить к изучению общих систем, рассмотрим некоторые специальные системы $Gu = l$.

25.1. Пусть матрица системы $Gu = l$ — треугольная, например правая, и сама система имеет вид

$$\begin{aligned} g_{11}u_1 + g_{12}u_2 + \ldots + g_{1n}u_n &= l_1, \\ g_{22}u_2 + \ldots + g_{2n}u_n &= l_2, \\ \cdot \quad \cdot \quad \cdot \quad \cdot & \\ g_{nn}u_n &= l_n. \end{aligned}$$

Для нахождения решения этой системы поступаем следующим образом. Определяем $u_n = l_n/g_{nn}$. Предположим, что уже вычислены $u_n, u_{n-1}, \ldots, u_{i+1}$ из последних $n-i$ уравнений. Из i-го уравнения находим

$$u_i = \frac{l_i - g_{i,i+1}u_{i+1} - \ldots - g_{i,n}u_n}{g_{ii}}.$$

Таким образом последовательно определяем все координаты $u_n, \ldots, u_1$ вектора u. Этот способ решения систем линейных алгебраических уравнений с невырожденной треугольной матрицей называется *обратной подстановкой*.

Иногда обратной подстановкой называется описанный способ решения системы только с правой треугольной матрицей. Аналогичный способ решения системы с левой треугольной матрицей в этом случае называется *прямой подстановкой*.

25.2. Пусть $\tilde{u}$ — реально вычисленное решение системы $Gu = l$ с невырожденной треугольной матрицей. Предположим, что при определении каждой из координат вектора $\tilde{u}$ используется режим накопления. Тогда вектор $\tilde{u}$ есть точное решение возмущенной системы $(G + \Delta)\tilde{u} = l + \nu$, причем возмущения Δ, ν обладают следующими свойствами:

— эквивалентное возмущение Δ есть диагональная матрица с элементами δ_{ii}, удовлетворяющими неравенствам $|\delta_{ii}| \lesssim 0{,}5 \cdot |g_{ii}| p^{-t+1}$;

— координаты вектора эквивалентного возмущения ν удовлетворяют неравенствам $|\nu_i| \lesssim |g_{ii}|\omega$, где ω — машинный нуль;

— выполняется равенство $\Delta\nu = 0$.

Решение систем линейных алгебраических уравнений с треугольной матрицей редко бывает самостоятельной задачей. Как правило, оно представляет собой лишь вспомогательную задачу в общем процессе решения систем. Поэтому очень важно обеспечить режим вычислений, при котором процесс не может остановиться из-за возникновения какой-либо непредвиденной ситуации, связанной с невозможностью выполнить ту или иную операцию. В обратной подстановке остановка процесса может произойти, в основном, из-за переполнения при делении на диагональный элемент. Для устранения таких ситуаций эффективным является введение в решение нормирующих множителей.

25.3. Пусть задана система 25.1 с треугольной матрицей; ставится задача найти вектор u_α и число α, удовлетворяющие услониям $\|u_\alpha\|_\infty = 1$, $|\alpha| \leqslant 1$ и такие, что $Gu_\alpha = \alpha l$. Если l_n и g_{nn} не обращаются в нуль одновременно, то находим

$$u_n^{(n)} = \begin{cases} l_n/g_n, & \text{если } |g_{nn}| > |l_n|, \\ 1, & \text{если } |g_{nn}| \leqslant |l_n|; \end{cases}$$

$$\alpha^{(n)} = \begin{cases} 1, & \text{если } |g_{nn}| > |l_n|, \\ g_{nn}/l_n, & \text{если } |g_{nn}| \leqslant |l_n|. \end{cases}$$

Если $l_n = g_{nn} = 0$, то положим $u_n^{(n)} = \alpha^{(n)} = 1$. Предположим, что уже вычислены числа $u_{i+1}^{(i+1)}, \ldots, u_n^{(i+1)}, \alpha^{(i+1)}$. Вычисляем

$$\gamma_i = \alpha^{(i+1)} l_i - \sum_{s=i+1}^{n} g_{is} u_s^{(i+1)}.$$

Если γ_i и g_{ii} не обращаются в нуль одновременно, то находим

$$u_s^{(i)} = \begin{cases} \gamma_i/g_{ii}, & \text{если } |g_{ii}| > |\gamma_i|, \quad s = i, \\ u_s^{(i+1)}, & \text{если } |g_{ii}| > |\gamma_i|, \quad s > i, \\ 1, & \text{если } |g_{ii}| \leqslant |\gamma_i|, \quad s = i, \\ (g_{ii}/\gamma_i)\, u_s^{(i+1)}, & \text{если } |g_{ii}| \leqslant |\gamma_i|, \quad s > i; \end{cases}$$

$$\alpha^{(i)} = \begin{cases} 1, & \text{если } |g_{ii}| > |\gamma_i|, \\ (g_{ii}/\gamma_i)\,\alpha^{(i+1)}, & \text{если } |g_{ii}| \leqslant |\gamma_i|. \end{cases}$$

Если $\gamma_i = g_{ii} = 0$, то считаем $g_{ii}/\gamma_i = 1$. Вектор u_α с координатами $u_1^{(1)}, \ldots, u_n^{(1)}$ и число $\alpha = \alpha^{(1)}$ являются искомыми. Этот способ решения систем с треугольной матрицей называется *обратной подстановкой с нормировкой*.

25.4. Пусть $\tilde{u}$, $\tilde{\alpha}^{(1)}$ — реально вычисленные величины процесса 25.3. Если при вычислении γ_i используется режим накопления, то имеет место равенство $(G + \Delta)\tilde{u}_\alpha = \tilde{\alpha}(l + \nu)$, где

$$\|\Delta\|_E \lessapprox np^{-t+1}\|G\|_E, \quad \|\nu\|_E \lessapprox p^{-t+1}\|l\|_E.$$

Число $\tilde{\alpha}$ совпадает с $\tilde{\alpha}^{(1)}$, если $\tilde{\alpha}^{(1)} \neq 0$. Если $\tilde{\alpha}^{(1)} = 0$, но все диагональные элементы матрицы G отличны от нуля, то $\tilde{\alpha}$ по модулю меньше машинного нуля. Всегда $\|\tilde{u}_\alpha\|_\infty = 1$.

25.5. После выполнения процесса 25.3 можно оценить снизу число обусловленности матрицы G, выраженное в ∞-норме. Именно:

$$\operatorname{cond} A \geqslant \|A\|_\infty/(\alpha\|l\|_\infty).$$

Процесс 25.3 имеет значительные преимущества перед процессом 25.1. Он может применяться к любой системе с невырожденной или вырожденной матрицей без предварительного анализа ситуации. Если после его выполнения окажется, что $\tilde{\alpha} \neq 0$, то это означает, что с точностью до множителя $\tilde{\alpha}^{-1}$ получено точное решение слабо возмущенной системы. Если же

$\tilde{\alpha}=0$, то это говорит о том, что в пределах заданной точности выполнения операций даже при использовании режима накопления в решении системы нельзя гарантировать ни одного верного знака из-за чрезвычайно плохой обусловленности матрицы системы. В дальнейшем мы увидим, что полезная информация в векторе $\tilde{u}_\alpha$ содержится и в том случае, когда $\tilde{\alpha}=0$. Конечно, процесс 25.3 логически несколько сложнее процесса 25.1, но этот фактор имеет какое-то значение только для ручного счета.

25.6. Пусть матрица системы $Gu=l$ имеет ненулевые ортогональные строки. Для нахождения решения такой системы можно разделить каждую координату вектора l на квадрат евклидовой нормы соответствующей строки матрицы G, а затем умножить полученный вектор на матрицу G'.

25.7. Пусть $\tilde{u}$ — реально вычисленное решение системы $Gu=l$ согласно 25.6. Если при вычислении скалярных произведений используется режим накопления, то $\tilde{u}$ является точным решением возмущенной системы $G\tilde{u}=l+\nu$, где $\|\nu\|_E \lessapprox (\sqrt{n}+0{,}5)\times \times p^{-t+1}\|l\|_E$.

Значительная часть наиболее известных численных методов решения систем линейных алгебраических уравнений $Ax=b$ основана на разложении матрицы A на множители. В зависимости от того, как связаны сомножители с матрицей A, различают две схемы построения методов.

25.8. Пусть матрица A разложена на множители, например: $A=BC$. Тогда решение системы $Ax=b$ сводится к решению двух систем: $By=b$, $Cx=y$.

25.9. Пусть найдены матрицы L, S, G, для которых выполняется соотношение $LAS=G$. Тогда решение системы $Ax=b$ сводится к решению системы $Gu=l$ с правой частью $l=Lb$ и вычислению вектора $x=Su$.

Все рассмотренные ранее разложения матрицы на множители имеют вид либо 25.8, либо 25.9. Первый вид имеют треугольные разложения, получаемые по компактной схеме и по методу квадратных корней. Разложения, осуществляемые по методу Гаусса и по методу ортогонализации, могут быть отнесены к обеим схемам. Все разложения, использующие преобразования вращения и отражения, представлены по второй схеме. Характерной особенностью второй схемы является то, что матрицы L и S задаются, как правило, в факторизованном виде, причем именно через те элементарные матрицы и матрицы перестановок, которые получаются в процессе нахождения представления 25.9.

Численные методы решения систем линейных алгебраических уравнений можно строить на основе любых исследованных разложений, решая вспомогательные системы с треугольной или ортогональной матрицей одним из описанных выше способов. Эти методы в отношении скорости и объема требуемой памяти ЭВМ обладают такими же характеристиками, что и соответствующие разложения матрицы. Главный член числа арифметических операций остается без изменения, так как при наличии разложений 25.8, 25.9 для решения вспомогательных систем с треугольной или ортогональной матрицей и на перестановку координат векторов нужно выполнить на порядок меньше вычислительной работы, чем для получения самих разложений. При этом, по существу, не требуется никакой дополнительной памяти ЭВМ по сравнению с той, которая уже была использована при разложении матрицы. Поэтому выбор вида разложения матрицы для построения численного метода решения систем линейных алгебраических уравнений, обладающего нужными характеристиками скорости и объема памяти ЭВМ, можно осуществлять, используя табл. 24.1. Опишем некоторые из таких методов.

25.10. (*Метод Гаусса.*) Метод основан на разложении 25.9. Матрица G — правая треугольная, матрица L представлена как произведение матриц типа N_r. Если используется стратегия выбора ведущего элемента, связанная с изменением порядка просмотра столбцов, то матрица S будет матрицей перестановок. При изменении порядка просмотра строк матрица L будет представлена как произведение матриц типа N_r и матриц перестановок. Матрица S будет единичной, если ведущий элемент не выбирается.

25.11. (*Компактная схема метода Гаусса.*) Метод основан на разложении 25.8. Матрица B — левая треугольная, матрица C — правая треугольная.

25.12. (*Метод квадратного корня.*) Метод основан на разложении 25.8 для положительно определенной матрицы A. Матрица C — правая треугольная, $B = C'$.

25.13. (*Метод отражений.*) Метод основан на разложении 25.9. Матрица G — правая треугольная, матрица L представлена как произведение матриц отражения, матрица S — единичная.

25.14. (*Нормализованный метод отражений.*) Метод основан на разложении 25.9. Матрица G — нормализованная левая треугольная, матрица S представлена как произведение матриц отражения, матрица L является матрицей перестановок.

25.15. (*Метод вращений.*) Метод основан на разложении 25.9. Матрица G — правая треугольная, матрица L представлена как произведение матриц вращения, матрица S — единичная.

25.16. (*Нормализованный метод вращений.*) Метод основан на разложении 25.9. Матрица G — нормализованная левая треугольная, матрица S представлена как произведение матриц вращения, матрица L является матрицей перестановок.

25.17. (*Метод ортогонализации.*) Метод основан на разложении 25.9. Матрица G имеет ортогональные строки, матрица L представлена как произведение матриц типа M_r, матрица S — единичная.

Все рассмотренные методы особенно удобны для решения систем линейных алгебраических уравнений с многими правыми частями и одной и той же матрицей. В этом случае соответствующие разложения 25.8, 25.9 находятся лишь один раз. Многократно приходится решать только простые системы с треугольными и ортогональными матрицами и умножать векторы на последовательности элементарных матриц. Важно отметить, что если разложение матрицы на множители требует выполнения порядка n^3 операций, то решение системы, основанное на этом разложении, дополнительно требует выполнения только порядка n^2 операций на каждую новую правую часть.

25.18. Пусть для эквивалентного возмущения M при разложении матрицы на множители имеет место оценка $\|M\|_E \lessapprox f(n)\,p^{-t+1}\|A\|_E$. Предположим, что системы с треугольной и ортогональной матрицами решаются в соответствии с 25.2, 25.4 или 25.6, а умножение вектора на последовательность элементарных матриц осуществляется так же, как при получении разложения. Тогда реально вычисленное по любому из методов 25.10—25.16 решение $\tilde{x}$ системы $Ax = b$ является точным решением возмущенной системы $(A + \mathscr{E})\tilde{x} = b + \varepsilon$. При этом

$$\|\mathscr{E}\|_E \leqslant \varphi(n)\,p^{-t+1}\|A\|_E, \quad \|\varepsilon\|_E \leqslant \psi(n)\,p^{-t+1}\|b\|_E,$$

где $\varphi(n) + \psi(n) \lessapprox 2f(n)$, если только в пределах таких возмущений матрица системы остается невырожденной.

25.19. В условиях и обозначениях 25.18 выполняется неравенство

$$\|\tilde{x} - x\|_E / \|x\|_E \lessapprox 2\operatorname{cond} f(n)\,p^{-t+1}.$$

25.20. В условиях и обозначениях 25.18 выполняется неравенство

$$\|A\tilde{x} - b\|_E \lessapprox 2f(n)\,p^{-t+1}\|A\|_E\|\tilde{x}\|_E.$$

Согласно 25.19 точность любого метода полностью определяется точностью разложения матрицы на множители. Но, как видно из табл. 24.1, в этом отношении различные разложения отличаются друг от друга не так уж сильно.

25.21. Если при правильной реализации какой-либо из методов не обеспечивает нужной точности решения системы линейных алгебраических уравнений, то нет никаких оснований надеяться на то, что другой метод будет давать для этой же системы существенно лучшие результаты.

Если при правильной реализации не удается получить в решении системы ни одного верного знака хотя бы одним из рассмотренных методов, то такую систему скорее всего следует рассматривать как *неустойчивую*, а не пытаться получить подходящее решение каким-нибудь другим аналогичным методом. Если же в решении имеются верные знаки, то в некоторых случаях можно увеличить точность приближенного решения.

Все рассмотренные численные методы решения систем линейных алгебраических уравнений обладают одним общим свойством. Именно, реально вычисленное решение является точным для некоторой возмущенной задачи. Выполненные исследования показывают, что эти возмущения весьма малы и нередко соизмеримы с ошибками округления входных данных. Если входные данные получены посредством каких-либо измерений или предварительных расчетов, то обычно они уже содержат значительно бóльшие ошибки. В этом случае всякая попытка улучшить приближенное решение без привлечения дополнительных сведений о точной задаче или ошибках входных данных окажется несостоятельной, ибо нет никакого критерия предпочтения одного приближенного решения другому.

Положение существенно изменяется, если входные данные заданы точно. Теперь среди всех приближенных решений, соответствующих определенному уровню эквивалентных возмущений, можно выбрать то, которое наиболее близко к точному. Как правило, это будет правильно округленное точное решение.

25.22. Пусть x — точное решение системы $Ax = b$, $x^{(k)}$ — некоторое приближение к нему, полученное любым способом. Если $x = x^{(k)} + \Delta^{(k)}$, то поправка $\Delta^{(k)}$ удовлетворяет системе $A\Delta^{(k)} = r_k$ с той же матрицей A и правой частью, совпадающей с невязкой $r_k = b - Ax^{(k)}$.

25.23. Возьмем произвольный вектор x_0 и построим последовательность векторов $x^{(k+1)} = \mathrm{fl}\,(x^{(k)} + \widetilde{\Delta}^{(k)})$, где $\widetilde{\Delta}^{(k)}$ есть поправка, реально вычисленная в соответствии с 25.22. Будем считать, что способ вычисления невязки и численный метод решения системы $A\Delta^{(k)} = r_k$ таковы, что

$$\|\Delta^{(k)} - \widetilde{\Delta}^{(k)}\|_E / \|\Delta^{(k)}\|_E \leqslant \theta < 1.$$

В этом случае выполняется следующее предельное соотношение:

$$\varlimsup_{k\to\infty} \frac{\|x - x^{(k)}\|_E}{\|x\|_E} \leqslant \frac{p^{-t+1}}{1-\theta},$$

где x есть точное решение системы $Ax = b$. Описанный процесс называется *процессом уточнения решения.*

Основная трудность в удовлетворении условия $\theta < 1$ связана с выбором соответствующего способа вычисления невязки. Если $x^{(k)}$ близко к точному решению, то невязка становится малой и прямое ее вычисление с одинарной точностью будет приводить к большим относительным ошибкам. Кроме того, абсолютная малость невязки может быть причиной значительных ошибок в поправке $\Delta^{(k)}$ из-за нерегулярности поведения ошибок округления вблизи машинного нуля.

25.24. Пусть процесс уточнения 25.22 осуществляется по следующему предписанию:

— вычисляется невязка в режиме накопления;

— невязка нормируется;

— решается система $A\Delta^{(k)} = r_k$ одним из методов, удовлетворяющих условию 25.18;

— вычисленная поправка умножается на обратную величину нормирующего множителя невязки.

В этом случае процесс уточнения реализуется на всех этапах и выполняется неравенство $\theta \lesssim (2f(n) + 3/2)\,\mathrm{cond}\,A p^{-t+1}$.

Таким образом, если входные данные системы с невырожденной матрицей заданы точно, то можно построить последовательность векторов $\{x^{(k)}\}$, определяющую исключительно точное приближение к точному решению. Процесс уточнения решения тем эффективнее, чем меньше θ. Обычно достаточно взять 2—3 приближения, чтобы получить вектор, который отличается от точного решения так же, как отличается от него правильно округленное точное решение. Но и построение большого числа векторов практически не приводит к заметному увеличению общего времени решения системы. Многократное решение систем $A\Delta^{(k)} = r_k$ может быть осуществлено весьма быстро, если разложение матрицы A на множители, выполненное при решении первой системы, использовать при решении всех последующих систем с другими правыми частями.

С процессом уточнения решения связан один интересный факт. Напомним, что уже первое приближение к решению является точным решением возмущенной системы. При этом возмущения не только малы, но и не зависят практически от обусловленности матрицы. Правильно округленное точное решение также является точным решением некоторой возмущенной системы. И снова возмущения малы и не зависят от обусловленности матрицы. Для наиболее точных методов решения систем эти возмущения соизмеримы по величине. Поэтому нет никакого основания ожидать существенного уменьшения норм невязок при последовательном выполнении процесса уточнения решения. Более того, на некоторых шагах нормы невязок могут даже несколько увеличиться. Описанный процесс уточнения связан не с уменьшением эквивалентных возмущений или невязок, а с уменьшением влияния обусловленности матрицы исходной системы на погрешность в решении.

§ 26. Особенности решения неустойчивых систем

Формулы 16.19 показывают, что для матрицы, близкой к вырожденной, возможны большие возмущения в решении системы линейных алгебраических уравнений даже при малых возмущениях в матрице и правой части. Если фиксированы уровень ошибок входных данных и точность вычислений, то всегда найдутся системы с настолько большими значениями чисел обусловленности, что для них нельзя гарантировать в решении никакой точности. Такие системы принято называть неустойчивыми или плохо обусловленными. Это понятие, в основном, отражает качественную сторону задачи, связанную с нарушением непрерывности решения системы в окрестности вырожденной матрицы или, в более общем случае, матрицы неполного ранга. Явление разрывности характерно не только для классического решения системы линейных алгебраических уравнений, но и для любого его обобщения, например нормального псевдорешения.

26.1. Если матрица системы имеет полный ранг, то в некоторой окрестности изменения коэффициентов матрицы нормальное псевдорешение непрерывно.

26.2. Какова бы ни была матрица системы, нормальное псевдорешение непрерывно по правой части всюду.

26.3. Если матрица системы не имеет полного ранга, то в любой окрестности изменения коэффициентов матрицы нормальное псевдорешение разрывно.

26.4. Если входные данные системы с матрицей неполного ранга заданы с ошибками, то никакое повышение точности вычислений и никакие преобразования системы не могут обеспечить гарантированной точности нормального псевдорешения без привлечения дополнительной информации о точной задаче или ошибках входных данных.

Факт разрывности псевдорешения настолько важен, что заставляет считать исследование зависимости погрешности псевдорешения от возмущения входных данных и ошибок округления неотъемлемой и обязательной частью любого численного метода решения систем с матрицами неполного ранга. Тем не менее такое исследование проводится не так уж часто. По-видимому, немалую роль в этом играет тот гипноз легкости, с которой математика точных вычислений предлагает «эффективные» методы решения систем линейных алгебраических уравнений.

Согласно многочисленным рецептам можно решать любую систему, например, методом Гаусса с выбором главного элемента по всей матрице. Если матрица имеет неполный ранг, то после выполнения определенного числа преобразований будет получена система с трапециевидной матрицей. Решение системы с такой матрицей не вызывает особых трудностей. В процессе преобразований легко устанавливается и факт совместности исходной системы.

Подобные рецепты выглядят весьма привлекательно. Особенно интересными они кажутся для методов, основанных на унитарных преобразованиях, так как в случаях, когда используются модификации, приводящие к левой нормализованной трапециевидной матрице, нормальное псевдорешение находится особенно просто. Почти не возникает сомнений в том, что требуется лишь незначительное изменение уже известных численных методов и можно будет решать системы общего вида, по крайней мере совместные системы. Кажется ясным и то, как нужно модифицировать методы.

В основе большинства модификаций лежит следующая идея. Ошибки округления малы. Как правило, малы и ошибки входных данных. Будем решать систему каким-либо подходящим прямым методом. Если точная матрица имеет неполный ранг, то в процессе реальных преобразований, по-видимому, получится простая матрица, у которой все элементы последних строк будут малы. Отбросим соответствующие уравнения и найдем решения полученной системы. Они будут служить достаточно хорошим приближением к решениям точной системы. Однако в действительности дело обстоит не так просто.

26.5. Если матрица точной системы имеет неполный ранг, то малость возмущений входных данных и ошибок округления совсем не обязательно приведет к появлению в процессе преобразования системы каких-либо строк или столбцов, целиком состоящих из таких же малых элементов.

В этом заключается основная, но не единственная трудность построения численных методов решения систем с матрицами неполного ранга, основанных на эквивалентных преобразованиях исходной системы. Еще одна трудность связана с обоснованием дальнейших преобразований тех систем, матрицы которых имеют строки или столбцы с малыми элементами.

Необходимость использования дополнительной информации при решении неустойчивых систем вызывает определенные трудности при конструировании соответствующих вычислительных алгоритмов. Эта информация весьма разнородна по своей природе. Наиболее простая возможность учесть ее для сколько-нибудь широкого класса задач состоит в параметризации вычислительного алгоритма. В этом случае получение достоверного приближения к нужному решению исходной задачи будет заключаться в многократном решении параметризованной задачи с целью подбора совокупности параметров согласно дополнительной информации.

Мы рассмотрим два вида параметризации — дискретную и непрерывную. Всюду в этом параграфе будем считать, что сингулярные числа ρ_i матрицы A занумерованы в порядке невозрастания.

26.6. Пусть $A = U\Lambda V$ — сингулярное разложение произвольной прямоугольной матрицы A, причем сингулярные числа ρ_i расположены на диагонали матрицы Λ в порядке невозрастания. Обозначим через A_k матрицу $U\Lambda_k V$, где Λ_k — диагональная матрица, первые k диагональных элементов которой ненулевые и совпадают с соответствующими диагональными элементами матрицы

Λ, а остальные равны нулю. Эти матрицы обладают следующими свойствами:

— матрицы A_k имеют ранг, равный k;

— для любой матрицы A и любого числа k справедливы равенства $\|A - A_k\|_2 = \rho_{k+1}$, $\|A - A_k\|_E = \left(\sum\limits_{i>k} \rho_i^2\right)^{1/2}$;

— на множестве матриц B, ранг которых равен k, величина $\|A - B\|_E$ достигает минимума на матрице A_k;

— на множестве матриц B, удовлетворяющих неравенству $\|A - B\|_E \leqslant \delta$, величина $\|A - B\|_E$ достигает максимума на одной из матриц A_k, причем k таково, что $\|A - A_k\|_E \leqslant \delta$, но $\|A - A_{k-1}\|_E > \delta$.

26.7. Пусть $A = U\Lambda V$ — сингулярное разложение матрицы A с упорядоченными сингулярными числами на диагонали Λ. Подпространство, совпадающее с линейной оболочкой первых k столбцов матрицы V^* (U), называется *главным правым* (*левым*) *сингулярным подпространством* матрицы A.

26.8. Нормальное псевдорешение x_k системы $A_k x = b$ есть ортогональная проекция нормального псевдорешения x_0 системы $Ax = b$ на главное правое сингулярное подпространство размерности k для матрицы A.

26.9. Пусть b_k есть ортогональная проекция правой части b системы $Ax = b$ на главное левое сингулярное подпространство для матрицы A размерности k. Система $A_k x = b_k$ является совместной, и ее нормальное решение есть x_k.

26.10. В обозначениях 26.8, 26.9 выполняется равенство $\|Ax_k - b\|_E = \|b - b_k\|_E$.

26.11. Для любой матрицы A величина $\|A\| \|A^+\|$ называется *числом обусловленности* матрицы A и обозначается $\operatorname{cond}^+ A$.

26.12. Рассмотрим систему $Ax = b$ и возмущенную систему $(A + \mathscr{E})\hat{x} = b + \varepsilon$. Пусть

$$\delta x_0^{(k)} = \frac{\|\hat{x}_k - x_0\|_E}{\|x_0\|_E}, \quad \delta A = \frac{\|\mathscr{E}\|_E}{\|A\|_E}, \quad \delta b = \frac{\|\varepsilon\|_E}{\|b\|_E}.$$

Предположим, что система $Ax = b$ совместна, матрица A имеет ранг r и величины δA, δb малы по сравнению с ρ_r. Тогда справедлива оценка $\delta x_0^{(r)} \lesssim \operatorname{cond}^+ A\,(\delta A + \delta b)$.

26.13. Пусть в обозначениях 26.10—26.12 система $Ax = b$ несовместна, матрица A имеет ранг r и величины δA, δb малы по сравнению с ρ_r. Имеет место оценка

$$\delta x_0^{(r)} \lesssim \operatorname{cond}^+ A\,(\delta A + \delta b) + \operatorname{cond}^+ A\left(\operatorname{cond}^+ A \cdot \delta A + \delta b\right) \frac{\|b - b_r\|_E}{\|b_r\|_E}.$$

Величина $\|b - b_r\|$ определяет меру несовместности системы $Ax = b$. Если система $Ax = b$ совместна или почти совместна, т. е. $\|b - b_r\|$ равна нулю или почти равна нулю, то согласно 26.13 нормальное решение x_0 точной системы можно определить с помощью проекции $\hat{x}_r$ возмущенной системы с такой же точностью, как и для системы с невырожденной матрицей. В общем случае точность нормального псевдорешения зависит от степени согласования матрицы и правой части исходной системы.

В связи со сказанным обратим внимание на следующее обстоятельство. Пусть входные данные системы заданы точно и сама система является совместной. Предположим, что эта система с помощью некоторых эквивалентных преобразований приводится к некоторой системе простого вида. За счет влияния ошибок округления новая система в точном смысле будет почти наверняка несовместной. Однако мера возникшей несовместности будет соизмерима с величиной ошибок округления и, следовательно, в соответствии с 26.13 не будет играть существенной роли.

Оценки 26.12, 26.13 получены для возмущений, достаточно малых по сравнению с минимальным ненулевым сингулярным числом. Такое предположение хотя и позволяет решать задачу и даже гарантировать оценку точности, но, по существу, требует, кроме малости ошибок, знания ранга точной матрицы. Совокупность условий, сформулированных в 26.12, 26.13,— это и есть в конкретном частном случае та дополнительная информация об исходной системе, которая позволяет по возмущенной системе высказать какое-то суждение о решении точной системы. Эти условия можно ослабить даже в случае систем самого общего вида.

26.14. Пусть входные данные системы $Ax = b$ заданы с абсолютной ошибкой порядка ε. Имеет место оценка

$$\delta x_0^{(k)} \lessapprox \alpha \left(\frac{\varepsilon}{\rho_k - \rho_{k+1}} + \rho_{k+1}^2 \right)$$

в случае совместности точной системы и оценка

$$\delta x_0^{(k)} \lessapprox \beta \left(\frac{\varepsilon}{(\rho_k - \rho_{k+1})\, \rho_k} + \rho_{k+1}^2 \right)$$

в противном случае. Коэффициенты α, β зависят только от параметров точной системы и способа измерения ошибок.

26.15. Для $0 \leqslant \varepsilon \leqslant (4n)^{-1}$ и любого набора чисел γ_k, где $1 = \gamma_1 \geqslant \gamma_2 \geqslant \ldots \geqslant \gamma_n \geqslant \gamma_{n+1} = 0$, справедливы соотношения

$$\min_{1 \leqslant k \leqslant n} \left(\frac{\varepsilon}{\gamma_k - \gamma_{k+1}} + \gamma_{k+1}^2 \right) \leqslant 3\,(n\varepsilon)^{2/3},$$

$$\min_{1 \leqslant k \leqslant n} \left(\frac{\varepsilon}{(\gamma_k - \gamma_{k+1})\, \gamma_k} + \gamma_{k+1}^2 \right) \leqslant 4\,(n\varepsilon)^{1/2}.$$

26.16. Пусть входные данные произвольной системы линейных алгебраических уравнений с прямоугольной матрицей заданы с точностью порядка ε. Существует такое k, что проекция $\hat{x}_k$ возмущенной системы приближает нормальное псевдорешение x_0 точной системы с точностью порядка ε^α. Если исходная система совместна, то $\alpha \geqslant 2/3$, и $\alpha \geqslant 1/2$ — в противном случае.

Заметим, что наличие малых сингулярных чисел матрицы не обязательно свидетельствует о невозможности вычислить псевдорешение с достаточно хорошей точностью. Если матрица имеет группу больших сингулярных чисел, а остальные сингулярные числа соизмеримы с точностью входных данных, то из 16.14 следует, что одна из проекций $\hat{x}_k$ приближает нормальное псевдорешение x_0 с точностью порядка ε как для совместной, так и для несовместной системы. Этот факт имеет исключительное значение для обоснования большинства численных методов решения систем линейных алгебраических уравнений с вырожденной матрицей.

Таким образом, в зависимости от свойств исходной системы ее нормальное псевдорешение можно определить по реально заданной системе с точностью порядка ε^{α}, где $1/2 \leqslant \alpha \leqslant 1$. Единственное, чего нам теперь не хватает,— это критерия выбора нужной проекции $\hat{x}_k$. Такой критерий может быть разработан на основе привлечения дополнительной информации о задаче. Номер проекции определяет дискретную параметризацию вычислительного алгоритма, о которой говорилось выше. Непрерывная параметризация устроена иначе.

26.17. Функционал $\Phi_\alpha(x) = \alpha \|x\|_E^2 + \|Ax - b\|_E^2$, где $\alpha \geqslant 0$, называется *регуляризирующим функционалом* системы $Ax = b$.

26.18. При $\alpha > 0$ регуляризирующий функционал имеет единственную точку минимума x_α.

26.19. Точка минимума x_α является решением системы линейных алгебраических уравнений $(A^*A + \alpha E)x_\alpha = A^*b$ и называется *регуляризованным решением* системы $Ax = b$.

26.20. Матрица $A^*A + \alpha E$ при $\alpha > 0$ является эрмитовой положительно определенной.

26.21. Для любой матрицы A и $\alpha > 0$ справедливы неравенства

$$\|(A^*A + \alpha E)^{-1}A^*\|_{2,E} \leqslant \|A^+\|_{2,E},$$
$$\|(A^*A + \alpha E)^{-1}A^*\|_{2,E} \leqslant \|E\|_{2,E}/(2\alpha)^{1/2}.$$

26.22. Разность $x_\alpha - x_\beta$ удовлетворяет уравнению

$$(A^*A + \alpha E)(A^*A + \beta E)(x_\alpha - x_\beta) = (\beta - \alpha)A^*b.$$

26.23. Если x_0 есть нормальное псевдорешение системы $Ax = b$, то всегда выполняется предельное соотношение

$$\lim_{\alpha \to +0} x_\alpha = x_0.$$

26.24. Имеют место оценки

$$\|x_\alpha\|_E^2 \leqslant \|x_0\|_E^2 \leqslant \|x_\alpha\|_E^2 + 2\alpha\eta^2, \quad \|x_0 - x_\alpha\|_E^2 \leqslant \alpha^2\gamma^2,$$

где η и γ — евклидовы нормы нормальных решений соответственно уравнений $(A^*A)^{3/2}x = A^*b$ и $(A^*A)^2 x = A^*b$.

26.25. Для любых матрицы A, вектора b и $\alpha > 0$ справедливо неравенство

$$\|(A^*A + \alpha E)A^*y\|_E \leqslant \|b\|/\alpha^{1/2}.$$

26.26. Пусть $Ax = b$ — точная система, $\hat{A}\hat{x} = \hat{b}$ — возмущенная. Предположим, что $\|A - \hat{A}\|_2 \leqslant \varepsilon_A$, $\|b - \hat{b}\|_E \leqslant \varepsilon_b$; тогда

$$\|x_0 - \hat{x}_\alpha\|_E \leqslant \alpha\gamma + \frac{\varepsilon_A}{\alpha}\left(\|Ax_0 - b\|_E^2 + 2\alpha^2\eta^2\right)^{1/2} + \frac{1}{\alpha^{1/2}}(\varepsilon_A\|x_0\|_E + \varepsilon_b),$$

где параметры η, γ определены в 26.24, а $\hat{x}_\alpha$ есть решение ситемы $(\hat{A}^*\hat{A} + \alpha E)\hat{x}_\alpha = \hat{A}^*\hat{b}$.

Правая часть неравенства 26.26 не содержит никакой информации, связанной с возмущенной системой. Поэтому существует такое α, при котором она достигает своего минимума. Это значение α обеспечивает почти наилучшее приближение $\hat{x}_\alpha$ к точному нормальному псевдорешению x_0.

26.27. Пусть входные данные системы $Ax = b$ заданы с абсолютной ошибкой порядка ε. Имеет место оценка

$$\|x_0 - \hat{x}_\alpha\|_E \leqslant \delta(\alpha + \varepsilon + \varepsilon/\alpha^{1/2})$$

в случае $Ax_0 - b = 0$, т. е. совместности точной системы, и оценка

$$\|x_0 - \hat{x}_\alpha\|_E \leqslant \tau(\alpha + \varepsilon + \varepsilon/\alpha + \varepsilon/\alpha^{1/2})$$

в случае $Ax_0 - b \neq 0$, т. е. несовместности точной системы. Коэффициенты δ, τ зависят только от параметров точной системы и меры измерения ошибок.

26.28. Пусть входные данные произвольной системы линейных алгебраических уравнений с прямоугольной матрицей заданы с точностью порядка ε. Существует такое α, что регуляризованное решение $\hat{x}_\alpha$ возмущенной системы приближает нормальное псевдорешение x_0 точной системы с точностью порядка ε^θ. Если исходная система совместна, то $\theta \geqslant 2/3$, и $\theta \geqslant 1/2$ — в противном случае.

Параметр α, обеспечивающий необходимое приближение $\hat{x}_\alpha$, снова не может быть найден лишь по возмущенной системе, и для его определения опять необходимо привлекать дополнительную информацию о задаче.

В некоторых случаях в регуляризирующем функционале бывает необходимо брать не евклидовы нормы вектора x и невязки $Ax - b$, а какие-либо другие нормы. В общем случае решение соответствующей задачи оказывается очень сложным. Но если взять нормы вида $\|z\|_D^2 = (Dz, z)$ для эрмитовых положительно определенных матриц D, то эта задача сводится к исследованной.

26.29. Имеет место тождество

$$\alpha\|x\|_D^2 + \|Ax - b\|_S = \alpha\|y\|_E^2 + \|S^{1/2}AD^{-1/2}y - S^{1/2}b\|_E^2,$$

где $y = D^{1/2}x$, при любых эрмитовых положительно определенных матрицах D, S.

26.30. Определение нормального псевдорешения с помощью регуляризованных решений называется *методом регуляризации*.

§ 27. Тактика решения систем общего вида

Приступая к решению систем линейных алгебраических уравнений, приходится отвечать на вопрос, какой численный метод взять за основу. Сложность ситуации заключается в том, что однозначный выбор метода удается осуществить довольно редко. Если матрица не симметричная, но имеет доминирующую диагональ, то нет никаких веских доводов против применения компактной схемы или метода Гаусса без выбора ведущих элементов. В случае симметричной положительно определенной матрицы вне всякой конкуренции находится метод квадратного корня. А дальше желание эффективно решить систему вступает в противоречие с желанием обеспечить регулярность вычислительного процесса. Если, например, известно, что матрица системы не очень плохо обусловлена, то можно применить метод Гаусса с выбором ведущего элемента по столбцу. Однако при этом есть риск, связанный с ростом элементов и потерей точности. Можно использовать метод вращений, в котором такого риска нет, но он более трудоемкий, и т. д. Если же в процессе решения выясняется, что матрица системы очень плохо обусловлена, то вся работа, затраченная на выяснение этого факта, может оказаться бесполезной для дальнейшего анализа задачи.

Опираясь на изложенные методы и факты, мы можем теперь разработать некоторую тактику действий по решению систем линейных алгебраических уравнений общего вида. Применение этой тактики целесообразно в тех случаях, когда имеющихся сведений о системе недостаточно для того, чтобы сделать выбор численного метода и гарантировать его устойчивость.

Мы не будем накладывать сейчас какие-либо ограничения на исходную систему. Она может быть как совместной, так и несовместной, как хорошо, так и плохо обусловленной. Ранг матрицы системы и ее размеры могут быть произвольными. Вычислительный процесс будет устроен таким образом, что чем «лучше» исходная система, тем раньше он прекратится, давая приближенное решение. Оценка точности будет зависеть от свойств системы, обнаруживаемых по ходу процесса, и от некоторой информации, дополнительно привлекаемой по мере необходимости.

Описываемая совокупность действий легко реализуется на ЭВМ и, по-видимому, является оптимальной как по объему вычислительной работы, так и по использованию памяти ЭВМ. В целом процесс состоит из следующих этапов:

— предварительное преобразование системы к простому виду с помощью унитарных преобразований;
— решение системы простого вида;
— оценка числа обусловленности и точности решения;
— уточнение решения;
— вычисление устойчивых проекций нормального псевдорешения;
— вычисление регуляризованных решений;
— общий анализ задачи.

Конечно, не все этапы обязательно присутствуют при решении каждой системы и не все этапы четко различаются друг от друга.

27.1. Матрица G называется *правой* (*левой*) *двухдиагональной*, если для ее элементов g_{ij} выполняются соотношения

$$g_{ij}=0,\quad j<i,\quad i<j-1 \qquad (j>i,\quad i>j+1).$$

Двухдиагональные матрицы обладают всеми свойствами треугольных матриц. Дополнительно для квадратных двухдиагональных матриц отметим, что теперь можно указать простые формулы для элементов обратной матрицы.

27.2. Пусть, например, G — квадратная правая двухдиагональная матрица и $g_{ij}^{(-1)}$ — элементы G^{-1}; тогда

$$g_{ij}^{(-1)}=\prod_{k=i}^{j-1}(-a_{k,k+1})\Big/\prod_{k=i}^{j}a_{kk}.$$

27.3. Пусть A — прямоугольная матрица размера $m\times n$, $m\geqslant n$. Умножим ее слева на последовательность матриц вращения T_{12}, T_{13}, ..., T_{1m}, исключая при этом внедиагональные элементы первого столбца. Умножим, далее, полученную матрицу справа на последовательность матриц вращения T_{32}, T_{42}, ..., T_{n2}, исключая внедиагональные элементы первой строки, лежащие правее элемента в позиции (1.2). Будем умножать поочередно на группы матриц вращения слева и справа и исключать элементы в таком порядке:

$$\begin{array}{r} (2.1),\ (3.1),\ (4.1),\ \ldots,\ (m.1), \\ (1.3),\ (1.4),\ \ldots,\ (1.n), \\ (3.2),\ (4.2),\ \ldots,\ (m.2), \\ (2.4),\ \ldots,\ (2.n), \\ \cdot \\ \cdot \\ \cdot \end{array}$$

Все шаги этого процесса могут быть реализованы, и после их выполнения исходная матрица A будет преобразована в правую двухдиагональную матрицу G.

27.4. Процесс 27.3 определяет соотношение $G = LAS$, где L и S — произведения соответствующих матриц вращения, участвующих в процессе.

27.5. Пусть $\tilde{T}_{ij}$, $\tilde{G}$ — реально вычисленные матрицы и $\tilde{L}$, $\tilde{S}$ — точные произведения соответствующих реально вычисленных матриц $\tilde{T}_{ij}$. Если $\tilde{G} = \tilde{L}(A + M)\tilde{S}$, то для эквивалентного возмущения M справедлива оценка

$$\|M\|_E \lessapprox 2{,}9 \cdot (m + n)\, p^{-t+1}\, \|A\|_E.$$

27.6. Если A — прямоугольная матрица размера $m \times n$, $m < n$, то, применяя процесс 27.3 к матрице A', можно преобразовать матрицу A в левую двухдиагональную матрицу. При этом оценка 27.5 снова имеет место.

27.7. Пусть в зависимости от выполнения условия $m \geqslant n$ или $m < n$ матрица G имеет блочный вид

$$G = \begin{bmatrix} \breve{G} \\ 0 \end{bmatrix} \quad \text{или} \quad G = [\hat{G} \ \ 0],$$

где матрицы $\breve{G}$, $\hat{G}$ квадратные и двухдиагональные; тогда

$$G^+ = [\breve{G}^+ \ \ 0] \quad \text{или} \quad G^+ = \begin{bmatrix} \hat{G}^+ \\ 0 \end{bmatrix}.$$

27.8. Если известно разложение 27.4, то нормальное псевдорешение системы $Ax = b$ равно $x_0 = SG^+Lb$ и может быть вычислено в соответствии с предписанием:

$$l = Lb, \quad u_0 = G^+l, \quad x_0 = Su_0.$$

Итак, при наличии разложения 27.4 решение системы $Ax = b$ с любой правой частью b, в том числе определение ее нормального псевдорешения, сводится к решению аналогичной задачи с квадратной двухдиагональной матрицей. Поэтому в дальнейшем мы будем заниматься в основном только такими задачами. Не ограничивая общности, можно считать, что двухдиагональная матрица является правой.

Решение систем с двухдиагональными матрицами можно осуществлять многими способами. Однако, по-видимому, проще всего попытаться сначала применить процесс 25.3 обратной подстановки с нормировкой. Этот процесс практически не вносит сколько-нибудь заметного вклада в общие затраты, но дает полезную дополнительную информацию.

27.9. Пусть $\tilde{G}$ — реально вычисленная матрица разложения 27.4, $\tilde{l}$ — реально вычисленный вектор из 27.8. Предположим, что система $\tilde{G}u = \tilde{l}$ с двухдиагональной матрицей $\tilde{G}$ решается с помощью обратной подстановки с нормировкой 25.3. Если реально вычисленный нормирующий множитель $\tilde{\alpha}$ равен нулю, то в пределах возмущений 27.5 матрица A меняет ранг, и исходную систему следует рассматривать как неустойчивую.

27.10. Если в условиях 27.9 реально вычисленный нормирующий множитель $\tilde{\alpha}$ не равен нулю, то реально вычисленное нормированное решение $\tilde{u}$ удовлетворяет возмущенной системе $(\tilde{G} + \Delta)\tilde{u} = \tilde{\alpha}(\tilde{l} + \varepsilon)$, где

$$\|\Delta\|_E \lessapprox p^{-t+1}\|\tilde{G}\|_E, \quad \|\varepsilon\|_E \lessapprox p^{-t+1}\|\tilde{l}\|_E.$$

Если $\tilde{\alpha} \neq 0$, то отсюда не обязательно следует, что матрица исходной системы является матрицей полного ранга в пределах допустимых возмущений. Этот факт в общем случае говорит только о том, что $\tilde{\alpha}^{-1}u$ есть какое-то решение слабо возмущенной системы. Возможность находить такие решения в какой-то мере оказывается полезной, и об этом мы будем говорить дальше. Для проверки матрицы на полноту ранга необходимо оценить число обусловленности. Если $\tilde{\alpha} = 0$, то эту оценку заведомо не надо делать.

27.11. Если имеет место разложение 27.4, то для евклидовой и спектральной норм выполняются равенства

$$\|A\| = \|G\|, \quad \|A^+\| = \|G^+\|, \quad \mathrm{cond}^+A = \mathrm{cond}^+G.$$

Евклидова норма матрицы G вычисляется без особого труда. Для вычисления евклидовой нормы матрицы G^{-1} можно воспользоваться формулами 27.2. Однако для оценки $\|G^{-1}\|_E$ оказывается удобным применение процесса 25.3.

27.12. Пусть задана произвольная невырожденная треугольная матрица R. Рассмотрим множество треугольных матриц, у которых модули всех элементов совпадают с модулями соответствующих элементов матрицы R. Максимальное значение евклидовой нормы обратной матрицы на данном множестве достигается на матрице $\hat{R}$ с положительными диагональными и отрицательными внедиагональными элементами.

27.13. Пусть заданы согласно 27.12 матрицы R и $\hat{R}$ порядка n. Возьмем вектор d, все координаты которого равны единице, и решим две системы:

$$Rz = d, \quad \hat{R}v = d.$$

Имеют место неравенства

$$\|z\|_E/\sqrt{n} \leqslant \|R^{-1}\|_E \leqslant \|v\|_1.$$

27.14. Если R — двухдиагональная матрица, то в обозначениях 27.12 всегда $\|R^{-1}\|_E = \|\widehat{R}^{-1}\|_E$.

27.15. Пусть R — двухдиагональная матрица порядка n с диагональными элементами, представимыми в ЭВМ ненулевыми числами. Решим в соответствии с обозначениями 27.12, 27.13 систему $\widehat{R}v = d$ с помощью процесса 25.3 обратной подстановки с нормировкой. Если $\widetilde{\alpha}$, $\widetilde{v}$ — реально вычисленные нормирующий множитель и нормированное решение, причем $\widetilde{\alpha} \neq 0$, то

$$\frac{\|\widetilde{v}\|_E}{\widetilde{\alpha} n^{1/2}}\left(1 - n^2 p^{-t+1}\right) \leqslant \|R^{-1}\|_E \leqslant \frac{\|\widetilde{v}\|_1}{\widetilde{\alpha}}\left(1 + n^2 p^{-t+1}\right).$$

Полученные оценки евклидовой нормы матрицы, обратной к двухдиагональной, исключительно точны, так как правая часть отличается от левой не более чем в n раз. Аналогичные оценки имеют место и в случае треугольной матрицы. При выводе оценок 27.15 учитывается тонкая структура ошибок процесса, а не только мажорантные оценки норм эквивалентных возмущений. Заметим, что для вычисления границ требуется выполнить не более $9n$ операций. Используя оценки 27.15, легко получить хорошие оценки числа обусловленности матрицы. Теперь снова вернемся к решению исходной системы линейных алгебраических уравнений.

27.16. Если для реально вычисленной матрицы G разложения 27.4 выполняется неравенство

$$2{,}9\beta(m+n)p^{-t+1}\|\widetilde{G}\|_E\|\widetilde{G}^+\|_E < 1,$$

то в пределах β-кратного возмущения 27.5 матрица A является матрицей полного ранга.

27.17. Пусть неравенство 27.16 выполняется при $\beta = 3$. Предположим, что в соответствии с 27.5—27.8 нормальное псевдорешение системы $Ax = b$ вычисляется в следующем порядке:

$$\widetilde{l} = \mathrm{fl}\,(\widetilde{L}b), \quad \widetilde{u}_0 = \mathrm{fl}\,(G^+\widetilde{l}), \quad \widetilde{x}_0 = \mathrm{fl}\,(\widetilde{S}u_0),$$

причем нормальное псевдорешение $\widetilde{u}_0$ находится с помощью обратной подстановки. Тогда ошибка реально вычисленного нормального псевдорешения $\widetilde{x}_0$ имеет оценки

$$\frac{\|\widetilde{x}_0 - x_0\|_E}{\|x_0\|_E} \lesssim$$

$$\lesssim \begin{cases} 5{,}8\cdot(m+n)\,p^{-t+1}\,\mathrm{cond}^+\,\widetilde{G}, & m = n,\\ 9{,}8\cdot(m+n)\,p^{-t+1}\,\mathrm{cond}^+\,\widetilde{G}, & m < n,\\ 9{,}8\cdot(m+n)\,p^{-t+1}\,\mathrm{cond}^+\,\widetilde{G} + 4{,}9\cdot(m+n)\,p^{-t+1}\left(\mathrm{cond}^+\,\widetilde{G}\right)^2 + \\ \qquad + 4{,}9\cdot(m+n)\,p^{-t+1}\dfrac{\|\widetilde{l}''\|_E}{\|\widetilde{l}'\|_E}\,\mathrm{cond}^+\,\widetilde{G}, & m > n. \end{cases}$$

В последней оценке $\widetilde{l}'$ есть вектор, содержащий первые n координат вектора $\widetilde{l}$, $\widetilde{l}''$ есть вектор, содержащий последние $m - n$

координат $\tilde{l}$. Число обусловленности матрицы $\tilde{G}$ выражено в евклидовой норме.

Если известно, что входные данные системы заданы точно, но оценки 27.17 не гарантируют нужный уровень погрешности, то нормальное псевдорешение можно уточнить. Уточнение осуществляется аналогично тому, как это было описано для систем с квадратными матрицами. На особенностях итерационного процесса для систем с прямоугольными матрицами мы останавливаться не будем. Если входные данные системы заданы с ошибками, то оценками 27.17 можно воспользоваться для оценки полной погрешности нормального псевдорешения. Для этого достаточно в оценках 27.17 к каждому множителю, стоящему перед числом обусловленности или его квадратом, добавить сумму относительных ошибок в матрице и правой части, выраженных в евклидовой норме.

27.18. Если матрица исходной системы не является патологически плохо обусловленной, т. е. ее число обусловленности не есть величина порядка p^t, то с помощью унитарного преобразования к двухдиагональному виду такую систему можно не только решить, но и достоверно оценить погрешность ее решения, а в случае необходимости и уточнить найденное решение. При этом достоверно устанавливается, что матрица системы в пределах допустимых возмущений является матрицей полного ранга.

Вообще говоря, все эти задачи можно решить, не прибегая к преобразованию матрицы к двухдиагональному виду, а воспользоваться, например, обычным или нормализованным методом вращений или отражений. Хотя при этом и потребуется выполнить примерно вдвое меньше вычислительной работы, трудно будет получить хорошие оценки числа обусловленности. К тому же, если в процессе решения системы выяснится ее неустойчивость, то в этом случае имеющееся разложение матрицы на множители оказывается почти бесполезным для дальнейшего анализа задачи. Разложение же 27.4 можно эффективно использовать и при решении неустойчивых систем.

Предположим, что каким-либо образом установлено, что матрица A системы $Ax = b$ является матрицей неполного ранга или очень близка к такой матрице. Будем считать также, что известно разложение 27.4. Любое суждение о решении системы теперь невозможно без привлечения дополнительной информации об исходной задаче. Простейшее предположение может заключаться в априорном знании того, что система $Ax = b$ совместна, матрица A имеет неполный ранг и ее минимальное ненулевое сингулярное число велико по сравнению с p^{-t}.

Снова для определения нормального псевдорешения системы $Ax = b$ нужно находить нормальное псевдорешение системы $\tilde{G}u = \tilde{l}$ с двухдиагональной матрицей, и опять полезным оказывается процесс 25.3 обратной подстановки с нормировкой. Не ограничивая общности, будем считать, что диагональные элементы матрицы $\tilde{G}$ отличны от нуля. В противном случае сделаем их такими, добавив к ним величины порядка p^{-t+1}. Это возмущение не приведет к потере информации, так как такие возмущения уже могут быть в этих элементах. Кроме того, предположим, что при реализации процесса 25.3 всегда $\tilde{\alpha} \neq 0$. Этого всегда можно добиться, если хранить $\tilde{\alpha}$ не как одно число, а как произведение ненулевых чисел.

27.19. Пусть матрица A совместной системы $Ax = b$ имеет неполный ранг и ее минимальное ненулевое сингулярное число равно ρ. Предположим, что $\tilde{x}$ есть решение слабо возмущенной системы и нормы относительных возмущений в матрице и правой части суть величины порядка ε. В этом случае вектор $\tilde{x}$ может быть представлен в виде суммы $\tilde{x} = x_0 + z + z(\varepsilon, \rho)$, где x_0 есть нормальное решение исходной системы, вектор z принадлежит ядру матрицы A, а вектор $z(\varepsilon, \rho)$ имеет длину порядка ε/ρ.

27.20. Пусть в условиях и обозначениях 27.19 известны два решения $\tilde{x}_1$, $\tilde{x}_2$ двух различных слабо возмущенных систем. С точностью до вектора длиной порядка ε/ρ вектор $\tilde{x}_1 - \tilde{x}_2$ принадлежит ядру матрицы A.

27.21. Пусть размерность ядра матрицы A равна r. Предположим, что выполнены условия 27.19 и найдены такие решения $\tilde{x}_0$, $\tilde{x}_1$, ..., $\tilde{x}_r$ слабо возмущенных систем, что векторы $z_i = \tilde{x}_i - \tilde{x}_0$, $1 \leqslant i \leqslant r$, с точностью до векторов длиной порядка ε/ρ являются линейно независимыми. Тогда перпендикуляр, опущенный из любого вектора $\tilde{x}_i$ на линейную оболочку векторов z_1, ..., z_r, с точностью до вектора длиной порядка ε/ρ совпадает с нормальным решением x_0 системы $Ax = b$.

Последнее утверждение лежит в основе следующего способа определения нормального псевдорешения при сформулированных ранее предположениях относительно исходной системы. Пусть система приведена к двухдиагональному виду. Будем возмущать случайным образом элементы обеих диагоналей и правой части на величины порядка p^{-t} и решать получающиеся системы с помощью процесса обратной подстановки с нормировкой. Построим согласно 27.21 разности решений и ортонормируем эту последовательность с помощью процесса ортогонализации Грама — Шмидта. Устойчивое появление малых векторов при ортогонализации говорит о том, что построенные ортонормированные векторы образуют базис ядра. Перпендикуляр на ядро при наличии ортонормированного базиса определяется довольно просто.

Описанный процесс реализуется очень быстро, так как приходится решать только двухдиагональные системы. Однако при этом требуется дополнительная память для хранения разностей решений слабо возмущенных систем. Можно обойтись хранением лишь 3—4 дополнительных векторов и даже меньше. Действительно, пусть $\tilde{x}_0$ — некоторое решение слабо возмущенной системы. Если получено другое решение $\tilde{x}_i$, то образуем разность $\tilde{x}_i - \tilde{x}_0$, заменим $\tilde{x}_0$ перпендикуляром, опущенным из $\tilde{x}_0$ на $x_i - \tilde{x}_0$, и повторим процесс. Вычисления прекращаются, как только вектор $\tilde{x}_0$ стабилизируется с нужной точностью.

Процесс особенно эффективен, когда размерность ядра равна 1. В этом случае два решения слабо возмущенных систем позволяют с высокой точностью определить собственный вектор матрицы A, соответствующий нулевому собственному значению. Решения слабо возмущенных систем можно получать и другими способами, например, решая одну и ту же систему принципиально различными численными методами.

Если относительно исходной системы известно, что ее матрица имеет «оторванную» группу малых сингулярных чисел, то мы опять находимся в условиях применения описанных процессов. Теперь с ε связывается величина малых сингулярных чисел, с величиной ρ — значение минимального из остальных сингулярных чисел. Конечно, вместо нормального псевдорешения можно определить только его проекцию на правое сингулярное пространство, соответствующее большим сингулярным числам.

Информация о существовании разделения сингулярных чисел матрицы системы на большие и малые настолько важна, что, если она отсутствует, целесообразно попытаться выяснить, имеет ли место такая ситуация.

27.22. Пусть G — правая двухдиагональная матрица порядка n. Построим последовательность двухдиагональных матриц $G_0 = G$, G_1, G_2, ..., где матрицы G_k с четными индексами k — правые, с нечетными — левые. Если k — четное, то

$$G_{k+1} = G_k T_{21}^{(k)} T_{32}^{(k)} \ldots T_{n,n-1}^{(k)};$$

при нечетном k

$$G_{k+1} = T_{n-1,n}^{(k)} \ldots T_{23}^{(k)} T_{12}^{(k)} G_k;$$

в обоих случаях матрицы $T_{ij}^{(k)}$ находятся из условия исключения элемента G_k в позиции (j, i). Для любой матрицы G последовательность G_k сходится к диагональной матрице, составленной из сингулярных чисел матрицы G.

27.23. Для процесса 27.22 исключения элементов в матрицах G_k выполняются следующие свойства:

— если все элементы обеих диагоналей матрицы G_0 отличны от нуля, то будут отличны от нуля все элементы обеих диагоналей у каждой из матриц G_k;

— если среди элементов обеих диагоналей матрицы G_0 имеются нулевые, то матрица G_2 будет квазидиагональной, клетки которой либо диагональные, либо двухдиагональные с ненулевыми элементами на обеих диагоналях; матрицы G_k имеют аналогичное с G_2 строение при всех $k \geqslant 2$.

Эти свойства позволяют без уменьшения общности считать, что все элементы на обеих диагоналях матриц G_k являются ненулевыми.

27.24. Пусть G — двухдиагональная матрица порядка n с ненулевыми элементами на обеих диагоналях. Обозначим диагональные элементы матрицы G_k через $\rho_1^{(k)}, \ldots, \rho_n^{(k)}$, внедиагональные — через $\varepsilon_1^{(k)}, \ldots, \varepsilon_{n-1}^{(k)}$. Предположим, что сингулярные числа ρ_i матрицы G упорядочены в порядке невозрастания, т. е. $\rho_1 \geqslant \rho_2 \geqslant \ldots \geqslant \rho_n$. Имеют место соотношения для всех i:

$$\lim_{k\to\infty} \rho_i^{(k)} = \rho_i, \quad \rho_i^{(k)} = \rho_i + O\left(\left(\frac{\rho_i}{\rho_{i-1}}\right)^k\right) + O\left(\left(\frac{\rho_{i+1}}{\rho_i}\right)^k\right),$$

$$\lim_{k\to\infty} \varepsilon_i^{(k)} = 0, \quad \varepsilon_i^{(k)} = O\left(\left(\frac{\rho_{i+1}}{\rho_i}\right)^k\right).$$

27.25. Реально вычисленная матрица $\tilde{G}_k$ унитарно эквивалентна матрице $G + \mathscr{E}_k$, при этом

$$\|\mathscr{E}_k\|_E \lesssim \frac{4\sqrt{2}+5}{2} k p^{-t+1} \|G\|_E.$$

27.26. Матрицы G_0, G_1, ..., построенные согласно 27.22, удовлетворяют соотношениям

$$G_0^* G_0 = G_1^* G_1, \quad G_1 G_1^* = G_2 G_2^*, \ \ldots,$$

причем произведения матриц по обеим сторонам всех равенств являются эрмитовыми трехдиагональными положительно определенными матрицами.

Если нет информации о разделении сингулярных чисел матрицы A, то проведем несколько шагов процесса 27.22 с матрицей $\widehat{G}$ в качестве начальной. Предположим, что матрица точной системы имеет «оторванную» группу из r малых сингулярных чисел. Тогда после выполнения небольшого числа шагов последние r строк и столбцов матрицы G_k становятся малыми. Обнаружив этот факт, мы сможем определить как величину малых сингулярных чисел, так и их число. После этого в соответствии с вышеизложенными рекомендациями можно определить главную проекцию нормального псевдорешения.

Если расщепление матрицы G_k на «большую» и «малую» не наступило после 8—10 шагов, то, скорее всего, это говорит о том, что сингулярные числа матрицы не имеют достаточного разделения. В этом случае приходится прибегать к построению регуляризованных решений при различных значениях регуляризирующего параметра. Опять наличие разложения 27.4 позволяет эффективно его использовать.

27.27. При наличии разложения 27.4 для всех α уравнения

$$(A^*A+\alpha E)x_\alpha = A^*b, \quad (G^*G+\alpha E)y_\alpha = G^*Lb,$$

где $x_\alpha = Sy_\alpha$, эквивалентны; при этом

$$\|x_\alpha\|_E = \|y_\alpha\|_E, \quad \|Ax_\alpha - b\|_E = \|Gy_\alpha - Lb\|_E.$$

Заметим, что если при каждом значении параметра α для решения первого уравнения надо выполнить порядка n^3 операций, то для решения уравнения с обратным преобразованием неизвестных — только порядка n^2 операций. Если же нужно определить не само решение x_α, а его евклидову норму и норму невязки, то обратное преобразование неизвестных не нужно делать. В этой ситуации один шаг по α для второго уравнения осуществляется по порядку в n^2 раз быстрее, чем один шаг по α для первого уравнения.

При малых значениях α оба уравнения будут плохо обусловленными. Для решения второго уравнения целесообразно применять метод квадратных корней и обратную подстановку с нормировкой.

§ 28. Методы сопряженных направлений

Построение ортогональных, псевдоортогональных и т. п. систем векторов, особенно на основе использования степенных последовательностей, дает большие возможности для создания различных численных методов решения систем линейных алгебраических уравнений $Ax = b$.

Мы уже неоднократно касались различных аспектов этой задачи. Теперь опишем большую группу численных методов решения систем, которые получили общее название *методов сопряженных направлений*. Все они основаны на ортогонализации степенных последовательностей. Будем предполагать, что матрица системы невырожденная, а пространство унитарное.

28.1. Пусть задана невырожденная матрица A порядка n. Возьмем любые невырожденные матрицы B, C и построим какимнибудь способом CAB-псевдоортогональную систему векторов $s_1, \ldots, s_n$, т. е. для нее

$$(CABs_i, s_i) \neq 0, \quad (CABs_i, s_k) = 0, \quad k < i,$$

при всех i. Пусть задан некоторый вектор x_0. Для любого вектора x существует единственное представление вида

$$x = x_0 + B \sum_{j=1}^{n} a_j s_j,$$

где a_j — числовые коэффициенты.

28.2. Обозначим в соответствии с 28.1

$$x_i = x_0 + B \sum_{j=1}^{i} a_j s_j.$$

Тогда для векторов x_i и невязок $r_i = Ax_i - b$ системы $Ax = b$ имеют место следующие рекуррентные соотношения:

$$x_i = x_{i-1} + a_i B s_i, \quad r_i = r_{i-1} + a_i A B s_i.$$

28.3. В условиях и обозначениях 28.1, 28.2 справедливы равенства $(Cr_i, s_k) = 0$, $1 \leqslant k \leqslant i$.

28.4. Пусть система векторов $s_1, \ldots, s_n$ строится параллельно с системой $r_0, \ldots, r_{n-1}$ с помощью процесса ее CAB-псевдоортогонализации. Положим $s_1 = r_0$ и для всех i будем иметь

$$s_{i+1} = r_i + \sum_{k=1}^{i} \beta_{k,i+1} s_k.$$

В этом случае справедливы равенства $(Cr_i, r_k) = 0$, $k < i$.

28.5. Если в условиях 28.4 $(Cr_i, r_i) \neq 0$ для всех i, то последовательность векторов $r_0, \ldots, r_{n-1}$ является C-псевдоортогональной.

28.6. В n-мерном пространстве C-псевдоортогональная система не может содержать более n ненулевых векторов.

28.7. В условиях 28.1—28.5 на некотором шаге процесса вычисления векторов x_i одна из невязок r_i, $i \leqslant n$, станет нулевой и будет получено точное решение системы $Ax = b$.

28.8. Для коэффициентов a_i справедливы формулы

$$a_i = -\frac{(Cr_{i-1}, r_{i-1})}{(CABs_i, r_{i-1})} = -\frac{(Cr_{i-1}, s_i)}{(CABs_i, s_i)}.$$

В общем случае коэффициенты $\beta_{k,i+1}$ определяются как решение некоторой системы линейных алгебраических уравнений с треугольной матрицей. Эта система строится из условий CAB-ортогональности вектора s_{i+1} слева к векторам $s_1, \ldots, s_i$.

28.9. В условиях и обозначениях 28.2—28.5 последовательности $s_1, \ldots, s_n$ и $r_0, \ldots, r_{n-1}$ с точностью до нормировок векторов являются результатами выполнения процессов соответственно CAB- и C-псевдоортогонализации одной и той же степенной последовательности r_0, ABr_0, $(AB)^2 r_0, \ldots$

28.10. Начальный вектор x_0, обеспечивающий выполнение условий $(CABs_i, s_i) \neq 0$, $(Cr_i, r_i) \neq 0$ при всех i таких, что $r_i \neq 0$, называется *подходящим*. Подходящие начальные векторы, и только они, дают возможность найти решение системы.

28.11. Для того чтобы любой вектор x_0 был подходящим, необходимо и достаточно, чтобы квадратичные формы $(CABx, x)$ и (Cx, x) не имели ненулевых изотропных векторов.

28.12. Для того чтобы любой вектор x_0 был подходящим, достаточно, чтобы матрицы CAB и C были строго знакоопределенными.

28.13. Для того чтобы вектор x_0 был подходящим, необходимо и достаточно, чтобы в матрицах Грама степенной последовательности r_0, ABr_0, $(AB)^2r_0$, ..., составленных по отношению к билинейным формам $(CABx, y)$ и (Cx, y), все ненулевые ведущие миноры были первыми, а нулевые — последними.

28.14. При невырожденных матрицах A, B, C для обеих матриц Грама из 28.13 число ненулевых ведущих миноров одинаково и на единицу меньше степени минимального аннулирующего вектор r_0 многочлена для матрицы AB.

28.15. Если вектор x_0 является подходящим и минимальный аннулирующий вектор r_0 многочлен для матрицы AB имеет степень p, то решение системы будет найдено на p-м шаге, и не раньше.

Описанный процесс вполне пригоден для нахождения решения системы линейных алгебраических уравнений $Ax = b$. В качестве начального вектора x_0 может быть взят почти любой вектор, причем утверждение 28.15 указывает принцип его выбора с целью получения решения за меньшее число шагов. В этом процессе, по существу, не накладывается никаких ограничений на матрицу системы, кроме ее невырожденности. Тем не менее обратим внимание на следующее обстоятельство.

Если матрицы B, C никак не связаны с матрицей A, то почти для всех векторов x_0 все коэффициенты $\beta_{k,i+1}$ из 28.4 будут отличны от нуля. Это означает, что для реализации процесса необходимо хранить все векторы $s_1, \ldots, s_n$ и иметь к ним оперативный доступ. Однако утверждение 23.22 показывает, что, по-видимому, при некоторых соотношениях между матрицами A, B, C можно надеяться на более простой вид рекуррентного соотношения 28.4 для векторов s_{i+1}.

28.16. Пусть матрицы A, B, C — невырожденные и удовлетворяют уравнению

$$(CABC^{-1})^* = \sum_{l=0}^{t} \alpha_l (AB)^l$$

при некотором $t \geqslant 0$. Тогда для коэффициентов $\beta_{k,i+1}$ из 28.4 выполняются соотношения $\beta_{k,i+1} = 0$ при $k \leqslant i - t$ для любого подходящего начального вектора x_0.

28.17. Пусть для всех подходящих начальных векторов x_0 выполняются соотношения $\beta_{k,i+1} = 0$ при $k \leqslant i - t$, где $t \geqslant 0$, и хотя бы для одного начального приближения процесс 28.1—28.5 заканчивается точно на n-м шаге. Предположим, что матрицы A, B, C невырожденные и $n \geqslant 2t + 3$. Тогда матрицы A, B, C удовлетворяют уравнению 28.16.

Если матрицы A, B, C связаны соотношением 28.16, то процесс 28.1—28.5 становится проще в части вычисления векторов $s_1, \ldots, s_n$. Наиболее простым и содержательным он оказывается при $t = 1$.

28.18. Пусть решается система линейных алгебраических уравнений $Ax = b$. Предположим, что матрицы A, B, C невырожденные порядка n и связаны между собой соотношением

$$(CABC^{-1})^* = \alpha_0 E + \alpha_1 AB$$

для некоторых чисел α_0, α_1. Возьмем любой подходящий начальный вектор x_0 и будем осуществлять процесс по следующему предписанию:

$$s_1 = r_0 = Ax_0 - b, \quad s_{i+1} = r_i + b_i s_i,$$
$$r_i = r_{i-1} + a_i ABs_i, \quad x_i = x_{i-1} + a_i Bs_i,$$

где

$$a_i = -\frac{(Cr_{i-1}, r_{i-1})}{(CABs_i, r_{i-1})} = -\frac{(Cr_{i-1}, s_i)}{(CABs_i, s_i)} = -\frac{(Cr_{i-1}, r_{i-1})}{(CABs_i, s_i)},$$
$$b_i = -\frac{(CABr_i, s_i)}{(CABs_i, s_i)} = -\bar{\alpha}_1 \frac{(Cr_i, ABs_i)}{(CABs_i, s_i)}.$$

Тогда при некотором $i \leqslant n$ будем иметь $r_i = 0$, и, следовательно, x_i будет решением системы $Ax = b$. Вычислительные процессы этого типа называются *методами сопряженных направлений.*

28.19. Векторы s_i, r_i из 28.18 связаны трехчленными соотношениями. Именно, в обозначениях 28.18

$$s_{i+1} = a_i ABs_i + (1 + b_i) s_i - b_{i-1} s_{i-1}, \qquad i > 1,$$
$$r_{i+1} = a_{i+1} ABr_i + \left(1 + \frac{ba_{i+1}}{a_i}\right) r_i - \frac{b_i a_{i+1}}{a_i} r_{i-1}, \qquad i \geqslant 1.$$

28.20. Векторы x_i из 28.18 связаны между собой соотношением

$$x_{i+1} = x_{i-1} + \omega_{i+1}(\beta_i Br_i + x_i - x_{i-1}),$$

где ω_{i+1}, β_i — соответствующим образом выбранные числа.

28.21. Если в соотношении 28.18, связывающем матрицы A, B, C, матрица AB имеет хотя бы два различных собственных значения, то $|\alpha_1| = 1$.

28.22. Пусть матрица AB имеет простую структуру. Если в разложении вектора r_0 по собственным векторам матрицы AB ненулевые составляющие соответствуют m попарно различным собственным значениям, то невязка r_m в методе сопряженных направлений окажется равной нулю, а вектор x_m будет решением системы.

28.23. Матрица AB при переходе к базису из векторов $s_1; \ldots, s_n$ или $r_0, \ldots, r_{n-1}$ становится трехдиагональной.

28.24. Если матрицы CAB и C эрмитовы, то

$$b_i = (Cr_i, r_i)/(Cr_{i-1}, r_{i-1}).$$

28.25. Если матрицы CAB и C эрмитовы и положительно определенные, то $a_i < 0$, $b_i > 0$ для всех i.

Для последовательностей векторов, участвующих в реализации методов сопряженных направлений, выполняются многочисленные условия ортогональности в отношении билинейных форм (Cx, y) и $(CABx, y)$. Эти условия могут быть указаны видом матриц билинейных форм в различных базисах.

28.26. Матрица A называется *правой* (*левой*) *почти треугольной* или матрицей *Хессенберга*, если для ее элементов a_{ij} выполняются соотношения

$$a_{ij} = 0, \quad i > j+1 \qquad (j > i+1).$$

28.27. Матрица билинейной формы $(CABx, y)$ является:

— правой треугольной в базисе $s_1, \ldots, s_n$;

— правой почти треугольной в базисе $r_0, \ldots, r_{n-1}$;

— правой треугольной в базисах $s_1, \ldots, s_n$ и $r_0, \ldots, r_{n-1}$.

28.28. Если матрица CAB эрмитова, то матрица билинейной формы $(CABx, y)$ является:

— диагональной в базисе $s_1, \ldots, s_n$;

— трехдиагональной эрмитовой в базисе $r_0, \ldots, r_{n-1}$;

— диагональной в базисах $s_1, \ldots, s_n$ и $r_0, \ldots, r_{n-1}$.

28.29. Матрица билинейной формы (Cx, y) является:

— правой треугольной в базисе $r_0, \ldots, r_{n-1}$;

— правой треугольной в базисах $r_0, \ldots, r_{n-1}$ и $s_1, \ldots, s_n$;

— правой треугольной в базисах $ABs_1, \ldots, ABs_n$ и $s_1, \ldots, s_n$;

— правой почти треугольной в базисах $ABr_0, \ldots, ABr_{n-1}$ и $r_0, \ldots, r_{n-1}$.

28.30. Если матрица C эрмитова, то матрица билинейной формы (Cx, y) является:

— диагональной в базисе $r_0, \ldots, r_{n-1}$;

— диагональной в базисах $r_0, \ldots, r_{n-1}$ и $s_1, \ldots, s_n$;

— правой двухдиагональной в базисах $ABs_1, \ldots, ABs_n$ и $s_1, \ldots, s_n$;

— трехдиагональной в базисах $ABr_0, \ldots, ABr_{n-1}$ и $r_0, \ldots, r_{n-1}$.

До сих пор построение и анализ методов сопряженных направлений осуществлялись только на основе идей, связанных с использованием последовательностей, ортогональных в специальной метрике. Однако, как отмечалось в 15.18—15.26, имеется очень тесная связь между такими последовательностями и процессом минимизации квадратичного функционала.

28.31. Если матрица CAB эрмитова и положительно определенная, то на каждом шаге метода сопряженных направлений минимизируется функционал ошибок

$$\psi(x) = ((B^{-1})^*CAx, x) - 2\mathrm{Re}((B^{-1})^*Cb, x).$$

28.32. В условиях 28.31 вектор x_i из 28.18 при любом подходящем векторе x_0 дает минимум функционала $\psi(x)$ среди всех векторов вида $x = x_0 + Bs$, где s принадлежит линейной оболочке векторов $s_1, \ldots, s_i$ или, что то же самое, векторов $r_0, ABr_0, \ldots \ldots, (AB)^{i-1}r_0$.

28.33. В условиях 28.31 функционал $\psi(x)$ достигает глобального минимума на решении системы $Ax = b$.

Несмотря на то, что уже исследованы многие свойства методов сопряженных направлений, не было ничего сказано о том, насколько содержательным является класс матриц A, B, C, связанных соотношением 28.18.

28.34. Соотношению 28.18 удовлетворяют, например, невырожденные матрицы A, B, C с такими свойствами:

— матрицы A, B, C эрмитовы, при этом $B = C$;

— матрицы CAB, C эрмитовы;

— матрица C перестановочна с матрицей AB, матрица AB нормальная, и ее спектр лежит на прямой линии;

— матрицы $2(A + A^*)^{-1}$, B и C совпадают;

— матрицы $2(A - A^*)^{-1}$, B, C совпадают.

Существует очень много эквивалентных форм методов сопряженных направлений. В основе построения любого из них лежит либо минимизация функционала ошибок, либо ортогонализация степенной последовательности. Первый принцип характерен для методов, возникающих как средство решения задач математической физики, второй — для методов решения общих задач. В конкретных методах оба принципа нередко в значительной мере смешиваются. Однако можно показать, что любой метод сопряженных направлений, основанный на минимизации функционала ошибок, в математическом отношении эквивалентен ортогонализации степенной последовательности в подходящим образом выбранной метрике. В конечном счете любой из вариантов метода сопряженных направлений эквивалентен 28.18. Мы рассмотрим здесь только основные варианты.

28.35. (*Метод сопряженных градиентов.*) В этом методе матрица A эрмитова положительно определенная, $B = C = E$, условие 28.18 выполняется при $\alpha_0 = 0$, $\alpha_1 = 1$. Положительная определенность матриц $CAB = A$ и $C = E$ гарантирует отсутствие вырождений в вычислительном процессе. На каждом шаге минимизируется функционал ошибок с матрицей A. Вычислительная схема метода имеет вид

$$\begin{aligned} s_1 &= r_0, \\ r_i &= r_{i-1} + a_i A s_i, \\ s_{i+1} &= r_i + b_i s_i, \\ x_i &= x_{i-1} + a_i s_i, \end{aligned}$$

где

$$a_i = -\frac{(r_{i-1}, r_{i-1})}{(As_i, r_{i-1})} = -\frac{(r_{i-1}, s_i)}{(As_i, s_i)} < 0, \quad b_i = -\frac{(r_i, As_i)}{(As_i, s_i)} = \frac{(r_i, r_i)}{(r_{i-1}, r_{i-1})} > 0.$$

В методе сопряженных градиентов векторы r_i образуют ортогональную систему, s_i — A-ортогональную.

28.36. (*Метод AA^*-минимальных итераций.*) В этом методе матрица A — произвольная невырожденная, $B = A^*$, $C = E$, условие 28.18 выполняется при $\alpha_0 = 0$, $\alpha_1 = 1$. Положительная определенность матриц $CAB = AA^*$ и $C = E$ гарантирует отсутствие вырождений в вычислительном процессе. На каждом шаге метода минимизируется функционал ошибок с матрицей E, т. е. квадрат евклидовой нормы самой ошибки. Вычислительная схема метода имеет вид

$$
\begin{aligned}
u_1 &= A^* r_0,\\
r_i &= r_{i-1} + a_i A u_i,\\
u_{i+1} &= A^* r_i + b_i u_i,\\
x_i &= x_{i-1} + a_i u_i,
\end{aligned}
$$

где

$$
a_i = -\frac{(r_{i-1}, r_{i-1})}{(Au_i, r_{i-1})} = -\frac{(r_{i-1}, r_{i-1})}{(u_i, u_i)} < 0,
$$

$$
b_i = -\frac{(r_i, Au_i)}{(u_i, u_i)} = \frac{(r_i, r_i)}{(r_{i-1}, r_{i-1})} > 0.
$$

В методе AA^*-минимальных итераций векторы $A^* r_i$ и Au_i образуют ортогональные системы.

28.37. (*Метод A^*A-минимальных итераций.*) В этом методе матрица A — произвольная невырожденная, $B = A^*$, $C = AA^*$, условие 28.18 выполняется при $\alpha_0 = 0$, $\alpha_1 = 1$. Положительная определенность матриц $CAB = (AA^*)^2$ и $C = AA^*$ гарантирует отсутствие вырождений в вычислительном процессе. На каждом шаге метода минимизируется функционал ошибок с матрицей A^*A, т. е. квадрат евклидовой нормы вектора невязки. Вычислительная схема метода имеет вид

$$
\begin{aligned}
u_1 &= A^* r_0,\\
r_i &= r_{i-1} + a_i A u_i,\\
u_{i+1} &= A^* r_i + b_i u_i,\\
x_i &= x_{i-1} + a_i u_i,
\end{aligned}
$$

где

$$
a_i = -\frac{(A^* r_{i-1}, A^* r_{i-1})}{(Au_i, Au_i)} < 0, \quad b_i = \frac{(A^* r_i, A^* r_i)}{(A^* r_{i-1}, A^* r_{i-1})} > 0.
$$

В методе A^*A-минимальных итераций векторы $A^* r_i$ и Au_i образуют ортогональные системы.

28.38. (*Метод эрмитова разложения.*) В этом методе матрица A — произвольная невырожденная. Представим ее согласно 10.62 в виде суммы $A = M + N$, где $M = M^*$, $N = -N^*$. В случае невырожденности матрицы M или N положим соответственно $B = C = M^{-1}$ или $B = C = N^{-1}$. Условие 28.18 выполняется при $\alpha_0 = 2$, $\alpha_1 = -1$. Если матрица M или iN является положительно определенной, процесс протекает без вырождения. Для $B = C =$

$= M^{-1}$ вычислительная схема метода имеет вид

$$\begin{aligned} Mu_1 &= r_0, \\ r_i &= r_{i-1} + a_i A u_i, \\ Mv_i &= r_i, \\ u_{i+1} &= v_i + b_i u_i, \\ x_i &= x_{i-1} + a_i u_i, \end{aligned}$$

где

$$a_i = -\frac{(r_i, u_i)}{(Au_i, u_i)}, \quad b_i = \frac{(v_i, Au_i)}{(Au_i, u_i)}.$$

Если $B = C = N^{-1}$, то вычислительная схема и формулы для коэффициентов a_i, b_i остаются такими же, кроме замены, конечно, M на N. В случае вещественной матрицы A нельзя брать $B = C = N^{-1}$, так как $(Cr, r) \equiv 0$, и для данного набора матриц A, B, C не существует ни одного подходящего начального вектора. Однако можно взять $B = C = iN^{-1}$ и перейти к комплексным вычислениям.

28.39. (*Метод неполного разложения.*) В этом методе матрица A эрмитова и положительно определенная. Представим ее в виде $A = M + N$, где $M = M^*$, $N = N^*$. Если M невырожденная, то положим $B = C = M^{-1}$. Условие 28.18 выполняется при $\alpha_0 = 0$, $\alpha_1 = 1$. Если матрица M положительно определенная, процесс протекает без вырождения. На каждом шаге метода минимизируется функционал ошибок с матрицей A. Вычислительная схема метода остается такой же, как в случае метода эрмитова разложения.

Как мы уже отмечали, методы сопряженных направлений позволяют определять решение особенно быстро, если матрица AB имеет простую структуру и мало попарно различных собственных значений. На использовании этой особенности построены различные приемы ускорения процесса решения системы $Ax = b$, в основе которых лежит следующая идея.

Предположим, что матрицу A можно представить в виде суммы $A = M + N$, где матрица M определяет «главную» часть матрицы A и при этом возможно простое решение системы с матрицей M. Теперь вместо системы $Ax = b$ будем решать систему $(E + NM^{-1})y = b$, где $Mx = y$. Если в каком-либо разумном смысле матрица M близка к матрице A, то большинство собственных значений матрицы N и, следовательно, матрицы NM^{-1} будут близки к нулю или равны нулю. Применение методов сопряженных направлений к новому уравнению в данном случае приводит к быстрому нахождению решения. Именно эта идея лежит в основе создания метода неполного разложения, который во многих случаях оказывается более эффективным, чем метод сопряженных градиентов в классическом варианте. Все зависит от того, насколько удачно осуществлено разложение матрицы A.

Методы сопряженных направлений являются не единственными методами решения системы линейных алгебраических уравнений $Ax = b$, основанными на использовании билинейных форм. Огромные возможности для создания методов дает построение систем векторов, двойственных или псевдодвойственных по отношению к некоторой билинейной форме, связанной с матрицей системы.

28.40. Пусть задана невырожденная матрица A порядка n. Предположим, что для нее каким-либо способом построены A-псевдодвойственные с точностью до нормировки системы векторов $u_1, \ldots, u_n$ и $v_1, \ldots, v_n$, т. е.

$$(Au_i, v_i) \neq 0, \quad (Au_i, v_k) = 0, \quad k < i,$$

при всех $i \leqslant n$. Пусть задан произвольный вектор x_0. Для любого вектора x существует единственное представление вида

$$x = x_0 + \sum_{j=1}^{n} a_j u_j.$$

28.41. Обозначим в соответствии с 28.40

$$x_i = x_0 + \sum_{j=1}^{i} a_j u_j.$$

Тогда для векторов x_i и невязок $r_i = Ax_i - b$ системы $Ax = b$ имеют место следующие рекуррентные соотношения:

$$x_i = x_{i-1} + a_i u_i, \quad r_i = r_{i-1} + a_i A u_i.$$

28.42. В условиях и обозначениях 28.40, 28.41 справедливы равенства $(r_i, v_k) = 0$ при $k \leqslant i$, причем для всех i

$$a_i = -(r_{i-1}, v_i)/(Au_i, v_i).$$

28.43. В условиях 28.40 – 28.42 невязка r_n ортогональна n линейно независимым векторам $v_1, \ldots, v_n$. Поэтому одна из невязок r_i, $i \leqslant n$, станет нулевой и будет получено точное решение системы $Ax = b$. Вычислительные процессы типа 28.41 – 28.43 называются *методами двойственных направлений*.

Число различных двойственных методов бесконечно в полном смысле этого слова, так как существует бесконечное число различных A-псевдодвойственных пар систем векторов. Рассмотренные методы сопряженных направлений, очевидно, входят в эту группу. Для методов двойственных направлений не существует в общем случае какого-либо аналога утверждения 28.31 даже при положительно определенной матрице A. Поведение ошибок $e_k = x - x_k$ в этих методах характеризуется лишь слабыми, но все же полезными следующими свойствами.

28.44. Пусть P_k — матрица проектирования на линейную оболочку векторов $u_1, \ldots, u_k$ параллельно линейной оболочке векторов $u_{k+1}, \ldots, u_n$. Тогда $e_k = (E - P_k)e_0$.

28.45. В методах двойственных направлений последовательные ошибки связаны друг с другом соотношением $e_k = (E - S_k)e_{k-1}$, где $S_k = (Au_k, v_k)^{-1}(u_k v_k^* A)$.

28.46. Матрицы S_k из 28.45 удовлетворяют равенствам

$$S_k^2 = S_k, \quad S_i S_k = 0,$$

$$S_i(E - S_k)(E - S_{k-1}) \ldots (E - S_1) = 0, \quad i < k.$$

28.47. Матрицы проектирования P_k из 28.44 и матрицы S_k из 28.45 связаны между собой соотношением

$$P_k = (E - S_k)(E - S_{k-1}) \ldots (E - S_1).$$

Интересные результаты, связанные с A-псевдодвойственными системами, можно получить, рассматривая матричную трактовку описанных методов.

28.48. Обозначим через U (V) матрицу, столбцами которой являются векторы $u_1, \ldots, u_n$ $(v_1, \ldots, v_n)$. Тот факт, что эти векторы являются A-псевдодвойственными, т. е. удовлетворяют соотношениям 28.40, означает, что матрица $C = V^*AU$ является невырожденной левой треугольной.

28.49. В соответствии с 28.48 любая пара A-псевдодвойственных систем векторов определяет разложение матрицы A на множители $A = (V^{-1})^*CU^{-1}$, среди которых матрица C — невырожденная левая треугольная.

28.50. Если каким-либо способом получено разложение $A = QCR$ матрицы A на невырожденные множители, причем матрица C — левая треугольная, то это разложение определяет пару A-псевдодвойственных систем векторов. В обозначениях 28.48 имеем $U = R^{-1}$, $V = (Q^{-1})^*$.

Большинство рассмотренных ранее численных методов решения систем линейных алгебраических уравнений основано на получении разложения матрицы системы на множители, среди которых обязательно есть либо левый, либо правый треугольный множитель. Поэтому существует очень тесная связь между разложением матрицы на множители, решением систем прямыми методами и построением A- и A'-псевдодвойственных систем векторов. Заметим, что A'-псевдодвойственная система получается из A-псевдодвойственной системы, и наоборот, с помощью всего лишь перестановки векторов в обратном порядке.

28.51. Прямые методы решения системы линейных алгебраических уравнений $Ax = b$, основанные на разложении матрицы на множители, являются методами построения A-псевдодвойственных систем векторов, и наоборот. Несмотря на разнообразие конкретных форм самих методов, все они могут быть исследованы с общих позиций на основе утверждения 28.44.

Основная трудность, связанная с практической реализацией описанных методов, вызвана их значительной неустойчивостью к ошибкам округления результатов промежуточных вычислений.

В общем случае явление неустойчивости в методах двойственных направлений выражено слабее, чем в методах сопряженных направлений. Объясняется это, в основном, тем, что двойственные системы векторов строятся почти независимо и, что более важно, строятся на основе использования всех предварительно построенных векторов, а не только последних из них, как в методах сопряженных направлений. Этот вывод вполне согласуется с относительно хорошей устойчивостью всех прямых методов решения систем, основанных на предварительном разложении матрицы на множители.

Неустойчивость к ошибкам округления в методах сопряженных направлений, как правило, оказывается достаточно большой, и вызвана она теми же причинами, которые порождают неортогональность реально вычисленных векторов в процессе ортогонализации Грама — Шмидта. К сожалению, в этих методах нельзя ориентироваться на использование переортогонализации всей системы векторов, так как это приведет к потере основных преимуществ методов сопряженных направлений, касающихся скорости и объема памяти.

Несмотря на неустойчивость, методы сопряженных направлений исключительно активно применяются при решении самых различных задач. Особенно часто они применяются при решении очень больших систем с сильно разреженными матрицами. Такие системы практически невозможно решать прямыми методами, основанными на разложении матрицы на множители, так как при этом разрушается структура матрицы. Разрушение структуры разреженной матрицы приводит, в свою очередь, к резкому увеличению объема вычислений и требуемой памяти. В методах сопряженных направлений основной операцией является умножение матрицы на вектор. При этом, конечно, могут быть эффективно учтены структуры всех используемых в процессе вычислений матриц.

Исследованию устойчивости и влияния ошибок округления в методах сопряженных направлений уделяется очень много внимания. Тем не менее в настоящее время нельзя утверждать, что решение данной проблемы существенно продвинулось вперед. Однако придуманы различные приемы, позволяющие в той или иной мере уменьшить и даже почти совсем исключить влияние ошибок округления в этих методах. Опишем один из таких приемов.

28.52. Пусть решается система линейных алгебраических уравнений $Ax = b$ одним из методов сопряженных направлений. Выберем некоторое число $0 < \alpha \ll 1$ и предположим, что имеется вектор x_0, который берется в качестве приближения к решению системы. Для получения лучшего приближения поступаем следующим образом:

— вычисляем невязку $r_0 = b - Ax_0$ в режиме накопления;

— вычисляем нормированную невязку $\widehat{r}_0$;

— для системы $A\Delta = \widehat{r}_0$ с помощью выбранного метода проводим процесс до тех пор, пока норма невязки не станет меньше $\alpha\|\widehat{r}_0\|$;

— вычисляем поправку к x_0, умножая найденное приближение к решению системы $A\Delta = r_0$ на нормирующий множитель;

— прибавляем вычисленную поправку к x_0;

— повторяем процесс до тех пор, пока не получится решение с заданной точностью.

Вообще говоря, число α нельзя брать не очень малым, так как придется часто прибегать к использованию режима накопления. Нельзя его брать и очень малым, так как из-за влияния ошибок округления можно не достичь нужного уровня уменьшения невязки. Из опытных расчетов известно, что хорошие результаты дает число α в пределах $p^{-t/3} \leqslant \alpha \leqslant p^{-t/6}$. При этом общее время решения системы почти не зависит от α, а само решение можно найти с относительной ошибкой существенно меньшей, чем p^{-t}.

§ 29. Другие методы

Рассмотренными методами, конечно, не ограничивается все многообразие приемов, связанных с разложением матрицы на множители и с решением систем линейных алгебраических уравнений. Желание эффективнее решить задачу приводит к необходимости более полно учитывать ее специфику, что, в свою очередь, порождает новые модификации методов. Мы опишем здесь кратко основные идеи построения некоторых других методов, довольно часто применяемых на практике и относящихся в основном к решению систем и обращению матриц. Все эти методы обладают какими-нибудь достоинствами (и недостатками тоже) по сравнению с ранее рассмотренными.

29.1. Пусть матрица A порядка n системы $Ax = b$ — невырожденная и имеет ненулевые ведущие миноры. Разделим все коэффициенты первого уравнения на элемент, стоящий в позиции (1.1). Это дает эквивалентную систему $A_1x = b_1$, у которой матрица в позиции (1.1) имеет 1 и все уравнения, начиная со второго, остаются неизменными. Пусть для $k \geqslant 2$ мы построим эквивалентную систему $A_{k-1}x = b_{k-1}$, у которой матрица ведущего минора порядка $k-1$ является единичной и все уравнения, начиная с k-го, остаются неизменными. Переход от системы $A_{k-1}x = b_{k-1}$ к системе $A_kx = b_k$ состоит из следующих этапов:

— для $1 \leqslant s \leqslant k-1$ умножим s-е уравнение на элемент в позиции (k, s) и вычтем его из k-го уравнения; после выполнения этого этапа все поддиагональные элементы в k-й строке матрицы будут нулевыми;

— разделим k-е уравнение на диагональный элемент; после выполнения этого этапа сохраняются все нулевые элементы k-й строки матрицы и ее диагональный элемент равен 1;

— для $1 \leqslant s \leqslant k-1$ умножим k-е уравнение на элемент в позиции (s, k) и вычтем из s-го уравнения; после выполнения этого этапа матрица ведущего минора порядка k является единичной и все уравнения, начиная с $k+1$-го, остаются неизменными.

Вектор b_n является решением системы $Ax = b$. Описанный процесс его нахождения называется *методом оптимального исключения.*

Изложенный вариант метода исключения по своей математической структуре близок к методу Гаусса. Однако по сравнению с ним метод оптимального исключения позволяет более эффективно использовать память машины. Как видно из самого процесса, после реализации k-го шага последние $n-k$ уравнений исходной системы остаются без изменения. Поэтому нет никакой необходимости держать их в памяти ЭВМ до тех пор, пока они не потребуются. Будем вводить в память ЭВМ коэффициенты уравнений не все сразу, а последовательно, по одной строке перед каждым шагом. Тогда легко показать, что для решения системы n-го порядка достаточно иметь память порядка $n^2/4$ слов. Это означает, что при одной и той же памяти ЭВМ метод оптимального исключения позволяет решать системы вдвое большего порядка, чем метод Гаусса.

Интересно отметить, что для своей реализации метод оптимального исключения требует выполнения такого же числа операций, как и метод Гаусса. Почти одинаково ведут себя оба метода и с точки зрения устойчивости к ошибкам округления. В методе оптимального исключения достаточно просто организовать выбор ведущего элемента по строке. При этом общая схема метода сохраняется, меняется лишь порядок исключения неизвестных.

Пожалуй, единственным серьезным недостатком метода оптимального исключения по сравнению с методом Гаусса является то, что он не позволяет решать системы с многими правыми частями, так как нельзя запомнить все преобразования с матрицей системы в памяти объемом $n^2/4$. Если же для реализации метода отводится достаточно большой объем памяти, то этот недостаток устраняется, но пропадает и основное преимущество.

29.2. Пусть A — невырожденная матрица. Обозначим $A_0 = A$, $B_0 = E$ и построим последовательности матриц

$$A_k = L_k A_{k-1}, \quad B_k = L_k B_{k-1}, \qquad k \geqslant 1,$$

для некоторых невырожденных матриц L_k. Если $A_k = E$, то $B_k = A^{-1}$.

На формулах 29.2 построен один из самых эффективных методов вычисления обратной матрицы. Ясно, что последовательности матриц A_k и B_k преобразуются одинаково. Будем выбирать L_k так, чтобы в матрицах A_k появлялось все больше и больше нулевых элементов, и будем размещать на их месте ненулевые элементы матрицы B. Тогда на каждом шаге, начиная с матрицы A, придется преобразовывать некоторую «сводную» матрицу C_k, содержащую A_k и B_k. В конце процесса получим на месте матрицы A матрицу A^{-1}.

29.3. Пусть L_k — элементарная неунитарная матрица вида 22.53, в которой в качестве вектора b взят координатный вектор e_k. Тогда матрица L_k отличается от единичной не более чем элементами k-го столбца.

29.4. Для матрицы L_k из 29.3 матрица L_k^{-1} отличается от единичной не более чем элементами k-го столбца.

29.5. Пусть L_k — матрицы вида 29.3. Тогда в матрице $L_k \ldots L_1$ только первые k столбцов могут отличаться от столбцов единичной матрицы.

29.6. Пусть $l_{1k}, \ldots, l_{nk}$ и $l_{1k}^{(-1)}, \ldots, l_{nk}^{(-1)}$ — элементы k-х столбцов матриц L_k и L_k^{-1}. Если $l_{kk} \neq 0$, то

$$l_{ik}^{(-1)} = \begin{cases} 1/l_{kk}, & i = k, \\ -\,l_{ik}/l_{kk}, & i \neq k. \end{cases}$$

29.7. Пусть задан вектор z с координатами $z_1, \ldots, z_n$. Если $z_k \neq 0$, то существует матрица L_k такая, что $L_k z = e_k$; при этом

$$l_{ik} = \begin{cases} 1/z_k, & i = k, \\ -\,z_i/z_k, & i \neq k. \end{cases}$$

29.8. Пусть невырожденная матрица A порядка n имеет ненулевые ведущие миноры. Положим $C_0 = A$, и пусть построена матрица C_{k-1}, $k \geqslant 1$. Переход от C_{k-1} к C_k состоит из следующих этапов:

— по элементам k-го столбца матрицы C_{k-1} вычисляются согласно 29.7 элементы k-го столбца матрицы L_k;

— k-й столбец матрицы C_{k-1} заменяется столбцом единичной матрицы;

— полученная матрица умножается слева на матрицу L_k.

Матрица C_n есть A^{-1}. Описанный процесс ее нахождения называется *методом Жордана*.

Для реализации описанной схемы необходимо отличие от нуля ведущих миноров. Если этот факт не известен заранее, то целесообразно применить схему с каким-либо выбором ведущего элемента. Нужный элемент выбирается как достаточно большой из последних $n-k+1$ строк и столбцов матрицы C_{k-1} и с помощью перестановки ее строк и столбцов помещается на место k-го диагонального элемента. Чтобы получить матрицу A^{-1}, необходимо выполнить в C_n все обратные перестановки. Метод Жордана может применяться для разложения матрицы на множители и решения систем. В целом он обладает такой же степенью устойчивости, как и метод Гаусса, но в три раза более трудоемкий.

29.9. Пусть A — невырожденная матрица порядка n, U, V' — прямоугольные матрицы размеров $n\times m$. Тогда

$$(A+UV)^{-1}=A^{-1}-A^{-1}U(E+VA^{-1}U)^{-1}VA^{-1},$$

если существуют все выписанные обратные матрицы. Матрица в левой части равенства называется *модифицированной.*

29.10. Пусть A_{k-1} — произвольная невырожденная матрица. Рассмотрим окаймленную матрицу

$$A_k=\begin{bmatrix} A_{k-1} & u_k \\ v_k & \alpha_k \end{bmatrix},$$

где u_k — вектор-столбец, v_k — вектор-строка, α_k — число. Если матрица A_k невырожденная, то

$$A_k^{-1}=\begin{bmatrix} A_{k-1}^{-1}+\dfrac{A_{k-1}^{-1}u_kv_kA_{k-1}^{-1}}{\beta_k} & -\dfrac{A_{k-1}^{-1}u_k}{\beta_k} \\ -\dfrac{v_kA_{k-1}^{-1}}{\beta_k} & \dfrac{1}{\beta_k} \end{bmatrix},$$

где $\beta_k=\alpha_k-v_kA_{k-1}^{-1}u_k$.

29.11. Если матрица A_{k-1} в 29.10 эрмитова, $u_k=v_k^*$ и α_k — вещественное число, то матрица A_k также эрмитова.

29.12. Пусть невырожденная матрица A порядка n имеет ненулевые ведущие миноры. Обратная матрица ведущего минора первого порядка имеет единственный элемент, равный обратной величине элемента в позиции (1.1). Предположим, что для $k\geqslant 2$ вычислена обратная матрица A_{k-1} ведущего минора порядка $k-1$. Переход от $k-1$-го шага к k-му состоит в вычислении матрицы A_k в соответствии с 29.10. Матрица A_n есть A^{-1}. Описанный процесс ее нахождения называется *методом окаймления.*

В случае произвольной матрицы A метод окаймления обладает примерно такими же характеристиками, как и метод Жордана. Если же матрица A эрмитова, то он оказывается вдвое эффективнее по объему вычислений и требуемой памяти, так как можно ограничиться нахождением лишь половины всех элементов A^{-1}. Метод окаймления позволяет эффективно обращать треугольные матрицы.

29.13. Пусть A — невырожденная эрмитова матрица и для нее известно разложение $A=LDL^*$, где D — эрмитова, а L — левая треугольная с единичными диагональными элементами. Рассмотрим эрмитову окаймленную матрицу B порядка $n+m$:

$$B = \begin{bmatrix} A & U \\ U^* & Z \end{bmatrix}.$$

Имеет место блочное разложение

$$B = \begin{bmatrix} L & 0 \\ V^* & E \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} L^* & V \\ 0 & E \end{bmatrix},$$

где $V = (LD)^{-1}U$, $Q = Z - V^*DV$.

29.14. Пусть A — произвольная невырожденная эрмитова матрица. Если наибольший по модулю ее диагональный элемент не равен нулю, то одноименной перестановкой строк и столбцов переставляем его в позицию (1.1). Предположим, что он равен нулю. Тогда существуют ненулевые внедиагональные элементы и любой из них одноименной перестановкой строк и столбцов можно переставить в позицию (2.1). Следовательно, всегда можно добиться того, что матрица ведущего минора 1-го или 2-го порядка будет невырожденной. Пусть для матрицы A_k ведущего минора порядка $k \geqslant 1$ известно разложение $A_k = L_k D_k L_k^*$, где матрица D_k эрмитова блочно диагональная, с блоками 1-го или 2-го порядка. Переход к следующему шагу состоит в одноименной перестановке последних строк и столбцов с целью добиться невырожденности матрицы Q первого или второго порядка из 29.13 и вычисления согласно 29.13 компонент треугольного разложения окаймленной матрицы. Этот способ получения разложения эрмитовой матрицы на множители называется *блочным методом квадратного корня* и осуществляет разложение 11.20, 11.21.

Описанный метод требует для своей реализации примерно таких же затрат по памяти и числу арифметических операций, как обычный метод квадратного корня. Что же касается устойчивости, то в этом отношении он сравним с методом Гаусса с перестановками.

29.15. Пусть A — теплицева матрица с комплексными или вещественными элементами:

$$A = \begin{bmatrix} a_0 & a_{-1} & & a_{-n} \\ a_1 & a_0 & \dots & a_{-n+1} \\ \cdot & \cdot & \cdot & \cdot \\ a_n & a_{n-1} & \dots & a_0 \end{bmatrix}.$$

Обозначим через u_k, v_k соответственно первый и последний столбцы обратной матрицы A_k^{-1} ведущего минора порядка $k+1$. Пусть они имеют вид

$$u_k = \begin{bmatrix} \hat{u}_{0,k} \\ \hat{u}_{1,k} \\ \dots \\ \hat{u}_{kk} \end{bmatrix} \theta_k, \quad v_k = \begin{bmatrix} \hat{v}_{0,k} \\ \hat{v}_{1,k} \\ \dots \\ \hat{v}_{kk} \end{bmatrix} \nu_k.$$

Если так же представлены векторы u_{k-1}, v_{k-1}, то

$$u_k = \left\{ \begin{bmatrix} \hat{u}_{0,k-1} \\ \hat{u}_{1,k-1} \\ \cdots \\ \hat{u}_{k-1,k-1} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \hat{v}_{0,k-1} \\ \cdots \\ \hat{v}_{k-2,k-1} \\ \hat{v}_{k-1,k-1} \end{bmatrix} \tau_k \right\} \theta_k,$$

$$v_k = \left\{ \begin{bmatrix} \hat{u}_{0,k-1} \\ \hat{u}_{1,k-1} \\ \cdots \\ \hat{u}_{k-1,k-1} \\ 0 \end{bmatrix} \delta_k + \begin{bmatrix} 0 \\ \hat{v}_{0,k-1} \\ \cdots \\ \hat{v}_{k-2,k-1} \\ \hat{v}_{k-1,k-1} \end{bmatrix} \right\} \nu_k,$$

где

$$\tau_k = -\nu_{k-1}\varphi_k, \qquad \delta_k = -\theta_{k-1}\psi_k,$$

$$\theta_k = (1-\delta_k\tau_k)^{-1}\theta_{k-1}, \quad \nu_k = (1-\tau_k\delta_k)^{-1}\nu_{k-1},$$

$$\varphi_k = \sum_{t=k}^{1} a_t\hat{u}_{k-t,k-1}, \qquad \psi_k = \sum_{t=1}^{k} a_{-t}\hat{v}_{t-1,k-1}.$$

29.16. Пусть A — теплицева матрица порядка $n+1$ с ненулевыми ведущими минорами. В обозначениях 29.15 первый и последний столбцы матрицы A_0^{-1} совпадают и содержат один элемент a_0^{-1}. Пусть для $k \geqslant 1$ вычислены столбцы u_{k-1}, v_{k-1} матрицы A_{k-1}^{-1}. Тогда столбцы u_k, v_k матрицы A_k^{-1} определяются согласно 29.15. При $k=n$ будут вычислены первый и последний столбцы матрицы A^{-1} и, следовательно, будут найдены представления 17.17 обратной матрицы A^{-1}.

29.17. Если при реализации процесса 29.16 умножение векторов на θ_k, ν_k в 29.15 осуществить лишь в конце, то представления 17.17 обратной матрицы могут быть найдены за $4n^2$ операций умножения и сложения.

29.18. Пусть решается система линейных алгебраических уравнений $Ax=b$ с теплицевой матрицей A. Рассмотрим усеченные системы $A_kx_k=b_k$, где A_k есть матрица углового минора порядка $k+1$, b_k — вектор, содержащий первые $k+1$ координат вектора b. Если $x_{0,k}, \ldots, x_{k,k}$ — координаты вектора x_k, то в обозначениях 29.15 для всех k выполняется соотношение

$$\begin{bmatrix} x_{0,k} \\ \cdots \\ x_{k-1,k} \\ x_{k,k} \end{bmatrix} = \begin{bmatrix} x_{0,k-1} \\ \cdots \\ x_{k-1,k-1} \\ 0 \end{bmatrix} + \begin{bmatrix} \hat{u}_{0,k} \\ \cdots \\ \hat{u}_{k-1,k} \\ \hat{u}_{k,k} \end{bmatrix} (\nu_k\alpha_k),$$

$$\alpha_k = b_k - \sum_{l=1}^{k} a_l x_{k-l,k-1}.$$

29.19. Если на основе соотношений 29.18 решать систему с теплицевой матрицей порядка n, то с учетом 29.15 — 29.17 решение может быть найдено за $6n^2$ операций умножения и сложения.

В случае, когда решается одна система с теплицевой матрицей, алгоритм 29.18 является одним из самых эффективных. Если же приходится решать много систем с одной и той же матрицей, но с различными правыми частями, то целесообразно спачала получить представление 17.17 для обратной матрицы. Умножение матрицы 17.17 на вектор при больших порядках согласно 17.41 можно осуществлять с малыми затратами. Как будет показано в § 30, этот прием позволяет довольно быстро находить собственные значения и собственные векторы теплицевых матриц.

29.20. Если матрица A является блочно теплицевой, все блоки которой являются комплексными или вещественными матрицами, то алгоритм 29.15, 29.18 решения системы сохраняется. При этом в формулах 29.15 необходимо число 1 заменить на единичную матрицу E соответствующего порядка.

Для матриц, обратных к блочно теплицевым, в общем случае имеют место представления 17.6. Для того чтобы получить первую и последнюю блочные строки обратной матрицы, нужно применить алгоритм 29.15 к транспонированной матрице.

29.21. Пусть с теплицевой матрицей A решается регуляризованная система $(A^*A+\alpha E)x_\alpha = A^*b$. Рассмотрим систему

$$\begin{bmatrix} \alpha E & A^* \\ A & -E \end{bmatrix} \begin{bmatrix} z_\alpha \\ w_\alpha \end{bmatrix} = \begin{bmatrix} A^*b \\ 0 \end{bmatrix}$$

вдвое большего порядка. Всегда $x_\alpha = z_\alpha$.

29.22. Матрица второй системы 29.21 обладает следующими свойствами:

— с помощью одноименной перестановки строк и столбцов она может быть преобразована в блочно теплицеву матрицу с блоками второго порядка;

— при $\alpha > 0$ все ведущие блочные подматрицы преобразованной матрицы являются невырожденными;

— при малых α минимальные сингулярные числа преобразованной матрицы и матрицы регуляризованной системы асимптотически совпадают.

Пример теплицевых матриц показывает, что возможен формальный перенос алгоритмов с числовых матриц на блочные. Такой перенос правомочен далеко не всегда. Однако для многих методов, например, Гаусса, Жордана, квадратных корней, ортогонализации, окаймления, отражения и ряда других, блочные аналоги построены. Они с успехом применяются для решения задач большой размерности.

§ 30. Прямые и обратные итерации

При изложении методов решения проблемы собственных значений произвольной матрицы A порядка n мы будем предполагать, что ее собственные значения занумерованы в порядке неубывания модулей, т. е.

$$|\lambda_1| = \ldots = |\lambda_{r_1}| < |\lambda_{r_1+1}| = \ldots = |\lambda_{r_2}| < \ldots$$
$$\ldots < |\lambda_{r_{m-1}+1}| = \ldots = |\lambda_{r_m}|.$$

Модуль собственных значений s-й группы будем обозначать через τ_s.

Вычисление собственных значений и собственных векторов матрицы является одной из самых сложных задач линейной алгебры. Численные методы решения проблемы собственных значений должны быть итерационными по существу, так как в конечном счете они связаны с определением корней алгебраического многочлена.

В этих методах собственные значения вычисляются как пределы некоторых числовых последовательностей, без предварительного определения коэффициентов характеристического многочлена. Как правило, одновременно находятся и собственные векторы или другие векторы, связанные с ними простыми соотношениями. Если же собственные векторы по каким-либо причинам не находятся в процессе отыскания собственных значений, то они могут быть легко определены после вычисления собственных значений (с помощью описываемых ниже методов).

30.1. Пусть A — произвольная матрица. Возьмем любой вектор u_0 и построим последовательность векторов $u_k = \alpha_k A u_{k-1}$ для $k \geqslant 1$ при некоторых ненулевых числах α_k. Этот процесс называется *прямыми итерациями* вектора u_0.

30.2. Для прямых итераций выполняются соотношения

$$u_k = \beta_k A^k u_0, \quad \beta_k = \prod_{l=1}^{k} \alpha_k.$$

Если $A \neq 0$, то возможны два случая: либо $A^k \neq 0$ при всех k, либо $A^k = 0$, начиная с некоторого k. Второй случай означает, что матрица A нильпотентная, все ее собственные значения равны нулю, а все собственные векторы являются решениями однородной системы линейных алгебраических уравнений $Ax = 0$. В дальнейшем будем считать, что матрица A не является нильпотентной.

30.3. Представим матрицу A в виде $A = Q\Lambda Q^{-1}$, где Λ — каноническая матрица Жордана, диагональные элементы которой упорядочены по невозрастанию модулей, и пусть $D = \operatorname{diag}(\lambda_n, \ldots, \lambda_1)$ — соответствующая диагональная матрица собственных значений. Предположим, что максимальный порядок канонического ящика Жордана матрицы Λ равен s. Тогда при $k \geqslant s$ имеет место следующее представление:

$$u_k = \beta_k Q(\Lambda^k (D^k)^+) D^k Q^{-1} u_0.$$

30.4. Корневое подпространство L матрицы A, соответствующее максимальным по модулю собственным значениям, совпадает с линейной оболочкой первых $r_m - r_{m-1}$ вектор-столбцов матрицы Q из 30.3. Корневое подпространство R, соответствующее остальным собственным значениям, совпадает с линейной оболочкой остальных вектор-столбцов матрицы Q.

30.5. Пусть $u_k = u_k^{(L)} + u_k^{(R)}$, где $u_k^{(L)} \in L$, $u_k^{(R)} \in R$. Если $u_0^{(L)} \neq 0$, то $u_k^{(L)} \neq 0$ при всех k.

30.6. В условиях и обозначениях 30.5 выполняется соотношение

$$\frac{\left\| u_k^{(R)} \right\|_E}{\left\| u_k^{(L)} \right\|_E} = O\left(k^{s-1}\left(\frac{\tau_{m-1}}{\tau_m}\right)^k\right).$$

Таким образом, при больших значениях k вектор u_k оказывается сколь угодно близким к корневому подпространству матрицы A, соответствующему максимальным по модулю собственным значениям. Скорость этого приближения зависит как от степени оторванности старших собственных значений, так и от соотношения между $\left\| u_0^{(L)} \right\|_E$ и $\left\| u_0^{(R)} \right\|_E$. Предельное поведение вектора u_k определяется величиной соответствующих ящиков Жордана матрицы A и арифметическими свойствами совокупности аргументов максимальных по модулю собственных значений. Это поведение удобнее изучать на нормированных векторах.

30.7. Для любого вектора $u_k \neq 0$ существует единственный вектор $v_k \neq 0$ такой, что:

— евклидова норма вектора v_k равна единице;

— векторы v_k и u_k коллинеарны, т. е. $u_k = \gamma_k v_k$ для некоторого числа $\gamma_k \neq 0$;

— среди всех максимальных по модулю координат вектора v_k координата с минимальным номером положительна.

30.8. Пусть все максимальные по модулю собственные значения матрицы A совпадают и ни одно из них не входит в канонический ящик Жордана выше первого порядка. Тогда со скоростью 30.6 при $s = 1$:

— последовательность нормированных векторов v_k сходится к одному из собственных векторов, соответствующих максимальному по модулю собственному значению;

— последовательность отношений $\gamma_k \alpha_k^{-1}$ сходится к максимальному по модулю собственному значению.

Положение существенно меняется, если максимальные по модулю собственные значения совпадают, но входят в жордановы ящики выше первого порядка. Сходимость последовательностей из 30.8 снова будет иметь место, однако практического значения она уже не имеет, так как скорость сходимости оказывается всего лишь порядка k^{-1}. Конечно, предельное соотношение 30.6 сохраняется и в этом случае.

30.9. Пусть все максимальные по модулю собственные значения матрицы A различаются только знаками и ни одно из них не входит в канонический ящик Жордана выше первого порядка. Тогда со скоростью 30.6 при $s = 1$:

— последовательность нормированных векторов v_k с четными номерами сходится к вектору, равному сумме некоторых двух собственных векторов, соответствующих собственным значениям с противоположными знаками;

— последовательность нормированных векторов v_k с нечетными номерами сходится к вектору, равному разности тех же собственных векторов;

— последовательность отношений $\gamma_k(\alpha_k\alpha_{k-1}\gamma_{k-2})^{-1}$ для четных и нечетных k сходятся к квадрату максимального по модулю собственного значения.

30.10. Пусть матрица A имеет два различных, но равных по модулю максимальных собственных значения. Если отношение разности аргументов этих собственных значений к числу π есть иррациональное число, то почти для всех векторов v_0 не сходится никакая подпоследовательность векторов v_{k_i}, у которой совокупность индексов k_i образует арифметическую прогрессию.

30.11. Пусть в условиях 30.10 v есть любой нормированный согласно 30.7 вектор корневого подпространства, соответствующего максимальным по модулю собственным значениям. Для любого $\varepsilon > 0$ найдется такое k, что $\|v_k - v\| \leqslant \varepsilon$.

Ситуации, описанные в утверждениях 30.8, 30.9, наиболее просты с точки зрения организации вычислительного процесса. Они часто встречаются на практике, например в случае эрмитовой матрицы A. Однако в случае произвольной матрицы A не менее типичной является ситуация, описанная в утверждениях 30.10, 30.11. Сложность ее анализа заключается в том, что предельные точки последовательности векторов $v_k^{(L)}$ равномерно покрывают некоторую область корневого подпространства, соответствующего максимальным по модулю собственным значениям. Это существенно затрудняет выделение информации о спектральных компонентах матрицы A из последовательности векторов v_k.

В некоторых простых ситуациях, когда, например, вещественная матрица имеет изолированную пару максимальных по модулю комплексно сопряженных собственных значений, по векторам v_k можно определить соответствующие собственные значения и собственные векторы. В более сложных случаях, особенно при наличии жордановых клеток, их определение становится очень громоздким. Во всех ситуациях основой исследования последовательности векторов v_k является представление 30.3. Мы не будем более детально описывать простые итерации, так как определение любых собственных значений и собственных векторов, как правило, может быть сведено к случаям, аналогичным 30.8.

30.12. Рассмотрим последовательность чисел $\sigma_0, \sigma_1, \ldots$, сходящуюся к некоторому числу σ. Возьмем любой вектор u_0 и построим последовательность векторов $u_k = \alpha_k(A - \sigma_{k-1}E)u_{k-1}$ для $k \geqslant 1$ при каких-то ненулевых числах α_k. Этот процесс называется *прямыми итерациями* вектора u_0 *со сдвигами.*

Прямые итерации применяются в основном для определения корневого базиса, соответствующего максимальным по модулю собственным значениям. Используя сдвиги, можно несколько увеличить скорость сходимости. Значительного же ускорения нельзя получить, так как обычно невозможно с помощью сдвигов сделать достаточно малым отношение τ_{m-1}/τ_m в 30.6. Если же матрица A имеет группу равных по модулю, но различных собственных значений, то с помощью сдвигов можно добиться разделения их модулей.

Прямые итерации можно использовать и для определения корневых векторов, соответствующих минимальным по модулю собственным значениям, если матрицу A в 30.1 заменить матрицей A^{-1}. Реальный процесс осуществляется по несколько иной схеме.

30.13. Пусть A — произвольная матрица. Возьмем любой вектор u_0 и построим последовательность векторов u_k, удовлетворяющих равенствам $Au_k = \alpha_k u_{k-1}$ для $k \geqslant 1$ при некоторых ненулевых числах α_k. Этот процесс называется *обратными итерациями* вектора u_0.

30.14. Обозначим через S корневое подпространство, соответствующее минимальным по модулю собственным значениям матрицы A, через Q — корневое подпространство, соответствующее остальным собственным значениям. Тогда, аналогично 30.5, 30.6, имеем

$$\frac{\| u_k^{(Q)} \|_E}{\| u_k^{(S)} \|_E} = O\left(k^{s-1}\left(\frac{\tau_1}{\tau_2}\right)^k\right).$$

30.15. Рассмотрим последовательность чисел $\sigma_0, \sigma_1, \ldots$, сходящуюся к некоторому числу σ. Процесс вида $(A - \sigma_{k-1}E)u_k = \alpha_k u_{k-1}$, $k \geqslant 1$, называется *обратными итерациями* вектора u_0 *со сдвигами*.

С точки зрения скорости сходимости обратные итерации принципиально отличаются от прямых итераций. Теперь с помощью сдвигов отношение τ_1/τ_2 можно сделать сколь угодно малым и даже нулем, если взять сдвиг, равный одному из собственных значений матрицы A. К сожалению, за это достоинство приходится платить, так как на каждом шаге обратных итераций надо решать систему линейных алгебраических уравнений с матрицей A, а не умножать матрицу на вектор, как в прямых итерациях. Однако, как будет показано, в действительности эта плата не так высока.

30.16. Пусть λ — собственное значение матрицы A. Предположим, что λ не входит ни в одну жорданову клетку и расстояние между λ и другими не равными ему собственными значениями равно a. Пусть известно приближенное значение $\tilde{\lambda}$ и $|\tilde{\lambda} - \lambda| = \varepsilon < a$. Выберем вектор u_0 и рассмотрим обратные итерации $(A - \tilde{\lambda}E)u_k = \alpha_k u_{k-1}$. Тогда почти для всех векторов u_0 со скоростью $(\varepsilon/a)^k$:

— последовательность нормированных векторов v_k сходится к одному из собственных векторов, соответствующих собственному значению λ;

— последовательность отношений $\gamma_k \alpha_k^{-1}$ сходится к числу $(\lambda - \tilde{\lambda})^{-1}$.

Если ε мало по сравнению с a, то скорость сходимости обратных итераций оказывается исключительно большой. Такая ситуация возникает, когда собственное значение λ вычислено каким-то другим способом и необходимо уточнить его, а также определить принадлежащие ему корневые векторы. В этом случае обратные итерации являются одним из самых эффективных численных методов. Выбирая различным образом начальные векторы u_0, можно быстро определить весь корневой базис.

Реализация обратных итераций с постоянным сдвигом сводится к многократному решению систем линейных алгебраических уравнений с одной и той же матрицей, но с различными правыми частями. Как уже отмечалось раньше, решение таких систем осуществляется с небольшими временными затратами. Необходимо один раз разложить матрицу на множители и использовать данное разложение при решении каждой системы. Если сдвиг близок к одному из собственных значений, то во избежание переполнений при решении треугольных систем целесообразно использовать процесс 25.3 обратной подстановки с нормировкой.

На первый взгляд кажется, что влияние ошибок округления в реальных вычислениях должно существенно изменить свойства обратных итераций при наличии сдвигов, близких к собственным значениям. Опасения обычно связываются с тем, что в этом случае системы оказываются плохо обусловленными и, следовательно, решения содержат большие ошибки. Однако в целом правильные аргументы приводят здесь к неправильному выводу. Если решение системы $(A-\tilde{\lambda}E)u_k = \alpha_k u_{k-1}$ содержит большую ошибку, то вектор ошибок будет в основном принадлежать именно тому подпространству, которое мы пытаемся определить. Чем больше ошибка в вычислительном векторе, тем с большей точностью этот вектор принадлежит нужному подпространству. Ситуация почти уникальная в численных методах.

30.17. Пусть собственное значение λ из 30.16 определяется с помощью обратных итераций с переменным сдвигом. Предположим, что уже найден сдвиг σ_{k-1}, и выполним следующие действия:

— решим систему $(A-\sigma_{k-1}E)u_k = \alpha_k u_{k-1}$;

— вычислим нормированный вектор v_k;

— вычислим следующий сдвиг $\sigma_k = \sigma_{k-1} + \gamma_k^{-1}\alpha_k$.

Если σ_{k-1} достаточно близко к λ, то с квадратичной скоростью почти для всех векторов u_0:

— последовательность нормированных векторов v_k сходится к одному из собственных векторов, соответствующих собственному значению λ;

— последовательность σ_k сходится к λ.

При переменных сдвигах на каждом шаге процесса приходится решать системы с различными матрицами. В этом случае предварительное разложение матрицы на множители не дает никаких преимуществ. Если не предпринимать дополнительных мер, то выигрыш во времени, получаемый за счет квадратичной сходимости, может быть значительно меньше, чем проигрыш за счет решения систем. Один из путей устранения этого недостатка основан на предварительном преобразовании исходной матрицы к подобной ей матрице, имеющей более простой вид.

30.18. Матрица A называется *правой* (*левой*) *почти треугольной* или *правой* (*левой*) *матрицей Хессенберга*, если для ее элементов a_{ij} выполняются соотношения

$$a_{ij}=0, \quad i>j+1 \qquad (j>i+1).$$

30.19. Множество одноименных почти треугольных матриц одного порядка есть линейное пространство.

Почти треугольные матрицы широко используются при решении различных спектральных задач. Объясняется это тем, что любая матрица унитарно подобна почти треугольной, и это преобразование эффективно реализуется. К тому же, почти треугольная матрица значительно проще раскладывается на множители, чем матрица общего вида, и с матрицами этого типа легко решаются системы линейных алгебраических уравнений.

30.20. Пусть A — произвольная квадратная матрица порядка n. Умножим ее слева на матрицы вращения T_{23}, T_{24}, ..., T_{2n}, выбирая их так, чтобы последовательно исключить все элементы первого столбца, кроме верхних двух. Умножим, далее, полученную матрицу справа на вычисленные матрицы T^*_{23}, T^*_{24}, ..., T^*_{2n}, при этом нулевые элементы первого столбца остаются нулевыми. Подберем T_{34}, T_{35}, ..., T_{3n} так, чтобы после умножения слева на эти матрицы были исключены все поддиагональные элементы второго столбца, кроме верхнего, и умножим справа на матрицы T^*_{34}, T^*_{35}, ..., T^*_{3n}. Переходим к исключению элементов третьего столбца и т. д. Все шаги этого процесса могут быть реализованы, и после их выполнения исходная матрица A будет преобразована в правую треугольную матрицу A_N, унитарно подобную матрице A. Описанный процесс называется *методом вращений* для унитарно подобного преобразования матрицы к почти треугольному виду.

30.21. Процесс 30.20 определяет соотношение $A_N = R_N A R^*_N$, где $R_N = (\ldots T_{35}T_{34}\ldots T_{24}T_{23})$ есть произведение всех участвующих в процессе матриц вращения.

30.22. Если матрица A эрмитова, то матрица A_N эрмитова трехдиагональная.

В случае эрмитовой матрицы A в процессе 30.20 возникают некоторые особенности. Влияние ошибок округления приводит к тому, что матрица A_N, вообще говоря, не будет эрмитовой. Для восстановления эрмитовости обычно вычисляют половину элементов матрицы, а остальным приписывают принудительные значения.

Хотя такая процедура несколько изменяет распределение ошибок по сравнению с процессами, рассмотренными ранее, общий их уровень остается малым.

30.23. Пусть $\tilde{T}_{ij}$, $\tilde{A}_N$ — реально вычисленные матрицы, $\tilde{R}_N$ — точное произведение реально вычисленных матриц $\tilde{T}_{ij}$. Если $\tilde{A}_N = \tilde{R}_N (A + M) \tilde{R}^*_N$, то для эквивалентного возмущения M выполняются неравенства

$$\| M \|_E \leqslant \begin{cases} 4\sqrt{2}\, n p^{-t+1} \| A \|_E, & \text{если } A \text{ произвольная}, \\ 8 n p^{-t+1} \| A \|_E, & \text{если } A \text{ симметричная}. \end{cases}$$

Отметим, что процесс 30.20 осуществляется весьма эффективно. Вся информация о сомножителях может быть размещена на месте матрицы A. Что касается трудоемкости, то в случае эрмитовой матрицы она такая же, как при получении разложения матрицы на треугольный и унитарный множители методом вращений; в случае произвольной матрицы она вдвое больше. Конечно, все выводы остаются в силе, если исходная матрица приводится к левой почти треугольной матрице.

30.24. Пусть A — правая почти треугольная матрица. Умножим ее слева на матрицы вращения T_{12}, T_{23}, ..., $T_{n-1,n}$, выбирая их так, чтобы последовательно исключить все поддиагональные элементы. Полученная матрица A_N будет правой треугольной. Если $R_N = (T_{n-1,n} \dots T_{23}T_{12})$, то равенство $A_N = R_N A$ определяет разложение почти треугольной матрицы на ортогональный и треугольный множители.

30.25. Обозначим через $\tilde{T}_{ij}$, $\tilde{A}_N$ реально вычисленные матрицы, через $\tilde{R}_N$ — точное произведение реально вычисленных матриц $\tilde{T}_{ij}$. Если $\tilde{A}_N = \tilde{R}_N(A + M)$, то для эквивалентного возмущения M процесса 30.24 выполняется неравенство $\|M\|_E \leqslant \sqrt{2}np^{-t+1}\|A\|_E$.

Предварительное преобразование к унитарно подобной почти треугольной матрице, по существу, снимает проблему потери времени при использовании обратных итераций с переменными сдвигами. Теперь на каждом шаге придется решать системы с почти треугольными матрицами.

Особенно эффективно описанное преобразование в случае эрмитовой матрицы. В этом случае и сами обратные итерации приобретают исключительно важные дополнительные свойства. Именно, оказывается возможным разработать стратегию сдвигов, обеспечивающую сходимость с кубической скоростью почти для всех начальных векторов.

30.26. Для любой эрмитовой матрицы A, любого числа σ и любого вектора $u \neq 0$ найдется такое собственное значение λ матрицы A, что

$$|\lambda - \sigma| \leqslant \frac{\|Au - \sigma u\|_E}{\|u\|_E}.$$

30.27. Среди всех чисел σ отношение Релея, вычисленное для матрицы A и вектора u, минимизирует правую часть неравенства 30.26.

30.28. Если σ есть отношение Релея для матрицы A и вектора u, r есть вектор невязки $Au - \sigma u$, то пара σ, u является точным собственным значением и точным собственным вектором для матрицы $A + M$, где $M = \|u\|_E^{-2}(ur^* + ru^*)$.

30.29. Матрица M из 30.28 либо нулевая, либо обладает следующими свойствами:

— эрмитова;

— имеет ранг 2;

— ненулевые собственные значения равны $\pm\|r\|_E/\|u\|_E$;

— собственными являются векторы $\|r\|u \pm \|u\|r$ и любой вектор, ортогональный u и r.

30.30. Пусть обратные итерации $(A - \sigma_{k-1}E)u_k = \alpha_k u_{k-1}$ осуществляются со сдвигами σ_{k-1}, равными отношениям Релея векторов u_{k-1}. В этом случае последовательность нормированных согласно 30.7 векторов v_0, v_1, ... называется *последовательностью Релея*.

30.31. Обозначим $r_k = Av_k - \sigma_k v_k$. При любом векторе v_0 для невязок векторов последовательности Релея при всех k выполняется неравенство $\|r_k\|_E \leqslant \|r_{k-1}\|_E$, причем равенство достигается тогда и только тогда, когда $\sigma_k = \sigma_{k-1}$, а вектор v_{k-1} — собственный для матрицы $(A - \sigma_{k-1}E)^2$.

30.32. Почти для всех векторов v_0 последовательность Релея сходится асимптотически с кубической скоростью к некоторому собственному вектору матрицы A.

30.33. Если $\tilde{v}_k$ — реально вычисленные векторы последовательности Релея, то найдется нормированный собственный вектор v матрицы A, для которого

$$\overline{\lim_{k\to\infty}} \|v_k - v\| \leqslant O\left(np^{-t+1} \|A\|_E\right).$$

Мы ничего не говорим о параметре α_k в прямых и обратных итерациях. Во всех алгоритмах он играет чисто служебную роль и появляется либо как нормирующий множитель вектора u_k, либо в процессе решения треугольных систем с помощью обратной подстановки с нормировкой.

§ 31. *QR*- и *QL*-алгоритмы

Прямые и обратные итерации в большей мере приспособлены для определения отдельных собственных значений и собственных векторов, чем для решения спектральной проблемы в целом. Конечно, можно проводить одновременные итерации с несколькими векторами и пытаться с помощью таких процессов выделять необходимую информацию. Собственно говоря, именно эта идея и лежит в основе рассматриваемых в данном параграфе методов решения полной проблемы собственных значений. Снова будем предполагать, что собственные значения матрицы занумерованы в порядке неубывания их модулей. Для простоты изложения будем всюду считать, что матрица A невырожденная.

31.1. Пусть A — произвольная матрица. Построим последовательность унитарных матриц Q_k и правых треугольных матриц R_k по следующим рекуррентным формулам:

$$\begin{array}{ll} A = Q_1R_1, & A_1 = R_1Q_1, \\ A_1 = Q_2R_2, & A_2 = R_2Q_2, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ A_{k-1} = Q_kR_k, & A_k = R_kQ_k, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot\ \cdot\ \cdot \end{array}$$

Процесс построения по матрице A последовательностей матриц Q_k, R_k называется *QR-алгоритмом*.

31.2. Матрицы A_k унитарно подобны матрице A. Если обозначить $P_k = Q_1 \ldots Q_k$, то $A_k = P_k^* A P_k$.

31.3. Обозначим $U_k = R_k \ldots R_1$. Для матриц A^k имеет место разложение $A^k = P_kU_k$. В соответствии с введенными обозначениями матрица P_k — унитарная, матрица U_k — правая треугольная.

31.4. Рассмотрим произвольную невырожденную матрицу A порядка n. Представим ее в виде $A = Q\Lambda Q^{-1}$, где Λ — каноническая матрица Жордана, диагональные элементы которой расположены в порядке невозрастания модулей. Пусть $D = \operatorname{diag}(\lambda_n, \ldots, \lambda_1)$ — диагональная матрица соответствующих собственных значений. Предположим, что существует разложение $Q^{-1}=LU$, где L — левая треугольная матрица с единичными элементами, U — невырожденная правая треугольная матрица. Имеет место представление

$$A_k=\Delta_k\{D^k(L^{-1}\Lambda L)D^{-k}\}\Delta_k^{-1},$$

где $\Delta_k=U_kUD^{-k}$ — правая треугольная матрица, причем

$$\Delta_k=P_k^*Q(\Lambda^k LD^{-k}),\quad \Delta_k^{-1}=(D^kL^{-1}\Lambda^{-k})Q^{-1}P_k.$$

31.5. Разобьем матрицы $C_k=D^k(L^{-1}\Lambda L)D^{-k}$ на блоки таким образом, чтобы диагональные блоки были квадратными и имели те же размеры, что и группы равных по модулю собственных значений. При таком разбиении матрицы C_k будут блочными левыми треугольными. Если

$$C_k=\begin{bmatrix} C_{11}^{(k)} & & & 0 \\ C_{21}^{(k)} & C_{22}^{(k)} & & \\ \cdot & \cdot & \cdot & \cdot \\ C_{m1}^{(k)} & C_{m2}^{(k)} & \dots & C_{mm}^{(k)} \end{bmatrix},$$

то из представления матрицы C_k следует, что элементы поддиагональных блоков $C_{ij}^{(k)}$ убывают до нуля, как величины

$$\gamma_{ij}^{(k)}=O\left(\left(\frac{\tau_{m-i+1}}{\tau_{m-j+1}}\right)^k\right),\qquad i>j.$$

Таким образом, при всех k элементы C_k остаются ограниченными, а сами матрицы C_k с ростом k приближаются к блочно диагональной матрице. Скорость этого приближения указана в 31.5. Если бы элементы Δ_k и Δ_k^{-1} оставались ограниченными при всех k, то из представления 31.4 матрицы A_k и вида матриц C_k сразу вытекало бы, что при неограниченном увеличении k матрицы A_k со скоростью 31.5 будут приближаться к блочной правой треугольной матрице с диагональными блоками таких же размеров, как у матриц C_k.

31.6. Если матрица A имеет простую структуру, то элементы матриц Δ_k и Δ_k^{-1} ограничены.

31.7. Если порядок жордановых ящиков матрицы A не превосходит s, то элементы матриц Δ_k и Δ_k^{-1} по порядку роста не превосходят k^{s-1}.

31.8. Представим матрицы A_k в блочном виде:

$$A_k=\begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} & \dots & A_{1m}^{(k)} \\ A_{21}^{(k)} & A_{22}^{(k)} & \dots & A_{2m}^{(k)} \\ \cdot & \cdot & \cdot & \cdot \\ A_{m1}^{(k)} & A_{m2}^{(k)} & \dots & A_{mm}^{(k)} \end{bmatrix},$$

где диагональные блоки $A_{ii}^{(k)}$ квадратные и имеют те же размеры, что и группы равных по модулю собственных значений, начиная со старших. Элементы матриц A_k всегда равномерно по k ограничены, а элементы поддиагональных блоков $A_{ij}^{(k)}$ почти всегда убывают до нуля, как величины

$$\nu_{ij}^{(k)} = O\left(k^{2(s-1)}\left(\frac{\tau_{m-i+1}}{\tau_{m-j+1}}\right)^k\right),$$

если порядок жордановых ящиков матрицы A не превосходит s. О последовательности матриц A_k говорят, что она *сходится по форме* к блочной правой треугольной матрице.

31.9. Для того чтобы последовательность матриц A_k из 31.8 сходилась по форме к блочной правой треугольной матрице, достаточно, чтобы в разложении $A = Q\Lambda Q^{-1}$ при упорядочении диагональных элементов матрицы Λ по невозрастанию модулей в матрице Q^{-1} были отличны от нуля все ведущие миноры.

31.10. Если все собственные значения матрицы A различны по модулю, то в условиях 31.9 последовательность матриц A_k сходится по форме к правой треугольной матрице.

31.11. В условиях 31.10 диагональные элементы матриц A_k сходятся к собственным значениям матрицы A.

31.12. Если матрица A вещественная и только комплексно сопряженные собственные значения могут иметь равные модули, то в условиях 31.9 последовательность матриц A_k сходится к блочной правой треугольной матрице с блоками первого и второго порядков на диагонали.

31.13. Обозначим через $l_i(\lambda)$ многочлен, корнями которого являются все собственные значения матрицы A из i-й по старшинству группы равных по модулю, через $l_i^{(k)}(\lambda)$ — характеристический многочлен блока $A_{ii}^{(k)}$ матрицы A_k из 31.8. Имеют место соотношения

$$l_1^{(k)}(\lambda) = l_1(\lambda) + O\left(k^{2(s-1)}\left(\frac{\tau_{m-1}}{\tau_m}\right)^k\right),$$

$$l_i^{(k)}(\lambda) = l_i(\lambda) + O\left(k^{2(s-1)}\left(\frac{\tau_{m-i+1}}{\tau_{m-i+2}}\right)^k\right) + O\left(k^{2(s-1)}\left(\frac{\tau_{m-i}}{\tau_{m-i+1}}\right)^k\right),$$

$$i \neq 1, m,$$

$$l_m^{(k)}(\lambda) = l_m(\lambda) + O\left(k^{2(s-1)}\left(\frac{\tau_1}{\tau_2}\right)^k\right).$$

31.14. Если матрица A нормальная, то в условиях 31.9 последовательность матриц A_k сходится по форме к блочно диагональной матрице.

31.15. Если A — нормальная матрица и все ее собственные значения различны по модулю, то последовательность матриц A_k сходится к диагональной матрице из собственных значений.

15*

31.16. В условиях 31.14 и обозначениях 31.13 имеют место соотношения

$$l_1^{(k)}(\lambda) = l(\lambda) + O\left(\left(\frac{\tau_{m-1}}{\tau_m}\right)^{2k}\right),$$

$$l_i^{(k)}(\lambda) = l(\lambda) + O\left(\left(\frac{\tau_{m-i+1}}{\tau_{m-i+2}}\right)^{2k}\right) + O\left(\left(\frac{\tau_{m-i}}{\tau_{m-i+1}}\right)^{2k}\right), \qquad i \neq 1, m,$$

$$l_m^{(k)}(\lambda) = l_m(\lambda) + O\left(\left(\frac{\tau_1}{\tau_2}\right)^{k}\right).$$

31.17. Если последовательности

$$s_1(\lambda),\ s_2(\lambda),\ \ldots,\ s_k(\lambda),\ \ldots;\quad t_1(\lambda),\ t_2(\lambda),\ \ldots,\ t_k(\lambda),\ \ldots$$

многочленов одинаковых степеней со старшими коэффициентами, равными единице, сходятся к взаимно простым многочленам $s(\lambda)$ и $t(\lambda)$, то скорость их сходимости не меньше, чем скорость сходимости последовательности

$$s_1(\lambda)t_1(\lambda),\quad s_2(\lambda)t_2(\lambda),\ \ldots,\ s_k(\lambda)t_k(\lambda),\ \ldots$$

31.18. Если последовательность корней многочленов $l_i^{(k)}(\lambda)$ сходится к простому собственному значению матрицы A, то скорость этой сходимости не меньше, чем скорость сходимости последовательности $l_i^{(k)}(\lambda)$ к $l_i(\lambda)$.

Проведем QR-алгоритм 31.1 настолько далеко, чтобы все элементы поддиагональных клеток $A_{ij}^{(k)}$ матрицы A_k стали малыми. Заменив эти элементы нулями, мы получим блочную правую треугольную матрицу. Для нее решение проблемы собственных значений осуществляется значительно проще, чем для матрицы A, так как обычно группы равных по модулю собственных значений не бывают большими. Утверждения 31.13, 31.16—31.18 показывают, какой при этом можно достичь точности. Особенно выгодно применять QR-алгоритм для нормальных матриц. Однако в таком виде QR-алгоритм применяется относительно редко из-за своей слабой сходимости. Обычно под QR-алгоритмом понимают нечто большее, включая в него всю совокупность приемов ускорения.

31.19. Если матрица A правая почти треугольная, то все матрицы A_k также будут правыми почти треугольными.

31.20. Если матрица A эрмитова трехдиагональная, то все матрицы A_k также будут эрмитовыми трехдиагональными.

Это утверждение характеризует одно из самых важных свойств QR-алгоритма — его инвариантность к правой почти треугольной форме. Предварительное преобразование матрицы A к почти треугольному виду делается только один раз, что требует выполнения порядка n^3 операций. Затем каждый шаг QR-алгоритма осуществляется с правой почти треугольной матрицей, что требует выполнения уже порядка n^2 операций. Как приведение матрицы к почти треугольному виду, так и разложение почти треугольной матрицы на множители легко реализуется с помощью методов, описанных в 30.20, 30.24. При большом числе шагов этот прием дает ускорение примерно в $n/2$ раз. Если же матрица A эрмитова, то общее ускорение увеличивается еще на один порядок по n.

31.21. Предположим, что элемент $a_{j+1,j}$ правой почти треугольной матрицы A равен нулю. Разобьем A на такие четыре блока, чтобы диагональные блоки были квадратными и блок в левом верхнем углу имел порядок j. При этом блок в нижнем левом углу будет нулевым, т. е.

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ 0 & \alpha_{22} \end{bmatrix}.$$

QR-алгоритм инвариантен к данной форме матрицы, и на всех его шагах диагональные блоки преобразуются независимо друг от друга.

При наличии нулевых поддиагональных элементов правой почти треугольной матрицы A задачу отыскания всех ее собственных значений можно свести к аналогичным задачам меньшего размера. Этот прием также позволяет достичь определенного ускорения.

31.22. Пусть матрица A — правая почти треугольная и все ее поддиагональные элементы отличны от нуля. Если какое-нибудь собственное значение такой матрицы является кратным, то оно должно входить только в один канонический ящик Жордана.

31.23. Если правая почти треугольная матрица с ненулевыми поддиагональными элементами имеет простую структуру, то все ее собственные значения различны.

К сожалению, мы не можем реально надеяться на то, что наличие близких собственных значений обязательно приведет к появлению малых поддиагональных элементов. Существуют эрмитовы трехдиагональные матрицы порядка n, элементы которых заключены в пределах от 1 до $n/2$, но у которых имеются собственные значения, отличающиеся от ближайших на величину порядка $(n!)^{-2}$. Тем не менее, если по каким-либо причинам все же появился нулевой или очень малый поддиагональный элемент, полезно иметь в виду возможность сведения исходной задачи к двум задачам меньшего размера.

31.24. Рассмотрим любую последовательность чисел $\sigma_1, \sigma_2, \ldots$ Снова построим унитарные матрицы Q_k и правые треугольные матрицы R_k по рекуррентным формулам

$$\begin{array}{ll} A - \sigma_1 E = Q_1 R_1, & A_1 = R_1 Q_1 + \sigma_1 E, \\ A_1 - \sigma_2 E = Q_2 R_2, & A_2 = R_2 Q_2 + \sigma_2 E, \\ \ldots\ldots\ldots & \ldots\ldots\ldots \\ A_{k-1} - \sigma_k E = Q_k R_k, & A_k = R_k Q_k + \sigma_k E, \\ \ldots & \ldots \end{array}$$

Этот процесс построения по матрице A последовательностей матриц Q_k, R_k называется *QR-алгоритмом со сдвигами*.

31.25. QR-алгоритм со сдвигами инвариантен правой почти треугольной форме.

31.26. Матрицы A_k унитарно подобны исходной матрице A. При этом

$$A_k = (Q_1 \ldots Q_k)^* A (Q_1 \ldots Q_k),$$
$$(Q_1 \ldots Q_k)(R_k \ldots R_1) = (A - \sigma_1 E)(A - \sigma_2 E) \ldots (A - \sigma_k E).$$

31.27. Пусть матрица A — правая почти треугольная порядка n и ее элемент в позиции $(j+1,\ j)$ равен по модулю малому числу ε. Разобьем, аналогично 31.21, матрицу A на блоки:

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix},$$

где блок α_{21} имеет лишь один ненулевой элемент величины ε в верхнем правом углу. Предположим, что вычислены каким-либо образом собственные значения $\sigma_1, \ldots, \sigma_{n-j}$ блока α_{22}. Тогда в матрице A_{n-j}, полученной согласно 31.24 с помощью таких сдвигов, элемент в позиции $(j+1,\ j)$ будет иметь величину порядка ε^2 или меньше.

31.28. Циклическое повторение процесса 31.27 асимптотически обеспечивает не менее чем квадратичную скорость уменьшения одного из поддиагональных элементов.

По существу, все известные стратегии выбора сдвигов в общем случае состоят из двух различных этапов. На первом этапе обеспечивается заметное уменьшение одного из поддиагональных элементов матрицы A_k. На втором этапе малость поддиагонального элемента позволяет начать целенаправленный выбор сдвигов, например, в соответствии с 31.27. Это позволяет очень быстро осуществить разбиение матрицы A_k на две матрицы меньших размеров.

Очень редко первый этап осуществляется в точном соответствии с процессом 31.1 даже для почти треугольной матрицы A. Основная причина — малая скорость убывания поддиагональных элементов в данном процессе. Как правило, на первом этапе пытаются сразу же применять сдвиги, надеясь на получение малого поддиагонального элемента либо в позиции $(n,\ n-1)$, либо в позиции $(n-1,\ n-2)$. Теоретическое обоснование этого этапа для матриц общей структуры, как правило, отсутствует, но проводится экспериментальное подтверждение его эффективности.

Если удалось получить малый элемент в позициях $(n,\ n-1)$ или $(n-1,\ n-2)$, то определение дальнейших сдвигов становится простой задачей.

Практическая реализация описанного процесса ускорения не всегда оказывается эффективной. Предположим, что вещественная матрица A имеет комплексные собственные значения. В этом случае для обеспечения квадратичной сходимости в соответствии со схемой 31.24 необходимо использовать комплексные сдвиги. Появление комплексных матриц весьма нежелательно как с точки зрения времени счета, так и с точки зрения памяти ЭВМ. Заметим, что если в обозначениях 31.27 матрица A является вещественной, то вещественной будет и матрица A_{n-j}, несмотря на то что промежуточные матрицы будут комплексными. Поэтому покажем, как можно вычислять матрицу A_{n-j}, минуя получение промежуточных матриц.

31.29. Пусть для матрицы A выполнены унитарно подобные преобразования $C_1 = T_1^* A T_1$, $C_2 = T_2^* A T_2$, причем матрицы C_1, C_2 — правые почти треугольные, с ненулевыми поддиагональными элементами. Если первые столбцы матриц T_1, T_2 совпадают, то существует такая диагональная матрица S с элементами, равными по модулю единице, что $T_2 = T_1S$, $C_2 = S^*C_1S$.

31.30. В условиях и обозначениях 31.27 выполняются соотношения

$$A_{n-j} = T^*_{n-j}AT_{n-j}, \quad T_{n-j}L_{n-j} = f_{n-j}(A).$$

Здесь T_{n-j} — унитарная матрица, L_{n-j} — правая треугольная, $f_{n-j}(\lambda)$ — характеристический многочлен блока α_{22} матрицы A.

Предположим, что мы каким-либо способом нашли такую ортогональную матрицу T, у которой первый столбец совпадает с первым столбцом T_{n-j} и при этом матрица $C = T^*AT$ является правой почти треугольной. Возможны две ситуации. Если все поддиагональные элементы матрицы C отличны от нуля, то согласно 31.29 $T = T_{n-j}S$, $C = S^*A_{n-j}S$, где S — диагональная матрица с элементами, равными по модулю единице. Безразлично, с какой из матриц: A_{n-j} или C — продолжать QR-алгоритм. Мы будем продолжать его с матрицей C. Если же какие-то поддиагональные элементы матрицы C равны нулю, то это более благоприятный случай, так как можно продолжать QR-алгоритм с матрицами меньших размеров.

31.31. Если матрица A правая почти треугольная, то в первом столбце матрицы $f_{n-j}(A)$ могут быть ненулевыми только первые $n-j+1$ элементов.

31.32. Если матрица L_{n-j} получена в результате исключения поддиагональных элементов матрицы $f_{n-j}(A)$ с помощью умножения слева на матрицы вращения $T_{12}, \ldots, T_{1n}, \ldots, T_{n-1,n}$, то первый столбец матрицы T_{n-j} из 31.30 совпадает с первым столбцом матрицы $T^*_{12} \ldots T^*_{1,n-j+1}$.

31.33. Вычислим первый столбец матрицы $f_{n-j}(A)$ и определим по нему соответствующие матрицы вращения $T_{12}, \ldots, T_{1,n-j+1}$. Получим, далее, матрицу

$$B = T_{1,n-j+1} \ldots T_{12}AT^*_{12} \ldots T_{1,n-j+1}$$

и приведем ее с помощью алгоритма 30.20 к правой почти треугольной матрице C. Тогда либо матрица C имеет нулевой поддиагональный элемент, либо $C = S^*A_{n-j}S$, где S — диагональная матрица с элементами, равными по модулю единице.

Матрица B из 31.33 имеет много нулевых элементов и отличается от правой почти треугольной тем, что почти все элементы ведущего минора порядка $n-j+2$ могут быть отличны от нуля. Специальный вид матрицы B легко учесть при ее приведении к почти треугольной форме. На всех этапах приведения промежуточные матрицы будут отличаться от правой почти треугольной тем, что почти все элементы некоторого минора порядка $n-j+2$, опирающегося на главную диагональ, могут быть отличны от нуля. По мере выполнения процесса этот минор будет перемещаться по диагонали вниз.

В целом прямое получение матрицы C из матрицы A согласно 31.33 требует выполнения примерно такого же объема вычислений, как и последовательное ее получение за $n-j$ шагов процесса 31.24 с вещественными сдвигами.

Описанное ускорение сходимости QR-алгоритма особенно эффективно, когда $j = n - 1$. В этом случае очередной сдвиг совпадает с последним диагональным элементом и заведомо будет вещественным, а квадратичное убывание последнего поддиагонального элемента будет происходить на каждом шаге процесса 30.24. Если же вещественная матрица A имеет комплексные собственные значения, то прямое получение матрицы C согласно 31.33 наиболее эффективно при $j = n - 2$. Обязательное появление малых поддиагональных элементов связано в основном с наличием у матрицы A канонических ящиков Жордана больших размеров.

31.34. Следующие приемы ускорения значительно повышают эффективность QR-алгоритма:

— предварительное приведение матрицы к правой почти треугольной форме; без этого приведения QR-алгоритм обычно не применяется;

— использование сдвигов для повышения скорости убывания поддиагональных элементов; при наличии комплексных собственных значений у вещественной матрицы наиболее эффективно прямое получение матриц A_k из 31.24;

— замена малых поддиагональных элементов нулями; это позволяет продолжать применение QR-алгоритма с матрицами меньших размеров.

В приемах ускорения остается неясным только один вопрос: как начинать выбор сдвигов, чтобы достаточно быстро получить малый поддиагональный элемент в позиции $(j+1, j)$ с наибольшим по возможности значением j? За исключением некоторых специальных классов матриц, пока на него не получено убедительного ответа. Мы опишем сейчас одну из практических процедур выбора сдвигов. Она эффективна, хотя ее применение также связано с некоторым риском.

31.35. Пусть процесс 31.24 начинается с вычисления матрицы A_1 при $\sigma_1 = 0$. Предположим, что уже получены матрицы A_{k-1} и A_k. Проверяем выполнение неравенства

$$\left| a_{nn}^{(k)} - a_{nn}^{(k-1)} \right| \leqslant \frac{1}{3} \left| a_{nn}^{(k-1)} \right|.$$

Если оно справедливо, то находим матрицу A_{k+1}, беря $\sigma_{k+1} = a_{nn}^{(k)}$. Если же это неравенство не имеет места, то проверяем выполнение другого неравенства:

$$\left\| \alpha^{(k)} - \alpha^{(k-1)} \right\| \leqslant \frac{1}{3} \left\| \alpha^{(k-1)} \right\|,$$

где $\alpha^{(k-1)}$, $\alpha^{(k)}$ — блоки второго порядка матриц A_{k-1}, A_k, находящиеся в нижнем правом углу. Если это неравенство справедливо, то вычисляем характеристический многочлен матрицы $\alpha^{(k)}$ и находим матрицу A_{k+2}, используя прямой способ ее получения. Если же и последнее неравенство не имеет места, то находим матрицу A_{k+1}, беря $\sigma_{k+1} = 0$.

Применение этой процедуры показывает, что среднее число итераций на каждое собственное значение, как правило, не превосходит 5.

31.36. Если считать, что на каждое собственное значение матрицы A требуется не более пяти итераций, то вычисленные собственные значения будут точными для некоторой возмущенной матрицы $A+\mathscr{E}$, где

$$\|\mathscr{E}\|_E \lessapprox \frac{20\sqrt{2}+25}{2} n^2 p^{-t+1} \|A\|_E.$$

QR-алгоритм лучше приспособлен для определения собственных значений, чем собственных векторов. При нахождении собственных векторов необходимо дополнительно вычислять и запоминать матрицу результирующего преобразования подобия. Эта операция оказывается очень невыгодной, если нужно определить лишь несколько векторов. К тому же вычисление матрицы преобразования усложняет численный метод, особенно в тех случаях, когда приходится переходить к матрицам меньших размеров при появлении нулевых поддиагональных элементов.

Нет никакой необходимости находить собственные векторы, используя QR-алгоритм. Значительно проще и быстрее определить с помощью QR-алгоритма только собственные значения, а собственные векторы вычислить, применив обратные итерации с постоянным сдвигом.

31.37. Пусть A — произвольная матрица. Построим последовательность унитарных матриц Q_k и левых треугольных матриц L_k по следующим рекуррентным формулам:

$$\begin{aligned} A &= Q_1L_1, & A_1 &= L_1Q_1, \\ A_1 &= Q_2L_2, & A_2 &= L_2Q_2, \\ &\cdots & &\cdots \\ A_{k-1} &= Q_kL_k, & A_k &= L_kQ_k, \\ &\cdots & &\cdots \end{aligned}$$

Процесс построения по матрице A последовательностей матриц Q_k, L_k называется *QL-алгоритмом.*

В теоретическом отношении QL- и QR-алгоритмы почти совпадают. Основное их различие состоит в том, что матрицы A_k в QL-алгоритме сходятся по форме к блочной левой треугольной матрице и собственные значения в диагональных блоках упорядочиваются по неубыванию их модулей. Для QL-алгоритма могут быть применены все те приемы ускорения, которые были описаны для QR-алгоритма, с заменой, конечно, правой почти треугольной формы на левую почти треугольную. QR-алгоритм выгоднее применять в тех случаях, когда большие элементы матрицы A сосредоточены в ее левом верхнем углу, QL-алгоритм — когда в нижнем правом.

§ 32. Эрмитовы матрицы

32.1. Если некоторые внедиагональные элементы эрмитовой трехдиагональной матрицы равны нулю, то эта матрица может быть представлена в виде прямой суммы диагональных матриц и эрмитовых трехдиагональных матриц с ненулевыми внедиагональными элементами.

32.2. Эрмитова трехдиагональная матрица с ненулевыми внедиагональными элементами является якобиевой матрицей.

32.3. Обозначим через $\Delta_1, \ldots, \Delta_n$ ведущие миноры эрмитовой якобиевой матрицы

$$A = \begin{bmatrix} b_1 & a_2 & & & & 0 \\ a_2 & b_2 & a_3 & & & \\ & & \ddots & & & \\ & & & a_{n-1} & b_{n-1} & a_n \\ 0 & & & & a_n & b_n \end{bmatrix}$$

порядка n. Выполняются рекуррентные соотношения

$$\Delta_0 = 1, \quad \Delta_1 = b_1, \ldots, \quad \Delta_r = b_r\Delta_{r-1} - a_r^2\Delta_{r-2}, \qquad 2 \leqslant r \leqslant n.$$

32.4. В условиях и обозначениях 32.3 имеют место следующие свойства:

— никакие два соседних ведущих минора матрицы A не могут одновременно равняться нулю;

— если минор Δ_r, $1 < r < n$, равен нулю, то соседние миноры Δ_{r-1}, Δ_{r+1} отличны от нуля и имеют противоположные знаки;

— все собственные значения матрицы A простые.

32.5. Вычислим каким-либо способом ведущие миноры матрицы A из 32.3 и припишем любые знаки нулевым минорам, если таковые имеются. Число перемен знаков в последовательности $\Delta_0, \Delta_1, \ldots, \Delta_n$ равно числу отрицательных собственных значений матрицы A; число нулевых собственных значений совпадает с числом нулевых миноров; число положительных собственных значений равно разности между порядком матрицы и суммой отрицательных и нулевых собственных значений.

При реальных вычислениях можно надеяться только на то, что будут правильно определены знаки ведущих миноров некоторой возмущенной матрицы $A + \mathscr{E}$. Конечно, вычислительный процесс должен обеспечивать малость возмущения $\mathscr{E}$. Но чтобы по знакам ведущих миноров матрицы $A + \mathscr{E}$ правильно определить число ее нулевых, положительных и отрицательных собственных значений, необходимо быть уверенным, что $A + \mathscr{E}$ является матрицей Якоби. Не каждый вычислительный процесс гарантирует выполнение обоих условий одновременно.

Формулы 32.3 в прямом виде нельзя использовать для вычислений на ЭВМ по многим причинам. Даже для самых простых матриц вычислительный процесс 32.3 может привести либо к переполнению, либо к неправильным выводам из-за появления машинных нулей. Последнее, например, эквивалентно замене некоторых внедиагональных элементов нулями, что, конечно, в общем случае недопустимо. Машинный нуль заставляет преодолевать гораздо больше трудностей при реализации формул 32.3, чем это может показаться на первый взгляд. Связано это прежде всего с тем, что мы должны гарантировать правильность знаков вычисляемых величин. Такая задача непроста, если сами величины близки к машинному нулю.

Соотношения 32.3 для $r \geqslant 2$ являются линейными и однородными по $\Delta_1, \ldots, \Delta_n$. Поэтому, если перед вычислением Δ_r мы умножим Δ_{r-1}, Δ_{r-2} на любое положительное число γ, то вместо $\Delta_{r-1}, \Delta_r, \ldots, \Delta_n$ получим в дальнейшем $\gamma\Delta_{r-1}, \ldots, \gamma\Delta_n$. Знаки новых величин будут такими же, как у ведущих миноров. Следовательно, число нулевых положительных и отрицательных собственных значений исходной матрицы будет определенно правильно, если на каждом шаге процесса 32.3 правые части при $r \geqslant 2$ умножать на произвольные положительные числа. Мы будем выбирать эти нормирующие числа таким образом, чтобы гарантировать выполнение необходимых требований в отношении точности.

32.6. Пусть $M = (4p\omega)^{-1}$. Рассмотрим типичный шаг процесса 32.3 с нормировкой, состоящий в нахождении величин

$$u = \gamma(\beta x - \alpha^2 y), \quad v = \gamma x$$

и замене x на v. Здесь числа x, y не равны нулю одновременно, параметр γ выбирается по ходу вычислений. Если $x = 0$, то положим

$$u = -\operatorname{sign} y, \quad v = 0.$$

Если $x \neq 0$, то вычисления осуществляются в такой последовательности:

$$s = \beta y, \quad z = \beta s, \quad \theta = \max\{|x|, |z|\},$$

$$v = \begin{cases} M/\theta, & \theta \geqslant 1, \\ (x/\theta)M, & \theta < 1, \end{cases}$$

$$q = \alpha v, \quad r = M\beta, \quad m = r\beta,$$

$$l = \begin{cases} (m/\theta)\,y, & m < 1, \quad \theta < 1 \text{ или } m \geqslant 1, \quad \theta \geqslant 1, \\ (y/\theta)\,m, & m < 1, \quad \theta \geqslant 1 \text{ или } m \geqslant 1, \quad \theta < 1, \end{cases}$$

$$u = q - l.$$

32.7. Если $|\beta| \leqslant 1$, $2p\omega < |\alpha| < 1/4$, где p есть основание системы счисления, на которой построена работа ЭВМ, ω — машинный нуль, то при реализации формул 32.6 никогда не происходит переполнение.

32.8. Пусть коэффициенты матрицы A из 32.3 удовлетворяют условиям $|b_i| \leqslant 1$, $2p\omega < |a_i| < 1/4$. Процесс 32.6 позволяет по знакам реально вычисленных величин $u_0 = 1$, $\tilde{u}_1, \ldots, \tilde{u}_n$ точно определить число нулевых, положительных и отрицательных собственных значений некоторой эрмитовой якобиевой матрицы $A + \mathscr{E}$. При этом модуль каждого диагонального (внедиагонального) элемента матрицы $\mathscr{E}$ не превосходит модуля соответствующего диагонального (внедиагонального) элемента матрицы A, умноженного на $p^{-t+1}((7/4)p^{-t+1})$.

Заметим, что в 32.8 гарантируется не только малость нормы эквивалентного возмущения, но даже малость относительного возмущения в каждом элементе матрицы.

32.9. С помощью нормировки матрицы A из 32.3 и замены ее нулевых внедиагональных элементов малыми числами всегда можно добиться выполнения условий

$$|b_i| \leqslant 1/4, \quad 2p\omega < |a_i| \leqslant 1/4.$$

32.10. Замена любой пары симметрично расположенных нулевых внедиагональных элементов числами ε изменяет собственные значения матрицы не более чем на $|\varepsilon|$.

32.11. Если матрица A удовлетворяет условиям 32.9, то все ее собственные значения не превосходят по модулю 3/4.

32.12. Пусть $|\lambda| \leqslant 3/4$ и матрица A удовлетворяет условиям 32.9. Тогда матрица $A-\lambda E$ удовлетворяет условиям 32.8.

32.13. Предположим, что в условиях 32.12 для матрицы $A-\lambda E$ согласно 32.6 реально вычисляются величины $u_0=1, \tilde{u}_1, \ldots, \tilde{u}_n$. По знакам этих величин можно точно определить число собственных значений, равных λ, бо́льших λ и меньших λ, для некоторой эрмитовой якобиевой матрицы $A+\mathscr{E}$, где $\|\mathscr{E}\|_1 \lessapprox (19/8)\,p^{-t+1}$.

32.14. Пусть эрмитова якобиева матрица A удовлетворяет условиям 32.9. Предположим, что ее собственные значения занумерованы в порядке алгебраического неубывания, т. е. $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n$, и ставится задача отыскания k-го по номеру собственного значения λ_k независимо от остальных. Пусть известно, что λ_k принадлежит полуинтервалу $(a_s, b_s]$. Возьмем $\lambda=(b_s-a_s)/2$ и в соответствии с 32.13 определим, в каком из полуинтервалов, $(a_s, \lambda]$ или $(\lambda, b_s]$, находится k-е собственное значение слабо возмущенной матрицы $A+\mathscr{E}$. Это позволяет локализовать λ_k в полуинтервале $(a_{s+1}, b_{s+1}]$ вдвое меньшей длины, чем $(a_s, b_s]$. Описанный процесс последовательного сужения полуинтервала, содержащего нужное собственное значение эрмитовой якобиевой матрицы, называется *методом бисекций.*

32.15. В условиях 32.9 в качестве начального полуинтервала $(a_0, b_0]$ для метода бисекций всегда можно взять полуинтервал $(-3/4, +3/4]$.

32.16. Если в условиях 32.9, 32.14, 32.15 в качестве приближения к λ_k брать середину c_s полуинтервала $(a_s, b_s]$, то

$$|c_s-\lambda_k| \leqslant \frac{19}{8}\,p^{-t+1}+\frac{3}{4}\,2^{-s}.$$

32.17. В условиях 32.9 каждое собственное значение можно определить методом бисекций с абсолютной точностью $(25/8)p^{-t+1}$ не более чем за $t\log_2 p$ шагов.

Рассмотренный метод определения собственных значений матрицы Якоби обладает исключительной универсальностью. Его можно использовать не только для нахождения заданного по номеру собственного значения, но и для вычисления всех (или части) собственных значений, принадлежащих любой области, для исследования общего распределения собственных значений и т. п. На его реализацию не оказывает никакого влияния наличие близких и кратных собственных значений, и даже очень большое их скопление. При этом достижимая точность 32.16 не зависит от размеров матрицы.

Все эти свойства кажутся особенно удивительными, если вспомнить, что в конечном счете метод связан с распознаванием нулевых и ненулевых чисел, причем распознавание осуществляется в условиях влияния ошибок округления.

Если необходимо определить собственные значения полной эрмитовой матрицы, то целесообразно вначале привести ее к трехдиагональному виду с помощью алгоритма 30.20. Все дальнейшие вычисления с трехдиагональной матрицей при больших порядках не вносят существенного вклада в общий объем вычислений даже при относительно сложном алгоритме 32.6. Если же исходная эрмитова матрица с самого начала является трехдиагональной, то в этом случае приходится думать о выборе более экономичного метода, чем метод бисекций, если нужно отыскать много собственных значений. В случае, когда необходимо определить лишь несколько собственных значений с заданными номерами, метод бисекций исключительно эффективен.

32.18. Пусть A — эрмитова матрица. Предположим, что, исходя из координатного вектора e_n, для нее выполняются обратные итерации со сдвигами, вычисляемыми по отношению Релея. Пусть также для матрицы A с этими же сдвигами проводится QR-алгоритм. Если ни один из сдвигов не совпадает с собственным значением матрицы A, то сдвиг, вычисляемый по отношению Релея на k-м шаге в обратных итерациях, совпадает с последним диагональным элементом матрицы A_{k-1} в QR-алгоритме.

Таким образом, если матрица A эрмитова, то, проводя для нее QR-алгоритм, можно использовать сдвиги, начиная с первого шага. При этом почти всегда будет иметь место сходимость с кубической скоростью. Маловероятную возможность, связанную с нарушением сходимости, на практике можно просто игнорировать. Конечно, прежде чем применять QR-алгоритм, необходимо исходную матрицу привести к трехдиагональному виду.

Сдвиги, вычисляемые по отношению Релея, оказываются исключительно эффективными. Тем не менее существуют другие стратегии выбора сдвигов, обеспечивающие более быструю сходимость.

32.19. Пусть для эрмитовой матрицы A проводится QR-алгоритм со сдвигами. Если вычислена матрица A_{k-1}, то очередной сдвиг σ_k можно определять следующим образом. Вычисляются собственные значения матрицы второго порядка, стоящей в нижнем правом углу матрицы A_{k-1}, и в качестве σ_k берется то из них, которое ближе к последнему диагональному элементу матрицы A_{k-1}. Если оба собственных значения находятся на одинаковом расстоянии от этого элемента, то в качестве σ_k берется меньшее из собственных значений.

32.20. Для любой эрмитовой трехдиагональной матрицы A с ненулевыми внедиагональными элементами QR-алгоритм со сдвигами 32.19 всегда сходится. При этом скорость сходимости всегда не хуже квадратичной и почти всегда лучше кубической.

32.21. В условиях 32.20 все собственные значения матрицы A обычно можно вычислить с максимальной точностью в среднем приблизительно за 1—7 QR-преобразований на одно собственное значение.

32.22. Пусть собственные значения $\lambda_1, \ldots, \lambda_n$ эрмитовой матрицы A находятся с помощью унитарно подобного ее преобразования к трехдиагональной форме и последующего применения QR-алгоритма со сдвигами 32.19. Обозначим через $\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_n$ реально вычисленные собственные значения. Если каждое из собственных значений вычислялось с максимально возможной точностью, то

$$\left(\sum_{i=1}^{n}|\lambda_i-\tilde{\lambda}_i|^2\Big/\sum_{i=1}^{n}|\lambda_i|^2\right)^{1/2}\lessapprox 10n^2p^{-t+1}.$$

Если находятся собственные значения эрмитовой трехдиагональной матрицы, то трудно построить какой-либо численный метод, обладающий существенно лучшими характеристиками, чем QR-алгоритм со сдвигами 32.19. Многочисленные его усовершенствования связаны только с уменьшением времени выполнения отдельных шагов и не носят принципиального характера. В частности, довольно много усилий было затрачено на то, чтобы разработать схему метода, не требующую вычисления квадратных корней при реализации преобразований вращения. Если же находятся собственные значения полной эрмитовой матрицы с помощью ее приведения к трехдиагональному виду, то все эти усовершенствования оказываются малозначимыми на фоне общего объема вычислений.

Эффективность QR-алгоритма существенно изменяется, если нужно находить не только собственные значения, но и собственные векторы. Основная трудность связана с тем, что теперь необходимо по мере выполнения QR-алгоритма накапливать произведение всех матриц Q_k. Это требует значительного увеличения числа выполняемых операций и объема требуемой памяти. Однако вычисленные собственные векторы всегда будут ортогональны с высокой точностью, даже если имеются очень близкие собственные значения. Выполнения данного условия трудно добиться, если, например, для определения собственных векторов использовать обратные итерации.

32.23. Пусть A — эрмитова матрица порядка n с элементами a_{ij}. Обозначим через $\omega(A)$ сумму квадратов модулей ее внедиагональных элементов. Существует такое упорядочение $\lambda_{p_1}, \ldots, \lambda_{n_p}$ собственных значений матрицы A, что

$$\sum_{i=1}^{n}|\lambda_{p_i}-a_{ii}|^2\leqslant\omega(A).$$

32.24. Пусть $B=T_{ij}^{*}AT_{ij}$, где T_{ij} — некоторая матрица вращения. Обозначим через b_{ij} элементы матрицы B. Имеет место соотношение

$$\omega(B)=\omega(A)-2(|a_{ij}|^2-|b_{ij}|^2).$$

32.25. Для любой матрицы A и любой пары индексов i, j, где $i\neq j$, всегда можно найти такую матрицу вращения T_{ij}, что $b_{ij}=0$ в обозначениях 32.24.

32.26. В условиях 32.25 выполняется равенство $\omega(B)=\omega(A)-2|a_{ij}|^2$.

32.27. Если a_{ij} есть максимальный по модулю внедиагональный элемент матрицы A, то в условиях 32.25 справедливо неравенство

$$\omega(B)\leqslant\left(1-\frac{2}{n(n-1)}\right)\omega(A).$$

32.28. Пусть матрицы — вещественные и в представлении 22.1 матрицы T_{ij} имеем $c = \cos\varphi_{ij}$, $s = \sin\varphi_{ij}$. Для выполнения утверждения 32.25 достаточно взять угол φ_{ij}, удовлетворяющий условиям:

$$\operatorname{tg} 2\varphi_{ij} = \frac{2a_{ij}}{a_{ii} - a_{jj}}, \quad |\varphi_{ij}| \leqslant \frac{\pi}{4}.$$

32.29. Пусть A — произвольная эрмитова матрица. Построим последовательность матриц

$$A_0 = A,\ A_1,\ \ldots,\ A_\nu,\ \ldots,$$

каждая из которых получается из предыдущей с помощью выполнения преобразования подобия, содержащего лишь одну матрицу вращения. Пусть $A_\nu = T^*_{i_\nu j_\nu} A_{\nu-1} T_{i_\nu j_\nu}$ и параметры матрицы вращения выбираются в соответствии с 32.25. Тогда на каждом шаге процесса сумма квадратов модулей внедиагональных элементов будет убывать, если только элементы матриц $A_{\nu-1}$, стоящие в позициях (i_ν, j_ν), отличны от нуля. Этот процесс называется *методом вращений* или *методом Якоби* для решения полной проблемы собственных значений эрмитовой матрицы.

Если последовательность матриц A_ν сходится к диагональной матрице, то диагональные элементы матрицы A_ν являются приближениями к собственным значениям матрицы A, а столбцы произведения $T_{i_1 j_1} \ldots T_{i_\nu j_\nu}$ — приближениями к соответствующим собственным векторам. Для того чтобы гарантировать эту сходимость, необходимо добиться выполнения двух условий. Во-первых, нужно обеспечить, чтобы $\omega(A_\nu)$ не просто убывала, а убывала до нуля, и, во-вторых, нельзя допускать появления матриц вращения, близких к матрицам перестановок. Если матрица A вещественная, то второе условие выполняется при введении ограничения на величину угла, например, согласно 32.28. Аналогичное ограничение нетрудно получить и в случае комплексных матриц. Что касается первого условия, то его выполнение зависит от величины элементов, исключаемых на каждом шаге.

32.30. Если матрица A вещественная и симметричная, индексы i_ν, j_ν соответствуют на каждом шаге максимальному по модулю внедиагональному элементу матрицы A_ν и параметры матрицы вращения $T_{i_\nu j_\nu}$ выбираются в соответствии с 32.28, то, независимо от наличия кратных собственных значений, метод вращений обеспечивает сходимость последовательности матриц A_ν к диагональной матрице из собственных значений, причем асимптотически с квадратичной скоростью.

Реализация данного варианта метода вращений требует выбора *максимального* по модулю внедиагонального элемента матрицы на каждом шаге. При выполнении этой операции на ЭВМ требуется значительная затрата машинного времени. Поэтому необходимость указанного выбора является существенным недостатком метода с точки зрения удобства его реализации на ЭВМ.

Более удобными оказываются *циклические процессы* и, в частности, циклические процессы с *барьерами*. При циклическом процессе выбирается определенная нумерация внедиагональных элементов матрицы и их исключение происходит по циклам. В течение каждого цикла исключаются по очереди все внедиагональные элементы в порядке их нумерации. Чаще всего элементы нумеруются подряд по строкам слева направо и сверху вниз или по столбцам сверху вниз и слева направо. При этом, конечно, нумеруются только наддиагональные или поддиагональные элементы.

Недостатком такого процесса является то, что приходится исключать малые внедиагональные элементы, в то время как в матрице еще присутствуют большие. Это обстоятельство приводит к значительному уменьшению скорости сходимости.

Отмеченный недостаток частично устраняется введением барьеров. Вводится монотонно убывающая к нулю последовательность положительных чисел, называемых *барьерами*, и при циклическом просмотре исключаются лишь те из внедиагональных элементов, которые по модулю не меньше α_1. После того, как все внедиагональные элементы станут по модулю меньше α_1, барьер α_1 заменяется на α_2, и процесс продолжается.

Этот процесс позволяет решать полную проблему собственных значений быстрее, чем процесс с выбором максимального элемента. Однако практическое его использование встречает ряд трудностей, связанных с оптимальным выбором барьеров. Если барьер выбрать очень большим, то будет затрачено много времени на просмотр малых элементов. Если же его выбрать очень малым, то будет затрачено много времени на исключение малых элементов, которые, по существу, не влияют на скорость сходимости.

Особого внимания заслуживает следующий способ выбора элемента, подлежащего исключению. Если в матрице A_ν исключается элемент, стоящий в позиции (i_ν, j_ν), то суммы квадратов модулей внедиагональных элементов в каждой строке матрицы $A_{\nu+1}$ будут такие же, как у матрицы A_ν, кроме строк с номерами i_ν, j_ν. Поэтому, если в начале процесса вычислить суммы квадратов модулей внедиагональных элементов строк, то в дальнейшем в полученной последовательности из n чисел нужно пересчитать на каждом шаге только два числа. *Оптимальным* исключаемым элементом будет максимальный по модулю элемент, находящийся в строке с максимальной суммой квадратов модулей внедиагональных элементов. Его можно найти путем просмотра всего лишь $2n-1$ чисел, при этом выполняется неравенство 32.27.

32.31. В различных вариантах метода вращений для максимального уменьшения суммы квадратов модулей внедиагональных элементов в пересчете на циклы обычно требуется выполнить 2—7 полных циклов, независимо от порядка матрицы.

32.32. Для циклических вариантов метода вращений точные и реально вычисленные собственные значения матрицы A связаны между собой соотношением

$$\left(\sum_{i=1}^{n} |\lambda_i - \tilde{\lambda}_i|^2 \Big/ \sum_{i=1}^{n} |\lambda_i|^2\right)^{1/2} \lesssim 48np^{-t+1}.$$

Для других вариантов метода вращений оценка примерно в n раз хуже. Это связано лишь с трудностью получения хорошей оценки, а не с тем, что другие варианты менее точны.

§ 33. Метод Ланцоша

Рассмотренные методы определения собственных значений и собственных векторов предполагают, что в процессе их реализации допустимы различные преобразования исходной матрицы. Такие преобразования широко используются с целью уменьшения общего объема вычислений. Однако они оказываются приемлемыми только в тех случаях, когда при этом существенно не возрастают размеры требуемой памяти ЭВМ. Подобная ситуация будет иметь место, если, например, решается проблема собственных значений для плотной матрицы, все элементы которой отличны от нуля и не связаны друг с другом большим числом соотношений. В этом случае вся информация о преобразовании матрицы к более простому виду может быть размещена, в основном, на месте исходной матрицы.

Решение проблемы собственных значений существенно усложняется, если матрица имеет очень большие размеры и либо является сильно разреженной, либо ее элементы определяются небольшим числом параметров. Теперь, как правило, нельзя применять преобразования матрицы, так как при этом нарушается ее структура и резко возрастает как объем вычислений, так и объем требуемой памяти ЭВМ. Единственной возможной операцией с матрицей остается операция умножения на вектор.

С такими трудностями мы уже встречались при решении систем линейных алгебраических уравнений, и с точки зрения их преодоления эффективными оказались методы сопряженных направлений. Подобные методы полезны и при решении спектральных задач. Как уже отмечалось в 23.22—23.27, на основе процессов ортогонализации степенных последовательностей, или, как их называют иначе, последовательностей Крылова, матрица A может быть приведена к подобной трехдиагональной матрице. Если это преобразование осуществлено, то решение спектральных задач для трехдиагональных матриц является вполне реализуемым делом, несмотря на то, что приходится преодолевать трудности, связанные с неустойчивостью процессов ортогонализации.

Метод Ланцоша — это совокупность аналогичных преобразований и различных вспомогательных приемов, предназначенных для решения спектральных задач эрмитовой матрицы. Интерес к нему в настоящее время определяется тем, что методы такого типа являются единственными среди известных, которые дают хотя бы какую-то возможность решать спектральные задачи для очень больших разреженных матриц.

Для простоты изложения мы будем здесь предполагать, если не сделано специальной оговорки, что матрица A порядка n вещественная, симметричная, все ее собственные значения λ_i различные и занумерованы в порядке алгебраического возрастания.

33.1. Пусть вектор x не ортогонален ни одному из собственных векторов матрицы A. Тогда последовательность векторов x, Ax, ..., $A^{n-1}x$ линейно независима.

33.2. Пусть векторы степенной последовательности 33.1 подвергаются процессу ортогонализации Грама — Шмидта с нормировкой получаемых векторов по евклидовой норме. Построенные таким способом ортонормированные векторы q_1, ..., q_n удовлетворяют соотношениям

$$\beta_0 q_1 = x \equiv r_0,$$
$$\beta_1 q_2 = Aq_1 - \alpha_1 q_1 \equiv r_1,$$
$$\beta_j q_{j+1} = Aq_j - \alpha_j q_j - \beta_{j-1} q_{j-1} \equiv r_j, \qquad j > 1,$$

где $\alpha_j = (Aq_j, q_j)$, $\beta_j = (q_{j+1}, Aq_j)$.

33.3. В условиях и обозначениях 33.2 для всех j выполняется равенство $|\beta_j| = \|r_j\|_E$.

Сравнивая соотношения 22.22 и 33.2, можно заметить определенное различие между ними. Рекуррентные формулы 22.22 содержат два независимых коэффициента α_i, β_{j-1}. Это вполне естественно, так как требуется удовлетворить два условия ортогональности вектора f_{i+1} к векторам f_i, f_{i-1}.

Формулы 33.2 также содержат два коэффициента, хотя к двум условиям ортогональности добавлено третье условие нормировки. Эта ситуация в действительности не противоречива. Более того, нормирующие коэффициенты β_j всегда можно брать положительными.

33.4. Пусть имеются векторы q_{j-1}, r_{j-1} и коэффициент $\beta_{j-1}>0$. Тогда в соответствии с 33.2 процесс можно осуществлять по следующему предписанию:

$$\begin{aligned} q_j &= r_{j-1}/\beta_{j-1}, & u_j &= Aq_j,\\ s_j &= u_j-\beta_{j-1}q_{j-1}, & \alpha_j &= (q_j, s_j),\\ r_j &= s_j-\alpha_j q_j, & \beta_j &= \|r_j\|_E. \end{aligned}$$

Чтобы начать вычисления, нужно положить $q_0=0$, $r_0=x$, $\beta_0=\|x\|$.

33.5. В базисе $q_1, \ldots, q_n$ матрица A имеет симметричную трехдиагональную форму:

$$T=\begin{bmatrix} \alpha_1 & \beta_1 & & & 0\\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \cdot & \cdot & \cdot & \\ & 0 & \beta_{n-2} & \alpha_{n-1} & \beta_{n-1}\\ & & & \beta_{n-1} & \alpha_n \end{bmatrix}.$$

33.6. Если Q есть матрица, столбцы которой совпадают с векторами $q_1, \ldots, q_n$, то $AQ=QT$.

33.7. Обозначим через T_m матрицу ведущего минора порядка m матрицы T из 33.5, через Q_m — прямоугольную матрицу размера $n\times m$, столбцы которой совпадают с векторами $q_1, \ldots, q_m$. Для всех $1\leqslant m\leqslant n$ имеет место соотношение $AQ_m=Q_mT_m+R_m$, где $R_m=r_me_m'$ и e_m есть координатный вектор размерности m, с единицей на последнем месте.

33.8. Всегда $\|R_m\|_E=\beta_m$.

33.9. В условиях и обозначениях 33.5—33.7 выполняются равенства

$$Q'AQ=T,\quad Q_m'AQ_m=T_m.$$

33.10. Пусть $\theta_1, \ldots, \theta_m$ — собственные значения матрицы T_m. Найдутся такие собственные значения $\lambda_{m_1}, \ldots, \lambda_{m_m}$ матрицы T, что для всех $1\leqslant j\leqslant m$ справедливы неравенства $|\theta_j-\lambda_{m_j}|\leqslant\beta_m$.

33.11. Пусть $T_ms_j=\theta_js_j$ для вектора s_j и числа θ_j. Найдется такое собственное число λ матрицы A, что

$$|\theta-\lambda|\leqslant\beta_{mj}/\|s_j\|_E,$$

где $\beta_{mj}=\beta_m|s_{mj}|$ и s_{mj} есть последняя координата вектора s_j.

33.12. Пусть в разложении вектора x по собственным векторам матрицы A только m координат отличны от нуля. Тогда процессы 33.2, 33.4 заканчиваются на m-м шаге и $\beta_m = 0$. При этом оценки 33.10, 33.11 остаются справедливыми.

33.13. Предположим, что в условиях 33.12 определены собственные значения $\theta_1, \ldots, \theta_m$ и соответствующие собственные векторы $s_1, \ldots, s_m$ матрицы T_m. Тогда числа $\theta_1, \ldots, \theta_m$ являются точными собственными значениями матрицы A, а векторы $Q_m s_1, \ldots, Q_m s_m$ — ее точными собственными векторами.

33.14. Процесс нахождения собственных значений и собственных векторов матрицы A, основанный на сведении этой матрицы к трехдиагональной форме согласно 33.2, 33.4, называется *методом Ланцоша*.

Прежде чем переходить к обсуждению метода Ланцоша, остановимся на его геометрической интерпретации. Пусть задано некоторое подпространство L размерности m. Если L инвариантно по отношению к эрмитовой матрице A, то в подпространстве L обязательно имеется ровно m собственных векторов матрицы A. Если же L не является инвариантным, то можно попытаться найти в этом подпространстве m векторов, в том или ином смысле близких к собственным векторам матрицы A.

Одна из форм близости может быть определена из следующих соображений. Если λ, x являются точными собственным значением и собственным вектором матрицы A, то выполняется равенство $Ax = \lambda x$. Следовательно, в данном случае невязка $r(\lambda, x) = Ax - \lambda x$ ортогональна ко всем векторам любого подпространства. Поэтому, если задано подпространство L и мы хотим найти такие вектор $y \in L$ и число θ, чтобы они «приближали» вектор x и число λ, то можем потребовать, чтобы невязка $r(\theta, y) = Ay - \theta y$ была ортогональна ко всем векторам подпространства L. Эта идея является всего лишь реализацией частного случая общего метода решения самых различных уравнений, называемого *методом Ритца*.

33.15. Пусть $q_1, \ldots, q_m$ — произвольный ортонормированный базис подпространства L размерности m. Обозначим через Q_m прямоугольную матрицу размера $n \times m$, столбцы которой совпадают с векторами $q_1, \ldots, q_m$. Если невязка $r(\theta, y) = Ay - \theta y$ ортогональна подпространству L, то:

— число θ — собственное значение матрицы $H = Q'_m A Q_m$;

— вектор y равен $Q_m s$, где s есть собственный вектор матрицы H, соответствующий собственному значению θ.

33.16. Пусть Q_m — любая матрица размера $n \times m$ с ортонормированными столбцами, $H = Q'_m A Q_m$. Для любой матрицы B

$$\|AQ - QH\|_E \leqslant \|AQ - QB\|_E.$$

33.17. Если в обозначениях 33.16 $\theta_1, \ldots, \theta_m$ — собственные значения матрицы H, то найдутся m собственных значений $\lambda_{m_1}, \ldots, \lambda_{m_m}$ матрицы A таких, что для всех $1 \leqslant j \leqslant m$ справедливы неравенства

$$|\theta_j - \lambda_{m_j}| \leqslant \|AQ - QH\|_E.$$

33.18. Пусть $\theta_1, \ldots, \theta_m$ и $s_1, \ldots, s_m$ — соответствующие друг другу собственные значения и собственные векторы матрицы H

из 33.16. Тогда для всех $1 \leqslant j \leqslant m$ число θ_j равно отношению Релея для вектора $Q_m s_j$.

33.19. При нумерации собственных значений в порядке неубывания для всех $1 \leqslant j \leqslant m$ выполняются неравенства

$$\lambda_j \leqslant \theta_j \leqslant \lambda_{n-m+j}.$$

33.20. Для каждого θ_j найдется такое собственное значение λ_{m_j} матрицы A, что

$$|\theta_j - \lambda_{m_j}| \leqslant \|(A - \theta E) Q_m s_i\|_E / \|s_i\|_E.$$

33.21. В обозначениях 33.16 спектральная задача $Hs = \theta s$, $y = Q_m s$, называется задачей, *редуцированной на подпространство* L по отношению к исходной спектральной задаче $Ax = \lambda x$. Вектор $Q_m s$ и число θ называются соответственно *вектором Ритца* и *числом Ритца*.

Возвращаясь снова к методу Ланцоша, мы можем теперь дать геометрическую интерпретацию связи спектральных задач с матрицами A и T_m.

33.22. Линейная оболочка векторов x, Ax, ..., $A^{m-1}x$ называется *подпространством Крылова*.

33.23. В обозначениях 33.7 спектральная задача $T_m s = \theta s$, $y = Q_m s$, является редуцированной на подпространство Крылова размерности m по отношению к исходной спектральной задаче $Ax = \lambda x$.

Таким образом, осуществляя метод Ланцоша, мы в действительности строим ортонормированные базисы вложенных друг в друга подпространств Крылова и получаем последовательность редуцированных на эти подпространства спектральных задач с симметричными трехдиагональными матрицами. Принимая во внимание, что редуцированные задачи аппроксимируют в смысле Ритца исходную спектральную задачу, мы можем надеяться на то, что при $m \ll n$ решение задачи $T_m s = \theta s$ даст полезную информацию о задаче $Ax = \lambda x$. Об этом говорят и оценки 33.10, 33.11. Если при каком-либо m окажется, что мало́ β_m или s_{mj}, то это означает, что все или некоторые из собственных чисел матрицы T_m являются хорошими приближениями к части собственных значений матрицы A.

Изящная теоретическая картина метода Ланцоша основательно нарушается при его практической реализации. Реальный процесс не обладает некоторыми свойствами, указанными точной теорией, но приобретает другие, не предсказанные ею.

33.24. Основные качественные черты практической реализации метода Ланцоша следующие:

— векторы q_j рано или поздно теряют даже приблизительную ортогональность к ранее вычисленным векторам; матрица Q_j практически становится вырожденной;

— среди чисел Ритца появляются повторные копии, не соответствующие кратным собственным значениям матрицы A; соответствующие векторы Ритца почти коллинеарны;

— чем дальше от краев спектра, тем медленнее сходимость чисел Ритца к собственным значениям матрицы A; обычно вначале сходятся крайние числа Ритца; один раз сошедшись к какому-нибудь собственному значению, числа Ритца, как правило, стабилизируются и в дальнейшем меняются мало;

— коэффициенты β_j редко бывают малыми; появляется возможность продолжать процесс неограниченно долго, не наталкиваясь на нулевые или очень малые коэффициенты β_j;

— в течение n шагов метода Ланцоша хотя бы одно число Ритца сойдется к собственному значению матрицы A;

— с ростом числа шагов $m \gg n$ в спектре трехдиагональной матрицы T_m появляются приближения ко всем собственным значениям матрицы A.

Не все перечисленные характеристики получили объяснение, подкрепленное хотя бы каким-нибудь теоретическим обоснованием. Поэтому схема практической реализации метода Ланцоша остается очень сложной. Тем не менее в большинстве случаев удается получить вполне приемлемый результат счета.

В конечном счете метод Ланцоша представляет собой некоторую разновидность процесса ортогонализации Грама — Шмидта. Все особенности, присущие методу ортогонализации, остались и в методе Ланцоша, в том числе главная из них — потеря ортогональности. В § 23 мы подробно рассмотрели как причины потери ортогональности, так и способ ее восстановления, связанный с полной переортогонализацией векторов. К сожалению, в отношении метода Ланцоша полная переортогонализация представляет только теоретический интерес, так как приводит к огромному росту объема вычислений и требуемой памяти ЭВМ. Предложены различные варианты выборочной переортогонализации. Установлено, что нарушение ортогональности связано во многом с появлением малых чисел s_{mi}, т. е. с факторами, свидетельствующими о сходимости чисел Ритца к собственным значениям матрицы A.

Однако в этом методе пока еще слишком много неясного, и его практическое использование скорее похоже на искусство, чем на обоснованные вычисления. Мы не можем здесь останавливаться на описании многочисленных деталей, которыми обросла вычислительная схема метода Ланцоша.

§ 34. Общие вопросы теории итерационных методов решения систем линейных уравнений

В этом и в последующих параграфах будут рассмотрены итерационные методы решения совместных систем линейных алгебраических уравнений $Au = f$ с квадратными матрицами A и векторами $f \in \operatorname{im} A$. При этом всюду, кроме § 40, будет предполагаться, что матрица A невырожденная и, следовательно, система имеет единственное решение $u = A^{-1}f$ при произвольном векторе $f \in \mathbf{R}_n$. В наиболее общей форме *итерационный метод* решения системы $Au = f$ мы запишем в виде вычислительной процедуры

$$u^k = F_k(A, f, u^{k-1}, \ldots, u^0), \qquad k = 1, 2, \ldots,$$

где F_k — некоторая последовательность операторов, действующих для заданных A и f из пространства $\mathbf{R}_n \times \ldots \times \mathbf{R}_n$ в пространство $\mathbf{R}_n$. Ниже мы ограничиваемся только теми итерационными методами, которые оставляют инвариантным решение u исходной системы, т. е. $u = F_k(A, f, u, \ldots, u)$ для всех $k \geqslant 1$.

34.1. Вектор u^0 итерационного метода называется *начальным*.

34.2. Векторы $z^k = u^k - u$, где $u = A^{-1}f$, называются ***векторами ошибок*** итерационного метода.

34.3. Векторы $\xi^k = Au^k - f$ называются ***векторами невязок*** итерационного метода.

34.4. Если для некоторого $k \geqslant 1$, используя соотношения

$$z^l = F_l(A,\ Au,\ z^{l-1} + u,\ \ldots,\ z^0 + u) - u,$$

выразить вектор ошибки z^k через вектор начальной ошибки z^0 по формуле

$$z^k = Z_k(A,\ u,\ z^0),$$

то возникающий в результате оператор Z_k называется *разрешающим* оператором k-го шага итерационного метода.

В дальнейшем всюду предполагается, что

$$Z_k(A,\ u,\ z) \equiv Z_k(A,\ z)$$

для всех $k \geqslant 1$.

34.5. Пусть p — некоторое положительное целое. Тогда итерационный метод называется *p-шаговым* (*$p+1$-слойным*), если

$$F_k(A,\ f,\ u^{k-1},\ \ldots,\ u^0) \equiv F_k(A,\ f,\ u^{k-1},\ \ldots,\ u^{k-p})$$

для всех $k \geqslant p$.

34.6. p-шаговый итерационный метод называется ***стационарным***, если для всех $k \geqslant p$ операторы F_k не зависят от номера итерации k, т. е.

$$F_k(A,\ f,\ u^{k-1},\ \ldots,\ u^{k-p}) \equiv F(A,\ f,\ u^{k-1},\ \ldots,\ u^{k-p}),$$

где F — некоторый заданный оператор.

34.7. Пусть s — некоторое положительное целое. Тогда итерационный метод называется *циклическим* с периодом s, если $F_k = F_{k-s}$ для всех $k > s$.

34.8. Циклический с периодом $s \geqslant 1$ итерационный метод может быть записан в виде

$$u^k = F_l(A,\ f,\ u^{k-1},\ \ldots,\ u^{k-l}),$$

где $l = k - s[k/s]$ ($[t]$ — целая часть числа t).

34.9. Циклический с периодом $s \geqslant 1$ итерационный метод всегда может быть записан в виде стационарного одношагового итерационного метода

$$u^k = F(A,\ f,\ u^{k-1}),$$

каждый шаг которого реализуется по формулам

$$u^{k-1+i/s} = F_i(A,\ f,\ u^{k-1+(i-1)/s},\ \ldots,\ u^{k-1});\qquad i = 1,\ 2,\ \ldots,\ s.$$

34.10. Итерационный метод называется ***линейным***, если операторы F_k являются линейными по отношению к переменным векторам $f,\ u^{k-1},\ \ldots,\ u^0$.

34.11. Линейный итерационный метод всегда представим в виде

$$u^k = \sum_{i=0}^{k-1} F_{ki} u^i + F_{kk} f,$$

где F_{ki}, $i = 0, 1, \ldots, k$,— некоторые постоянные (т. е. зависящие только от значений индексов k и i) матрицы.

34.12. В силу инвариантности решения системы относительно итерационного метода имеем

$$F_{kk} = \left(E - \sum_{i=0}^{k-1} F_{ki}\right) A^{-1}.$$

34.13. Разрешающие операторы линейного итерационного метода являются матрицами и представимы в виде

$$Z_k = E - M_k A,$$

где M_k — некоторая последовательность матриц.

34.14. Линейный итерационный метод для всех $k \geqslant 1$ может быть записан в виде

$$u^k = u^{k-1} - H_k (Au^{k-1} - f) + \sum_{i=0}^{k-1} S_{ki} u^i,$$

где H_k, S_{ki}, $i = 0, 1, \ldots, k-1$,— некоторые последовательности матриц, причем

$$\sum_{i=0}^{k-1} S_{ki} = 0.$$

34.15. Линейный двухшаговый итерационный метод может быть записан в виде

$$u^k = u^{k-1} - H_k (Au^{k-1} - f) - S_k (u^{k-1} - u^{k-2}),$$

где H_k, S_k — некоторые последовательности матриц.

34.16. Линейный одношаговый итерационный метод всегда может быть записан в виде

$$u^k = u^{k-1} - H_k (Au^{k-1} - f),$$

где H_k — некоторая последовательность матриц. Если матрицы H_k невырожденные, то метод часто записывают в другом, эквивалентном, универсальном виде:

$$B_k (u^k - u^{k-1}) = -\tau_k (Au^{k-1} - f),$$

где τ_k — некоторая последовательность ненулевых чисел (параметров итерационного метода), $B_k = \tau_k H_k^{-1}$.

34.17. Линейный стационарный одношаговый итерационный метод всегда может быть записан либо в виде

$$u^k = u^{k-1} - H(Au^{k-1} - f),$$

где H — некоторая постоянная матрица, либо, если матрица H невырожденная, в виде

$$B(u^k - u^{k-1}) = -\tau(Au^{k-1} - f),$$

где τ — некоторый итерационный параметр, $B = \tau H^{-1}$.

34.18. Итерационный метод называется *сходящимся*, если для любого начального приближения $u^0 \in \mathbf{R}_n$ последовательность векторов u^k этого метода сходится к решению u исходной системы $Au = f$.

Здесь и далее в определениях и формулировках различных результатов (неявно или явно) часто используется свойство эквивалентности норм в конечномерных пространствах. Более конкретно, если последовательность векторов (матриц) сходится к какому-либо вектору (матрице) в одной выбранной норме, то она сходится и в любой другой норме. Поэтому при изучении сходимости нет необходимости фиксировать вид используемой нормы. В то же время, если нас интересует вопрос о близости векторов (матриц) или о скорости сходимости последовательности за конечное число шагов, то способ задания нормы может уже играть очень важную роль.

34.19. Пусть $\|\cdot\|$ — некоторая векторная норма и Z — некоторый оператор, действующий из $\mathbf{R}_n$ в $\mathbf{R}_n$. Тогда величина

$$L(Z) = \sup_{v\neq 0} \frac{\|Z(v)\|}{\|v\|}$$

называется *константой Липшица* оператора Z.

34.20. Если Z — матрица порядка n и $\|\cdot\|$ — некоторая векторная норма, то

$$L(Z) = \|Z\|,$$

где $\|\cdot\|$ — матричная норма, подчиненная соответствующей исходной векторной норме.

34.21. Для любой векторной нормы условие

$$\lim_{k\to\infty} L(Z_k) = 0$$

достаточно для сходимости итерационного метода.

34.22. Для любой матричной нормы $\|\cdot\|$ условие

$$\lim_{k\to\infty} \|Z_k\| = 0$$

необходимо и достаточно для сходимости линейного итерационного метода.

34.23. Для любого $k \geqslant 1$ оператор T_k, определяемый соотношением

$$T_k(z) \equiv T_k(A, z) \equiv F_k(A, Au, z+u) - u,$$

называется *оператором перехода* k-го шага соответствующего одношагового итерационного метода.

34.24. Для линейных одношаговых методов матрицы

$$T_k = E - H_k A$$

называются *матрицами перехода k-го шага*, а в случае стационарных методов, когда $T_k \equiv T = E - HA$, матрица T называется просто *матрицей перехода*.

34.25. Для произвольной векторной нормы условие

$$\sup_{k \geqslant 1} L(T_k) < 1$$

достаточно для сходимости одношагового итерационного метода.

34.26. Пусть $\|\cdot\|$ — некоторая матричная норма. Тогда условие

$$\|T\| < 1$$

достаточно для сходимости линейного стационарного одношагового итерационного метода.

34.27. Пусть оператор перехода T стационарного одношагового итерационного метода определяется формулой

$$T(z) = z - \mathscr{H}(Az)Az,$$

где $\mathscr{H}$ — некоторый оператор, действующий из пространства $\mathbf{R}_n$ в пространство матриц порядка n и удовлетворяющий следующим условиям:

— $\mathscr{H}$ определен и непрерывен на всех ненулевых векторах из $\mathbf{R}_n$;

— $\mathscr{H}$ является однородным оператором нулевой степени, т. е.

$$\mathscr{H}(\lambda\xi) = \mathscr{H}(\xi)$$

для всех ненулевых векторов ξ и чисел λ.

Тогда, если существует такая векторная норма $\|\cdot\|$, что для любого ненулевого $z \in \mathbf{R}_n$ выполняется неравенство

$$\|T(z)\| < \|z\|,$$

то

$$L(T) = \sup_{\substack{z \in \mathbf{R}_n \\ \|z\| = 1}} \| T(z) \| < 1,$$

и, следовательно, соответствующий итерационный метод

$$u^k = u^{k-1} - \mathscr{H}(\xi^{k-1})(Au^{k-1} - f)$$

сходится.

34.28. Пусть λ — некоторое отличное от нуля число и

$$J = \begin{bmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{bmatrix}$$

— матрица порядка s. Тогда

$$\lim_{k\to\infty} \frac{1}{\lambda^{k-s+1}C_k^{s-1}} J^k = \Lambda_0,$$

где $\Lambda_0 = e_s' \otimes e_s$, $e_s = (0, \ldots, 0, 1)' \in \mathbf{R}_s$, и, следовательно, для любой матричной нормы $\|\cdot\|$ выполняется равенство

$$\lim_{k\to\infty} \|J^k\|^{1/k} = |\lambda|.$$

34.29. Пусть $T = QJQ^{-1}$ — некоторая квадратная матрица, где J — ее нормальная жорданова форма, а Q — матрица, столбцами которой являются собственные и корневые векторы T, и пусть s — максимальный порядок жордановых клеток, соответствующих собственным числам T, равным по модулю $\rho(T) > 0$. Тогда существует такая унитарная диагональная матрица U, что

$$\lim_{k\to\infty} \frac{1}{[\rho(T)]^{k-s+1}C_k^{s-1}} U^{k-s+1}Q^{-1}T^kQ = \Lambda,$$

где Λ — блочно диагональная матрица, диагональные блоки которой либо являются нулевыми матрицами, либо имеют порядок s и совпадают с матрицей Λ_0 из 34.28.

34.30. Для любой матрицы T и любой матричной нормы $\|\cdot\|$ выполняется равенство

$$\lim_{k\to\infty} \|T^k\|^{1/k} = \rho(T).$$

34.31. Условие

$$\lim_{k\to\infty} \|T^k\| = 0$$

необходимо и достаточно для сходимости линейного стационарного одношагового итерационного метода.

34.32. Условие

$$\rho(T) < 1$$

необходимо и достаточно для сходимости линейного стационарного одношагового итерационного метода.

34.33. Пусть $\|\cdot\|$ — заданная векторная норма. Тогда, если для конкретного $k \geqslant 1$ константа Липшица L разрешающего оператора Z_k соответствующего итерационного метода меньше единицы, то величина

$$r(Z_k) = -\frac{1}{k}\ln L(Z_k)$$

называется *средней скоростью сходимости* этого итерационного метода за k шагов в заданной норме $\|\cdot\|$.

34.34. Пусть в заданной векторной норме, начиная с некоторого номера $k_0 \geqslant 1$, для разрешающих операторов Z_k соответствующего итерационного метода выполняются неравенства

$$L(Z_k) < 1, \quad k = k_0, k_0 + 1, \ldots$$

Тогда величина

$$r(Z_\infty) = \varliminf_{k\to\infty} r(Z_k) = -\varlimsup_{k\to\infty} \frac{1}{k} \ln L(Z_k)$$

называется *асимптотической скоростью сходимости* этого итерационного метода.

34.35. Асимптотическая скорость сходимости итерационного метода не зависит от способа задания векторной нормы.

34.36. Средняя за k шагов скорость сходимости линейного итерационного метода по отношению к заданной векторной норме вычисляется по формуле

$$r(Z_k) = -\frac{1}{k} \ln \| Z_k \|,$$

где $\|\cdot\|$ — некоторая произвольная матричная норма, подчиненная соответствующей векторной норме.

34.37. Асимптотическая скорость сходимости линейного итерационного метода вычисляется по формуле

$$r(Z_\infty) = -\varlimsup_{k\to\infty} \frac{1}{k} \ln \| Z_k \|.$$

34.38. Средняя за k шагов скорость сходимости стационарного одношагового итерационного метода с оператором перехода T вычисляется по формуле

$$r_k(T) \equiv r(T) = -\frac{1}{k} \ln L(T^k),$$

а в случае линейного оператора T (постоянной матрицы) — по формуле

$$r_k(T) = -\frac{1}{k} \ln \| T^k \|.$$

34.39. Пусть T — оператор перехода стационарного одношагового итерационного метода, и пусть для некоторой векторной нормы $L(T) < 1$. Тогда для асимптотической скорости сходимости этого метода справедлива оценка

$$r_\infty(T) \equiv \varliminf_{k\to\infty} r_k(T) \geqslant -\ln L(T).$$

34.40. Асимптотическая скорость сходимости стационарного линейного одношагового итерационного метода с матрицей перехода T, для которой $\rho(T) < 1$, вычисляется по формуле

$$r_\infty(T) = -\ln \rho(T).$$

34.41. Пусть оператор перехода T стационарного одношагового итерационного метода ограничен, т. е. $L(T) < +\infty$, и существует такая векторная норма $\|\cdot\|$, по отношению к которой для некоторого целого $s \geqslant 1$ выполняется неравенство

$$L(T^s) < 1.$$

Тогда асимптотическая скорость сходимости этого метода положительна и

$$r_\infty(T) > r_k(T)$$

для всех k, для которых определена его средняя за k шагов скорость сходимости.

Последнее утверждение может показаться читателю несколько неожиданным и даже противоречащим здравому смыслу, поскольку на практике обычно встречается обратная ситуация: наблюдаемая на первых шагах скорость сходимости итерационного метода значительно выше, чем на последующих шагах. Фактически же никакого противоречия нет. Дело в том, что проведенные теоретические рассуждения относятся, так как мы имеем дело с константами Липшица или нормами операторов, к самым плохим из возможных ситуаций. Для стационарных линейных одношаговых итерационных методов эта плохая ситуация соответствует, например, случаю, когда норма матрицы перехода очень сильно превосходит ее спектральный радиус, а вектор начальной ошибки почти совпадает с вектором, на котором достигается максимум отношения $\|Tz\|/\|z\|$.

34.42. Пусть для стационарного одношагового итерационного метода с оператором перехода T по отношению к заданной векторной норме $L(T) < 1$, и пусть существуют такое целое положительное s и такая величина a, равная по модулю $L(T)$, что для некоторого ненулевого вектора z выполняется равенство

$$T^s z = az.$$

Тогда

$$r_\infty(T) = r_{ls}(T) = -\ln L(T)$$

для всех целых $l \geqslant 1$.

34.43. Пусть по отношению к некоторой заданной векторной норме $\|\cdot\|$ константы Липшица разрешающих операторов Z_k итерационного метода для всех $k \geqslant k_0$, где k_0 — некоторое положительное целое, удовлетворяют неравенствам

$$L(Z_k) \leqslant L(Z_{k-1}) \leqslant \ldots \leqslant L(Z_{k_0}) < 1.$$

Тогда для любого положительного $\varepsilon < 1$ наименьшее целое число $k = k_\varepsilon \geqslant k_0$, удовлетворяющее неравенству

$$k \geqslant |\ln \varepsilon|/r(Z_k),$$

задает число шагов этого итерационного метода, достаточное для уменьшения заданной нормы вектора начальной ошибки z^0 в $1/\varepsilon$

раз, т. е. для выполнения неравенства

$$\left\| z^{k_\varepsilon} \right\| \leqslant \varepsilon \| z^0 \|.$$

34.44. Пусть для стационарного одношагового итерационного метода с оператором (матрицей) перехода T по отношению к заданной векторной норме выполняется неравенство $L(T) < 1$ ($\|T\| < 1$). Тогда величина

$$k = \left[\frac{1}{r_1(T)}\right] + 1 = \left[-\frac{1}{\ln L(T)}\right] + 1 \qquad \left(k = \left[-\frac{1}{\ln \| T \|}\right] + 1\right),$$

где $[t]$ — целая часть числа t, задает число шагов этого метода, достаточное для уменьшения нормы начальной ошибки z^0 в e раз (e — основание натуральных логарифмов). Соответственно для любого положительного $\varepsilon < 1$ величина

$$k_\varepsilon = \left[\frac{|\ln \varepsilon|}{r_1(T)}\right] + 1$$

задает число шагов метода, достаточное для уменьшения заданной нормы вектора ошибки z^0 в $1/\varepsilon$ раз, т. е. достаточное (при условии $\|z^0\| = 1$) для решения исходной системы $Au = f$ в этой норме с точностью ε.

Эффективность применения конкретного итерационного метода для решения того или иного класса систем определяется многими факторами. Основными из них являются:

— минимальность количества арифметических действий, требуемых для решения системы с заданной точностью;

— минимальность объема памяти ЭВМ, требуемой для хранения всех используемых величин;

— устойчивость реализации метода в условиях приближенного выполнения различных, в том числе арифметических, операций;

— логическая простота метода с точки зрения составления программы, реализующей его на ЭВМ.

Создание метода, удовлетворяющего наилучшим образом перечисленным требованиям для конкретной системы, которую требуется решить,— мечта каждого математика. Даже при самых идеальных условиях она, как правило, оказывается неосуществимой. Поэтому задачу о выборе из заданного множества методов наиболее эффективного метода решают при различных, часто весьма жестких, ограничениях. В настоящей книге в качестве основной характеристики итерационного метода будет использоваться (весьма условно) величина w_ε, равная количеству арифметических действий, достаточному для уменьшения заданной нормы вектора начальной ошибки в $1/\varepsilon$ раз ($\varepsilon < 1$). При этом неявно предполагается, что имеется устойчивый алгоритм реализации рассматриваемого метода.

Таким образом, при сделанном предположении под проблемой оптимизации конкретного итерационного метода мы будем понимать задачу выбора операторов (или матриц), определяющих этот метод, из некоторого заданного множества с целью минимизации либо самой величины w_ε, либо величины, достаточно хорошо ее приближающей. Конечно, при более полном теоретическом или практическом анализе проблемы оптимизации все сделанные замечания нуждаются в серьезном уточнении. Заметим также, что все дальнейшие рассуждения будут связаны только с p-шаговыми итерационными методами при некоторых значениях $p \geqslant 1$.

34.45. Пусть относительно p-шагового итерационного метода (p — некоторое положительное целое) выполнены следующие предположения:

— операторы F_k, определяющие итерационный метод, выбираются из некоторого множества операторов G_F (разрешающие операторы Z_k принадлежат соответствующему множеству операторов G_Z);

— для любого оператора $\widehat{F} \in G_F$ для нахождения по произвольным векторам $f, v_1, \ldots, v_p \in \mathbf{R}_n$ (при заданной фиксированной матрице A) вектора $v = \widehat{F}(A, f, v_1, \ldots, v_p)$ требуется одно и то же (не зависящее от $f, v_1, \ldots, v_p$) количество w_0 арифметических действий;

— последовательность операторов Z_k удовлетворяет условиям 34.43.

Тогда $w_\varepsilon = k_\varepsilon w_0$, где k_ε определено в 34.43, и проблема оптимизации этого итерационного метода по отношению к множеству операторов G_F одновременно для всех значений $\varepsilon < 1$ заключается для каждого $k \geqslant 1$ в выборе такой последовательности операторов $F_i \in G_F$ (оператора $Z_k \in G_Z$), $i = 1, \ldots, k$, чтобы константа Липшица оператора Z_k была минимальна, или, эквивалентно, средняя за k шагов скорость сходимости была максимальна.

34.46. Пусть матрица перехода T линейного стационарного одношагового итерационного метода зависит (как от параметров) от компонент вектора $\gamma = (\gamma_1, \ldots, \gamma_p)' \in \mathbf{R}_p$ при некотором $p \geqslant 1$, т. е. $T = T(\gamma)$. Тогда, если $\|\cdot\|$ — некоторая норма и $G \subseteq \mathbf{R}_p$ — множество таких векторов γ, для которых $\|T(\gamma)\| < 1$, то задача оптимизации этого итерационного метода за один шаг (в данной норме) заключается в нахождении такого вектора $\gamma_{\text{опт}} \in G$, чтобы

$$\| T(\gamma_{\text{опт}}) \| = \min_{\gamma \in G} \| T(\gamma) \|.$$

34.47. Пусть выполнено предположение 34.46 и $G \subseteq \mathbf{R}_p$ — множество векторов $\gamma \in \mathbf{R}_p$, для которых $\rho(T) < 1$. Тогда задача асимптотической оптимизации соответствующего итерационного метода эквивалентна максимизации его асимптотической скорости сходимости и заключается в нахождении такого вектора $\gamma_{\text{опт}} \in G$, чтобы

$$\rho(T(\gamma_{\text{опт}})) = \min_{\gamma \in G} \rho(T(\gamma)).$$

34.48. Множество из n собственных чисел матрицы T порядка n (каждое отличное от других собственное число берется столько раз, какова его кратность) называется ее *спектром* и обозначается через $\sigma(T)$.

34.49. Задача асимптотической оптимизации линейного стационарного одношагового итерационного метода из 34.47 является минимаксной задачей вида

$$\max_{\lambda \in \sigma(T(\gamma))} |\lambda| = \min_{\gamma \in G}$$

На практике вывод явной формулы для нормы или спектрального радиуса матрицы перехода линейного стационарного одношагового итерационного метода как функции параметров $\gamma_1, \ldots, \gamma_p$ из 34.47, 34.49 возможен лишь в исключительных случаях. Как правило, это связано с тем, что исходной информацией о матрицах, участвующих в данном итерационном методе, являются только некоторые достаточно общие сведения, например, об областях комплексной плоскости, которым принадлежат их спектры. Поэтому с практической точки зрения реально говорить лишь о приближенной оптимизации конкретных итерационных методов. В дальнейшем, если задача оптимизации итерационного метода будет заключаться в минимизации какой-то функции $g(\gamma)$, $\gamma \in G \subseteq \mathbf{R}_p$, то, как правило, мы будем переходить к некоторому множеству $\widehat{G} \subseteq G$ и к некоторой функции $\hat{g}(\gamma)$ такой, что $\widehat{g}(\gamma) \geqslant g(\gamma)$ для $\gamma \in \widehat{G}$, а вместо исходной задачи минимизации — решать задачу нахождения такого $\widehat{\gamma}_{\text{опт}}$, чтобы

$$\widehat{g}\left(\widehat{\gamma}_{\text{опт}}\right) = \min_{\gamma \in \widehat{G}} \widehat{g}(\gamma).$$

Компоненты вектора $\widehat{\gamma}_{\text{опт}}$ мы и будем принимать за приближенное решение исходной задачи оптимизации. Конечно, всегда возможны ситуации, когда найденные значения параметров весьма далеки от искомых оптимальных значений.

34.50. Итерационный метод

$$u^k = u^{k-1} - \tau_k(Au^{k-1} - f),$$

где τ_k — некоторая последовательность параметров, называется *одношаговым методом Ричардсона* (*методом Ричардсона первого порядка*), а в случае, когда $\tau_k \equiv \tau$, т. е. параметры принимают постоянное значение τ, он называется *стационарным* одношаговым методом Ричардсона или *методом простой итерации.*

34.51. Пусть $\tau > 0$ $(\tau < 0)$. Тогда выполнение условия

$$\operatorname{Re} \lambda > 0 \quad (< 0) \qquad \forall \lambda \in \sigma(A)$$

необходимо для сходимости метода простой итерации.

34.52. Пусть $\operatorname{Re} \lambda > 0$ $\forall \lambda \in \sigma(A)$. Тогда существует такое $\hat{\tau} > 0$, что метод простой итерации сходится для всех $\tau \in (0, \hat{\tau})$.

34.53. Пусть все собственные числа матрицы A положительные и $m = \min_{\lambda \in \sigma(A)} \lambda \leqslant M = \max_{\lambda \in \sigma(A)} \lambda$. Тогда

$$\rho(E - \tau A) = \max_{\lambda \in \sigma(A)} |1 - \tau\lambda| = \max\{|1 - \tau m|, |1 - \tau M|\},$$

соответствующий метод простой итерации сходится для любого положительного $\tau < 2/M$, а задача его асимптотической оптимизации решается выбором

$$\tau = \tau_{\text{опт}} = 2/(m + M).$$

34.54. Пусть $\sigma(A) \subset [\delta, \Delta]$, где $\delta \leqslant \Delta$ — положительные числа. Тогда

$$\rho(E - \tau A) \leqslant \max_{\delta \leqslant \lambda \leqslant \Delta} |1 - \tau\lambda| = \max\{|1 - \tau\delta|, |1 - \tau\Delta|\},$$

соответствующий метод простой итерации сходится для любого положительного $\tau < 2/\Delta$, а задача его приближенной асимптотической оптимизации решается выбором

$$\tau = \hat{\tau}_{\text{опт}} = 2/(\delta + \Delta).$$

Таким образом, поскольку $\delta \leqslant m \leqslant M \leqslant \Delta$, в случае метода простой итерации имеем

$$g(\tau) = \max\{|1 - \tau m|, |1 - \tau M|\} \leqslant \hat{g}(\tau) = \max\{|1 - \tau\delta|, |1 - \delta\Delta|\};$$

$$\hat{G} = (0, 2/\Delta) \subseteq G = (0, 2/M).$$

34.55. Пусть все собственные числа матрицы A положительные и она является матрицей простой структуры (т. е. существует такая вещественная матрица Q, что QAQ^{-1} — диагональная матрица с положительными диагональными элементами). Тогда существует такая норма $\|\cdot\|$ (например, D — норма, порождаемая матрицей $D = Q^*Q$), что

$$\|E - \tau A\| = \rho(E - \tau A)$$

для всех положительных значений τ, и, следовательно, задача асимптотической оптимизации соответствующего метода простой итерации совпадает с задачей его оптимизации за один шаг в этой норме.

34.56. Пусть H — некоторая постоянная матрица. Тогда итерационный метод

$$u^k = u^{k-1} - \tau_k H(Au^k - f),$$

где τ_k — некоторая последовательность параметров, называется *обобщенным одношаговым методом Ричардсона* (*обобщенным методом Ричардсона первого порядка*), а в случае, когда $\tau_k \equiv \tau$ (постоянному параметру), он называется *обобщенным стационарным* одношаговым методом Ричардсона или *обобщенным методом простой итерации*.

34.57. Обобщенный одношаговый метод Ричардсона является обычным одношаговым методом Ричардсона, примененным к системе

$$\tilde{A}u = \tilde{f},$$

где $\tilde{A} = HA$, $\tilde{f} = Hf$, т. е. полученной из исходной системы $Au = f$ путем ее умножения на соответствующую матрицу H.

34.58. Процедура умножения исходной системы на некоторую матрицу называется *преобусловливанием* этой системы.

34.59. Условие

$$\ker H \cap \operatorname{im} A = \varnothing$$

необходимо для сходимости обобщенного одношагового метода Ричардсона.

34.60. Пусть выполнено условие 34.59. Тогда все утверждения 34.51—34.55 переносятся на обобщенный метод простой итерации с соответствующей заменой матрицы A на матрицу HA.

34.61. Пусть матрица H — симметричная и положительно определенная, а матрица A — положительно определенная. Тогда $\operatorname{Re}\lambda > 0$ $\forall \lambda \in \sigma(HA)$, и, следовательно, существует такое $\hat{\tau}$, что соответствующий обобщенный метод простой итерации сходится для всех $\tau \in (0, \hat{\tau})$.

34.62. Пусть матрицы A и H — симметричные и положительно определенные. Тогда соответствующий обобщенный метод простой итерации сходится для всех положительных τ, для которых

$$\frac{\tau}{2} A \leqslant B,$$

т. е. при $\tau \in (0, 2/M)$, где $B = H^{-1}$ и $M = \rho(HA)$. Кроме того, задача асимптотической оптимизации этого метода совпадает с задачей его оптимизации за один шаг в B-норме, и их решение достигается выбором $\tau = \tau_{\text{опт}} = 2/(m + M)$, где

$$m = \min_{\lambda \in \sigma(HA)} \lambda \leqslant M = \max_{\lambda \in \sigma(HA)} \lambda.$$

§ 35. Методы релаксации

35.1. Пусть $A = (a_{ij})$ — некоторая матрица с ненулевыми диагональными элементами и $\Lambda = \operatorname{diag}(a_{11}, \ldots, a_{nn})$. Тогда итерационный метод

$$\Lambda u^k = (\Lambda - A) u^{k-1} + f$$

или, эквивалентно,

$$\Lambda(u^k - u^{k-1}) = -(Au^{k-1} - f)$$

называется *точечным методом Якоби* или *методом одновременных смещений.*

35.2. Пусть

$$A = \begin{bmatrix} A_{11} & \ldots & A_{1s} \\ \cdot & \cdot & \cdot \\ A_{s1} & \ldots & A_{ss} \end{bmatrix}$$

— блочная матрица с квадратными невырожденными блоками A_{ii}, $i = 1, \ldots, s$, и

$$\Lambda = A_{11} \oplus A_{22} \oplus \ldots \oplus A_{ss}.$$

Тогда итерационный метод из 35.1 называется *блочным методом Якоби* или *блочным методом последовательных смещений.*

35.3. Если A — матрица со *строгим диагональным преобладанием*, т. е.

$$|a_{ii}| > \sum_{\substack{j=1 \\ i \neq j}}^{n} |a_{ij}|, \qquad i = 1, \ldots, n,$$

то точечный и любой блочный методы Якоби сходятся.

35.4. Матрица A называется матрицей со *слабым диагональным преобладанием*; если

$$|a_{ii}| \geqslant \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|$$

для всех $i = 1, \ldots, n$, причем хотя бы для одного значения i имеет место строгое неравенство.

35.5. Пусть A — неразложимая матрица со слабым диагональным преобладанием. Тогда точечный и любой блочный методы Якоби сходятся.

35.6. Пусть Λ — матрица какого-либо из вариантов метода Якоби и ω — некоторый числовой параметр. Тогда итерационный метод

$$\Lambda u^{k-1/2} = (\Lambda - A) u^{k-1} + f, \quad u^k = \omega u^{k-1/2} + (1 - \omega) u^{k-1}$$

или, эквивалентно,

$$\Lambda (u^k - u^{k-1}) = -\omega (A u^{k-1} - f)$$

называется *экстраполированным методом Якоби.*

35.7. Экстраполированный метод Якоби является частным случаем обобщенного стационарного одношагового метода Ричардсона.

35.8. Пусть A — симметричная положительно определенная матрица. Экстраполированный метод Якоби сходится тогда и только тогда, когда $\omega \in (0, 2/\rho(\Lambda^{-1}A))$. При этом оптимальное значение параметра (в смысле максимума асимптотической скорости сходимости или средней за один шаг в соответствующей Λ-норме) вычисляется по формуле

$$\omega_{\text{опт}} = 2/(m + M),$$

где $m = 1/\rho(A^{-1}\Lambda)$ — минимальное собственное число матрицы $\Lambda^{-1}A$ и $M = \rho(\Lambda^{-1}A)$.

35.9. Пусть $A = (a_{ij})$ — некоторая матрица с ненулевыми диагональными элементами, $\Lambda = \operatorname{diag}(a_{11}, \ldots, a_{nn})$, L — строго нижняя треугольная матрица, U — строго верхняя треугольная матрица и $A = \Lambda - L - U$. Тогда итерационный метод

$$(\Lambda - L)u^k = Uu^{k-1} + f$$

или, эквивалентно,

$$(\Lambda - L)(u^k - u^{k-1}) = -(Au^{k-1} - f)$$

называется *точечным методом Гаусса — Зейделя*, или *точечным методом Зейделя*, или *точечным методом последовательных смещений*.

35.10. Пусть

$$A = \begin{bmatrix} A_{11} & \ldots & A_{1s} \\ \cdot & \cdot & \cdot \\ A_{s1} & \ldots & A_{ss} \end{bmatrix}$$

— некоторая блочная матрица с квадратными невырожденными диагональными блоками A_{ii}, $i = 1, \ldots, s$, и $\Lambda = A_{11} \oplus A_{22} \oplus \ldots \ldots \oplus A_{ss}$. Тогда итерационный метод из 35.9 называется *блочным методом Гаусса — Зейделя*, или *блочным методом Зейделя*, или *блочным методом последовательных смещений*.

35.11. Если A — некоторая матрица со строгим диагональным преобладанием или неразложимая матрица со слабым диагональным преобладанием, то соответствующие точечный и любой блочный методы Гаусса — Зейделя сходятся.

35.12. Пусть выполнены предположения из 35.10, векторы u и f системы $Au = f$ разбиты на блоки в соответствии с разбиением на блоки матрицы A (т. е. $u' = [u_1', u_2', \ldots, u_s']'$ и $f' = [f_1', f_2', \ldots, f_s']'$, где размерности векторов u_i и f_i совпадают с порядками блоков A_{ii}, $i = 1, \ldots, s$) и ω — некоторый отличный от нуля числовой параметр. Тогда итерационный метод

$$A_{ii}u_i^{k-1/2} = -\sum_{j=1}^{i-1} A_{ij}u_j^{k-1/2} - \sum_{j=i+1}^{s} A_{ij}u_j^{k-1} + f_i,$$

$$u_i^k = \omega u_i^{k-1/2} + (1-\omega)u_i^{k-1}, \qquad i = 1, \ldots, s,$$

или, эквивалентно,

$$\left(\frac{1}{\omega}\Lambda - L\right)(u^k - u^{k-1}) = -(Au^{k-1} - f)$$

называется *методом последовательной верхней релаксации* (в зарубежной литературе для метода последовательной верхней релаксации принято обозначение SOR — successive overrelaxation method), а параметр ω — *релаксационным параметром*. Иногда, если все блоки матрицы имеют порядок единица, метод последовательной верхней релаксации называется *точечным*, в противном случае — *блочным*.

Метод Гаусса — Зейделя является частным случаем метода последовательной верхней релаксации при $\omega = 1$. В свою очередь, метод последовательной верхней релаксации иногда называли *экстраполированным методом Гаусса — Зейделя* (или *экстраполированным методом Зейделя*).

35.13. Пусть A — матрица со строгим диагональным преобладанием или неразложимая матрица со слабым диагональным преобладанием. Тогда существует такое $\hat{\omega} > 1$, что метод последовательной верхней релаксации сходится для любого $\omega \in (0, \hat{\omega})$.

Раньше в литературе метод 35.12 для положительных значений параметра $\omega < 1$ называли методом *нижней релаксации*, для значений $\omega > 1$ — методом *верхней релаксации*, а для значения $\omega = 1$ — методом *полной релаксации*. Однако впоследствии, так как для подавляющего числа практически важных задач оказалось (в ряде случаев это строго доказано), что для определенных значений $\omega > 1$ метод сходится быстрее, чем для любого из значений $\omega < 1$, за методом для любых значений ω закрепилось название метода последовательной верхней релаксации.

35.14. Пусть A — симметричная положительно определенная матрица. Тогда векторы ошибок z^k метода последовательной верхней релаксации удовлетворяют соотношениям

$$\|z^k\|_A^2 = \|z^{k-1}\|_A^2 - \left(\frac{2}{\omega} - 1\right)\left\|\left(\frac{1}{\omega}\Lambda - L\right)^{-1} Az^{k-1}\right\|_\Lambda^2.$$

35.15. Если матрица A симметричная и положительно определенная, то согласно 34.27 и 35.14 соответствующий метод последовательной верхней релаксации из 35.12 сходится тогда и только тогда, когда $\omega \in (0, 2)$.

35.16. Пусть симметричная матрица A представлена в виде

$$A = \Lambda - L - L',$$

где Λ — симметричная положительно определенная матрица, L — некоторая, вообще говоря, произвольная матрица, и пусть ω — некоторый ненулевой числовой параметр, $\det\left(\frac{1}{\omega}\Lambda - L\right) \neq 0$. Тогда итерационный метод

$$\left(\frac{1}{\omega}\Lambda - L\right)(u^k - u^{k-1}) = -(Au^{k-1} - f)$$

называется *обобщенным методом последовательной верхней релаксации* (соответственно при $\omega = 1$ он называется *обобщенным методом Гаусса — Зейделя*).

35.17. Векторы ошибок z^k обобщенного метода последовательной верхней релаксации удовлетворяют соотношениям

$$(Az^k, z^k) = (Az^{k-1}, z^{k-1}) - \left(\frac{2}{\omega} - 1\right)\left\|\left(\frac{1}{\omega}\Lambda - L\right)^{-1} Az^{k-1}\right\|_\Lambda^2.$$

35.18. Согласно 34.27 и 35.17 обобщенный метод последовательной верхней релаксации сходится тогда и только тогда, когда либо матрица A положительно определена и $\omega \in (0, 2)$, либо матрица A отрицательно определена и $\omega \notin [0, 2]$.

35.19. В обоих случаях, когда выполняются условия сходимости обобщенного метода последовательной верхней релаксации из 35.18, $\det\left(\frac{1}{\omega}\Lambda - \dot{L}\right) \neq 0$.

35.20. Если матрица A отрицательно определенная и $\omega \notin [0, 2]$, то обобщенный метод последовательной верхней релаксации может быть переписан в виде

$$\left(\frac{1}{\tilde{\omega}}\Lambda - \tilde{L}\right)(u^k - u^{k-1}) = -(\tilde{A}u^{k-1} - \tilde{f}),$$

где $\tilde{A} = -A$, $\tilde{f} = f$ и $\tilde{L} = \Lambda - L$, а релаксационный параметр $\tilde{\omega}$ связан с релаксационным параметром ω соотношением $1/\tilde{\omega} = 1 - 1/\omega$, и, следовательно, $\omega \in (0, 2)$.

Таким образом, обобщенный метод последовательной верхней релаксации с параметром $\omega \notin [0, 2]$ для системы $Au = f$ с отрицательно определенной матрицей A эквивалентен обобщенному методу последовательной верхней релаксации с параметром $\tilde{\omega} \in (0, 2)$ применительно к преобразованной системе, но уже с положительно определенной матрицей $\tilde{A} = -A$.

35.21. Матрица перехода $T = T_\omega$ обобщенного метода последовательной верхней релаксации может быть записана в виде

$$T_\omega = (E - \omega\tilde{L})^{-1}[(1 - \omega)E + \omega\tilde{U}],$$

где $\tilde{L} = \Lambda^{-1}L$, $\tilde{U} = \Lambda^{-1}U$, $U = \dot{L}'$. Соответственно матрица перехода $T = T_1$ обобщенного метода Гаусса — Зейделя имеет вид

$$T_1 = (\Lambda - L)^{-1}U = (E - \tilde{L})\tilde{U}.$$

35.22. Пусть матрицы $\tilde{L}$ и $\tilde{U}$ из 35.21 являются соответственно строго нижней треугольной и строго верхней треугольной и $\lambda_i = \lambda_i(\omega)$ — собственные числа матрицы T_ω. Тогда

$$\prod_{i=1}^{n} \lambda_i(\omega) = \det T_\omega = (1 - \omega)^n,$$

и, следовательно, для любого $\omega \notin (0, 2)$ выполняется неравенство

$$\rho(T_\omega) \geqslant |1 - \omega|.$$

35.23. Если матрицы $\tilde{L}$ и $\tilde{U}$ из 35.21 являются соответственно строго нижней треугольной и строго верхней треугольной, то обобщенный метод последовательной верхней релаксации не сходится для значений $\omega \notin (0, 2)$.

35.24. Если матрица A симметричная и отрицательно определенная, то, поскольку выводы из 35.18 и 35.23 вступают в противоречие, для нее невозможно построить разложение

$$A = \Lambda - L - L'$$

с симметричной и положительно определенной матрицей Λ так, чтобы матрицы $\Lambda^{-1}L$ и $\Lambda^{-1}L'$ были соответственно строго нижней и строго верхней треугольными.

35.25. Матрица перехода обобщенного метода последовательной верхней релаксации при $\omega = 2$ является матрицей простой структуры, и все ее собственные числа равны по модулю единице.

35.26. Пусть $\tilde{L}$ и $\tilde{U}$ — соответственно строго нижняя и строго верхняя треугольные матрицы. Тогда матрица

$$T_0 = \tilde{L} + \tilde{U}$$

называется *согласованно упорядоченной*, если собственные числа матрицы

$$T_{0\alpha} = \alpha\tilde{L} + \frac{1}{\alpha}\tilde{U}$$

не зависят от числового параметра α, т. е.

$$\det (T_{0\alpha} - \lambda E) = \det (T_0 - \lambda E)$$

для любого $\alpha \neq 0$.

35.27. Пусть T — блочно трехдиагональная матрица вида (s — некоторое положительное целое)

$$T = \begin{bmatrix} \theta_1 & U_1 & & 0 \\ L_2 & \cdot & & \\ & & \cdot & \\ & & \cdot & U_{s-1} \\ 0 & & L_s & \theta_s \end{bmatrix},$$

где θ_i — квадратные матрицы порядка n_i, $i = 1, \ldots, s$, L_i и U_{i-1} — соответственно матрицы размеров $n_i \times n_{i+1}$ и $n_{i-1} \times n_i$, $i = 2, \ldots$ $\ldots, s$, и $n_1 + \ldots + n_s = n$. Тогда матрица T является согласованно упорядоченной.

35.28. Пусть T — циклическая индекса 2 матрица. Тогда существует такая матрица перестановок P, что матрица PTP' является согласованно упорядоченной.

35.29. Говорят, что матрица B обладает свойством «А» Янга, если существует такая матрица перестановок P, что матрица PBP' является блочно трехдиагональной матрицей вида (s — некоторое положительное целое)

$$PBP' = \begin{bmatrix} B_1 & -U_1 & & 0 \\ -L_2 & \cdot & & \\ & & \cdot & \\ & & \cdot & -U_{s-1} \\ 0 & & -L_s & B_s \end{bmatrix},$$

где B_i — диагональные матрицы порядка n_i, а размеры матриц L_i и U_i те же, что и в 35.27, $i = 1, \ldots, s$.

35.30. Пусть Ω — некоторое множество n различных пар индексов (i, j) и задана система пятиточечных уравнений

$$u_{ij} - b_{ij}u_{i-1,j} - c_{ij}u_{i+1,j} - d_{ij}u_{i,j-1} - e_{ij}u_{i,j+1} = f_{ij}, \quad (i, j) \in \Omega,$$

причем $u_{ij} = 0$ для $(i, j) \notin \Omega$. Тогда, если объединить в первую группу неизвестные, сумма индексов которых четна, а во вторую — сумма индексов которых нечетна, то мы приходим к системе $Au = f$ с блочной матрицей

$$A = \begin{bmatrix} E_1 & A_{12} \\ A_{21} & E_2 \end{bmatrix},$$

которая обладает свойством «А» Янга. Здесь E_1 — единичная матрица порядка n_1 (n_1 равно числу пар $(i, j) \in \Omega$ с четной суммой $i + j$) и E_2 — единичная матрица порядка $n_2 = n - n_1$, а A_{12} и A_{21} — некоторые матрицы соответственно размеров $n_1 \times n_2$ и $n_2 \times n_1$, имеющие в каждой строке и каждом столбце не более четырех ненулевых элементов.

35.31. Пусть выполнены предположения 35.30. Тогда, если распределить все неизвестные по s группам (s — некоторое положительное целое) так, чтобы в каждую группу входили неизвестные только с одной и той же суммой индексов (для разных групп различной), а группы расположить в порядке возрастания значений этих сумм, то мы придем к системе $Au = f$ с блочно трехдиагональной матрицей

$$A = E - T$$

(где T — матрица из 35.27), обладающей свойством «А» Янга.

35.32. Если A — блочно трехдиагональная матрица с невырожденными диагональными блоками, то матрица перехода T_0 соответствующего блочного метода Якоби является согласованно упорядоченной.

35.33. Пусть матрицы $\tilde{L} = \Lambda^{-1}L$ и $\tilde{U} = \Lambda^{-1}U$ обобщенного метода последовательной верхней релаксации являются соответственно строго нижней и строго верхней треугольными и $T_0 = \tilde{L} + \tilde{U}$ — согласованно упорядоченная матрица. Тогда

$$\det(T_\omega + \lambda E) = \det[(1 - \omega + \lambda)E - \omega\lambda^{1/2}(\tilde{L} + \tilde{U})],$$

и, следовательно,

$$\lambda_i - \omega + 1 = \omega\lambda_i^{1/2}\mu_i,$$

где μ_i — собственные числа матрицы T_0, а $\lambda_i = \lambda_i(\omega)$ — соответствующие собственные числа матрицы T_ω, $i = 1, \ldots, n$.

35.34. Если величина μ является собственным числом согласованно упорядоченной матрицы T_0, то величина $-\mu$ также является собственным числом T_0.

35.35. При выполнении предположений 35.33 собственные числа μ матрицы T_0 и собственные числа λ матрицы T_ω связаны соотношением

$$(1-\omega-\lambda)^2=\lambda\mu^2\omega^2,$$

причем если μ является собственным числом матрицы T_0, то оба значения λ, удовлетворяющие этому соотношению, являются собственными числами T_ω, а если λ является собственным числом T_ω, то величины $\mu=\pm\frac{1-\omega-\lambda}{\lambda^{1/2}\omega}$ являются собственными числами матрицы T_0.

35.36. Если матрица перехода T_0 блочного метода Якоби согласованно упорядоченная, то

$$\rho^2(T_0)=\rho(T_1),$$

где T_1 — матрица перехода соответствующего блочного метода Гаусса — Зейделя, т. е. оба метода сходятся только одновременно. Более того, в данном случае

$$r_\infty(T_1)=2r_\infty(T_0),$$

т. е. метод Гаусса — Зейделя сходится в два раза асимптотически быстрее, чем метод Якоби.

35.37. Пусть выполнены предположения 35.33, все собственные числа матрицы T_0 вещественны и $\rho(T_0)<1$. Тогда значение ω, минимизирующее $\rho(T_\omega)$ и, следовательно, являющееся оптимальным с точки зрения максимальности асимптотической скорости сходимости метода последовательной верхней релаксации, вычисляется по формуле

$$\omega_{\text{опт}}=2/(1+\sqrt{1-\rho^2(T_0)}).$$

При этом $\rho\left(T_{\omega_{\text{опт}}}\right)=\omega_{\text{опт}}-1$ и, соответственно,

$$r_\infty\left(T_{\omega_{\text{опт}}}\right)=-\ln(\omega_{\text{опт}}-1).$$

35.38. Пусть выполнены предположения 35.33, $\rho(T_0)<1$ и $\omega_{\text{опт}}$ вычисляется по формуле 35.37. Тогда

$$\frac{d}{d\omega}\rho(T_\omega)<0,\quad \omega\in(0,\omega_{\text{опт}}),$$

$$\frac{d}{d\omega}\rho(T_\omega)=1,\quad \omega\in(\omega_{\text{опт}},2),$$

и, следовательно, для любых ω_2 и ω_1, удовлетворяющих либо неравенствам $0<\omega_1<\omega_2\leqslant\omega_{\text{опт}}$, либо $\omega_{\text{опт}}\leqslant\omega_2<\omega_1<2$, выполняется неравенство

$$r_\infty\left(T_{\omega_1}\right)<r_\infty\left(T_{\omega_2}\right).$$

35.39. Пусть выполнены предположения 35.33 и $\rho(T_0)=1-h$, где $0<h\ll 1$. Тогда

$$\rho\left(T_{\omega_{\text{опт}}}\right)=1-2\sqrt{2}\,h^{1/2}+O\left(h^{3/2}\right).$$

В частности,

$$\lim_{h\to 0} \frac{r_\infty\left(T_{\omega_{\text{опт}}}\right)}{r_\infty^{1/2}\left(T_0\right)} = 2\sqrt{2}.$$

35.40. Пусть A — симметричная положительно определенная блочно трехдиагональная матрица. Тогда для соответствующего блочного метода последовательной верхней релаксации справедливы все утверждения 35.33—35.39.

35.41. Пусть задана система уравнений

$$A_{11}u_1 + A_{12}u_2 = f_1,$$
$$A_{21}u_1 + A_{22}u_2 = f_2,$$

где A_{ij} — матрицы размеров $n_i \times n_j$, а $u_i, f_i \in \mathbf{R}_{n_i}$, $i = 1, 2$, $(n_1 + n_2) = n$, и заданы некоторые вещественные параметры ω и ω'. Тогда, если матрицы A_{11} и A_{22} невырожденные, то итерационный метод

$$A_{11}u_1^k = \omega\left(-A_{12}u_2^{k-1} + f_1\right) + (1-\omega)A_{11}u_1^{k-1},$$
$$A_{22}u_2^k = \omega'\left(-A_{21}u_1^k + f_2\right) + (1-\omega')A_{22}u_2^{k-1}$$

или, эквивалентно,

$$\frac{1}{\omega}A_{11}\left(u_1^k - u_1^{k-1}\right) = -\left(A_{11}u_1^{k-1} + A_{12}u_2^{k-1} - f_1\right),$$
$$\frac{1}{\omega'}A_{22}\left(u_2^k - u_2^{k-1}\right) = -\left(A_{21}u_1^{k-1} + A_{22}u_2^k - f_2\right)$$

называется *модифицированным методом последовательной верхней релаксации.*

35.42. Матрица перехода $T = T_{\omega,\omega'}$ модифицированного метода последовательной верхней релаксации имеет вид

$$T_{\omega,\omega'} = \begin{bmatrix} E_1 & 0 \\ -\omega'\widetilde{L}_2 & E_2 \end{bmatrix}^{-1} \begin{bmatrix} (1-\omega)E_1 & \omega\widetilde{U}_1 \\ 0 & (1-\omega')E_2 \end{bmatrix},$$

где $\widetilde{L}_2 = A_{22}^{-1}A_{21}$ и $\widetilde{U}_1 = A_{11}^{-1}A_{12}$, а E_1 и E_2 — единичные матрицы соответственно порядков n_1 и n_2.

35.43. Собственные числа матрицы $T_{\omega,\omega'}$ из 35.42 и собственные числа μ матрицы перехода

$$T_0 = -\begin{bmatrix} 0 & \widetilde{U}_1 \\ \widetilde{L}_2 & 0 \end{bmatrix}$$

соответствующего блочного метода Якоби связаны соотношением

$$(\lambda + \omega - 1)(\lambda + \omega' - 1) = \omega\omega'\mu^2\lambda.$$

Более того, если $\mu \neq 0$ является собственным числом матрицы T_0, то оба корня λ этого уравнения являются собственными числами $T_{\omega,\omega'}$ (для $\mu = 0$ либо одна из величин $\omega - 1$ и $\omega' - 1$ является собственным числом $T_{\omega,\omega'}$, либо обе величины); если же λ является собственным числом матрицы $T_{\omega,\omega'}$, то величины

$$\mu = \pm \lambda^{1/2} \sqrt{\frac{(\lambda + \omega - 1)(\lambda + \omega' - 1)}{\omega \omega'}}$$

одновременно являются собственными числами матрицы T_0.

35.44. Если все собственные числа матрицы T_0 из 35.43 вещественные и $\rho(T_0) < 1$, то модифицированный метод последовательной верхней релаксации сходится для любых значений ω, $\omega' \in (0, 2)$.

35.45. Если матрица

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

из 35.41 симметричная и положительно определенная, то соответствующий модифицированный метод последовательной верхней релаксации сходится для любых ω, $\omega' \in (0, 2)$.

35.46. Если собственные числа матрицы T_0 вещественные и $\rho(T_0) < 1$, то

$$\rho(T_{\omega,\omega'}) > \rho(T_{\omega_{\text{опт}}})$$

для любых значений ω и ω', где T_ω — матрица перехода соответствующего блочного метода последовательной верхней релаксации, а $\omega_{\text{опт}}$ вычисляется по формуле из 35.37.

В этом случае оптимальным (в смысле максимума асимптотической скорости сходимости) вариантом модифицированного метода последовательной верхней релаксации является выбор $\omega = \omega' = \omega_{\text{опт}}$, и, следовательно, нет смысла вводить в метод последовательной верхней релаксации различные параметры.

35.47. Пусть $A = \Lambda - L - L'$ — симметричная положительно определенная матрица, где Λ также симметричная и положительно определенная. Тогда итерационный метод

$$\left(\frac{1}{\omega}\Lambda - L\right)(u^{k-1/2} - u^{k-1}) = -(Au^{k-1} - f),$$

$$\left(\frac{1}{\omega}\Lambda - L'\right)(u^{k} - u^{k-1/2}) = -(Au^{k-1/2} - f)$$

или, эквивалентно,

$$\left(\frac{1}{\omega}\Lambda - L\right)\Lambda^{-1}\left(\frac{1}{\omega}\Lambda - L'\right)(u^{k} - u^{k-1}) = -\left(\frac{2}{\omega} - 1\right)(Au^{k-1} - f)$$

называется *методом симметричной последовательной верхней релаксации* (в зарубежной литературе для него используется обозначение SSOR).

35.48. Если ввести обозначения $A_1 = \frac{1}{2}\Lambda - L$, $A_2 = A_1'$ и $\tau = 2\omega/(2-\omega)$, то метод симметричной последовательной верхней релаксации можно переписать в виде

$$(\Lambda - \tau A_1)\Lambda^{-1}(\Lambda + \tau A_2)(u^k - u^{k-1}) = -2\tau(Au^{k-1} - f).$$

35.49. Для векторов ошибок z^k метода симметричной последовательной верхней релаксации справедливо соотношение

$$\|z^k\|_A^2 = \|z^{k-1}\|_A^2 - \left(\frac{2}{\omega} - 1\right)\left\|\left(\frac{1}{\omega}\Lambda - L\right)^{-1} Az^{k-1}\right\|_\Lambda^2 - \\ - \left(\frac{2}{\omega} - 1\right)\left\|\left(\frac{1}{\omega}\Lambda - L'\right)^{-1} Az^{k-1/2}\right\|_\Lambda^2,$$

и, следовательно, метод сходится тогда и только тогда, когда $\omega \in (0, 2)$. Соответственно метод из 35.48 сходится тогда и только тогда, когда $\tau > 0$.

35.50. Матрица перехода $T(\omega) = E - H(\omega)A$ метода симметричной последовательной верхней релаксации, где

$$H(\omega) = \left(\frac{2}{\omega} - 1\right)\left(\frac{1}{\omega}\Lambda - L'\right)^{-1}\Lambda\left(\frac{1}{\omega}\Lambda - L\right)^{-1}$$

— симметричная матрица, подобна симметричной матрице $\tilde{T}(\omega) = E - A^{1/2}H(\omega)A^{1/2}$, т. е. является самосопряженной в смысле скалярного произведения, порождаемого матрицей A. Следовательно, при сделанных ранее предположениях

$$\|T(\omega)\|_A = \rho(\tilde{T}(\omega)).$$

35.51. Все собственные числа матрицы $T(\omega)$ из 35.50 вещественные, и она обладает полной A-ортонормированной системой собственных векторов.

35.52. Так как матрица

$$T(\omega)|_{\omega=1} = (\Lambda - L')^{-1}\Lambda(\Lambda - L)^{-1}L\Lambda^{-1}L'$$

подобна симметричной положительно полуопределенной (или положительно определенной, если $\det L \neq 0$) матрице

$$\Lambda^{1/2}(\Lambda - L)^{-1}L\Lambda^{-1}L'(\Lambda - L')^{-1}\Lambda^{1/2},$$

то все ее собственные числа неотрицательны (положительны). Соответственно собственные числа матрицы $H(\omega)A$ при $\omega = 1$ положительны и не превосходят единицы.

35.53. Асимптотическая оптимизация метода симметричной последовательной верхней релаксации эквивалентна минимизации A-нормы его матрицы перехода $T(\omega)$.

35.54. Пусть

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

— симметричная положительно определенная матрица с квадратными блоками A_{11} и A_{22}, $\Lambda = A_{11} \oplus A_{22}$ и

$$L = \begin{bmatrix} 0 & 0 \\ -A_{21} & 0 \end{bmatrix}.$$

Тогда для соответствующего метода симметричной последовательной верхней релаксации

$$\min_{\omega \in (0,2)} \| T(\omega) \|_\Lambda = \| T(1) \|_\Lambda,$$

т. е. оптимальное значение параметра ω равно единице.

Таким образом, для матриц A из 35.54 оптимальный метод симметричной последовательной верхней релаксации имеет скорость сходимости (правда, среднюю за один шаг), совпадающую с асимптотической скоростью сходимости блочного метода Гаусса — Зейделя. В то же время он требует на каждый шаг в два раза больше арифметических действий, чем метод Гаусса — Зейделя.

35.55. Пусть выполнены предположения 35.47,

$$\tilde{L} = \Lambda^{-1} L, \quad \tilde{U} = \Lambda^{-1} L', \quad T_1 = (\Lambda - L)^{-1} U,$$
$$\nu^2 = \rho(T_1) < 1, \ \gamma = \max(\rho(\tilde{L}\tilde{U}), 1/4),$$

и пусть

$$\hat{\omega} = \frac{2}{1 + (1 - 2\nu + 4\gamma)^{1/2}} \in (0, 2).$$

Тогда для соответствующего метода симметричной последовательной верхней релаксации справедлива оценка

$$\rho(T(\hat{\omega})) \leqslant \frac{1 - q^{1/2}}{1 + q^{1/2}},$$

где $q = (1 - \nu)^2/(1 - 2\nu + 4\gamma) < 1$.

35.56. Пусть выполнены предположения 35.55 и $\rho(\tilde{L}\tilde{U}) \leqslant 1/4$. Тогда

$$\hat{\omega} = \frac{2}{1 + 2(1 - \nu)^{1/2}}, \quad q = \frac{1 - \nu}{2}.$$

Кроме того, если $\nu = 1 - h$, где $0 < h \ll 1$, то

$$\lim_{h \to 0} \frac{r_\infty(T(\hat{\omega}))}{\sqrt{2}\, h^{1/2}} = 1.$$

35.57. Пусть выполнены все предположения 35.56, матрица $T_0 = \tilde{L} + \tilde{U}$ согласованно упорядоченная и параметр $\omega_{\text{опт}}$ соответствующего обобщенного метода последовательной верхней релаксации вычисляется по формуле из 35.37, т. е. $\omega_{\text{опт}} = 2/\left(1 + \sqrt{1 - \nu^2}\right)$. Тогда

$$\lim \frac{r_\infty(T_{\omega_{\text{опт}}})}{r_\infty(T(\hat{\omega}))} < 2.$$

§ 36. Итерационные методы для систем с монотонными матрицами

36.1. Матрица A называется *монотонной*, если любой вектор x, для которого $Ax \geqslant 0$, является неотрицательным.

36.2. Квадратная матрица A монотонна тогда и только тогда, когда она невырожденная и $A^{-1} \geqslant 0$.

36.3. Пусть A — некоторая монотонная матрица и для нее имеет место *расщепление* $A = B - C$, где B — неособенная матрица, $B^{-1} \geqslant 0$ и $C \geqslant 0$. Тогда это расщепление матрицы A называется *регулярным*.

36.4. Пусть A — монотонная матрица и $A = B - C$ — ее регулярное расщепление. Тогда

$$\rho(B^{-1}C) = \rho(A^{-1}C)/(1 + \rho(A^{-1}C)) < 1.$$

36.5. Пусть выполнены предположения 36.4. Тогда итерационный метод

$$Bu^k = Cu^{k-1} + f$$

или, эквивалентно,

$$B(u^k - u^{k-1}) = -(Au^{k-1} - f),$$

применяемый для решения системы $Au = f$, сходится.

36.6. Пусть $A = B_1 - C_1$ и $A = B_2 - C_2$ — два регулярных расщепления монотонной матрицы A, причем $C_1 \leqslant C_2$. Тогда

$$\rho(B_1^{-1}C_1) \leqslant \rho(B_2^{-1}C_2) < 1,$$

и, следовательно, метод из 36.5 сходится для первого расщепления асимптотически не медленнее, чем для второго.

36.7. Пусть выполнены предположения 36.4 и матрицы A и C симметричны. Тогда

$$\rho(B^{-1}C) \leqslant \rho(A^{-1})\rho(C)/(1 + \rho(A^{-1})\rho(C)) < 1.$$

36.8. Матрица $A = (a_{ij})$ называется *М-матрицей*, если она невырожденная, $a_{ij} \leqslant 0$ для всех $i \neq j$ и $A^{-1} \geqslant 0$.

36.9. M-матрица является монотонной матрицей.

36.10. Все диагональные элементы M-матрицы положительны.

36.11. Треугольная матрица, у которой все диагональные элементы положительные, а недиагональные элементы неположительные, является M-матрицей.

36.12. Пусть A — разложимая матрица и

$$PAP' = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1s} \\ & A_{22} & \dots & A_{2s} \\ & 0 & \ddots & \\ & & & A_{ss} \end{bmatrix}$$

— ее нормальная форма из 18.23, причем $0 \leqslant A_{ij}$ для всех $i \neq j$. Тогда A в том и только в том случае является M-матрицей, если блоки A_{ii}, $i = 1, \dots, s$, являются M-матрицами.

36.13. Пусть $A = aE - S$, где $S \geqslant 0$ и $a > \rho(S)$. Тогда

$$A^{-1} = \frac{1}{a}\sum_{i=0}^{\infty}\left(\frac{1}{a}S\right)^i \geqslant 0,$$

и, следовательно, A является M-матрицей.

36.14. Пусть $A = aE - S$, где $S \geqslant 0$ и a — некоторое положительное число. Тогда A в том и только в том случае является M-матрицей, если $a > \rho(S)$.

36.15. Пусть недиагональные элементы матрицы A — неположительные. Тогда любые два из следующих утверждений эквивалентны:

а) A является M-матрицей;

б) для A существует представление $A = aE - S$, где $S \geqslant 0$ и $a > \rho(S)$;

в) $\operatorname{Re}\lambda > 0$ для любого $\lambda \in \sigma(A)$;

г) все главные миноры матрицы A положительные;

д) все ведущие миноры матрицы A положительные;

е) существует такой положительный вектор v, что вектор $A^{-1}v$ положительный;

ж) существуют такие нижняя треугольная матрица L и верхняя треугольная матрица U, являющиеся M-матрицами, что $A = LU$.

36.16. Матрица A со строгим диагональным преобладанием и неположительными недиагональными элементами является M-матрицей.

36.17. Неразложимая матрица A со слабым диагональным преобладанием и неположительными недиагональными элементами является M-матрицей. Более того, все элементы матрицы A^{-1} положительные.

36.18. Пусть A является неразложимой M-матрицей. Тогда, если $A = B_1 - C_1$ и $A = B_2 - C_2$ — два регулярных разложения A и $C_1 \leqslant C_2$, причем $C_1 \neq C_2$, то

$$\rho\left(B_1^{-1}C_1\right) < \rho\left(B_2^{-1}C_2\right) < 1,$$

т. е. итерационный метод из 36.5 для первого расщепления асимптотически сходится быстрее, чем для второго.

36.19. Пусть A_1 — некоторая M-матрица, матрица $A_2 \geqslant A_1$ и ее недиагональные элементы неположительны. Тогда A_2 также является M-матрицей и $A_2^{-1} \leqslant A_1^{-1}$.

36.20. Пусть A — некоторая M-матрица и матрица B получена из A путем замены некоторых ее ненулевых недиагональных элементов нулями. Тогда B также является M-матрицей и $B^{-1} \leqslant A^{-1}$

36.21. Пусть матрица B получена из M-матрицы A путем замены некоторых ее недиагональных элементов нулями и $C = B - A$. Тогда расщепление $A = B - C$ матрицы A является регулярным и, следовательно, соответствующий итерационный метод из 36.5 сходится.

36.22. Точечные и любые блочные методы Якоби и Гаусса — Зейделя для систем с M-матрицами A сходятся. При этом асимптотические скорости сходимости методов Якоби не выше, а в случае, когда матрица A неразложима, ниже, чем асимптотические скорости сходимости соответствующих методов Гаусса — Зейделя.

36.23. Пусть матрица $A = E - L - U$ является M-матрицей, где L и U — строго нижняя и строго верхняя треугольные матрицы соответственно. Тогда расщепление $A = B_\omega - C_\omega$, где

$$B_\omega = \frac{1}{\omega} E - L, \quad C_\omega = \left(\frac{1}{\omega} - 1\right) E + U,$$

матрицы A является регулярным для любого $\omega \in (0, 1]$, и, следовательно, для этих значений ω

$$\rho\left(B_\omega^{-1} C_\omega\right) < 1.$$

36.24. Пусть выполнены предположения 36.23. Тогда существует такое значение $\tilde{\omega} > 1$, что соответствующий метод последовательной верхней релаксации сходится для всех значений $\omega \in (0, \tilde{\omega})$. Кроме того, если $0 < \omega_1 < \omega_2 \leqslant 1$, то, поскольку $C_{\omega_1} \geqslant C_{\omega_2}$, имеем $\rho\left(A^{-1} C_{\omega_1}\right) \geqslant \rho\left(A^{-1} C_{\omega_2}\right)$, и, следовательно,

$$\rho\left(B_{\omega_2}^{-1} C_{\omega_2}\right) < \rho\left(B_{\omega_1}^{-1} C_{\omega_1}\right) < 1,$$

т. е. метод для значения $\omega = \omega_2$ асимптотически сходится быстрее, чем для значения $\omega = \omega_1$.

36.25. Пусть выполнены предположения 36.23 и матрица A неразложимая. Тогда

$$\frac{d}{d\omega} \rho\left(B_\omega^{-1} C_\omega\right) < 0,$$

и, следовательно, значение $\omega_{\text{опт}}$, которое максимизирует асимптотическую скорость сходимости соответствующего метода последовательной верхней релаксации, больше единицы.

36.26. Пусть $A = B - C$ — регулярное расщепление M-матрицы A. Тогда матрица

$$\tilde{A} = B^{-1} A = E - B^{-1} C$$

также является M-матрицей.

36.27. Пусть $A = \Lambda - C$ — регулярное расщепление M-матрицы A, $C = L + U$, а $\tilde{L} = \Lambda^{-1} L$ и $\tilde{U} = \Lambda^{-1} U$ являются строго нижней и строго верхней треугольными матрицами соответственно. Тогда обобщенные методы Якоби

$$\Lambda u^k = (L + U) u^{k-1} + f$$

и Гаусса — Зейделя

$$(\Lambda - L)u^k = Uu^{k-1} + f$$

сходятся, причем обобщенный метод Гаусса — Зейделя сходится не медленнее, а в случае неразложимости матрицы $\Lambda^{-1}C$ — быстрее, чем обобщенный метод Якоби.

36.28. Пусть выполнены предположения 36.27. Тогда существует такое $\tilde{\omega} > 1$, что соответствующий обобщенный метод последовательной верхней релаксации

$$B_\omega u^k = C_\omega u^{k-1} + f,$$

где $B_\omega = \frac{1}{\omega}\Lambda - L$, $C_\omega = \left(\frac{1}{\omega} - 1\right)\Lambda - U$, сходится для любого $\omega \in (0, \tilde{\omega})$. Кроме того, если $\omega_1 < \omega_2 \leqslant 1$, то

$$\rho\left(B_{\omega_1}^{-1}C_{\omega_1}\right) > \rho\left(B_{\omega_2}^{-1}C_{\omega_2}\right),$$

и, следовательно, при $\omega = \omega_2$ метод асимптотически сходится быстрее, чем при $\omega = \omega_1$, а значение параметра $\omega_{\text{опт}}$ больше или равно единице.

36.29. Если выполнены предположения 36.27 и матрица $\Lambda^{-1}C$ неразложимая, то значение параметра $\omega_{\text{опт}}$, максимизирующего асимптотическую скорость сходимости соответствующего обобщенного метода последовательной верхней релаксации, больше единицы.

36.30. Пусть A — блочная M-матрица из 35.2. Тогда для соответствующего блочного метода последовательной верхней релаксации справедливы все выводы 36.27 и 36.28, а если матрица A неразложимая, то — и вывод предыдущего пункта.

36.31. Симметричная M-матрица A называется *матрицей Стилтьеса.*

36.32. Симметричная матрица, у которой недиагональные элементы неположительные, тогда и только тогда является матрицей Стилтьеса, когда она положительно определенная.

36.33. Если симметричная матрица B получена из матрицы Стилтьеса A путем замены нулями некоторых ее недиагональных элементов, то она тоже является матрицей Стилтьеса.

§ 37. Методы расщепления (методы переменных направлений)

37.1. Пусть $A = \sum_{i=1}^{p} A_i$, где $p \geqslant 2$, A_i — некоторые квадратные матрицы, $i = 1, \ldots, p$. Тогда итерационный метод

$$B_{\tau_k}\left(u^k - u^{k-1}\right) = -\alpha_k\left(Au^{k-1} - f\right)$$

с матрицей $B_{\tau_k} = \prod\limits_{i=1}^{p} \left(E + \tau_i^{(k)} A_i\right)$ и последовательностями параметров $\tau_i^{(k)}$, $i = 1, \ldots, p$, α_k называется *методом расщепления* (*методом переменных направлений*).

Методы переменных направлений были предложены в пятидесятые годы для решения систем линейных уравнений, возникающих при дискретизации уравнений математической физики методом сеток (в зарубежной литературе для них используется обозначение ADI — Alternating Direction Implicite). В дальнейшем эти методы интенсивно развивались применительно к самым различным задачам. При этом многие авторы стали использовать термин «методы расщепления», отталкиваясь от идеи, что матрицы B_τ этих методов представляются в виде произведения более простых матриц (расщепляются), например, ленточных или блочно треугольных.

37.2. Пусть выполнены предположения 37.1, $\tau_i^{(k)} \equiv \tau_i$, $\alpha_k \equiv \alpha$, где τ_i — некоторые неотрицательные числа, $i = 1, \ldots, r$, α — некоторое положительное число. Тогда, если

$$\operatorname{Re} \lambda > 0 \qquad \forall \lambda \in \sigma(A),$$

то существуют такие положительные числа $\hat{\tau}$, $\hat{\alpha}$, что соответствующий стационарный метод расщепления

$$B_\tau(u^k - u^{k-1}) = -\alpha(Au^{k-1} - f)$$

сходится для любых значений параметров $\tau_i \in [0, \hat{\tau})$, $i = 1, \ldots, r$, $\alpha \in (0, \hat{\alpha})$.

37.3. Пусть $\tau_i \equiv \tau$ и $\alpha = \beta\tau$, где τ, β — некоторые положительные числа, причем β фиксировано. Тогда, если $\operatorname{Re} \lambda > 0$ $\forall \lambda \in \sigma(A)$, то существует такое $\hat{\tau} > 0$, что соответствующий стационарный метод расщепления из 37.2 сходится для произвольного $\tau \in (0, \hat{\tau})$.

37.4. Метод расщепления называется *коммутативным* (коммутативный случай), когда матрицы $A_1, \ldots, A_r$ попарно перестановочны между собой.

37.5. Если матрицы простой структуры $A_1, \ldots, A_r$ попарно перестановочны между собой, то они и матрица $A = \sum\limits_{i=1}^{r} A_i$ обладают общей полной в $\mathbf{R}_n$ системой собственных векторов, т. е. существует матрица Ψ, столбцами которой являются векторы $\psi_1, \ldots, \psi_n$, образующие полную в $\mathbf{R}_n$ общую систему собственных векторов матриц $A_1, \ldots, A_r, A$:

$$A_i\psi_j = \lambda_j^{(i)}\psi_j, \ i = 1, \ldots, r, \ A\psi_j = \lambda_j\psi_j, \ j = 1, \ldots, n, \ \lambda_j = \sum_{i=1}^{r} \lambda_j^{(i)}.$$

37.6. Если симметричные матрицы $A_1, \ldots, A_r$ попарно перестановочны между собой, то они и матрица $A = \sum_{i=1}^{r} A_i$ обладают общей полной в $\mathbf{R}_n$ ортонормированной системой собственных векторов.

37.7. Пусть $T_{\tau_k,\alpha_k} = E - \alpha_k B_{\tau_k}^{-1} A$ — матрица перехода и $Z_k = T_{\tau_k,\alpha_k} \ldots T_{\tau_1,\alpha_1}$ — разрешающий оператор метода расщепления из 37.1, и выполнены условия из 37.5. Тогда, используя обозначения из 37.5, имеем

$$\rho\left(T_{\tau_k,\alpha_k}\right) = \max_{1 \leqslant j \leqslant n} \left| 1 - \alpha_k \lambda_j \left[\prod_{i=1}^{r} \left(1 + \tau_i^{(k)} \lambda_j^{(i)}\right) \right]^{-1} \right|,$$

$$\rho(Z_k) = \max_{1 \leqslant j \leqslant n} \prod_{t=1}^{k} \left| 1 - \alpha_t \lambda_j \left[\prod_{i=1}^{r} \left(1 + \tau_i^{(t)} \lambda_j^{(i)}\right) \right]^{-1} \right|.$$

Более того, существует такая норма $\|\cdot\|_*$ (например, норма $\|\cdot\|_D$, порождаемая матрицей $D = Q^* Q$), что

$$\|T_{\tau_k,\alpha_k}\|_* = \rho\left(T_{\tau_k,\alpha_k}\right), \quad \|Z_k\|_* = \rho(Z_k).$$

37.8. Пусть выполнены предположения 37.7. Тогда

$$\rho(Z_k) \leqslant \prod_{t=1}^{k} \rho\left(T_{\tau_t,\alpha_t}\right) \leqslant [\max_{1 \leqslant t \leqslant k} \rho\left(T_{\tau_t,\alpha_t}\right)]^k).$$

37.9. Пусть выполнены предположения из 37.3 и 37.7, $\operatorname{Re} \lambda_i \geqslant 0$ для всех $\lambda^{(i)} \in \sigma(A_i)$, $i = 1, \ldots, r$, и $\operatorname{Re} \lambda > 0$ для всех $\lambda \in \sigma(A)$. Тогда существует такое $\hat{\beta} > 0$, что соответствующий стационарный метод расщепления сходится для любых значений $\beta \in (0, \hat{\beta})$ и $\tau > 0$.

37.10. Пусть выполнены предположения из 37.7, $\tau_i^{(k)} = \tau_k$, $i = 1, \ldots, r$, где τ_k — последовательность параметров, для которой

$$0 < \tau_{\min} = \min_k \tau_k \leqslant \max_k \tau_k < +\infty,$$

и $\alpha_k = \beta_k \tau_k$. Тогда существует такое $\hat{\beta} > 0$, что соответствующий метод расщепления сходится для любой последовательности параметров $\beta_k \in (0, \hat{\beta})$, удовлетворяющих неравенству

$$\inf_k \beta_k > 0.$$

37.11. При $r = 2$ и $\alpha_k = \tau_1^{(k)} + \tau_2^{(k)}$ метод расщепления из 37.1 называется *методом Писмана — Рэкфорда* и может быть записан в виде

$$\left(E + \tau_1^{(k)} A_1\right) u^{k-1/2} = \left(E - \tau_1^{(k)} A_2\right) u^{k-1} + \tau_1^{(k)} f,$$

$$\left(E + \tau_2^{(k)} A_2\right) u^{k} = \left(E - \tau_2^{(k)} A_1\right) u^{k-1/2} + \tau_2^{(k)} f.$$

37.12. Матрица перехода k-го шага метода Писмапа — Рэкфорда может быть записана в виде

$$T_{\tau_k} = (E + \tau_2^{(k)}A_2)^{-1}(E - \tau_2^{(k)}A_1)(E + \tau_1^{(k)}A_1)^{-1}(E - \tau_1^{(k)}A_2).$$

Последующие несколько пунктов настоящего параграфа посвящены коммутативному случаю метода Писмана — Рэкфорда. При этом мы будем предполагать, что A_1 и A_2 являются матрицами простой структуры с вещественными и неотрицательными собственными числами, а параметры $\tau_1^{(k)}$ и $\tau_2^{(k)}$ для всех $k \geqslant 1$ положительны.

37.13. При сделанных предположениях для коммутативного случая метода Писмана — Рэкфорда (в обозначениях из 37.7)

$$\rho(T_\tau) = \max_{1 \leqslant j \leqslant n} \left| \frac{1 - \tau_2\lambda_j^{(1)}}{1 + \tau_1\lambda_j^{(1)}} \cdot \frac{1 - \tau_1\lambda_j^{(2)}}{1 + \tau_2\lambda_j^{(2)}} \right|.$$

При этом, если ввести величины $m_i = \min\limits_{1 \leqslant j \leqslant n} \lambda_j^{(i)}$, $M_i = \max\limits_{1 \leqslant j \leqslant n} \lambda_j^{(i)}$, $i = 1, 2$, то

$$\rho(T_\tau) \leqslant \max_{\substack{m_1 \leqslant \lambda \leqslant M_1 \\ m_2 \leqslant \mu \leqslant M_2}} |\Phi_\tau(\lambda, \mu)|,$$

где $\Phi_\tau(\lambda, \mu) = \dfrac{1 - \tau_2\lambda}{1 + \tau_1\lambda} \, \dfrac{1 - \tau_1\mu}{1 + \tau_2\mu}$.

37.14. Существуют дробно-линейные преобразования

$$\lambda = \frac{a\tilde{\lambda} - b}{1 - c\tilde{\lambda}}, \quad \mu = \frac{a\tilde{\mu} - b}{1 - c\tilde{\mu}},$$

где a, b, c — некоторые вещественные числа, отображающие для любых положительных $m < M$ квадрат $[m, M] \times [m, M]$ на прямоугольник $[m_1, M_1] \times [m_2, M_2]$, где $m_1 < M_1$ и $m_2 < M_2$, при котором функция Φ из 37.13 не изменяет своего вида, т. е. $\Phi_\tau(\lambda, \mu) = \Phi_\gamma(\tilde{\lambda}, \tilde{\mu})$, где

$$\gamma_1 = \frac{1 - b\tau_1}{a\tau_1 - c}, \quad \gamma_2 = \frac{1 + b\tau_2}{a\tau_2 + c}.$$

В дальнейшем без потери общности мы будем предполагать, что $m_1 = m_2 = m$ и $M_1 = M_2 = M$ для некоторых положительных $m < M$. Заметим, что случай $m_1 = M_1$ (или $m_2 = M_2$) для нас не представляет никакого теоретического интереса, так как при $\tau_2 = 1/m_1$ (или $\tau_1 = 1/m_2$) $\rho(T_\tau) = 0$, и соответствующий метод расщепления сходится за одну итерацию. Этот случай не представляет и практического интереса, поскольку матрица A_1 (или A_2) будет скалярной матрицей, и применять здесь метод расщепления просто не имеет смысла.

37.15. Пусть относительно матриц A_1 и A_2 коммутативного случая метода расщепления выполнены все сделанные ранее предположения и существует такая норма $\|\cdot\|_*$, что $\|Z_k\|_* = \rho(Z_k)$ для всех $k \geqslant 1$. Тогда проблема приближенной оптимизации этого метода за любые k шагов в указанной норме $\|\cdot\|$ заключается в нахождении последовательности параметров $\tau_1^{(t)}, \tau_2^{(t)}$, $t = 1, \ldots, k$, являющихся решением задачи

$$\max_{\lambda,\mu\in[m,M]} \prod_{t=1}^{k} |\Phi_{\tau_t}| = \min_{\tau_1^{(t)},\tau_2^{(t)},\ t=1,\ldots,k}.$$

37.16. Проблема приближенной асимптотической оптимизации циклического с периодом $k \geqslant 1$ коммутативного метода расщепления заключается в нахождении параметров $\tau_1^{(t)}, \tau_2^{(t)}$, $t = 1, \ldots, k$, являющихся решением минимаксной задачи 37.15.

37.17. Сформулированная в 37.15—37.16 задача приближенной оптимизации метода расщепления имеет точное решение при помощи эллиптических функций Якоби.

Мы не приводим, в силу громоздкости, явное решение задачи из 37.15 (оно содержится в ряде хорошо известных монографий по вычислительным методам), а ограничимся только одним приближенным способом ее решения.

37.18. Решение задачи

$$\max_{\lambda,\mu\in[m,M]} |\Phi_\tau(\lambda, \mu)| = \min_{\tau_1,\tau_2}$$

достигается выбором $\tau_1 = \tau_2 = \tau_{\text{опт}} = 1/\sqrt{mM}$, причем

$$\max_{\lambda,\mu\in[m,M]} |\Phi_{\tau_{\text{опт}}}(\lambda, \mu)| = \max_{m\leqslant\lambda\leqslant M} \left|\frac{1-\tau_{\text{опт}}\lambda}{1-\tau_{\text{опт}}\lambda}\right|^2 = \left[\frac{\sqrt{M}-\sqrt{m}}{\sqrt{M}+\sqrt{m}}\right]^2.$$

37.19. Пусть $m = m_0 < m_1 < \ldots < m_k = M$ — некоторые положительные числа и $\tau_1^{(t)} = \tau_2^{(t)} = \tau_t$ для $t = 1, \ldots, k$. Тогда

$$\max_{\lambda,\mu\in[m,M]} \prod_{t=1}^{k} |\Phi_{\tau_t}(\lambda, \mu)| \leqslant \hat{g}(\tau_1, \ldots, \tau_k, m_1, \ldots, m_{k-1}) =$$
$$= \max_{1\leqslant t\leqslant k} \max_{\lambda\in[m_{t-1},m_t]} \left|\frac{1-\tau_t\lambda}{1+\tau_t\lambda}\right|^2,$$

и решение проблемы приближенной оптимизации соответствующего метода расщепления из 37.15 и 37.16 заключается в минимизации функции $\hat{g}$ по переменным $\tau_1, \ldots, \tau_k, m_1, \ldots, m_{k-1}$.

37.20. Решение задачи приближенной оптимизации 37.19 достигается выбором

$$m_t = mh^{-t/k}, \qquad t = 1, \ldots, k-1,$$
$$\tau_{t,\text{опт}} = \frac{1}{\sqrt{m_{t-1}m_t}} = \frac{1}{m} h^{t(t-1)/(2k)}, \qquad t = 1, \ldots, k,$$

где $h = m/M$. При этом $\rho(Z_k) \leqslant [(1 - h^{1/k})/(1 + h^{1/k})]^2$.

37.21. Для асимптотической скорости сходимости циклического с фиксированным периодом $s \geqslant 1$ коммутативного метода при сделанных ранее предположениях и $h \ll 1$ справедлива оценка

$$r(Z_\infty) \geqslant \frac{2}{s} \ln\left[1 - 2h^{1/s}\right] \geqslant \frac{4}{s} h^{1/s}.$$

37.22. При $s = |\ln h|/\ln 2$ правая часть неравенства 37.21 достигает наибольшего значения, а само неравенство принимает вид

$$r(Z_\infty) \geqslant 2 \ln^2 2/|\ln h|.$$

37.23. Для асимптотической скорости сходимости стационарного коммутативного метода расщепления при сделанных ранее предположениях и $\tau_1 = \tau_2 = 1/\sqrt{h}$ при $h \ll 1$ справедлива оценка

$$r_\infty(T_\tau) \geqslant 4h^{1/2}.$$

37.24. Пусть $n = l \times s$, где l, s — положительные числа, $A_1 = K_1 \times E_s$ и $A_2 = E_l \times K_2$, где K_1 и K_2 — матрицы простой структуры порядка l и s соответственно, все собственные числа которых неотрицательны, через E_t обозначается единичная матрица порядка t, $A = A_1 + A_2$. Тогда соответствующий метод Писмана — Рэкфорда является коммутативным и

$$\rho(T_\tau) = \max_{\substack{\lambda \in \sigma(K_1) \\ \mu \in \sigma(K_2)}} \left| \frac{1 - \tau_2\lambda}{1 + \tau_1\lambda} \, \frac{1 - \tau_1\mu}{1 + \tau_2\mu} \right|.$$

37.25. Пусть B — положительно определенная (полуопределенная) матрица, τ — положительный параметр $S_\tau = (E + \tau B)^{-1} \times (E - \tau B)$ и $D = (E + \tau B)'(E + \tau B)$. Тогда

$$\|S_\tau\|_D^2 = \sup_{v \in R_n} \frac{\|(E - \tau B)v\|^2}{\|(E + \tau B)v\|^2} = \sup \frac{\|v\|^2 - 2\tau(Bv, v) + \tau^2 \|Bv\|^2}{\|v\|^2 + 2\tau(Bv, v) + \tau^2 \|Bv\|^2} < 1$$

для любого $\tau > 0$, где $\|\cdot\|$ — обычная сферическая норма.

В остальной части параграфа будут рассматриваться методы расщепления при сохранении следующих предположений: $r = 2$; $\tau_1^{(k)} = \tau_2^{(k)} = \tau_k > 0$; матрицы A_1 и A_2, по крайней мере, положительно полуопределенные, матрица A положительно определенная. Перестановочность матриц A_1 и A_2 нигде больше предполагаться не будет.

37.26. При сделанном предположении положительной полуопределенности матриц A_1 и A_2 матрицы $E + \tau A_i$, $i = 1, 2$, невырожденные для любого $\tau > 0$.

37.27. Пусть z^k — векторы ошибок метода расщепления, $\psi^k = (E + \tau A_2) z^k$, хотя бы одна из матриц A_i, $i = 1, 2$ положительно определенная (например, A_1) и $S_\tau^{(i)} = (E + \tau A_i)^{-1}(E - \tau A_i)$.

Тогда для любого $\psi^{k-1} \neq 0$

$$\psi^k = S_\tau^{(1)} S_\tau^{(2)} \psi^{k-1}, \quad \|\psi^k\| \leqslant \|S_\tau^{(1)}\| \cdot \|\psi^{k-1}\| < \|\psi^{k-1}\|$$

и, следовательно, метод расщепления сходится для любого $\tau > 0$, причем справедлива оценка

$$r_\infty(T_\tau) \geqslant -\ln \|S_\tau^{(1)}\|.$$

37.28. Пусть матрицы A_i, $i=1, 2$, симметричные и положительно полуопределенные. Тогда метод расщепления сходится для любого $\tau > 0$.

37.29. Пусть B — симметричная положительно определенная матрица и $S_\tau = (E + \tau B)^{-1}(E - \tau B)$. Тогда

$$\|S_\tau\| = \rho(S_\tau) = \max_{\lambda \in \sigma(S_\tau)} \left| \frac{1-\tau\lambda}{1+\tau\lambda} \right|.$$

37.30. Пусть A_i — симметричные положительно определенные матрицы и $\sigma(A_i) \in [m, M]$, где $m \leqslant M$ — положительные числа, $i = 1, 2$. Тогда (в обозначениях 37.27)

$$\|S_\tau^{(1)} S_\tau^{(2)}\| \leqslant \max_{\lambda \in [m, M]} \left| \frac{1-\tau\lambda}{1+\tau\lambda} \right|^2.$$

37.31. Пусть выполнены предположения 37.30. Тогда для приближенной оптимизации соответствующего стационарного метода расщепления достаточно минимизировать правую часть неравенства из 37.30, что достигается выбором $\tau_{\text{опт}} = 1/\sqrt{mM}$. При этом для асимптотической скорости сходимости в случае $h = m/M \ll 1$ будет справедлива оценка

$$r_\infty(T_\tau) \geqslant 4h^{1/2}.$$

37.32. Пусть матрица B положительно определенная

$$\|Bv\|^2 \leqslant \Delta(Bv, v) \quad \forall v \in R_n$$

и $S_\tau = (E + \tau B)^{-1}(E - \tau B)$, где Δ и τ — положительные числа. Тогда

$$\|S_\tau\|^2 \leqslant \max_{\lambda \in \sigma(B+B')} \frac{1 - \tau(1 - \tau\Delta/2)\lambda}{1 + \tau(1 + \tau\Delta/2)\lambda} < 1$$

для любого $\tau > 0$.

37.33. Пусть A_i — положительно определенные матрицы, $\|A_i v\|^2 \leqslant \Delta(A_i v, v)$ и $\delta(v, v) \leqslant (A_i v, v)$ для любого $v \in \mathbf{R}_n$, где $\delta \leqslant \Delta$ — положительные числа, и $S_\tau^{(i)} = (E + \tau A_i)^{-1}(E - \tau A_i)$. Тогда для любого $\tau > 0$

$$\|S_\tau^{(1)} S_\tau^{(2)}\| \leqslant \max_{\lambda \in [\delta, \Delta]} \left| \frac{1 - 2\tau(1 - \tau\Delta/2)\lambda}{1 + 2\tau(1 + \tau\Delta/2)\lambda} \right|$$

37.34. Пусть выполнены предположения 37.33. Тогда для приближенной оптимизации соответствующего стационарного метода расщепления достаточно минимизировать правую часть неравенства из 37.33, что достигается выбором $\tau_{\text{опт}} = 1/\sqrt{\delta\Delta}$. При этом $\| S_\tau^{(1)} S_\tau^{(2)} \| \leqslant (\sqrt{\Delta} - \sqrt{\delta})/(\sqrt{\Delta} + \sqrt{\delta})$, и для асимптотической скорости сходимости метода при $h = \delta/\Delta \ll 1$ справедлива оценка

$$r_\infty(T_\tau) \geqslant 2h^{1/2}.$$

37.35. Пусть A — симметричная положительно определенная матрица, $A = A_1 + A_2$ и $A_1 = A_2'$. Тогда матрицы A_1 и A_2 положительно определенные.

37.36. Пусть выполнены предположения 37.35. Тогда соответствующий вариант метода расщепления называется *попеременно треугольным методом.*

37.37. Матрица $B_\tau = (E + \tau A_1)(E + \tau A_2)$ попеременно треугольного метода является симметричной и положительно определенной.

37.38. Матрица $B_\tau^{-1}A$ попеременно треугольного метода является A-самосопряженной и A-положительно определенной матрицей. В частности, она является матрицей простой структуры и все ее собственные числа положительны.

37.39. Пусть $A = A_1 + A_2$ и D — некоторая симметричная положительно определенная матрица. Тогда итерационный метод

$$\left(D + \tau_1^{(k)} A_1\right) D^{-1} \left(D + \tau_2^{(k)} A_2\right)\left(u^k - u^{k-1}\right) = -\alpha_k \left(A u^{k-1} - f\right)$$

с последовательностями параметров $\tau_1^{(k)}, \tau_2^{(k)}$ и α_k эквивалентен методу расщепления

$$\left(E + \tau_1^{(k)} \tilde{A}_1\right)\left(E + \tau_2^{(k)} \tilde{A}_2\right)\left(\tilde{u}^k - \tilde{u}^{k-1}\right) = -\alpha_k \left(\tilde{A}\tilde{u}^{k-1} - \tilde{f}\right),$$

примененному к преобразованной системе $\tilde{A}\tilde{u} = \tilde{f}$, где либо $\tilde{A} = D^{-1}A$, $\tilde{u} = u$, $\tilde{f} = D^{-1}f$, $\tilde{A}_i = D^{-1}A_i$, $i = 1, 2$, либо $\tilde{A} = D^{-1/2}AD^{-1/2}$, $\tilde{u} = D^{1/2}u$, $\tilde{f} = D^{-1/2}f$, $\tilde{A}_i = D^{-1/2}A_iD^{-1/2}$, $i = 1, 2$.

§ 38. Чебышевские итерационные методы

В настоящем параграфе будут рассматриваться *линейные итерационные методы* с разрешающими операторами Z_k, являющимися для каждого $k \geqslant 1$ многочленами степени не выше k от матрицы HA (H — некоторая невырожденная матрица):

$$Z_k = Z_k(HA) = E - \sum_{i=1}^{k} \gamma_{i,k} (HA)^i$$

с некоторыми вещественными коэффициентами $\gamma_{i,k}$, $i = 1, \ldots, k$. На протяжении всего параграфа (за исключением последних пунктов) мы будем предполагать, что матрица HA обладает полной системой собственных векторов, т. е. является матрицей простой структуры, и все ее собственные числа λ_i, $i = 1, \ldots, n$, вещественны и положительны. Кроме того, будем обозначать через D симметричную положительно определенную матрицу, для которой матрица HA является D-самосопряженным и D-положительно определенным оператором в пространстве R_n или, эквивалентно, $(DHA)' =$

$= DHA > 0$. В качестве D всегда можно выбрать матрицу $[Q^{-1}]'Q^{-1}$, где Q — матрица порядка n, столбцы которой образуют полную систему собственных векторов матрицы HA, т. е. $HA = Q\Lambda Q^{-1}$, где $\Lambda = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$.

38.1. Пусть Ω — некоторое замкнутое множество вещественной положительной полуоси, содержащее спектр $\sigma(HA)$ матрицы HA. Тогда для рассматриваемых итерационных методов

$$\|Z_k\|_D = \rho(Z_k) \leqslant \max_{\lambda\in\Omega}\left|1 - \sum_{i=1}^{k}\gamma_{i,k}\lambda^i\right|.$$

38.2. Для множества Ω из 38.1 задача приближенной оптимизации рассматриваемых итерационных методов в D-норме за k шагов заключается в построении метода с такими значениями $\gamma_{i,k} = \overset{*}{\gamma}_{i,k}$, чтобы выполнялось равенство

$$\max_{\lambda\in\Omega}\left|1 - \sum_{i=1}^{k}\overset{*}{\gamma}_{i,k}\lambda^i\right| = \min_{\gamma_1,\ldots,\gamma_k}\max_{\lambda\in\Omega}\left|1 - \sum_{i=1}^{k}\gamma_i\lambda^i\right|,$$

т. е. задача эквивалентна нахождению многочлена степени не выше k, наименее уклоняющегося от нуля на множестве Ω. Итерационные методы с такими разрешающими операторами мы будем называть *чебышевскими*.

В дальнейшем мы будем предполагать, что $\Omega = [a, b]$, где a, b, $a < b$,— положительные числа.

38.3. При сделанном предположении задача из 38.2 о построении чебышевского итерационного метода с разрешающими операторами

$$\overset{*}{Z}_k(HA) = E - \sum_{i=1}^{k}\overset{*}{\gamma}_{i,k}(HA)^i$$

принимает вид

$$\max_{a\leqslant\lambda\leqslant b}\left|\overset{*}{Z}_k(\lambda)\right| = \min_{\gamma_1,\ldots,\gamma_k}\max_{a\leqslant\lambda\leqslant b}\left|1 - \sum_{i=1}^{k}\gamma_i\lambda^i\right|, \qquad k = 1, 2, \ldots.$$

38.4. Многочлены

$$C_k(t) = \frac{1}{2}\left[\left(t + \sqrt{t^2-1}\right)^k + \left(t - \sqrt{t^2-1}\right)^k\right]$$

или, эквивалентно,

$$C_k(t) = \begin{cases}\cos(k\arccos t), & |t| \leqslant 1,\\ \mathrm{ch}\,(k\,\mathrm{arcch}\,t), & |t| > 1,\end{cases}$$

называются *многочленами Чебышева* k-го порядка первого рода, $k = 0, 1, \ldots$.

38.5. Корни многочлена Чебышева $C_k(t)$ вычисляются по формуле

$$t_i = \cos\frac{(2i-1)\pi}{2k}, \qquad i = 1, \ldots, k;$$

максимальное по модулю значение на отрезке $[-1, 1]$, равное единице, $C_k(t)$ достигает в точках $\hat{t}_i = \cos(i\pi/k)$, $i = 0, 1, \ldots, k$. Более точно,

$$C_k(\hat{t}_i) = (-1)^i, \qquad i = 0, 1, \ldots, k.$$

38.6. Многочлены Чебышева $C_k(t)$ удовлетворяют рекуррентным соотношениям

$$C_{k+1}(t) = 2tC_k(t) - C_{k-1}(t), \qquad k = 1, 2, \ldots .$$

38.7. Пусть V_k — множество многочленов $P_k(t)$ степени k (с ненулевыми коэффициентами при t^k), удовлетворяющих условию $P_k(0) = 1$. Тогда многочлен Чебышева $C_k(t) \in V_k(t)$ является единственным решением минимаксной задачи

$$\max_{t\in[-1,1]} |P_k(t)| = \min_{P_k\in V_k},$$

т. е. $C_k(t)$ наименее отклоняется от нуля на отрезке $[-1, 1]$ среди всех многочленов заданной степени $k \geqslant 1$.

38.8. Задача из 38.3 о построении чебышевского итерационного метода имеет единственное решение

$$\overset{*}{Z}_k(\lambda) = \left[C_k\left(\frac{b+a}{b-a}\right)\right]^{-1} C_k\left(\frac{b+a}{b-a} - 2\frac{\lambda}{b-a}\right)$$

для любого $k \geqslant 1$.

38.9. Разрешающие операторы $\overset{*}{Z}_k(HA)$ чебышевского итерационного метода связаны для $k \geqslant 1$ соотношениями

$$\overset{*}{Z}_k = E - \alpha_k HA\overset{*}{Z}_{k-1} + \beta_k(\overset{*}{Z}_{k-1} - \overset{*}{Z}_{k-2}),$$

где

$$\alpha_k = \frac{4}{b-a}\frac{C_{k-1}(\eta)}{C_k(\eta)}, \qquad \beta_k = \begin{cases} 0, & k = 1, \\ \dfrac{C_{k-2}(\eta)}{C_k(\eta)}, & k > 1, \end{cases}$$

$$\eta = (b+a)/(b-a).$$

38.10. Чебышевский итерационный метод для $k \geqslant 1$ может быть реализован по формулам

$$u^k = u^{k-1} - \begin{cases} \alpha_1 H(Au^0 - f), & k = 1, \\ \alpha_k H(Au^{k-1} - f) - \beta_k(u^k - u^{k-1}), & k > 1, \end{cases}$$

где α_k и β_k определены в 38.9.

38.11. Линейный итерационный метод вида

$$u^k = u^{k-1} - \begin{cases} \alpha_1 H\,(Au^0 - f), & k = 1, \\ \alpha_k H\left(Au^{k-1} - f\right) - \beta_k\left(u^{k-1} - u^{k-2}\right), & k > 1, \end{cases}$$

с некоторой невырожденной матрицей H и некоторыми последовательностями параметров α_k, β_k называется *обобщенным двухшаговым методом Ричардсона* или *обобщенным методом Ричардсона второго порядка.* В случае $H \equiv E$ он соответственно называется *двухшаговым методом Ричардсона* или *методом Ричардсона второго порядка.*

38.12. Рассматриваемый чебышевский итерационный метод является двухшаговым методом Ричардсона со специальными последовательностями параметров α_k и β_k.

38.13. Для чебышевского итерационного метода при $k \geqslant 1$ справедлива оценка

$$\|z^k\|_D \leqslant \frac{1}{C_k(\eta)} \|z^0\|_D,$$

или, с учетом представления

$$C_k(\eta) = \frac{1}{2}\left[\left(\eta + \sqrt{\eta^2 - 1}\right)^k + \left(\eta + \sqrt{\eta^2 - 1}\right)^{-k}\right],$$

оценка

$$\|z^k\|_D \leqslant 2\left[\eta + \sqrt{\eta^2 - 1}\right]^{-k} \|z^0\|_D = 2\left[\frac{\sqrt{b} - \sqrt{a}}{\sqrt{b} + \sqrt{a}}\right]^k \|z^0\|_D.$$

38.14. Для любого фиксированного $k \geqslant 1$ на классе матриц HA, удовлетворяющих наложенным выше требованиям, первая оценка из 38.13 неулучшаема. Более конкретно, для любого $k \geqslant 1$ существует матрица простой структуры HA со спектром, принадлежащим отрезку $[a, b]$, для которой

$$\|\overset{*}{Z}_k\|_D = 1/C_k(\eta).$$

Для нахождения такой матрицы достаточно потребовать, чтобы величины (см. 38.5) $\frac{1}{2}\left[b + a - (b - a)\cos\frac{i\pi}{k}\right]$, $i = 0, 1, \ldots, k$, принадлежали спектру матрицы HA.

38.15. Для асимптотической скорости сходимости чебышевского итерационного метода имеет место оценка

$$r(Z_\infty) \geqslant \lim_{k\to\infty} \frac{1}{k} \ln C_k(\eta) = \ln\left(\eta + \sqrt{\eta^2 - 1}\right) = \ln\frac{\sqrt{b} - \sqrt{a}}{\sqrt{b} + \sqrt{a}}.$$

38.16. Для уменьшения D-нормы вектора начальной ошибки в $1/\varepsilon$ раз ($\varepsilon < 1$) достаточно k_ε шагов чебышевского итерационного метода из 38.10, где k_ε — минимальное целое, удовлетворяющее неравенству

$$\frac{1}{C_{k_\varepsilon}(\eta)}=\frac{2q^{k_\varepsilon}}{q^{2k_\varepsilon}+1}\leqslant\varepsilon,\quad q=\eta+\sqrt{\eta^2-1},$$

или эквивалентному неравенству

$$k_\varepsilon\geqslant\frac{\ln\left(1/\varepsilon+\sqrt{1/\varepsilon^2-1}\right)}{\ln\left(\eta+\sqrt{\eta^2-1}\right)},$$

или упрощенному неравенству

$$k_\varepsilon\geqslant\ln\frac{2}{\varepsilon}\Big/\ln\left(\eta+\sqrt{\eta^2-1}\right).$$

38.17. Пусть A, B — симметричные положительно определенные матрицы. Тогда для решения системы $Au=f$ применим чебышевский итерационный метод из 38.10 с матрицей $H=B^{-1}$, причем в качестве матрицы D можно выбрать либо матрицу A, либо матрицу B.

38.18. Пусть A, Λ — симметричные положительно определенные матрицы блочного метода Якоби из 35.2. Тогда для решения системы $Au=f$ применим чебышевский итерационный метод с матрицей $H=\Lambda^{-1}$.

Из 38.18 видно, что для построения конкретного варианта чебышевского итерационного метода, а именно, для выбора матрицы H мы обратились к идее построения блочных методов Якоби. Аналогичная ситуация будет возникать и в дальнейшем, когда мы обратимся к блочному методу Гаусса — Зейделя и методу переменных направлений. Поэтому в ряде случаев представляется целесообразным рассматривать чебышевские итерационные методы как способ ускорения сходимости (оптимизации) соответствующих одно- или двухшаговых линейных стационарных итерационных методов за счет использования последовательностей параметров.

38.19. Пусть $A=\Lambda-L-L'$ и Λ — симметричные положительно определенные матрицы и матрица L такова, что матрицы $\tilde{L}=\Lambda^{-1}L$ и $\tilde{U}=\Lambda^{-1}L'$ являются строго нижней и строго верхней треугольными соответственно. Тогда, если матрица $T_0=\tilde{L}+\tilde{U}$ является согласованно упорядоченной (отсюда согласно 35.36 следует, что $\rho(T_0)<1$), то для чебышевского итерационного метода с матрицей $H=\Lambda^{-1}$ выполняются равенства $a=1-\rho(T_0)$, $b=1+\rho(T_0)$, а для его асимптотической скорости сходимости имеет место оценка

$$r(\overset{*}{Z}_\infty)\geqslant\frac{1}{2}\ln\frac{1+\sqrt{1-\rho^2(T_0)}}{1-\sqrt{1-\rho^2(T_0)}}.$$

Отсюда, учитывая, что для соответствующего обобщенного метода последовательной верхней релаксации

$$\left(\frac{1}{\omega}\Lambda-L\right)\left(u^k-u^{k-1}\right)=-\left(Au^{k-1}-f\right)$$

с параметром $\omega = \omega_{\text{опт}} = 2/(1 + \sqrt{1 - \rho^2(T_0)})$ асимптотическая скорость сходимости вычисляется по формулам

$$r_\infty\left(T_{\omega_{\text{опт}}}\right) = -\ln(\omega_{\text{опт}} - 1) = \ln \frac{1 + \sqrt{1 - \rho^2(T_0)}}{1 - \sqrt{1 - \rho^2(T_0)}},$$

получаем неравенство

$$r(\overset{*}{Z}_\infty) / r_\infty\left(T_{\omega_{\text{опт}}}\right) \geqslant 1/2,$$

38.20. Пусть

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} A_{11} & 0 \\ A_{12} & A_{22} \end{bmatrix},$$

где A — симметричная положительно определенная блочная матрица, A_{ij} — матрица порядка $n_i \times n_j$, $i, j = 1, 2$; $n_1 + n_2 = n$, и пусть $C = B - A$. Тогда матрица перехода

$$T_1 = B^{-1}C = -\begin{bmatrix} 0 & A_{11}^{-1}A_{12} \\ 0 & A_{22}^{-1}A_{21}A_{11}^{-1}A_{12} \end{bmatrix}$$

блочного метода Гаусса — Зейделя

$$Bu^k = Cu^{k-1} + f$$

является матрицей простой структуры, все ее собственные числа вещественны и принадлежат отрезку $[0, \rho(T_1)]$, где $\rho(T_1) = \rho^2(T_0) < 1$, а T_0 — матрица перехода соответствующего блочного метода Якоби. Поэтому для решения системы $Au = f$ применим чебышевский итерационный метод из 38.10 с матрицей $H = B^{-1}$, для которого $a = 1 - \rho(T_1)$, $b = 1$, $\eta = (2 - \rho^2(T_0))/\rho^2(T_0)$ и для асимптотической скорости сходимости которого выполняется оценка

$$r(\overset{*}{Z}_\infty) \geqslant \ln \frac{1 + \sqrt{1 - \rho^2(T_0)}}{1 - \sqrt{1 - \rho^2(T_0)}} = r_\infty\left(T_{\omega_{\text{опт}}}\right),$$

где $T_{\omega_{\text{опт}}}$ — матрица перехода соответствующего блочного метода последовательной верхней релаксации с оптимальным значением параметра.

Таким образом, сравнительная оценка скоростей сходимости оптимального обобщенного метода последовательной верхней релаксации с соответствующим чебышевским итерационным методом не дает нам четкого ответа, какой же из методов асимптотически быстрее, она лишь указывает, что первый не может быть более чем в два раза асимптотически быстрее по сравнению со вторым методом. В то же время, оценка последнего пункта показывает, что чебышевский итерационный метод, основанный на блочном методе Гаусса — Зейделя с циклической индекса 2 матрицей

$$T_0 = -\begin{bmatrix} 0 & A_{11}^{-1}A_{12} \\ A_{22}^{-1}A_{21} & 0 \end{bmatrix},$$

всегда асимптотически быстрее соответствующего оптимального блочного метода последовательной верхней релаксации.

38.21. Для чебышевского итерационного метода

$$\alpha = \lim_{k\to\infty} \alpha_k = \frac{4}{b-a}\cdot\frac{1}{\eta+\sqrt{\eta^2-1}} = \left(\frac{2}{\sqrt{b}+\sqrt{a}}\right)^2,$$

$$\beta = \lim_{k\to\infty} \beta_k = \frac{1}{(\eta+\sqrt{\eta^2-1})^2} = \left(\frac{\sqrt{b}-\sqrt{a}}{\sqrt{b}+\sqrt{a}}\right)^2.$$

38.22. Для двухшагового метода Ричардсона второго порядка

$$u^k = u^{k-1} - \begin{cases} \dfrac{2}{b+a} H(Au^0 - f), & k = 1, \\ \alpha H(Au^{k-1} - f) - \beta(u^{k-1} - u^{k-2}), & k > 1, \end{cases}$$

с параметрами α и β из 38.21 разрешающие операторы $\widehat{Z}_k$ для всех $k \geqslant 1$ представлены в виде

$$\widehat{Z}_k \equiv \widehat{Z}_k(HA) = \frac{b-a}{b+a}\cdot\left(\frac{\sqrt{b}-\sqrt{a}}{\sqrt{b}+\sqrt{a}}\right)^{k-1} C_k\left(\frac{b+a}{b-a}E - \frac{2}{b-a}HA\right);$$

для D-норм этих операторов справедлива оценка

$$\|\widehat{Z}_k\|_D \leqslant \frac{b-a}{b+a}\left(\frac{\sqrt{b}-\sqrt{a}}{\sqrt{b}+\sqrt{a}}\right)^{k-1}$$

и, следовательно, для его асимптотической скорости сходимости имеет место оценка

$$r(\widehat{Z}_\infty) \geqslant \ln\frac{\sqrt{b}+\sqrt{a}}{\sqrt{b}-\sqrt{a}}.$$

38.23. Пусть $A = B - C$, B — симметричные положительно определенные матрицы порядка n и $\rho(B^{-1}C) < 1$. Тогда для решения системы $\widetilde{A}u = \widetilde{f}$ с симметричной положительно определенной блочной матрицей $\widetilde{A} = \begin{bmatrix} B & -C \\ -C & B \end{bmatrix}$ порядка $2n$ и вектором $\widetilde{f} = \begin{bmatrix} f \\ f \end{bmatrix} \in \mathbf{R}_{2n}$ применим соответствующий блочный метод последовательной верхней релаксации с оптимальным значением релаксационного параметра, для которого

$$r_\infty(T_{\omega_{\text{опт}}}) = \ln\frac{1+\sqrt{1-\rho^2(B^{-1}C)}}{1-\sqrt{1-\rho^2(B^{-1}C)}}.$$

При этом для итерационного метода из 38.22 с матрицей $H = B^{-1}$ имеем

$$a = 1 - \rho(B^{-1}C), \quad b = 1 + \rho(B^{-1}C),$$

$$\alpha = 1 + \beta = 2/(1 + \sqrt{1 - \rho^2(B^{-1}C)}) = \omega_{\text{опт}},$$

$$r(\hat{Z}_\infty) \geqslant \frac{1}{2} r_\infty (T_{\omega_{\text{опт}}}).$$

Нужно заметить, что реализация каждого шага блочного метода последовательной верхней релаксации из 38.23 требует ровно вдвое больше арифметических операций, чем для соответствующего чебышевского итерационного метода из 38.22.

38.24. В случае, когда для чебышевского итерационного метода $h = a/b \ll 1$, имеем оценку

$$r(\overset{*}{Z}_\infty) \geqslant 2h^{1/2},$$

а для величины k_ε из 38.15 может быть использовано неравенство

$$k_\varepsilon \geqslant \frac{1}{2} h^{-1/2} \ln \frac{2}{\varepsilon}.$$

38.25. Пусть A — симметричная положительно определенная матрица, для которой имеет место представление $A = A_1 + A_1'$ с некоторой матрицей A_1. Тогда для решения системы $Au = f$ может быть применен чебышевский итерационный метод с матрицей $H = B_\tau^{-1}$, где τ — некоторое, вообще говоря, произвольное положительное число и $B_\tau = (E + \tau A_1)(E + \tau A_1')$ — симметричная положительно определенная матрица попеременнотреугольного метода из 37.36—37.40. При этом, если для любого $v \in \mathbf{R}_n$ одновременно выполняются неравенства

$$\delta(u, u) \leqslant (A_1 v, v), \quad \|A_1 v\|^2 \leqslant \Delta(A_1 v, v),$$

где δ, Δ, $\delta < \Delta$,— некоторые положительные числа, и параметр τ выбирается по формуле $\tau = 1/\sqrt{\delta\Delta}$, то для соответствующего чебышевского итерационного метода имеем $a = \delta\sqrt{\Delta}/(\sqrt{\Delta} + \sqrt{\delta})$, $b = \Delta\sqrt{\delta}/(\sqrt{\Delta} + \sqrt{\delta})$ и, следовательно, справедлива оценка

$$r(\overset{*}{Z}_\infty) \geqslant \frac{1}{2} \ln \frac{\sqrt[4]{\Delta} + \sqrt[4]{\delta}}{\sqrt[4]{\Delta} - \sqrt[4]{\delta}}.$$

38.26. Если в 38.25 $\tilde{h} = \delta/\Delta \ll 1$, то для соответствующего чебышевского итерационного метода справедлива оценка

$$r(\overset{*}{Z}_\infty) \geqslant \tilde{h}^{-1/4}.$$

38.27. Для любого $k \geqslant 1$ разрешающий оператор $\overset{*}{Z}_k$ чебышевского итерационного метода представим в виде

$$\overset{*}{Z}_k = \prod_{i=1}^{k} \left(E - \frac{1}{\lambda_{k,i}} HA\right),$$

где $\lambda_{k,i} = \frac{1}{2}\left[b + a - (b - a)\cos\frac{(2i-1)\pi}{k}\right]$, $i = 1, \ldots, k$. Следовательно, каждые заданные k шагов чебышевского итерационного метода могут быть реализованы по двучленным формулам

$$u^i = u^{i-1} - \tau_{\varkappa_i} H(Au^{i-1} - f),$$

где $\tau_{\varkappa_i} = 1/\lambda_{k,\varkappa_i}$, $i = 1, \ldots, k$, и $\varkappa = (\varkappa_1, \ldots, \varkappa_k)$ — некоторая перестановка чисел $1, 2, \ldots, k$.

При практической реализации чебышевских итерационных методов по двучленным формулам оказалось, что далеко не каждая перестановка $\varkappa$ обеспечивает устойчивость процесса вычислений. К настоящему времени проблема определения перестановок, обеспечивающих вычислительную устойчивость, решена как для конечных значений k, так и для бесконечно продолжаемых значений. Здесь мы укажем алгоритм построения таких перестановок только для конечных значений k, являющихся целой степенью двойки, т. е. $k = 2^s$ для некоторого целого $s \geqslant 1$.

38.28. Пусть нам известна перестановка $\tilde{\varkappa} = (\tilde{\varkappa}_1, \ldots, \tilde{\varkappa}_{k_1})$ для значения $k_1 = 2^{s-1}$, обеспечивающая устойчивую реализацию k_1 шагов одношагового чебышевского итерационного метода 38.27. Тогда аналогичная перестановка для значения $k_2 = 2^s$ определяется формулой

$$\varkappa = (\tilde{\varkappa}_1, k_2 + 1 - \tilde{\varkappa}_1, \tilde{\varkappa}_2, \ldots, k_2 + 1 - \tilde{\varkappa}_{k_1}).$$

38.29. Пусть p — некоторое положительное целое, H — некоторая невырожденная матрица и $\sigma(HA) \in [a, b]$, где a, b, $a < b$, — некоторые положительные числа. Тогда итерационный метод с разрешающими операторами

$$\tilde{Z}_{ps+t} = \overset{*}{Z}_t(\overset{*}{Z}_p)^s, \quad t = 1, \ldots, p, \quad s = 0, 1, \ldots,$$

называется *циклическим с периодом p чебышевским итерационным методом*. Здесь

$$\overset{*}{Z}_t \equiv \overset{*}{Z}_t(HA) = E - \sum_{i=1}^{t} \overset{*}{\gamma}_{t,i}(HA)^i$$

— многочлены от матрицы HA, коэффициенты которых являются решениями минимаксных задач

$$\max_{\lambda\in[a,b]}\left|1 - \sum_{i=1}^{t} \overset{*}{\gamma}_{t,i}\lambda^i\right| = \min_{\gamma_1,\ldots,\gamma_t}\max_{\lambda\in[a,b]}\left|1 - \sum_{i=1}^{t}\gamma_i\lambda^i\right|$$

для значений $t = 1, \ldots, p$. Заметим, что матрица HA, вообще говоря, не является матрицей простой структуры.

38.30. Решение минимаксной задачи предыдущего пункта всегда существует, единственно и имеет вид

$$\overset{*}{Z}_t(\lambda) = \frac{1}{C_t(\eta)} C_t\left(\frac{b+a}{b-a} - \frac{2}{b-a}\lambda\right),$$

где $\eta = (b+a)/(b-a)$.

38.31. Для асимптотической скорости сходимости циклического с периодом p чебышевского итерационного метода выполняется оценка

$$r(\widetilde{Z}_\infty) \doteq -\frac{1}{p}\ln\rho(\overset{*}{Z}_p) = \frac{1}{p}\ln C_p(\eta),$$

которая при фиксированном заданном $p \geqslant 1$ и $h = a/b \ll 1$ принимает вид

$$r(\widetilde{Z}_\infty) \geqslant 2ph.$$

38.32. Пусть p — минимальное целое, для которого выполняется неравенство

$$C_p\left(\frac{b+a}{b-a}\right) \geqslant e,$$

где e — основание натуральных логарифмов, или эквивалентное ему неравенство

$$p \geqslant \ln\left(\frac{1}{e} + \sqrt{\frac{1}{e^2} - 1}\right) \Big/ \ln\frac{\sqrt{b}+\sqrt{a}}{\sqrt{b}-\sqrt{a}},$$

или упрощенное неравенство

$$p \geqslant \left(\ln\frac{\sqrt{b}+\sqrt{a}}{\sqrt{b}-\sqrt{a}}\right)^{-1}.$$

Тогда для асимптотической скорости сходимости соответствующего циклического с периодом p чебышевского итерационного метода справедлива оценка

$$r(\widetilde{Z}_\infty) \geqslant 1/p.$$

38.33. Пусть p — минимальное целое, удовлетворяющее неравенству

$$\frac{1}{p} \leqslant \ln\frac{\sqrt{b}-\sqrt{a}}{\sqrt{b}+\sqrt{a}},$$

и $h = a/b \ll 1$. Тогда для соответствующего циклического с периодом p чебышевского итерационного метода выполняется оценка

$$r(\widetilde{Z}_\infty) \geqslant 2h^{1/2}.$$

38.34. Циклический с периодом $p = 2^s$ (s — положительное целое) чебышевский итерационный метод из 38.29 допускает устойчивую реализацию по двучленным формулам

$$u^k = u^{k-1} - \tau_k H(Au^{k-1} - f).$$

Здесь $\tau_k = \tau_{\varkappa_i}$ для $k = i(\mathrm{mod}\, p)$, $\varkappa = (\varkappa_1, \ldots, \varkappa_p)$ — перестановка из 38.28 и $\frac{1}{\tau_j} = \lambda_j = \frac{1}{2}\left[b + a - (b - a)\cos\frac{(2j-1)\pi}{k}\right]$, $j = 1, \ldots, k$, — корни многочлена $C_p\left(\frac{b+a}{b-a} - \frac{2}{b-a}\lambda\right)$.

§ 39. Нелинейные итерационные методы

В этом параграфе мы рассмотрим один класс нелинейных итерационных методов, в основу которых положены идеи методов спуска для минимизации квадратичных функционалов. В формировании таких методов участвуют либо в качестве разрешающих операторов, либо в качестве операторов перехода определяемые ниже нелинейные операторы K_s, где s — некоторое положительное целое число.

39.1. Пусть заданы симметричная положительно определенная матрица D и невырожденная матрица H. Тогда для любого $z \in \mathbf{R}_n$ оператор $K_s : \mathbf{R}_n \to \mathbf{R}_n$ определяется с помощью равенства

$$\|AK_s(z)\|_D = \min_{\gamma_1, \ldots, \gamma_s} \left\| Az - \sum_{i=1}^{s} \gamma_i (AH)^i Az \right\|_D .$$

39.9. Для любого $z \in \mathbf{R}_n$ вектор $\tilde{z} = K_s(z)$ определяется однозначно, т. е. оператор K_s из предыдущего пункта определен однозначно.

39.3. Если система векторов $z, HAz, \ldots, (HA)^s z$ линейно зависима, то $K_s(z) = 0$.

39.4. Если для заданного $z \in \mathbf{R}_n$ система векторов $z, HAz, \ldots$ $\ldots, (HA)^{s-1} z$ линейно независима, то

$$K_s(z) = z - \sum_{i=1}^{s} \overset{*}{\gamma}_i (HA)^i z,$$

где вектор $\overset{*}{\gamma} = (\overset{*}{\gamma}_1, \ldots, \overset{*}{\gamma}_s)'$ является решением системы $G\gamma = g$ с невырожденной матрицей $G = (g_{ij})$ порядка s, являющейся матрицей Грама системы векторов $AH\xi, \ldots, (AH)^s\xi$, где $\xi = Az$, в скалярном произведении, порождаемом матрицей D, т. е.

$$g_{ij} = ((AH)^i\xi, (AH)^j\xi)_D, \qquad i, j = 1, \ldots, s,$$

и вектором $g \in \mathbf{R}_s$ с компонентами

$$g_i = ((AH)^i\xi, \xi)_D, \qquad i = 1, \ldots, s.$$

39.5. Если для заданного $z \in \mathbf{R}_n$ система векторов z, HAz, … …, $(HA)^{s-1}z$ линейно зависима, то

$$K_s(z) = z - \sum_{i=1}^{s} \overset{*}{\gamma}_i (HA)^i z,$$

где вектор $\overset{*}{\gamma} = (\overset{*}{\gamma}_1, \ldots, \overset{*}{\gamma}_s)'$ является некоторым, вообще говоря, произвольным решением системы $G\gamma = b$ с вырожденной матрицей G и вектором $g \in \mathbf{R}_s$ из 39.4, которая всегда совместна, т. е. $g \in \operatorname{im} G$.

39.6. Оператор K_s является непрерывным и однородным первой степени, т. е. $K_s(\lambda z) = \lambda K_s(z)$ для любого $z \in \mathbf{R}_n$.

39.7. Если вектор $g = g(z)$ из 39.4 является ненулевым для любого ненулевого $z \in \mathbf{R}_n$, то

$$L(K_s) = \sup_{\substack{z \in \mathbf{R}_n \\ \|Az\|D=1}} \|AK_s(z)\|_D < 1.$$

Величина $L(K_s)$ называется *константой Липшица* оператора K_s.

39.8. Если для некоторого положительного целого $t \leqslant s$ матрица $D(AH)^t$ является положительно определенной, то $L(K_s) < 1$.

39.9. Пусть DAH — симметричная положительно определенная матрица. Тогда

$$L(K_s) \leqslant \|\overset{*}{Z}_s\|_{A'DA} \leqslant \left(C_s\left(\frac{M+m}{M-m}\right)\right)^{-1},$$

где $m = \rho(HA)$, $M = 1/\rho(H^{-1}A^{-1})$, $\overset{*}{Z}_s$ — разрешающий оператор соответствующего чебышевского итерационного метода при $a = m$ и $b = M$. Здесь и всюду в дальнейшем предполагается, что матрица HA не является скалярной и, следовательно, $m < M$.

39.10. Пусть выполнены предположения 39.9 и величины $\lambda_i = \frac{1}{2}\left[M + m - (M - m)\cos\frac{i\pi}{s}\right]$, $i = 0, 1, \ldots, s$, принадлежат спектру матрицы HA. Тогда

$$L(K_s) = \|\overset{*}{Z}\|_{A'DA} = \left(C_s\left(\frac{M+m}{M-m}\right)\right)^{-1},$$

т. е. оценка 39.9 является неулучшаемой для класса матриц, спектр которых принадлежит заданному отрезку $[m, M]$. Если ввести положительные величины

$$w_i = \det\begin{bmatrix} 1 & \ldots & 1 & 1 & \ldots & 1 \\ \lambda_0 & \ldots & \lambda_{i-1} & \lambda_{i+1} & \ldots & \lambda_s \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \lambda_0^{s-1} & \ldots & \lambda_{i-1}^{s-1} & \lambda_{i+1}^{s-1} & \ldots & \lambda_s^{s-1} \end{bmatrix}, \qquad i = 0, 1, \ldots, s,$$

являющиеся определителями Вандермонда, то для любого вектора $z = \sum_{i=0}^{s} c_i \psi_i$, где ψ_i — ортонормированные (в смысле скалярного произведения, порождаемого матрицей $A'DA$) собственные векторы матрицы HA, соответствующие ее собственным числам λ_i, $i=1, \ldots, s$, коэффициент c_0 отличен от нуля, а остальные коэффициенты связаны соотношениями $c_i^2 = \frac{\lambda_0}{\lambda_i}\frac{w_i}{w_0}c_0^2$, $i=1, \ldots, s$, выполняется равенство

$$K_s^2(z) = z\Big/\left(C_s\left(\frac{M+m}{M-m}\right)\right)^2,$$

т. е. вектор z является собственным вектором оператора K_s^2, соответствующим его собственному числу $L(K_s^2) = L^2(K_s)$.

39.11. Пусть DAH — симметричная матрица. Тогда имеет место оценка

$$L(K_s) \leqslant \left(C_{[s/2]}\left(\frac{\widetilde{M}^2+\widetilde{m}^2}{\widetilde{M}^2-\widetilde{m}^2}\right)\right)^{-1},$$

где $[s/2]$ — целая часть числа $s/2$ и $0 < \widetilde{m} = (\rho(H^{-1}A^{-1}))^{-1} < \widetilde{M} = \rho(HA)$, причем эта оценка неулучшаема в смысле замечания предыдущего пункта.

39.12. Пусть DAH — симметричная положительно определенная матрица. Тогда

$$L(K_1) = \left(C_1\left(\frac{M+m}{M-m}\right)\right)^{-1} = \frac{M-m}{M+m}.$$

Кроме того, для векторов

$$z^{\pm} = \psi_m/\sqrt{m} \pm \psi_M/\sqrt{M},$$

где ψ_m и ψ_M являются $A'DA$ — ортонормированными собственными векторами матрицы HA, соответствующими ее собственным числам m и M, имеем

$$K_1(z^{\pm}) = z^{\pm} - \overset{*}{\gamma}_1 HAz^{\pm} = \left(\frac{M-m}{M+m}\right)z^{\mp}$$

(здесь $\overset{*}{\gamma}_1 = 2/(M+m)$) и, следовательно,

$$K_1^2(z^{\pm}) = \left(\frac{M-m}{M+m}\right)^2 z^{\pm},$$

т. е. z^+ и z^- являются собственными векторами оператора K_1^2, соответствующими его собственному числу $L(K_1^2) = L^2(K_1) = ((M-m)/(M+m))^2$.

39.13. Пусть выполнены предположения 39.12. Тогда для любого $z^0 \in \mathbf{R}_n$ такого, что $(z^0, \psi_m) \neq 0$ и $(z^0, \psi_M) \neq 0$, подпоследовательности z^{2k} и z^{2k+1}, $k=0, 1, \ldots$, итерационного процесса

$$y^k = K_1(z^k), \quad z^k = y^k/\|Ay^k\|_D$$

сходятся соответственно к векторам v^∞ и w^∞, каждый из которых пропорционален либо вектору z^+ из 39.12 либо вектору z^-.

39.14. Пусть DAH — симметричная положительно определенная матрица. Тогда

$$\lim_{s\to\infty} \frac{1}{s} \ln L\left(K_1^s\right) = \ln L\left(K_1\right) = \ln \frac{M-m}{M+m}.$$

39.15. Пусть $s = 1$. Тогда для любого ненулевого $z \in \mathbf{R}_n$ имеем

$$K_s(z) \equiv K_1(z) = z - \frac{(AH\xi, \xi)_D}{\| AH\xi \|_D^2} H\xi,$$

где $\xi = Az$.

39.16. Стационарный одношаговый итерационный метод

$$u^k = u^{k-1} - \tau_k H\left(Au^{k-1} - f\right), \quad \tau_k = \left(AH\xi^{k-1}, \xi^{k-1}\right)_D / \| AH\xi^{k-1} \|_D^2$$

тогда и только тогда сходится в пространстве $\mathbf{R}_n$, т. е. для любого $u^0 \in \mathbf{R}_n$, когда матрица DAH положительно (или отрицательно) определенная.

Действительно, если матрица DAH не является положительно или отрицательно определенной, то существует такой вектор начальной ошибки z^0, что $(AH\xi^0, \xi^0)_D = 0$, и, следовательно, $\tau_1 = 0$, $z^1 = z^0$, $\tau_2 = 0$, $z^2 = z^1$ и т. д.

39.17. Если матрица DAH положительно определенная и $Z_s = [K_1]^s$ — разрешающие операторы итерационного метода предыдущего пункта, то

$$r(Z_\infty) \geqslant -\ln L(K_1) > 0.$$

39.18. Пусть A — симметричная положительно определенная матрица, $H = E$ и $D = A^{-1}$. Тогда соответствующий итерационный метод из 39.16:

$$u^k = u^{k-1} - \tau_k\left(Au^{k-1} - f\right), \quad \tau_k = \| \xi^{k-1} \|^2 / \| \xi^{k-1} \|_A^2,$$

называется *методом наискорейшего спуска.*

39.19. Метод наискорейшего спуска осуществляет последовательную минимизацию функционала ошибок $\Phi(u) = (Au, u) - 2(u, f)$ в направлении его наискорейшего убывания:

$$-\operatorname{grad} \Phi(u)\,|_{u=u^{k-1}} = -2\left(Au^{k-1} - f\right)$$

по формуле

$$\Phi\left(u^k\right) = \Phi\left(u^{k-1}\right) - \| \xi^{k-1} \|^4 / \| \xi^{k-1} \|_A^2.$$

39.20. Пусть A, H — симметричные положительно определенные матрицы и $D = A^{-1}$. Тогда $DAH = H$ и, следовательно, итерационный метод

$$u^k = u^{k-1} - \tau_k H\left(Au^{k-1} - f\right), \quad \tau_k = \left\|\xi^{k-1}\right\|_H^2 / \left\|H\xi^{k-1}\right\|_A^2,$$

частным случаем которого при $H = E$ является метод наискорейшего спуска, сходится, и его асимптотическая скорость сходимости вычисляется по формуле

$$r\,(Z_\infty) = \ln\frac{M+m}{M-m},$$

где $M = \rho(HA)$ и $1/m = \rho(H^{-1}A^{-1})$.

39.21. Итерационный метод из 39.20 является методом наискорейшего спуска, примененным к преобразованной системе $\tilde{A}v = \tilde{f}$ с матрицей $\tilde{A} = H^{1/2}AH^{1/2}$ и векторами $v = H^{-1/2}u$, $\tilde{f} = H^{1/2}f$.

39.22. Пусть $H = D = E$. Тогда соответствующий итерационный метод из 39.16:

$$u^k = u^{k-1} - \tau_k\left(Au^{k-1} - f\right), \quad \tau_k = \left(A\xi^{k-1}, \xi^{k-1}\right) / \left\|A\xi^{k-1}\right\|^2,$$

называется *методом минимальных невязок*.

39.23. Метод минимальных невязок сходится тогда и только тогда, когда матрица A является положительно (или отрицательно) определенной.

39.24. Пусть H — симметричная положительно определенная матрица и $D = H$. Тогда итерационный метод

$$u^k = u^{k-1} - \tau_k H\left(Au^{k-1} - f\right), \quad \tau_k = \left(AH\xi^{k-1}, \xi^{k-1}\right)_H / \left\|AH\xi^{k-1}\right\|_H^2,$$

сходится тогда и только тогда, когда матрица A положительно (или отрицательно) определенная.

39.25. Пусть A и H — симметричные положительно определенные матрицы и $D = H$. Тогда итерационный метод из 39.24

$$u^k = u^{k-1} - \tau_k H\left(Au^{k-1} - f\right), \quad \tau_k = \left\|H\xi^{k-1}\right\|_A^2 / \left\|AH\xi^{k-1}\right\|_H^2,$$

частным случаем которого при $H = E$ является метод минимальных невязок, сходится, и для его асимптотической скорости сходимости справедлива (поскольку $DAH = HAH = (DAH)'$) формула

$$r(Z_\infty) = (M - m)/(M + m),$$

где величины M и m определены в 39.20.

39.26. Итерационный метод из 39.25 является методом наискорейшего спуска, примененным к преобразованной системе $\tilde{A}v = \tilde{f}$ с матрицей $\tilde{A} = A^{1/2}HA^{1/2}$ и векторами $v = A^{1/2}u$, $\tilde{f} = A^{1/2}Hf$.

39.27. Пусть $H = A'$, $D = (AA')^{-1}$. Тогда соответствующий итерационный метод из 39.16:

$$u^k = u^{k-1} - \tau_k A'\left(Au^{k-1} - f\right), \quad \tau_k = \left\|\xi^{k-1}\right\|^2 / \left\|A'\xi^{k-1}\right\|^2,$$

называется *методом минимальных ошибок*.

39.28. Метод минимальных ошибок осуществляет последовательную минимизацию квадрата нормы вектора ошибки $z = u - u^*$ по формуле

$$\|z^k\|^2 = \|z^{k-1}\|^2 - \|\xi^{k-1}\|^4 / \|A'\xi^{k-1}\|^2.$$

39.29. Пусть H_1, H_2 — симметричные положительно определенные матрицы, $H = H_1 A' H_2$, $D = (AH_1A')^{-1}$. Тогда соответствующий итерационный метод из 39.16

$$u^k = u^{k-1} - \tau_k H_1 A' H_2 (Au^{k-1} - f),$$
$$\tau_k = \|\xi^{k-1}\|^2_{H_2} / \|A'H_2\xi^{k-1}\|^2_{H_1},$$

частным случаем которого при $H_1 = H_2 = E$ является метод минимальных ошибок, сходится и его асимптотическая скорость сходимости (поскольку $DAH = H_2$) вычисляется по формуле

$$r(Z_\infty) = \ln\frac{M - m}{M + m},$$

где величины M, m определены в 39.20 при указанном выборе матрицы H. В частности, для метода минимальных невязок имеем

$$M = \rho(A'A), \quad 1/m = \rho((A'A)^{-1}).$$

39.30. Итерационный метод из 39.29 является методом наискорейшего спуска, примененным к преобразованной системе $\tilde{A}v = \tilde{f}$ с матрицей $\tilde{A} = H_2^{1/2} A H_1 A' H_2^{1/2}$ и векторами $v = H_2^{-1/2}[A']^{-1}H_1^{-1}$ и $\tilde{f} = H_2^{1/2} f$.

39.31. Стационарный итерационный метод

$$u^k = u^{k-1} - H\sum_{i=1}^{s} \overset{*}{\gamma}_{k,i}(AH)^{i-1}\xi^{k-1}$$

с операторами перехода K_s из 39.1, т. е. параметрами $\overset{*}{\gamma}_{k,i}$, $i = 1, \ldots, s$, выбираемыми из условия

$$\|\xi^k\|_D = \min_{\gamma_1,\ldots,\gamma_s}\left\|\xi^{k-1} - \sum_{i=1}^{s}\gamma_i(AH)^i\xi^{k-1}\right\|_D,$$

сходится тогда и только тогда, когда для любого ненулевого $\xi \in \mathbf{R}_n$ величины $g_i = ((AH)^i\xi, \xi)_D$, $i = 1, \ldots, s$, не равны одновременно нулю. Если последнее требование выполнено, то для асимптотической скорости сходимости этого метода справедлива оценка

$$r(Z_\infty) \geqslant -\ln L(K_s) > 0.$$

39.32. Если для некоторого положительного целого $t \leqslant s$ матрица $D(AH)^t$ положительно определенная, то итерационный метод из 39.31 сходится.

Начиная с данного момента, мы будем рассматривать вместо стационарных одношаговых итерационных методов с операторами перехода K_s итерационные методы с разрешающими операторами $Z_s = K_s$ для всех $s \geqslant 1$. Число шагов этих методов не превосходит порядка n матрицы A и, следовательно, они могут быть отнесены и к прямым методам решения линейных систем (см. § 28).

39.33. Пусть векторы $g_1, \ldots, g_{k-1}$ образуют $A'DA$-ортогональный базис в подпространстве, являющемся линейной оболочкой системы векторов $H\xi^0, \ldots, H(AH)^{k-2}\xi^0$. Тогда k-й шаг итерационного метода, порождаемого разрешающими операторами $K_1, K_2, \ldots, K_k$, может быть реализован по формулам

$$g_k = \begin{cases} H\xi^0, & k=1, \\ HAg_{k-1} - \sum\limits_{i=1}^{k-1} \alpha_{k,i} g_i, & k>1, \end{cases} \qquad u^k = u^{k-1} - \beta_k g_k,$$

$$\alpha_{k,i} = (AHAg_{k-1}, Ag_i)_D / \| Ag_i \|_D^2, \qquad i = 1, \ldots, k-1,$$

$$\beta_k = (\xi^{k-1}, Ag_k)_D / \| Ag_k \|_D^2.$$

При этом система векторов $g_1, \ldots, g_k$ образует $A'DA$ — ортогональный базис в подпространстве, натянутом на векторы $H\xi^0, \ldots$ $\ldots, H(AH)^{k-1}\xi^0$.

39.34. Вычислительный процесс из 39.33 заканчивается нахождением точного решения системы $Au = f$ не более чем через n шагов.

39.35. Пусть DAH — симметричная матрица, n_1 — число различных собственных чисел матрицы HA. Тогда $L(K_{n_1}) = 0$ и, следовательно, вычислительный процесс из 39.33 заканчивается нахождением точного решения системы $Au = f$ не более чем через n_1 шагов.

39.36. Пусть A — симметричная положительно определенная матрица, $H = E$, $D = A^{-1}$. Тогда для вычислительного процесса из 39.33 имеем $\alpha_{k,i} = 0$, $i = 1, \ldots, k-1$, а вычислительный процесс реализуется по формулам

$$g_k = \begin{cases} \xi^0, & k=1, \\ Ag_1 - \alpha_2 g_1, & k=2, \\ Ag_{k-1} - \alpha_k g_{k-1} - \gamma_k g_{k-2}, & k>2, \end{cases} \qquad u^k = u^{k-1} - \beta_k g_k,$$

$$\alpha_k = \| Ag_{k-1} \|^2 / \| g_{k-1} \|_A^2, \qquad \beta_k = (\xi^{k-1}, g_k) / \| g_k \|_A^2,$$

$$\gamma_k = \| g_{k-1} \|_A^2 / \| g_{k-2} \|_A^2,$$

и называется *методом минимальных итераций Ланцоша.*

39.37. Пусть DAH — симметричная матрица. Тогда для вычислительного процесса из 39.33 имеем $\alpha_{k,i} = 0$, $i = 1, \ldots, k-2$, а вычислительный процесс реализуется по формулам

$$g_k = \begin{cases} H\xi^0, & k=1, \\ HAg_1 - \alpha_2 g_1, & k=2, \\ HAg_{k-1} - \alpha_k g_{k-1} - \gamma_k g_{k-2}, & k>2, \end{cases} \qquad u^k = u^{k-1} - \beta_k g_k,$$

$$\alpha_k = (AHAg_{k-1}, Ag_{k-1})/\| Ag_{k-1}\|_D^2, \quad \beta_k = (AH\xi^{k-1}, Ag_k)_D/\| Ag_k\|_D^2,$$

$$\gamma_k = \| Ag_{k-1}\|_D^2/\| Ag_{k-2}\|_D^2,$$

и называется *обобщенным методом минимальных итераций Ланцоша.*

39.38. Для последовательности D-норм векторов невязок обобщенного метода минимальных итераций Ланцоша справедлива оценка

$$\|\xi^k\|_D \leqslant \left(C_{[k/2]}\left(\frac{\widetilde{M}^2+\widetilde{m}^2}{\widetilde{M}^2-\widetilde{m}^2}\right)\right)^{-1}\|\xi^0\|_D,$$

где $[k/2]$ — целая часть числа $k/2$, $\widetilde{M} = \rho(HA)$ и $1/\widetilde{m} = \rho(H^{-1}A^{-1})$.

39.39. Пусть A — симметричная матрица, H — симметричная положительно определенная матрица и $D = H$. Тогда, поскольку $DAH = HAH$ — симметричная матрица, то для решения системы $Au = f$ применим соответствующий обобщенный метод минимальных итераций Ланцоша, коэффициенты которого будут вычисляться по формулам

$$\alpha_k = (AHAg_{k-1}, HAg_{k-1})/\| Ag_{k-1}\|_H^2, \quad \beta_k = (\xi^{k-1}, Ag_k)_H/\| Ag_k\|_H^2,$$

$$\gamma_k = \| Ag_{k-1}\|_H^2/\| Ag_{k-2}\|_H^2.$$

39.40. Если A и H — симметричные положительно определенные матрицы и $D = A^{-1}$, то коэффициенты соответствующего обобщенного метода минимальных итераций Ланцоша вычисляются по формулам

$$\alpha_k = \| Ag_{k-1}\|_H^2/\| g_{k-1}\|_A^2, \quad \beta_k = (\xi^{k-1}, g_k)/\| g_k\|_A^2,$$

$$\gamma_k = \| g_{k-1}\|_A^2/\| g_{k-2}\|_A^2.$$

39.41. Пусть DAH — симметричная положительно определенная матрица. Тогда вычислительный процесс из 39.33 может быть реализован по двучленным формулам

$$g_k = \begin{cases} H\xi^0, & k=1, \\ H\xi^{k-1} - \alpha_k g_k, & k>1, \end{cases} \qquad u^k = u^{k-1} - \beta_k g_k,$$

$$\alpha_k = (AH\xi^{k-1}, Ag_{k-1})_D/\| Ag_{k-1}\|_D^2, \quad \beta_k = (\xi^{k-1}, Ag_k)_D/\| Ag_k\|_D^2.$$

Он называется *обобщенным методом сопряженных градиентов.*

39.42. Обобщенный метод сопряженных градиентов является частным случаем метода сопряженных направлений из 28.18 при $H = B^{-1}$ и $D = B^{-1}A^{-1}C^*$.

39.43. Для любого $\xi^0 \in \mathbf{R}_n$ векторы g_i, $\xi^i = Au^i - f$ обобщенного метода сопряженных градиентов удовлетворяют следующим соотношениям ортогональности ($i = 1, \ldots, s$):

$$(Ag_i, Ag_j)_D = \delta_{ij} \| Ag_i \|_D, \qquad j = 1, \ldots, i,$$
$$(\xi^i, Ag_j)_D = 0, \ j = 1, \ldots, i, \quad (\xi^i, \xi^j)_{DAH} = 0, \quad j = 0, 1, \ldots, i-1,$$

где s — минимальное целое, для которого $K_{s-1}(A^{-1}\xi^0) \neq 0$.

39.44. Коэффициенты двучленного обобщенного метода сопряженных градиентов можно вычислять по формулам

$$\alpha_k = -\| \xi^{k-1} \|^2_{DAH} / \| \xi^{k-2} \|^2_{DAH}, \quad \beta_k = \| \xi^{k-1} \|^2_{DAH} / \| Ag_k \|^2_D.$$

39.45. Обобщенный метод сопряженных градиентов может быть реализован по следующим трехчленным формулам:

$$u^{k+1} = u^k - \frac{1}{q_k} \left[H\xi^k - e_{k-1} \left(u^k - u^{k-1} \right) \right],$$

где $e_{-1} = 0$,

$$q_k = \| AH\xi^k \|^2_D / \| \xi^k \|^2_{DAH} - e_{k-1}, \quad e_k = q_k \| \xi^{k+1} \|^2_{DAH} / \| \xi^k \|^2_{DAH}.$$

39.46. Пусть A и H — симметричные положительно определенные матрицы и $D = A^{-1}$. Тогда коэффициенты двучленного и трехчленного вариантов соответствующего обобщенного метода сопряженных градиентов вычисляются по формулам

$$\alpha_k = -\| \xi^{k-1} \|^2_H / \| \xi^{k-2} \|^2_H, \quad \beta_k = \| \xi^{k-1} \|^2_H / \| g_k \|^2_A,$$
$$q_k = \| H\xi^k \|^2_A / \| \xi^k \|^2_H - e_{k-1}, \quad e_k = q_k \| \xi^{k+1} \|^2_H / \| \xi^k \|^2_H.$$

39.47. Для последовательности D-норм векторов ошибок обобщенного метода сопряженных градиентов справедлива оценка

$$\| \xi^k \|_D \leqslant \left(C_k \left(\frac{M+m}{M-m} \right) \right)^{-1} \| \xi^0 \|_D,$$

где $M = \rho(HA)$, $1/m = \rho(H^{-1}A^{-1})$.

39.48. Обобщенный метод сопряженных градиентов из 39.46 является методом сопряженных градиентов:

$$\tilde{g}_k = \begin{cases} \tilde{\xi}^0, & k = 1, \\ \tilde{\xi}^{k-1} - \alpha_k \tilde{g}_{k-1}, & k > 1, \end{cases} \qquad v^k = v^{k-1} - \beta_k \tilde{g}_k,$$
$$\alpha_k = -\| \tilde{\xi}^{k-1} \|^2 / \| \tilde{\xi}^{k-2} \|^2, \quad \beta_k = \| \tilde{\xi}^{k-1} \|^2 / \| g_k \|^2_{\tilde{A}},$$

где $\tilde{\xi}^k = \tilde{A}v^k - \tilde{f}$, примененным к преобразованной системе $\tilde{A}v = \tilde{f}$ с матрицей $\tilde{A} = H^{1/2}AH^{1/2}$ и векторами $v = H^{-1/2}u$, $\tilde{f} = H^{1/2}f$.

§ 40. Итерационные методы решения систем с вырожденными матрицами

40.1. Пусть задана система линейных алгебраических уравнений $Au = f$ с квадратной вырожденной матрицей A порядка n и некоторым вектором $f \in \mathbf{R}_n$. Тогда применяемый для ее решения итерационный метод

$$u^k = F_k(A,\ f,\ u^{k-1},\ \ldots,\ u^0)$$

из § 34 называется *сходящимся*, если для любого начального вектора $u^0 \in \mathbf{R}_n$ последовательность векторов u^k сходится к некоторому обобщенному решению этой системы.

На протяжении всего параграфа, за исключением последних пунктов 40.33—40.37, мы будем предполагать, что исходная система $Au = f$ совместна, т. е. $f \in \operatorname{im} A$, и, следовательно, любое обобщенное решение этой системы является ее обычным решением.

40.2. Пусть T — некоторая матрица. Тогда выполнение условий

$$\rho^*(T) = \max_{\substack{\lambda \in \sigma(T)\\ \lambda \neq 1}} |\lambda| < 1, \quad \operatorname{rank}(E - T) = \operatorname{rank}(E - T)^2$$

необходимо и достаточно для существования матрицы $T_\infty = \lim\limits_{k\to\infty} T^k$. Если эти условия выполнены, то матрица T_∞ является проектором со следующими свойствами:

$$\operatorname{im} T_\infty = \ker(E - T), \quad \ker T_\infty = \operatorname{im}(E - T).$$

40.3. Каждое из следующих условий:

1) собственному числу $\lambda = 1$ матрицы T (если такое существует) соответствует столько линейно независимых собственных векторов, какова его кратность;

2) $\ker(E - T) \cap \operatorname{im}(E - T) = 0$;

3) $\mathbf{R}_n = \operatorname{im}(E - T) \oplus \ker(E - T)$

эквивалентно второму условию из 40.2.

40.4. Пусть относительно матрицы $T = E - HA$, где A — матрица исходной системы и H — некоторая, вообще говоря, произвольная матрица, выполнены оба условия из 40.2. Тогда $\ker A \subseteq \operatorname{im} T_\infty$, а для выполнения равенства $\ker A = \operatorname{im} T_\infty$ необходимо и достаточно, чтобы было выполнено дополнительное условие $\ker H \cap \operatorname{im} A = 0$.

40.5. Для сходимости (как уже отмечалось, мы везде предполагаем совместность исходной системы) линейного стационарного одношагового итерационного метода

$$u^k = u^{k-1} - H(Au^{k-1} - f)$$

с некоторой заданной матрицей H необходимо и достаточно, чтобы существовала матрица $T_\infty = \lim\limits_{k\to\infty} T^k$, где $T = E - HA$, и выполнялось включение $\operatorname{im} T_\infty \subseteq \ker A$.

40.6. Для сходимости итерационного метода из 40.5 необходимо и достаточно выполнение следующих трех условий:

1) $\rho^*(T) = \max\limits_{\substack{\lambda\in\sigma(T)\\ \lambda\neq 1}} |\lambda| < 1$;

2) $\operatorname{rank}(HA) = \operatorname{rank}(HA)^2$;

3) $\ker H \cap \operatorname{im} A = 0$.

Если эти условия выполняются, то матрица T_∞ является проектором со следующими свойствами:

$$\operatorname{im} T_\infty = \ker(HA) = \ker A, \quad \ker T_\infty = \operatorname{im}(HA).$$

В частности, пространство $\mathbf{R}_n$ является прямой суммой своих подпространств $\operatorname{im}(HA)$ и $\ker A$, т. е.

$$\mathbf{R}_n = \operatorname{im}(HA) \oplus \ker A.$$

40.7. Если итерационный метод из 40.5 сходится, то для любого начального приближения

$$u^0 = u_1^0 + u_2^0 \in \mathbf{R}_n, \quad u_1^0 \in \operatorname{im}(HA), \quad u_2^0 \in \ker A,$$

вектор $u^\infty = \lim\limits_{k\to\infty} u^k$ является решением исходной системы $Au = f$ и представим в виде $u^\infty = u_1^\infty + u_2^\infty$, где

$$u_1^\infty = \sum_{i=0}^{\infty} (E - HA)^i Hf \in \operatorname{im}(HA), \quad u_2^\infty = u_2^0 \in \ker A.$$

Таким образом, для сходящегося линейного стационарного итерационного метода из 40.5 предельный вектор u^∞ зависит от начального приближения u^0 или, точнее говоря, от второй составляющей этого вектора в разложении по подпространствам $\operatorname{im}(HA)$ и $\ker A$. Если же $u^0 \in \operatorname{im}(HA)$, то эта составляющая равна нулю, и, следовательно, предельный вектор u^∞ всегда равен вектору u_1^∞ из 40.6.

40.8. Пусть итерационный метод из 40.1 сходится, и, следовательно, для каждого $u^0 \in \mathbf{R}_n$ однозначно определен вектор $u^\infty = \lim\limits_{k\to\infty} u^k$, являющийся решением исходной системы. Тогда векторы $z^k = u^k - u^\infty$ называются *векторами ошибок* этого итерационного метода, а величина

$$r_\infty = \lim_{k\to\infty} \ln \left[\frac{1}{k} \sup_{z^0\in\mathbf{R}_n} \frac{\|z^k\|}{\|z^0\|} \right]$$

называется его *асимптотической скоростью сходимости*.

Таким образом, в отличие от итерационных методов решения систем с невырожденными матрицами, такие понятия, как векторы ошибок, разрешающие операторы, операторы перехода, средние и асимптотические скорости сходимости и другие, которые используют при своем определении векторы u^∞, в случае вырожденной матрицы A могут быть введены (для произвольного итерационного метода) только после того, как установлен факт сходимости этого метода. В частном случае, когда, например, итерационный метод имеет вид

$$u^k = u^{k-1} - H_k(\xi^{k-1}),$$

где $\xi^{k-1} = Au^{k-1} - f$ — векторы невязок и H_k: $\operatorname{im} A \to \mathbf{R}_n$ — некоторая последовательность операторов, действующих только на векторы невязок, для любого решения u^* системы $Au = f$ имеем

$$\psi^k = \psi^{k-1} - H_k(A\psi^{k-1}),$$

где $\psi^k = u^k - u^*$. Поэтому перечисленные понятия можно определять сразу и затем использовать их при изучении различных вопросов, связанных со сходимостью итерационного метода.

40.9. Для линейного стационарного одношагового итерационного метода из 40.5

$$r_\infty \equiv r_\infty(T) = -\ln \rho^*(T),$$

где величина $\rho^*(T)$ определена в 40.6.

40.10. Пусть A, H — симметричные положительно полуопределенные матрицы, $\ker H \cap \operatorname{im} A = 0$ и

$$0 < m = \min_{\substack{\lambda \in \sigma(HA) \\ \lambda \neq 0}} \lambda < M = \max_{\lambda \in \sigma(HA)} \lambda.$$

Тогда итерационный метод

$$u^k = u^{k-1} - \tau H(Au^{k-1} - f)$$

сходится для любого положительного $\tau < 2/M$. Кроме того, оптимальное значение параметра τ, максимизирующего асимптотическую скорость сходимости этого метода и соответствующее значение скорости, находится по формулам

$$\tau_{\text{опт}} = 2/(M + m), \quad r_\infty\left(T_{\tau_{\text{опт}}}\right) = (M - m)/(M + m).$$

40.11. Пусть для решения системы $Au = f$ применяется стационарный одношаговый итерационный метод

$$u^k = u^{k-1} - H(\xi^{k-1})$$

с непрерывным и однородным первой степени оператором H: $\operatorname{im} A \to \mathbf{R}_n$. Тогда, если в подпространстве $\operatorname{im} A$ может быть задана такая норма $\|\cdot\|_*$, что для любого ненулевого $\xi \in \operatorname{im} A$ выполняется неравенство

$$\|\xi - AH(\xi)\|_* < \|\xi\|_*,$$

то этот итерационный метод сходится. Более того, существует такое положительное $q < 1$, что

$$\|u^k - u^{k-1}\| \leqslant L(H)\, q^{k-1} \|\xi^0\|_*,$$

и, следовательно,

$$\|u^k - u^\infty\| \leqslant \frac{L(H)}{1-q} q^k \|\xi^0\|_*,$$

где $L(H)$ — константа Липшица оператора H по отношению к векторной норме $\|\cdot\|_*$.

40.12. Пусть $A = \Lambda - F - F'$ — симметричная положительно полуопределенная матрица, где Λ — симметричная положительно определенная матрица, F — некоторая, вообще говоря, произвольная матрица. Тогда, задавая в подпространстве $\operatorname{im} A$ норму $\|\xi\|_* = (A^+\xi, \xi)^{1/2}$, $\xi \in \operatorname{im} A$, для любого ω, при котором матрица $B_\omega = \frac{1}{\omega}\Lambda - F$ невырожденная, и любого вектора $\xi \in \operatorname{im} A$ имеем

$$\|\xi - AB_\omega^{-1}\xi\|_*^2 = \|\xi\|_*^2 - \left(\frac{2}{\omega} - 1\right)\|B_\omega^{-1}\xi\|_\Lambda^2.$$

Далее, для любого $\omega \in (0, 2)$ матрица B_ω невырожденная и выполняется неравенство ($\xi \neq 0$)

$$\|\xi - AB_\omega^{-1}\xi\|_* < \|\xi\|_*.$$

40.13. Пусть выполнены предположения предыдущего пункта. Тогда обобщенный метод последовательной верхней релаксации

$$B_\omega(u^k - u^{k-1}) = -(Au^{k-1} - f)$$

сходится тогда и только тогда, когда $\omega \in (0, 2)$.

40.14. Пусть A — симметричная положительно полуопределенная матрица. Тогда любой блочный и точечный методы Гаусса — Зейделя сходятся.

40.15. Пусть выполняются предположения из 40.12, матрицы $\tilde{L} = \Lambda^{-1}F$ и $\tilde{U} = \Lambda^{-1}F'$ являются строго нижней и верхней треугольными и матрица $S = \tilde{L} + \tilde{U}$ согласованно упорядоченная. Тогда оптимальное значение параметра релаксации ω, максимизирующее асимптотическую скорость сходимости соответствующего обобщенного метода последовательной верхней релаксации, находится по формуле

$$\omega_{\text{опт}} = 2\left(1 + \sqrt{1 - \rho^*(S^2)}\right)^{-1},$$

где $\rho^*(S^2) = \max\limits_{\substack{\lambda \in \sigma(S) \\ |\lambda| \neq 1}} \lambda^2$. При этом $r_\infty(T_{\omega_{\text{опт}}}) = \omega_{\text{опт}} - 1$, где $T_\omega = E - B_\omega^{-1}A$.

40.16. Если выполнены предположения из 40.15, то

$$\rho^*(S^2) < [\rho^*(S)] = 1,$$

а соответствующий итерационный метод

$$\Lambda u^k = (F + F')u^{k-1} + f$$

с матрицей перехода S сходится не для любого $u^0 \in \mathbf{R}_n$. В то же время он сходится для всех начальных приближений, для которых векторы невязок $\xi^0 = Au^0 - f$ ортогональны к собственным векторам матрицы $(F+F')\Lambda^{-1}$, соответствующим ее собственному числу $\lambda = -1$.

40.17. Пусть A — вырожденная блочная матрица из 35.2, такая, что для любого $\alpha > 0$ матрица $A + \alpha E$ является M-матрицей, и пусть $\Lambda = A_{11} \oplus \ldots \oplus A_{ss}$. Тогда, если матрица $S = \Lambda^{-1}(\Lambda - A)$ примитивная, то соответствующий блочный метод Якоби

$$\Lambda(u^k - u^{k-1}) = -(Au^{k-1} - f)$$

сходится.

40.18. Пусть выполняются предположения предыдущего пункта, $A = \Lambda - L - U$, где L и U — строго нижняя и верхняя треугольные матрицы, и $T = (\Lambda - L)^{-1}U$ — блочная матрица вида

$$T = \begin{bmatrix} 0 & T_{12} & \ldots & T_{1s} \\ 0 & T_{22} & \ldots & T_{2s} \\ \cdot & \cdot & \cdot & \cdot \\ 0 & T_{s2} & \ldots & T_{ss} \end{bmatrix},$$

который согласован с блочным видом матрицы A из 35.2. Тогда, если матрица

$$\begin{bmatrix} T_{22} & \ldots & T_{2s} \\ \cdot & & \cdot \\ T_{s2} & \ldots & T_{ss} \end{bmatrix}$$

примитивная, то блочный метод Гаусса — Зейделя

$$(\Lambda - L)(u^k - u^{k-1}) = -(Au^{k-1} - f)$$

сходится.

40.19. Пусть $A = A_1 + A_2$, где A_1, A_2 — симметричные положительно полуопределенные матрицы. Тогда соответствующий метод расщепления

$$(E + \tau A_1)(E + \tau A_2)(u^k - u^{k-1}) = -2\tau(Au^{k-1} - f)$$

сходится для любого положительного значения параметра τ.

40.20. Если выполняются предположения из 40.19 и $\ker A_1 = \ker A_2$, то при $\tau = (\delta\Delta)^{-1/2}$, где

$$\delta = \min_{i=1,2}\Big\{\min_{\substack{\lambda\in\sigma(A_i)\\ \lambda\neq 0}} \lambda\Big\} < \Delta = \max_{i=1,2}\Big\{\max_{\lambda\in\sigma(A_i)} \lambda\Big\},$$

для асимптотической скорости сходимости метода расщепления выполняется оценка

$$r_\infty(T_\tau) \geqslant 2 \ln \frac{\sqrt{\Delta} + \sqrt{\delta}}{\sqrt{\Delta} - \sqrt{\delta}}.$$

40.21. Пусть $A = A_1 + A_1'$ — положительно полуопределенная матрица и $B_\tau = (E + \tau A_1)(E + \tau A_1')$, где τ — положительный параметр. Тогда существует отрицательное $\alpha > -1$ такое, что

$$\sigma(E - 2\tau B_\tau^{-1}A) \subset [\alpha, 1].$$

40.22. Попеременно треугольный метод

$$(E + \tau A_1)(E + \tau A_1')(u^k - u^{k-1}) = -2\tau(Au^{k-1} - f),$$

где $A = A_1 + A_1'$ — положительно полуопределенная матрица, сходится для любого $\tau > 0$.

40.23. Пусть $A = A_1 + A_2$, где A_1, A_2 — кососимметрические матрицы, т. е. $(A_i v, v) = 0$ для любого $v \in \mathbf{R}_n$, $i = 1, 2$. Тогда для метода расщепления из 40.19 для любого $\xi^{k-1} \in \mathbf{R}_n$ имеем

$$\|(E + \tau A_1)^{-1}\xi^k\| = \|(E + \tau A_1)^{-1}\xi^{k-1}\|,$$

и, следовательно, он не сходится ни для какого значения параметра τ.

40.24. Пусть для заданной матрицы T в пространстве $\mathbf{R}_n$ можно определить такую норму $\|\cdot\|_*$, что $\rho(T) = \|T\|_*$. Тогда любому собственному числу λ, равному по модулю $\rho(T)$, соответствует столько линейно независимых собственных векторов, какова его кратность.

40.25. Пусть для матрицы перехода $T = E - HA$ итерационного метода из 40.5 в пространстве $\mathbf{R}_n$ может быть определена такая норма $\|\cdot\|_*$, что $\|T\|_* \leqslant 1$. Тогда матрица

$$\widehat{T} = \lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} T^i$$

существует и является проектором со следующими свойствами:

$$\operatorname{im} \widehat{T} = \ker(HA), \quad \ker \widehat{T} = \operatorname{im}(HA).$$

40.26. Пусть выполнены предположения из 40.25 и $\ker H \cap \operatorname{im} A = 0$. Тогда для любого $u^0 \in \mathbf{R}_n$ последовательность векторов

$$w^k = \frac{1}{k} \sum_{i=0}^{k-1} u^i$$

итерационного метода из 40.5 сходится к некоторому решению w^∞ системы $Aw = f$.

40.27. Пусть $A = A_1 + A_2$, где A_1 и A_2 — положительно полуопределенные матрицы. Тогда матрица перехода метода расщепления

$$(E+\tau A_1)(E+\tau A_2)(u^k-u^{k-1})=-2\tau(Au^{k-1}-f)$$

при любом $\tau>0$ удовлетворяет требованиям из 40.27, причем $\|\cdot\|_*=\|\cdot\|_D$, где $D=(E+\tau A_2)'(E+\tau A_2)$.

40.28. Пусть A_1 — заданная симметричная положительно полуопределенная матрица порядка n_1, H — заданная симметричная положительно определенная матрица порядка $n>n_1$, A_2 — некоторая симметричная положительно полуопределенная матрица порядка $n_2=n-n_1$ и

$$A=A_1\oplus A_2=\begin{bmatrix}A_1 & 0\\ 0 & A_2\end{bmatrix},\quad A_0=\begin{bmatrix}A_1 & 0\\ 0 & 0\end{bmatrix}$$

— блочно диагональные матрицы порядка n. Тогда

$$\max_{A_2}\left\{\min_{\substack{\lambda\in\sigma(HA)\\ \lambda\neq 0}}\lambda\right\}=\min_{\substack{\lambda\in\sigma(HA_0)\\ \lambda\neq 0}}\lambda=m_0,$$

$$\min_{A_2}\rho(HA)=\rho(HA_0).$$

40.29. Пусть относительно матриц A_1, A_2, H, A и A_0 выполняются предположения из 40.28 и задана система $A_1u_1=f_1$ с вектором $f_1\in\operatorname{im}A_1$. Тогда для нахождения решения этой системы может быть применен, например, итерационный метод

$$u^k=u^{k-1}-\tau_0H(A_0u^{k-1}-f)$$

с вектором $f=\begin{bmatrix}f_1\\ 0\end{bmatrix}$ и параметром $\tau_0=\dfrac{2}{m_0+\rho(HA_0)}$ (m_0 определено в 40,28), который асимптотически не медленнее итерационного метода

$$u^k=u^{k-1}-\tau H\left(Au^{k-1}-f\right)$$

с параметром $\tau=\dfrac{2}{m+\rho(HA)}$, где $m=\min\limits_{\lambda\in\sigma(HA),\lambda\neq 0}\lambda$, при любой матрице A_2. Изложенная процедура перехода к системе с матрицей большого порядка называется *методом фиктивных компонент.*

Из 40.28 и 40.29 следует, что при сделанных относительно матриц A_1, A_2 и H предположениях оптимальным вариантом метода фиктивных компонент является случай $A_2=0$. В последние годы метод фиктивных компонент широко используется для решения систем уравнений с большими разреженными матрицами, возникающими при дискретизации уравнений математической физики. Его использование открывает хорошие возможности для выбора легко обратимых матриц $B=H^{-1}$, порядок которых является произведением целых чисел. Развиваются и другие способы построения матриц A, например, типа

$$A=\begin{bmatrix}A_1 & A_{12}\\ 0 & A_2\end{bmatrix}.$$

40.30. Пусть HA является матрицей простой структуры, все ее собственные числа неотрицательны и $\ker H \cap \operatorname{im} A = 0$. Тогда для решения системы $Au = f$ применим любой из чебышевских итерационных методов § 38 и имеют место соответствующие оценки для норм векторов ошибок. При этом в качестве величин a и b могут быть выбраны произвольные положительные числа, удовлетворяющие неравенствам

$$a \leqslant \min_{\substack{\lambda \in \sigma(HA) \\ \lambda \neq 0}} \lambda, \quad b \geqslant \rho(HA).$$

40.31. Пусть A, H — симметричные положительно полуопределенные матрицы, удовлетворяющие условию $\ker H \cap \operatorname{im} A = 0$. Тогда для решения системы $Au = f$ может быть применен обобщенный метод сопряженных градиентов, двучленные формулы которого имеют вид

$$g_k = \begin{cases} H\xi^0, & k = 1, \\ H\xi^{k-1} - \alpha_k g_{k-1}, & k > 1, \end{cases} \qquad u^k = u^{k-1} - \beta_k g_k,$$

$$\alpha_k = -\frac{(H\xi^{k-1}, \xi^{k-1})}{(H\xi^{k-2}, \xi^{k-2})}, \quad \beta_k = \frac{(H\xi^{k-1}, \xi^{k-1})}{(Ag_k, g_k)}.$$

40.32. Пусть заданы вырожденная матрица A и некоторая матрица H, удовлетворяющая условию $\ker H \cap \operatorname{im} A = 0$. Тогда системы $Au = f$ и $HAv = Hf$ эквивалентны, т. е. множества их обобщенных решений совпадают при любом $f \in \mathbf{R}_n$ в том и только том случае, если подпространство $\ker A'$ инвариантно относительно матрицы $H'H$.

40.33. Пусть A, H — симметричные положительно полуопределенные матрицы, удовлетворяющие условиям $\rho^*(E - HA) < 1$, $\ker H \cap \operatorname{im} A = 0$ и $H \ker A \subseteq \ker A$. Тогда последовательность векторов

$$v^k = u^k - k(u^k - u^{k-1})$$

итерационного метода

$$u^k = u^{k-1} - H(Au^{k-1} - f)$$

сходится при любом начальном приближении u^0 при $k \to \infty$ к некоторому обобщенному решению системы $Av = f$.

40.34. Пусть A, H — симметричные положительно полуопределенные матрицы, удовлетворяющие условиям $\rho^*(E - HA) < 1$ и $\ker H \cap \operatorname{im} A = 0$. Тогда в случае несовместности системы $Au = f$, т. е. $f \notin \operatorname{im} A$, ее обобщенное нормальное решение может быть найдено с помощью следующего вычислительного процесса, где используется обозначение $P = \lim\limits_{k \to \infty} (E - HA)^k$.

1. Сведение исходной системы к эквивалентной совместной системе путем ортогонализации правой части f к ядру

матрицы A:

$$g_k = \begin{cases} Pf, & k=1, \\ Pf^{k-1} - \sum\limits_{i=1}^{k-1} \alpha_{ki} g_i, & k>1, \end{cases} \qquad f^k = f^{k-1} - \beta_k g_k,$$

$$\alpha_{ki} = \frac{(Pf^{k-1}, g_i)}{\|g_i\|^2}, \qquad i = 1, \ldots, k-1,$$

$$\beta_k = \frac{(f^{k-1}, g_k)}{\|g_k\|^2}, \qquad k = 1, \ldots, s \leqslant \dim \ker A,$$

где $f^0 = f$.

2. Нахождение обобщенного решения u^* системы $Au = f$, являющегося решением совместной системы $Av = \hat{f}$, где $\hat{f} = f^s$:

$$v^k = v^{k-1} - H(Av^{k-1} - f), \quad u^* = \lim_{k\to\infty} v^k.$$

3. Нахождение обобщенного нормального решения исходной системы путем ортогонализации вектора u^* к ядру матрицы A:

$$g_k = \begin{cases} Pu^*, & k=1, \\ Pu^{k-1} - \sum\limits_{i=1}^{k-1} \alpha_{ki} g_i, & k>1, \end{cases} \qquad u^k = u^{k-1} - \beta_k g_k,$$

$$\alpha_{ki} = \frac{(Pu^{k-1}, g_i)}{\|g_i\|^2}, \qquad i = 1, \ldots, k-1,$$

$$\beta_k = \frac{(u^{k-1}, g_k)}{\|g_k\|^2}, \qquad k = 1, \ldots, l \leqslant \dim \ker A,$$

где $u^0 = u^*$ и $\hat{u} = u^l$ будет искомым обобщенным нормальным решением системы $Au = f$.

Заметим, что значения s и l процесса из 40.34 являются минимальными целыми, для которых соответственно векторы g^{s+1} и g^{l+1} оказываются нулевыми.

40.35. В случае, когда выполнены предположения из 40.34 и $\dim \ker A = 1$, метод нахождения обобщенного нормального решения $\hat{u}$ системы $Au = f$ принимает следующий вид:

1) $\hat{f} = f - \dfrac{(Pf, f)}{\|Pf\|^2} Pf$;

2) $u^* = Pu^0 + \lim\limits_{k\to\infty} \sum\limits_{i=0}^{k} (E - HA)^i H\hat{f}, \; u^0 \in \mathbf{R}_n$;

3) $\hat{u} = u^* - \dfrac{(Pu^*, u^*)}{\|Pu^*\|^2} Pu^*$.

40.36. Пусть выполнены предположения из 40.34 и последовательность векторов $\widehat{f}_\varepsilon$ сходится при $\varepsilon \to 0$ к вектору $\widehat{f}$. Тогда последовательность векторов

$$u_\varepsilon^* = \lim_{k \to \infty} u_\varepsilon^k,$$

где u_ε^k находятся с помощью итерационного метода

$$v_\varepsilon^k = v_\varepsilon^{k-1} - H\left(Av_\varepsilon^{k-1} - \widehat{f}_\varepsilon\right), \quad u_\varepsilon^k = v_\varepsilon^k - k\left(v_\varepsilon^k - v_\varepsilon^{k-1}\right),$$

сходится при $\varepsilon \to 0$ к вектору u^*.

ЛИТЕРАТУРА

Бахвалов Н. С. Численные методы.— М.: Наука, 1975.

Беллман Р. Введение в теорию матриц.— М.: Мир, 1969.

Березин И. С., Жидков Н. П. Методы вычислений.— Т. 1. Изд. 3-е. М.: Наука, 1966; Т. 2. Изд. 2-е. М.: Физматгиз, 1962.

Вазов В., Форсайт Дж. Разностные методы решения дифференциальных уравнений в частных производных.— М.: ИЛ, 1963.

Воеводин В. В. Численные методы алгебры. Теория и алгорифмы.— М.: Наука, 1966.

Воеводин В. В. Ошибки округления и устойчивость в прямых методах линейной алгебры.— М.: Изд-во МГУ, 1969.

Воеводин В. В. Линейная алгебра.— М.: Наука, 1980.

Воеводин В. В. Вычислительные основы линейной алгебры.— М.: Наука, 1977.

Вычислительные методы линейной алгебры. Тр. I Всесоюзной конференции.— Новосибирск, ВЦ СО АН СССР, 1969.

Вычислительные методы линейной алгебры. Тр. II Всесоюзной конференции.— Новосибирск, ВЦ СО АН СССР, 1972.

Вычислительные методы линейной алгебры. Тр. III Всесоюзной конференции.— Новосибирск, ВЦ СО АН СССР, 1974.

Вычислительные методы линейной алгебры. Тр. IV Всесоюзной конференции.— Новосибирск, ВЦ СО АН СССР, 1977.

Вычислительные методы линейной алгебры. Тр. V Всесоюзной конференции.— Новосибирск, ВЦ СО АН СССР, 1980.

Гантмахер Ф. Р. Теория матриц.— М.: Наука, 1967.

Гельфанд И. М. Лекции по линейной алгебре.— М.: Наука, 1971.

Годунов С. К. Решение систем линейных уравнений.— Новосибирск: Наука, 1980.

Дьяконов Е. Г. Разностные методы решения краевых задач.— М.: Изд-во МГУ, 1971.

Икрамов Х. Д. Задачник по линейной алгебре.— М.: Наука, 1975.

Икрамов Х. Д. Разреженные матрицы.— В кн.: Математический анализ. Итоги науки и техники. М.: ВИНИТИ, 1982.

Коллатц Л. Задачи на собственные значения с техническими приложениями.— М.: Наука, 1968.

Коллатц Л. Функциональный анализ и вычислительная математика.— М.: Мир, 1969.

Ланкастер П. Теория матриц.— М.: Наука, 1978.

Ланцош К. Практические методы прикладного анализа.— М.: Физматгиз, 1961.

Мальцев А. И. Основы линейной алгебры.— М.: Наука, 1970.

Маркус М., Минк Х. Обзор по теории матриц и матричных неравенств.— М.: Наука, 1972.

Марчук Г. И. Методы вычислительной математики.— М.: Наука, 1980.

Марчук Г. И., Кузнецов Ю. А. Итерационные методы и квадратичные функционалы.— Новосибирск: Наука, 1972.

Марчук Г. И., Лебедев В. И. Численные методы переноса нейтронов.— М.: Атомиздат, 1981.
Милн В. Э. Численный анализ.— М.: ИЛ, 1951.
Островский А. М. Решение уравнений и систем уравнений.— М.: ИЛ, 1963.
Парлетт Б. Симметричная проблема собственных значений.— М.: Мир, 1983.
Пароди М. Локализация характеристических чисел матриц и ее приложения.— М.: ИЛ, 1960.
Проскуряков И. В. Сборник задач по линейной алгебре.— М.: Наука, 1967.
Самарский А. А. Введение в теорию разностных схем.— М.: Наука, 1971.
Самарский А. А. Теория разностных схем.— М.: Наука, 1977.
Самарский А. А., Николаев Е. С. Методы решения сеточных уравнений.— М.: Наука, 1978.
Стренг Г. Линейная алгебра и ее применения.— М.: Мир, 1980.
Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач.— М.: Наука, 1979.
Тьюарсон Р. Разреженные матрицы.— М.: Мир, 1977.
Уилкинсон Дж., Райнш К. Справочник алгоритмов на языке Алгол. Линейная алгебра.— М.: Машиностроение, 1976.
Уилкинсон Дж. Х. Алгебраическая проблема собственных значений.— М.: Наука, 1970.
Фаддеев Д. К., Соминский И. Я. Сборник задач по высшей алгебре.— М.: Наука, 1972.
Фаддеев Д. К., Фаддеева В. Н. Вычислительные методы линейной алгебры.— М.— Л.: Физматгиз, 1963.
Фаддеев Д. К., Фаддеева В. Н. Вычислительные методы линейной алгебры. Зап. научн. семин.— Л.: ЛОМИ АН СССР, 1975.
Фаддеева В. Н., Кузнецов Ю. А. и др. Вычислительные методы линейной алгебры. Библиографический указатель 1828—1974.— Новосибирск: Наука. 1976.
Фаддеева В. Н., Икрамов Х. Д. и др. Вычислительные методы линейной алгебры. Библиографический указатель 1975—1980.— Л.: Наука, 1982.
Фаддеева В. Н., Колотилина Л. Ю. Вычислительные методы линейной алгебры. Набор матриц для тестирования.— Л.: Наука. Ч. I, 1982; Ч. II, III, 1983.
Форсайт Дж., Моллер К. Численное решение систем линейных алгебраических уравнений.— М.: Мир, 1969.
Фрезер Р., Дункан В., Коллар А. Теория матриц и ее приложения к дифференциальным уравнениям и динамике.— М.: ИЛ, 1950.
Хаусхолдер А. С. Основы численного анализа.— М.: ИЛ, 1956.
Яненко Н. Н. Метод дробных шагов решения многомерных задач математической физики.— Новосибирск: Наука, 1967.
Aitken A. C. Determinants and matrices.— Edinburgh — London: Oliver and Boyd, 1956.
Alefeld G., Herzberger J. Einführung in die Intervallrechnung. Manheim u. a.: B. I.— Wissenschaftverl., 1974.
Allen D. N. de G. Relaxation methods.— New York — Toronto — London: McGraw-Hill, 1954.
Bauer F. L., Heinhold J., Samelson K., Sauer R. Moderne Rehenanlagen; eine Einführung.— Stuttgart: B. G. Teubner, 1965.
Bjerhammar A. Theory of errors and generalized matrix inverses.— Amsterdam: Elsevier, 1973.
Blumental B. Einführung in die Matrizenrechnung, Allgemeinverständl. Darst. für Nichtmathematiker.— Berlin; Verl. Tchnik, 1960.
Bunch J., Rose D. (Eds.). Sparse matrix computations.— New York — San Francisco — London: Academic Press, 1976.

Cullen C. G. Matrices and linear transformations.— Reading — London: Addison-Wesley, 1966.
Dekker T. J. Evaluation of determinants, solution of systems of linear equations and matrix inversion.— Amsterdam: Bekenafdeling, 1963.
Durand E. Solutions numériques des équations algébriques. T. 2.— Paris: Masson, 1961.
Forsythe G. E. A numerical analyst's fifteen-root shelf.— Math. Tables and Other Aids Comput., 1953.
Forsythe G. E. Tentative classification of methods and bibliography on solving systems of linear equations.— Nat. Bur. Stand. Appl. Math. Ser., 1953, 29, p. 1—28.
Fox L. An introduction to numerical linear algebra.— Oxford: Clarendon Press, 1964.
Fox L., Parker I. B. Chebyshev polynomials in numerical analysis.— London: Oxford Univ. Press, 1968.
Fröberg G.-E. Introduction to numerical analysis.— Reading — London: Addison-Wesley, 1965.
Gastinel N. Analyse numèrique linéaire.— Paris: Hermann, 1966.
Gourlav A. R., Watson G. A. Computational methods for matrix eigenproblems.— London: J. Wiley, 1973.
Graeub W. Lineare Algebra.— Berlin; Springer, 1958.
Hammerling S. J. Latent roots and latent vectors.— London — Toronto — Buffulo: Higler and Watts — Univ. Toronto Press, 1970.
Hamming R. W. Numerical methods for scientists and engineers.— New York: McGraw-Hill, 1962.
Henrici P. Elements of numerical analysis.— New York: J. Wiley, 1964.
Householder A. S. The theory of matrices in numerical analysis.— New York: Blaisdell, 1964.
Householder A. S. Kwic index for numerical algebra.— Oak Ridge: Oak Ridge National Laboratory, 1972.
Householder A. S. Kwic index for numerical algebra.— Oak Ridge: Oak Ridge National Laboratory, 1973.
Householder A. S. Kwic index for numerical algebra.— Oak Ridge, Oak Ridge National Laboratory, 1975.
Khabaza I. M. Numerical analysis.— Oxford — London: Pergamon Press, 1965.
Korganoff A., Pavel-Parvu M. Méthodes de calcul numérique. T. 2. Elements de théorie des matrices carrées et rectangles en analyse en analyse numérique.— Paris: Dunod, 1967.
Lancaster P. Lambda matrices and vibrating systems.— Oxford: Pergamon Press, 1966.
Lewis T. O., Boullion T. L., Odell P. L. A bibliography on generalized matrix inverses.— Proc. Symp. on Theory and Appl. on Generalized Inverses of Matrices, Texas tech. college.— Lubhock: Texas, 1968, 283—315.
Marcus M., Minc H. Introduction to linear algebra.— New York: McMillan; London: Collier — McMillan, 1965.
Moore R. E. Interval analysis.— Englewood Cliffs, N. Y.: Prentice — Hall, 1966.
Parker W. V., Eaves J. C. Matrices.— New York: Ronald Press, 1960.
Rao C. R., Mitra S. K. Generalized inverse of matrices and its applications.— New York: J. Wiley and Sons, 1971.
Reid J. K. (Ed.) Large sparse sets of linear equations.— London — New York: Acad. Press, 1971.
Rose D. J., Willoughby R. A. (Eds). Sparce matrices and their applications.— New York — London: Plenum Press, 1972.
Schwarz H. R. Numerik symmetrischer Matrizen.— Stuttgart: Teubner, 1968.
Schwarz H. R., Rutishauser H., Stiefel E. Numerik symmetrischer Matrizen.— Stuttgart: Teubner, 1972.

Stanton R. G. Numerical methods for science and engineering.— Englewood Cliffs, N. J.: Prentice-Hall, 1961.
Stewart G. W. Introduction to matrix computations.— New York: Acad. Press, 1973.
Stiefel E. Einführung in die numerische Mathematik.— Stuttgart; Teubner, 1961.
Stoer J. Einführung in die Numerische Mathematik. Bd. I.— Berlin: Springer, 1972.
Stoer J., Bulirsch R. Einfüchrung in die Numerische Mathematik. Bd. 2.— Berlin: Springer, 1973.
Turnbull H. W. The theory of determinants, matrices and invariants.— New York: Dover, 1960.
Turnbull H. W., Aitken A. C. An introduction to the theory of canonical matrices.— New York: Dover, 1961.
Varga R. S. Matrix iterative analysis.— Englewood Cliffs, N. J.: Prentice-Hall, 1962.
Werner H. Praktische Mathematik. I. Methoden der Linearen Algebra.— Berlin: Springer, 1970.
Westlake J. R. A handbook of numerical matrix inversion and solution of linear equations.— New York: J. Wiley, 1968.
Wilkinson J. H. Rounding errors in algebraic processes.— London: H. M. Stat. Off., 1963.
Wilkinson J. H., Reinsch C. Linear Algebra.— Berlin: Springer, 1971.
Young D. M. Iterative solution of large linear systems.— New York — London: Acad. Press, 1971.
Zurmühl R. Matrizen und ihre technischen Anwendungen.— Berlin: Springer, 1961.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Валентин Васильевич Воеводин,
Юрий Алексеевич Кузнецов

МАТРИЦЫ И ВЫЧИСЛЕНИЯ

(серия: «Справочная математическая библиотека»)

Редактор *Е. Е. Тыртышников*
Техн. редактор *В. Н. Кондакова*
Корректоры *Т. С. Плетнева, Т. С. Вайсберг*

ИБ № 12063

Сдано в набор 16.09.83. Подписано к печати 19.01.84. Формат $60\times90^1/_{16}$. Бумага для глубокой печати. Обыкновенная гарнитура. Высокая печать. Условн. печ. л. 20. Усл. кр.-отт. 20. Уч.-изд. л. 21,69. Тираж 30 000 экз. Заказ № 805. Цена 1 р. 40 к.

Издательство «Наука»
Главная редакция физико-математической литературы
117071, Москва, В-71, Ленинский проспект, 15

4-я типография издательства «Наука»
630077, Новосибирск, 77, Станиславского, 25